Person Re-identification Using Multiple First-Person-Views on Wearable Devices

Anirban Chakraborty
Nanyang Technological University, Singapore 639798
a.chakraborty@ntu.edu.sg

Bappaditya Mandal
Institute for Infocomm Research (I^2R), A*STAR, Singapore 138632
bmandal@i2r.a-star.edu.sg

Hamed Kiani Galoogahi
Istituto Italiano di Tecnologia (IIT), Genova, 16163, Italy
kiani.galoogahi@iit.it

Abstract

The rise of wearable devices has led to many new ways of re-identifying an individual. Unlike static cameras, where the views are often restricted or zoomed out and occlusions are common scenarios, first-person-views (FPVs) or ego-centric views see people closely and mostly get unoccluded face images. In this paper, we propose a face re-identification framework designed for a network of multiple wearable devices. This framework utilizes a global data association method termed as Network Consistent Re-identification (NCR) that not only helps in maintaining consistency in association results across the network, but also improves the pair-wise face re-identification accuracy. To test the proposed pipeline, we collected a database of FPV videos of 72 persons using multiple wearable devices (such as Google Glasses) in a multi-storied office environment. Experimental results indicate that NCR is able to consistently achieve large performance gains when compared to the state-of-the-art methodologies.

1. Introduction

During the past few years there has been an exponential increase in the development of microelectronics and computer systems, enabling wearable sensors and mobile devices with unprecedented characteristics. To name a few, Google Glass (GG) [14] and GoPro [15] are such devices. These wearable devices can capture, record and analyze video data in the areas of human identification [21, 36, 35], which is of paramount interest within the field, especially for surveillance or monitoring, visual assistance to elderly, social interactions and security applications. These wearable devices (such as GG) can easily be networked and they can communicate and share information among each other. For example, when a thief steals a product in a large multistoried shopping mall, using multiple networked cameras (both static as well as wearable glasses) the security personnel can locate and catch the thief easily as compared to our earlier days of only standalone static cameras, mostly aided by availability of unconstrained high quality face shots at any location in the mall (Fig. 1).

Figure 1. Illustrative diagram for person re-identification using multiple first-person-view cameras. Five wearable devices (Google Glasses, Cam 1-5), interconnected and worn by security personnels at different levels in a multi-stored congested shopping mall, supply uncluttered face shots of the target at any location.

Most of the recent works focus on person re-identification problem taking into account the color/texture features associated mostly with the clothing of individuals [19, 4] and sometimes their pattern of movements using surveillance cameras. However, typically the surveillance cameras are set up to capture wide area videos and hence the individual targets are often few pixels in size in these cam-

era Field-of-Views (FoVs). Naturally, capturing face information of individuals has been very challenging for these surveillance cameras because of long distances (very small faces) and heavily occluded (by self human pose and other objects) regions. Hence, these traditional surveillance cameras stick to the analysis of color/textures of the observed targets, which are often non-discriminative and heavily affected by clutter, occlusion and wide illumination variation across camera FoVs. With the advancement of wearable devices (like GG), a network of multiple first-person-view (FPV) cameras is a good solution to alleviate the aforementioned challenges in person re-identification, as they can supply zoomed in, uncluttered face shots of targets. Besides, this network of FPVs can help in sharing information of the captured face, get relevant feedback from the caregivers and thereby prevent undesirable admission of the unknown people in private premises and improve safety.

In this paper, we present a framework for performing person re-identification using multiple wearable cameras supplying first-person-view facial images of targets. For this, we successfully combine the state-of-the-art holistic discriminative feature computation methods from the FPV face recognition literature with the robust data association techniques reported in the person re-identification community. To the best of our knowledge, this is the first work to perform person re-identification using FPV face images from a network of wearable devices. We also collect a wearable device re-id database where first person videos of 72 targets are captured using 4 GGs in the most realistic set up. In the network of more than two wearable cameras, multiple paths of association may exist between observation of the same target that often gives rise to network inconsistency [12]. Moreover, unlike classic person re-id problem, not all the persons are observed in all the camera. We ensure that the proposed pipeline can handle both of these *real-world* challenges and show that it achieves high accuracy in the collected FPV video dataset.

1.1. Related Work

Classic person re-identification. Person re-identification using multiple FPVs or egocentric views is a relatively new approach. In the classical person re-identification problem, typically the camera field-of-views are wide and whole targets are observed at a distance. Hence, the low resolution of the targets is often the main source of challenge in person re-identification. Existing camera pairwise person re-identification approaches can be roughly divided into three categories- (i) discriminative signature based methods [2, 4, 19, 27, 18, 40], (ii) metric learning based methods [6, 1, 38, 13], and (iii) transformation learning based methods [16, 29]. However, all of these methods suffer from the inherent challenges in person re-id datasets, viz., weakly discriminative features caused by low resolution, occlusion and dependence on color/texture features because of inability of capturing high-resolution, discriminative facial images.

First-person-views for face identification. Face identification (FI) in unconstrained environment has remained a dominating research area in the recent years owing to its countless useful practical applications [32]. For humans, face recognition (FR) is the most natural and common way to identify and/or verify individuals. This involves recognizing individuals based on what we see at a distance, e.g., first-person-view videos or images. Numerous researchers have begun collecting FPV videos for FR, involving two main tasks FI and face verification (FV) [21]. Capturing faces in non-occluded conditions together with sufficient face size (resolutions) have been a challenging task for static cameras (such as in cases of surveillance and monitoring). A human face not only defines identity but many more other attributes of the owner as well, such as personality, intension, trustworthiness, aggressiveness etc. [3]. Hence, analyzing faces with these wearable devices in cases where static cameras fails, is of paramount importance [23, 33].

The recent popularity of high quality wearable cameras such as GG and GoPro have created an opportunity to revisit the problem of capturing faces in partially occluded condition with sufficient face sizes. As compared to the static or fixed cameras, wearable devices have advantage of capturing the faces in much less non-occluded conditions. Mandal *et al.* in [25] have evaluated large number of local features with many distance measures on a wearable device database and have shown that features like binarized statistical image features (BSIF) and histogram of oriented gradients (HOG) tend to outperform other local features such as local binary patterns (LBP), local phase quantization (LPQ), local intensity order pattern (LOIP) for FR task. Moreover, features like scale invariant feature transform (SIFT) performs well when the number of persons in the database is small. When large number of images is available in the gallery, BSIF outperforms all other local features [25].

Extracting the above mentioned local features from face images are time consuming and they are of typically > 250 dimensions, making it unattractive for wearable devices that has limited computational resources [11]. To overcome these limitations, we use the entire face image based holistic features extracted using the recently proposed whole space subclass discriminant analysis (WSSDA) method for FR [26]. It is reported to be a good performer using lower dimensions (typical < 80 features) among many related methodologies [22] and is also suitable for wearable devices [21]. We use these robust low-dimensional features which are reported to be highly discriminative for FPV face videos. For comparison purpose we use the popular holistic features for FR, such as the principal component analysis (PCA) [34] and FisherFaces [5] and show that using the

proposed technique, we can have large improvement in the person re-identification accuracy over the baselines.

Consistent data association. Although the high quality facial features captured using wearable devices ([21, 23]) are more discriminative in general than the typical color/texture based features used in person re-id, they are still camera pairwise and has to be processed by a global data association method for generating consistent and improved results at the network level. Some recent work aims to find point correspondences in monocular image sequences [30] or links detections in a tracking scenario by solving a constrained flow optimization [8], or by using sparse appearance preserving tracklets [7]. However, all these flow based methods need temporal order information of observations to be known a-priori, which is not available in most re-identification problems. Recently, in [12], the authors have presented a network-consistent re-identification (NCR) method that does not require time order information of observations and proposes a scalable optimization framework for yielding globally consistent association results with high accuracy. However, [12] shows experiments on a wide area database and does not utilize face as an important cue for re-identification.

2. Person Re-identification From Multiple First Person Views

The proposed re-identification pipeline has two distinct parts cascaded to one another -

1. Computation of features from acquired first person view images in each device and subsequent estimation of feature similarity/distance scores between all pairs of observations in each camera pair. Following the general and widely accepted assumption in person re-identification problem set up, we assume that the observations from the same target in the same camera field-of-view can be clustered a-priori and hence intra-camera similarity score computation is not required in this problem.

2. When observations are acquired using more than two wearable devices/cameras, *network consistency* is enforced using network consistent re-identification framework. The inter-camera similarity scores computed in step 1 are used as inputs to this system and outputs are the final association labels between pairs of observations across any two camera.

2.1. Normalization and Feature Extraction

We use the implementation of FR on GG proposed by Mandal *et al.* in [21] and adopted a client-server architecture, where all the FR processes are performed on the server for less power consumption [21]. Our method uses the OpenCV face detector [28] to find faces in the incoming images. OpenCV eye detector [28] and integration of sketch and graph patterns (ISG) [39] based eye detector are fused together to locate the pair of eyes in oblique and frontal view faces. Through the integration of both eye detectors, high success rate of eye localization in the face images of FPV for both frontal and non-frontal faces at various scales (sizes) are achieved [21]. Using the detected eye coordinates, faces are aligned, cropped and resized to 67×75 pixels. Same normalization procedure is followed as described in [21]. Using the normalized face images, discriminative face features are extracted using popular FR algorithms like eigenfaces using principal component analysis (PCA) [34] and fisherfaces using PCA+linear discriminant analysis [31]. We also explore features from the recently proposed within-subclass subspace learning for FR in [26].

2.1.1 Within-Subclass Subspace Learning for Face Recognition

Traditional discriminant analysis techniques that employ between-class and within-class scatter information, when applied to FPV face images (with unconstrained lighting, expression and pose conditions), may lose crucial discriminant information [20, 9, 24]. Mandal *et al.* utilized the subclass discriminant analysis [41] and 'eigefeatures' feature regularization methodology [17] to alleviate the problems of modeling large variances appearing in within-class face images (images of an individual) and proposed the within-subclass subspace discriminant analysis (WSSDA) in [26]. On these regularized features, total-subclass and between-subclass scatter matrices (depending on the clusters for each person and the number of people in the database) are computed. Dimensionality reduction is performed and features are extracted after performing discriminant evaluation in the entire within-subclass eigenspace.

When training is complete only the gallery features and transformation matrix are stored in the system. When more people have to be enrolled in the database, the incoming face images are transformed using the above generated training module (transformation matrix) and only the gallery features are stored. In the recognition stage, any incoming face image vector is converted into a feature vector using the transformation matrix learned by WSSDA method. The feature vector is used to perform recognition by matching it with the gallery features. Using cosine distance measures with 1-nearest neighbor (NN) as the classifier, WSSDA was shown to be the best performer in [26] among many methods for FR on the challenging unconstrained YouTube face image database [37].

2.2. Estimating the Final Associations: Network Consistent Re-identification

The problem of network inconsistency in classic person re-identification tasks was introduced in [12] and later expanded in [10]. A binary integer program for establishing

consistency in re-identification and thereby improving association accuracy was proposed in these works and termed as Network Consistent Re-identification (NCR) [12] or Network Consistent Data Association (NCDA) [10].

We denote an observation i in camera/device g as $\mathcal{P}_i^g$. In previous section, we estimate feature similarity/distance between pairs of observations across cameras and let $c_{i,j}^{p,q}$ denote the similarity score estimated between features from observations $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$, observed in camera p and q respectively. The expected output of the NCR framework is a set of association labels between each of these pairs of observations. Thus, if each of the observations is considered a node in a network, clusters of nodes observed in the same camera can be termed as 'groups' and edges can be constructed between pairs of nodes belonging to different groups. The goal is to estimate a label $x_{i,j}^{p,q}$ for each such edge that will denote whether the two nodes associated with this edge are from the same target, i.e., $x_{i,j}^{p,q} = 1$, if $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ are the same targets and 0, otherwise.

A 'path' between two nodes $(\mathcal{P}_i^p, \mathcal{P}_j^q)$ is a set of edges that connect the nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ without traveling through a node twice. Moreover, each node on a path belongs to a different group. A path between $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ can be represented as the set of edges $e(\mathcal{P}_i^p, \mathcal{P}_j^q) = \{(\mathcal{P}_i^p, \mathcal{P}_a^r), (\mathcal{P}_a^r, \mathcal{P}_b^s), \cdots (\mathcal{P}_c^t, \mathcal{P}_j^q)\}$, where $\{\mathcal{P}_a^r, \mathcal{P}_b^s, \cdots \mathcal{P}_c^t\}$ are the set of intermediate nodes on a path between $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$.

2.2.1 Constraints in Data Association

Now, as there can be only one observation from the same target (clustered a-priori if multiple observations) in one camera FoV, an observation $\mathcal{P}_i^p$ in camera p may have at most one matching observation in any other camera q. If the same set of targets appear in all the camera FoVs, there is an exact one-to-one match between observations across any two camera pairs. However, in a realistic scenario, a target may or may not appear in every camera FoV and hence,

$$\sum_{j=1}^{n_q} x_{i,j}^{p,q} \leq 1 \; \forall i = 1 \text{ to } n_p, \; \sum_{i=1}^{n_p} x_{i,j}^{p,q} \leq 1 \; \forall j = 1 \text{ to } n_q \quad (1)$$

where, $x_{i,j}^{p,q} \in \{0, 1\} \; \forall i, j, p, q$. This is referred to as the 'pairwise association constraint' in NCR. Now, pairwise associations must also be consistent over the network of cameras. This set of conditions is important when there are three or more cameras/wearable devices to capture FPV images. The consistency condition simply states that if two nodes (observations) are indirectly associated via nodes in other groups, then these two nodes must also be directly associated. Therefore, given two nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$, it can be noted that for consistency, a logical 'AND' relationship between the association value $x_{i,j}^{p,q}$ and the set of association values $\{x_{i,a}^{p,r}, x_{a,b}^{r,s}, \cdots x_{c,j}^{t,q}\}$ of any possible path between the nodes has to be maintained. The association value between the two nodes $\mathcal{P}_i^p$ and $\mathcal{P}_j^q$ has to be 1 if the association values corresponding to all the edges of any possible path between these two nodes are 1. Keeping the binary nature of the association variables and the pairwise association constraint in mind, the relationship can be compactly expressed as,

$$x_{i,j}^{p,q} \geq \left(\sum_{(\mathcal{P}_k^r, \mathcal{P}_l^s) \in e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)} x_{k,l}^{r,s} \right) - |e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)| + 1 \quad (2)$$

$\forall$ paths $e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q) \in \mathcal{E}(\mathcal{P}_i^p, \mathcal{P}_j^q)$, where $|e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)|$ denotes the cardinality of the path $e^{(z)}(\mathcal{P}_i^p, \mathcal{P}_j^q)$, *i.e.*, the number of edges in the z^{th} path. The relationship holds true for all i and all j. Now, any network containing even a large number of wearable cameras can be exhaustively expressed as a collection of non-overlapping triplets of cameras. For such a triplet, the constraint in Eqn. (2) simplifies to,

$$x_{i,j}^{p,q} \geq x_{i,k}^{p,r} + x_{k,j}^{r,q} - 2 + 1 = x_{i,k}^{p,r} + x_{k,j}^{r,q} - 1 \quad (3)$$

2.2.2 Re-identification as an Optimization Problem

Under the constraints expressed by Eqn. 1 and Eqn. 3, the objective is to maximize the utility $\mathbf{C} = \sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n} c_{i,j}^{p,q} x_{i,j}^{p,q}$. However, this utility function is only valid for one-to-one re-identification case, as this may reward both true positive and false positive associations (for example, when $c_{i,j}^{p,q} \in [0, 1]$), and hence the optimal solution will try to assign as many positive associations as possible across the network. This will yield many false positive associations. One way of avoiding such a situation in the current framework is to modify the utility function as $\sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n_p, n_q} (c_{i,j}^{p,q} - k) x_{i,j}^{p,q}$, where there are m cameras in the network and k is any value within the range of $c_{i,j}^{p,q} \; \forall i, j, p, q$. The value of k can be learned from the training data (see Sec. 3.3.1) so that the true-positives are rewarded and false-positives are penalized as much as possible. Therefore, by combining the utility function with the constraints in Eqn. 1 and Eqn. 3, the overall optimization problem for m wearable devices with variable number of observations is written as,

$$\underset{\substack{x_{i,j}^{p,q} \\ i=[1,\cdots,n_p] \\ j=[1,\cdots,n_q] \\ p,q=[1,\cdots,m]}}{\text{argmax}} \left(\sum_{\substack{p,q=1 \\ p<q}}^{m} \sum_{i,j=1}^{n_p, n_q} (c_{i,j}^{p,q} - k) x_{i,j}^{p,q} \right)$$

$$\text{subject to } \sum_{j=1}^{n_q} x_{i,j}^{p,q} \leq 1 \; \forall i = [1, \cdots, n_p] \; \forall p, q = [1, \cdots, m],$$

$$\sum_{i=1}^{n_p} x_{i,j}^{p,q} \leq 1 \; \forall j = [1, \cdots, n_q] \; \forall p, q = [1, \cdots, m], \; p < q$$

Figure 2. (Left) Original image captured by Google Glass. (Right) Three rows show three persons with four images each, captured using different Google Glasses under different scenarios.

$$x_{i,j}^{p,q} \geq x_{i,k}^{p,r} + x_{k,j}^{r,q} - 1$$
$$\forall\, i = [1, \cdots, n_p],\ j = [1, \cdots, n_q],\ k = [1, \cdots, n_r],$$
$$\forall\, p, q, r = [1, \cdots m],\ \text{and}\ p < r < q$$
$$x_{i,j}^{p,q} \in \{0, 1\}\ \forall i = [1, \cdots, n_p],\ j = [1, \cdots, n_q],$$
$$\forall p, q = [1, \cdots, m],\ p < q \tag{4}$$

This is a binary integer linear program (ILP) and exact solution can be efficiently computed using methods such as 'branch and bound'/'branch and cut' etc.

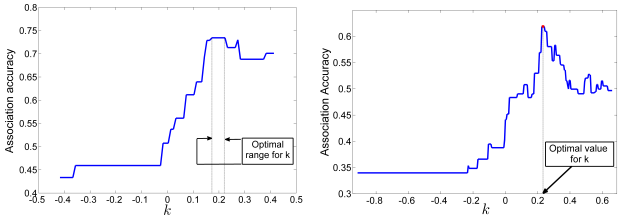
3. Experimental Results and Analysis

3.1. Database

We have collected a database consisting of FPV videos of 72 people comprising of 37 male and 35 females using 4 GGs (resulting in about 7077 images). The videos are captured using egocentric views at different levels in a large multi-storied office environment, in corridors, lifts, escalators, pantries, downstairs eateries, passage ways etc. Cameras 1, 2, 3 and 4 (corresponding to the 4 wearable devices) observe 52, 40, 43 and 50 persons in their respective FoVs. Unique target IDs in each camera FoV are given in the suppl. materials. The images collected using GG are often blurry in nature as the person wearing the GG moves his/her head quite frequently. Also, sometimes the images are out of camera focus. The face and eye detectors as described in section 2.1 serve as filters to remove images with large motion blur or poor image quality. Some selected images from the database are shown in Fig. 2. The database, protocol and experimental codes would be made available to public.

3.2. Pairwise Similarity Score Generation

Using the normalized images as described in section 2.1, we extract features applying various FR algorithms as described in section 2.1.1. We perform the training for FR algorithm using WSSDA on the FPV face image database provided in [21]. We obtain the transformation matrix using the same 42 people for training comprising of 305 images. We limit the dimensionality of the final transformation matrix to 80 features ($\times$ the dimensionality of face image vector [21]) for all the methods, so that the final features obtained are of 80 dimensions for each of the normalized face images. We use cosine distance measures with 1-nearest neighbor (NN) as the best match for each of the faces in a frame to generate pairwise scores between the persons observed in each of cameras FoVs.



(a) (b)

Figure 3. Estimation of optimal k (Eqn. 4) from an annotated training set. k is varied over the range of pairwise similarity scores in the training set and the overall association accuracy is computed for each value of k. (a) shows the plot of variation of accuracy with k for a training set of WSSDA similarity scores and (b) shows the same for PCA based pairwise measures.

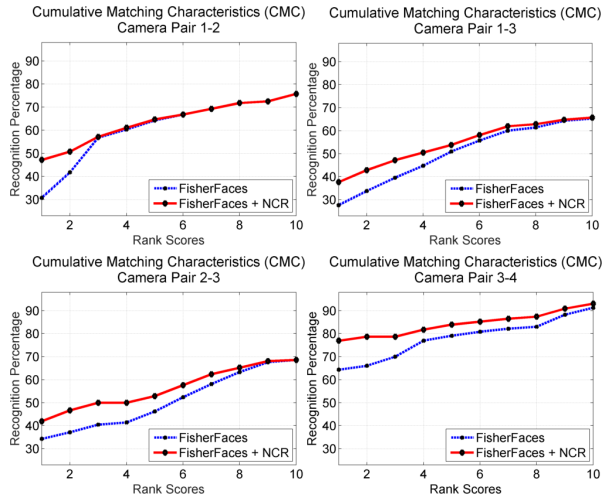
3.3. Network Consistent Re-identification

3.3.1 Test-Train Partitions: Learning k From Training Data

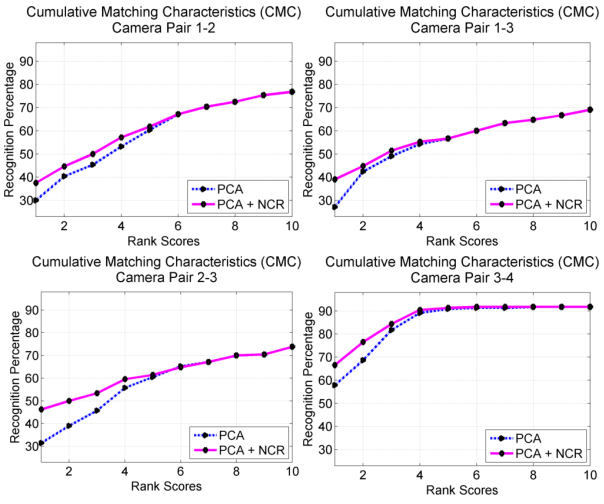
With the pairwise similarity scores generated (as explained in the previous section), the next step is to optimally combine them using the aforementioned Network Consistent Data Association (NCR) method, which yields the final association results. As shown in Eqn. 4, the value of k in the objective function of the NCR integer program is specific to the distribution of the pairwise similarity scores and hence k has to be learned from a training set before solving for the association labels.

Since we used three different methods, viz., PCA, FisherFaces and WSSDA (see Sec. 3.2 for details) for pairwise similarity score generation, we generate three separate sets of consistent association results - one for each of these baseline methods. We refer to them as (PCA + NCR), (FisherFaces + NCR) and (WSSDA + NCR) throughout the rest of the paper. For each of these three methods, we generate 10 sets of exhaustive training-testing partitions (non-overlapping) from the collected dataset. Each set contains 24 randomly selected targets (a third of the dataset) in the training set and the remaining 48 (two thirds of the dataset) are used for testing. The final test results including re-identification accuracy for each method are averaged over these 10 test sets.

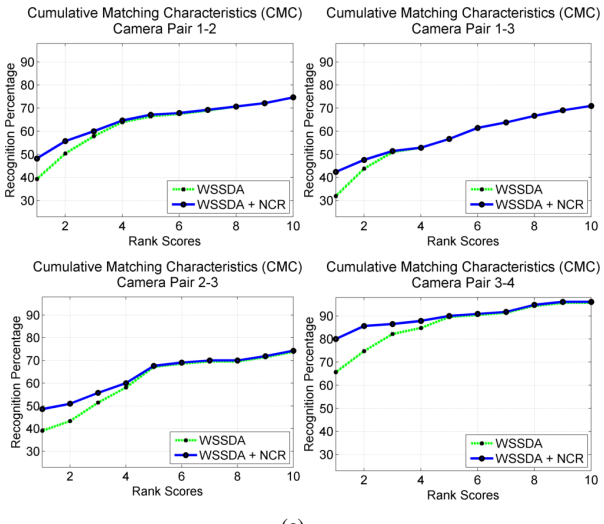
To learn k for each of the training sets, first the range of the pairwise similarity scores is identified. As the optimum value of k must lie within this interval, we vary k and compare the accuracy of data association against the



(a)



(b)



(c)

Figure 4. CMC curves comparing methods (a) FisherFaces, (b) PCA and (c) WSSDA respectively, both before and after NCR.

Table 1. Comparison of re-id performance of PCA, FisherFaces (FF) and WSSDA with their NCR counterparts based on nAUC values (computed using CMC values upto rank 10).

Cam pair	PCA	FF	WSSDA	PCA + NCR	FF + NCR	WSSDA + NCR
1-2	0.5978	0.6187	0.6387	0.6179	0.6393	**0.6544**
1-3	0.5614	0.5077	0.5741	0.5743	0.5484	**0.5847**
1-4	0.5349	0.5183	0.6508	0.5521	0.5410	**0.6717**
2-3	0.5849	0.5090	0.6172	0.6185	0.5646	**0.6407**
2-4	0.6455	0.5571	0.6717	0.6513	0.5817	**0.6950**
3-4	0.8570	0.7826	0.8708	0.8763	0.8423	**0.9017**

ground truth on the annotated training data. The accuracy is computed as $\frac{(\#\text{ true positive}+\#\text{ true negative})}{\#\text{ of unique people in the trainset}}$ and the value of k corresponding to the maximum association accuracy is estimated as the optimal of k and fixed during testing. We show examples of variation of training accuracy with k in Fig. 3. If the maximum accuracy is observed over a range of k (as seen in Fig. 3(a) for WSSDA + NCR case), the mean k over that range is taken as the optimum value. Fig. 3(b) shows another similar plot for learning optimum k for the PCA + NCR experiments.

3.3.2 Re-identification Performance Comparisons: Before and After NCR

The re-identification performances of the individual pairwise methods (PCA, Fisherfaces and WSSDA) are presented and compared - both before and after enforcing the network consistency. First, comparative evaluations are shown in terms of recognition rate as Cumulative Matching Characteristic (CMC) curves and normalized Area Under Curve (nAUC) values, which are the common practice in the literature. The CMC curve is a plot of the recognition percentage versus the ranking score and represents the expectation of finding the correct match inside top t matches. nAUC gives an overall score of how well a re-identification method performs irrespective of the dataset size. Please note that, we are presenting our results in the most generalized test setup where targets may not be visible in all the camera FoVs. Hence, while estimating the CMC and nAUC values between any pair of cameras i and j, only those targets in camera i are considered that are also observed in camera j's FoV.

Figs. 4(a), 4(b), 4(c) present the CMC curves for FisherFaces, PCA and WSSDA respectively and in each plot, comparisons in the recognition performances are shown before and after application of NCR (e.g., PCA and PCA + NCR in Fig. 4(b)). As we have 4 wearable devices in our dataset, there are 6 possible camera pairs and the plots are shown for camera pairs 1-2, 1-3, 2-3 and 2-4 for every feature computation method (See suppl. materials for all 6 camera pairs and 18 CMCs). Each CMC is plotted upto rank 10. As observed, amongst the three pairwise re-

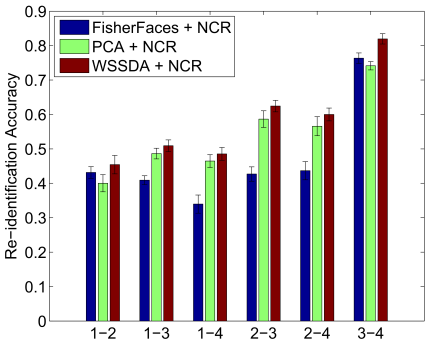


Figure 5. Comparison of overall re-identification accuracies (combining both true-positives and false-positives).

identification methods, WSSDA is superior to both PCA and FisherFaces. Moreover, for each of the features and every camera pair, individual pairwise methods are substantially outperformed by their respective NCR counterparts. In particular, WSSDA + NCR achieves the highest rank-1 performances across all the camera pairs, such as $\sim$49% in camera pairs 1-2 and 2-3 and $\sim$80% in camera pair 3-4.

These observations are further established by the nAUC values (computed from CMC until rank 10), as shown in Table 1. PCA+NCR, FisherFaces+NCR and WSSDA+NCR individually perform better than the pairwise methods PCA, FisherFaces and WSSDA respectively with WSSDA+NCR showing the best nAUC scores across all 6 camera pairs.

3.3.3 Overall Re-identification Accuracy by Combining Both True Positive and False Positive

A correct re-identification result in a realistic dataset such as ours not only contains correct matches (true positives) but also contains true negatives, when a target is only observed in a subset of cameras. Hence, the overall accuracy of person re-identification across any pair in the network of wearable devices should be estimated as $\frac{(\#\text{ true positive}+\#\text{ true negative})}{\#\text{ of unique targets in the testset}}$. We compare these accuracy values obtained by NCR when applied on each of PCA, FisherFaces and WSSDA similarity measures. From Fig. 5, it can be observed that NCR on WSSDA is more accurate than both of PCA+NCR and FisherFaces+NCR across all 6 camera pairs, with the best accuracy of more than 80% observed in camera pair 3-4.

4. Conclusions and Future Work

In this paper, we have introduced the problem of re-identification from first-person-view (FPV) videos collected using multiple wearable devices such as Google Glasses. We presented a pipeline for solving this re-identification problem by combining robust feature extraction methods for FPV face recognition with global data association techniques for network-consistent person re-identification (NCR). To test the proposed pipeline, we collected a large FPV video database using 4 Google Glasses and head mounted cameras that consists of 72 targets in a complex office environment. Our results indicate that the NCR based pipeline achieves high accuracy for re-identification across all camera pairs and show substantial improvement over the camera pairwise state-of-the-art methods. The future work would include development of an online network consistent face re-identification method and performing real-time on field testing.

Acknowledgment

The authors would like to thank Dr. Amit K. Roy-Chowdhury from University of California, Riverside for his helpful advice on the work and constructive feedback during preparation of the manuscript.

References

[1] A. Alavi, Y. Yang, M. Harandi, and C. Sanderson. Multi-shot person re-identification via relational stein divergence. In *IEEE International Conference on Image Processing*, 2013.

[2] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch. Learning implicit transfer for person re-identification. In *European Conference on Computer Vision, Workshops and Demonstrations*, pages 381–390, 2012.

[3] W. A. Bainbridge, P. Isola, and A. Oliva. The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, 4(142):1323–1334, 2013.

[4] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144, Nov. 2013.

[5] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[6] A. Bellet, A. Habrard, and M. Sebban. A survey on metric learning for feature vectors and structured data. *ArXiv e-prints*, 2013.

[7] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua. Tracking multiple people under global appearance constraints. In *IEEE International Conference on Computer Vision*, pages 137–144, 2011.

[8] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1806–1819, 2011.

[9] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(1):4–13, January 2005.

[10] A. Chakraborty, A. Das, and A. Roy-Chowdhury. Network consistent data association. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2015.

[11] S. Ching, B. Mandal, Q. Xu, L. Liyuan, and J.-H. Lim. Enhancing social interaction with seamless face recognition on google glass: Leveraging opportunistic multi-tasking on smart phones. In 17^{th} *International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileCHI)*, pages 750–757, Copenhagen, Denmark, Aug 2015.

[12] A. Das, A. Chakraborty, and A. Roy-Chowdhury. Consistent re-identification in a camera network. In *European Conference on Computer Vision*, pages 330–345, 2014.

[13] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. In *Asian conference on Computer vision*, pages 501–512, 2010.

[14] Google. Google glass. *http://www.google.com/glass/start/*, 2015.

[15] GoPro. *http://gopro.com/*, 2015.

[16] O. Javed, K. Shafique, Z. Rasheed, and M. Shah. Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views. *Computer Vision and Image Understanding*, 109(2):146–162, Feb. 2008.

[17] X. D. Jiang, B. Mandal, and A. Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 30(3):383–394, Mar 2008.

[18] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1622–1634, 2013.

[19] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification : What features are important ? In *European Conference on Computer Vision, Workshops and Demonstrations*, pages 391–401. Springer Berlin Heidelberg, 2012.

[20] W. Liu, Y. Wang, S. Z. Li, and T. N. Tan. Null space approach of fisher discriminant analysis for face recognition. In *European Conference on Computer Vision*, pages 32–44, 2004.

[21] B. Mandal, S. Ching, L. Li, V. Chandrasekha, C. Tan, and J.-H. Lim. A wearable face recognition system on google glass for assisting social interactions. In 3^{rd} *International Workshop on Intelligent Mobile and Egocentric Vision, Asian Conference on Computer Vision*, pages 419–433, 2014.

[22] B. Mandal and H.-L. Eng. Regularized discriminant analysis for holistic human activity recognition. *IEEE Intelligent Systems*, 27(1):21–31, 2012.

[23] B. Mandal, X. D. Jiang, and A. Kot. Multi-scale feature extraction for face recognition. In *IEEE International Conference on Industrial Electronics and Applications (ICIEA)*, pages 1–6, Singapore, May 2006.

[24] B. Mandal, X. D. Jiang, and A. Kot. Dimensionality reduction in subspace face recognition. In *IEEE* 6^{th} *International Conference on Information, Communications and Signal Processing (ICICS)*, pages 1–5, Singapore, Dec 2007.

[25] B. Mandal, W. Zhikai, L. Li, and A. Kassim. Evaluation of descriptors and distance measures on benchmarks and first-person-view videos for face identification. In *International Workshop on Robust Local Descriptors for Computer Vision, Asian Conference on Computer Vision*, pages 585–599, 2014.

[26] B. Mandal, W. Zhikai, L. Li, and A. Kassim. Whole space subclass discriminant analysis for face recognition. In *IEEE International Conference on Image Processing (ICIP)*, pages 329–333, Quebec city, Canada, Sep 2015.

[27] N. Martinel and C. Micheloni. Re-identify people in wide area camera network. In *International Conference on Computer Vision and Pattern Recognition Workshops*, pages 31–36, Providence, RI, June 2012. IEEE.

[28] Open source computer vision,(http://opencv.org/), 2015.

[29] F. Porikli and M. Hill. Inter-camera color calibration using cross-correlation model function. In *IEEE International Conference on Image Processing (ICIP)*, pages 133–136, 2003.

[30] K. Shafique and M. Shah. A noniterative greedy algorithm for multiframe point correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):51–65, 2005.

[31] D. L. Swets and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(8):831–836, August 1996.

[32] TedBlog. The future of facial recognition: 7 fascinating facts. *http://blog.ted.com/2013/10/17/the-future-of-facial-recognition-7-fascinating-facts/*, 2014.

[33] G. Tian, Y. Wong, B. Mandal, V. Chandrasekha, and M. Kankanhalli. Multi-sensor self-quantification of presentations. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 601–610, Brisbane, Australia, Oct 2015.

[34] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, January 1991.

[35] Y. Utsumi, Y. Kato, K. Kunze, M. Iwamura, and K. Kise. Who are you?: A wearable face recognition system to support human memory. In *ACM Proceedings of the 4th Augmented Human International Conference*, pages 150–153, 2013.

[36] X. Wang, X. Zhao, V. Prakash, W. Shi, and O. Gnawali. Computerized-eyewear based face recognition system for improving social lives of prosopagnosics. *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pages 77–80, 2013.

[37] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained video with matched background similarity. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 529–534, Jun 2011.

[38] L. Yang and R. Jin. Distance metric learning : A comprehensive survey. Technical report, Michigan State University, 2006.

[39] X. Yu, W. Han, L. Li, J. Shi, and G. Wang. An eye detection and localization system for natural human and robot interaction without face detection. *TAROS*, pages 54–65, 2011.

[40] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *International Conference on Computer Vision and Pattern Recognition*, 2013.

[41] M. Zhu and A. Martinez. Subclass discriminant analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, August 2006.