Local Fisher Discriminant Analysis for Pedestrian Re-identification

Sateesh Pedagadi, James Orwell
Kingston University London

Sergio Velastin
Universidad de Santiago de Chile

Boghos Boghossian
Ipsotek Ltd, UK

Abstract

*Metric learning methods, for person re-identification, estimate a scaling for distances in a vector space that is optimized for picking out observations of the same individual. This paper presents a novel approach to the pedestrian re-identification problem that uses metric learning to improve the state-of-the-art performance on standard public datasets. Very high dimensional features are extracted from the source color image. A first processing stage performs unsupervised PCA dimensionality reduction, constrained to maintain the redundancy in color-space representation. A second stage further reduces the dimensionality, using a Local Fisher Discriminant Analysis defined by a training set. A regularization step is introduced to avoid singular matrices during this stage. The experiments conducted on three publicly available datasets confirm that the proposed method outperforms the state-of-the-art performance, including all other known metric learning methods. Furthermore, the method is an effective way to process observations comprising multiple shots, and is non-iterative: the computation times are relatively modest. Finally, a novel statistic is derived to characterize the Match Characteristic: the normalized entropy reduction can be used to define the 'Proportion of Uncertainty Removed' (*PUR*). This measure is invariant to test set size and provides an intuitive indication of performance.*

1. Introduction

In recent years, the re-identification of people from medium-range CCTV observations has attracted considerable attention, particular with multiple non-overlapping source cameras. The main challenges in person re-identification can be attributed to the variations in pose, illumination and viewpoint. We briefly review the existing methods in literature below.

Many researchers have approached the problem by proposing *feature-based* methods which estimate a reliable and distinctive signature per person regardless of the scene by combining one or more extracted feature types. Swain and Ballard [30] first proposed the use of histograms as stable representations of objects, for identification and recognition. A recent study [4] evaluated the performance of various local features for person re-identification: it has been proposed [10] to augment maximally stable color regions (MSCR) with histograms and recurrent local color patches. A similar proposal [5] is to use histograms along with 'epitome' - a collection of recurrent stable color patches extracted over a series of frames. Another approach [34] is the computation of region covariance descriptors combining Gabor and local binary patterns from several non-overlapping regions. More complex models have been proposed, *e.g.* a decomposable triangulated graph to be fitted on a over segmented image [12]. Feature-based methods can suffer from illumination variations and human shape deformations. To address the illumination variation across cameras, a learned inter-camera Brightness Transfer function (BTF) has been proposed [19].

Some methods focus on feature extraction for *fast matching*, *e.g.* per-person signatures using Camillia key-points [15], stored in a KD-tree; a code-book using a Global color context [15]; or Random forests to weight the most informative features [25]. Code-book representations and the addition of spatial information in a layered framework were proposed in [22].

Several researchers [1, 21] investigate the possibility of acquiring more distinctiveness by determining the pose prior to feature extraction. Also, the use of '*multi-shot*' data (from multiple observations) to estimate appearance models may be advantageous, *e.g.* using a histogram of gradients (HoG) detector to extract different body parts over a series of frames [7]. Semantic segmentation is the basis of works [10, 6] in which multiple observations are analysed.

Researchers have also learning framework to iteratively select the most reliable subset of features, using Adaboost to select the most discriminative *e.g.* Haar-like features [2]. Adaboost has also been used to learn discriminative features and an ensemble of weak classifiers for a given dataset [14]; and a set of weak RankSVMs have been boosted [28] by making various overlapping partitions of the training data. Model estimation using Adaboost to improve ranking results has also been proposed [18].

Person re-identification can be formulated as a data asso-

ciation problem, to match observations obtained from pairs of cameras. In this case, it is assumed that a mapping space common to both sets of observations can be found, in which a nearest neighbor classification solves the data association problem. Various methods have been proposed to learning this mapping space: the 'Large Margin' Nearest Neighbor (with Rejection) – LMNN(-R) algorithm has been used [8] to learn the most effective metric to match data from two arbitrary cameras. To account for the non-linear nature of the feature space, it has been proposed to use a kernel [32] as part of the PCA to find an appropriate low dimensional manifold. Other learned metrics include the PRDC [35], estimated by measuring the probabilistic distance between features belonging to same and different persons; or one learnt from a number of SVMs trained on random subsets of data [24], similar to the work in [28].

'Rank-loss optimization' has been proposed, to improve the re-identification accuracy [31], whereas RLPP, a variant of Locality Preserving Projections (LPP) formulated over Riemannian manifolds, has also been introduced [16]. Recently, metric learning has been proposed from equivalence constraints in pairwise difference space. PCCA was proposed in [27] the focus of which was to address the situations when only a small set of limited examples [23], whereas Zheng em et al [36] reformulate the person re-identification problem as a set-based verification task and show that discriminant information can be learnt from unlabelled data. Out of the various methods listed above, metric learning methods can be identified as providing a significant improvement in performance.

Dimensionality reduction and distance metric learning are known to have a significant contextual dependency. In the case of unsupervised distance metric learning, a low-dimensional manifold space is estimated for which the objective is to maximize the preservation of geometric relationships (*e.g.* distances) between the data. As an example, the principle Eigenvectors φ_i of the covariance matrix can be used to define a distance metric

$$D_m = \sum_i \varphi_i \varphi_i' \tag{1}$$

The distance between two samples, $\mathbf{y}$ and $\mathbf{x}$ is then

$$d_{\mathbf{yx}} = (\mathbf{y} - \mathbf{x})' D_m (\mathbf{y} - \mathbf{x}) \tag{2}$$

Typically, a single embedding space based on the above approximation may not be adequate for this class of nearest neighbor classification problems. Hence, the idea of projecting simple features into different, local, embedding spaces is explored in this work. This is refered to as LF: **Local Fisher** Discriminant Analysis.

The main contributions of this work are i) to apply the Local Fisher Discriminant analysis to the pedestrian re-identification problem; ii) to propose that the PCA-reduced representation of the features should maintain the dual color-space representation; iii) to apply a regularization process to the Discriminant Analysis to achieve a working solution and iv) propose a characterization of the Match Characteristic curve that in essence reports the percentage of the re-identification problem solved, from 0% to 100%.

The rest of the paper is organised as follows. Section 2.1 proposes the features to extract. The first and second stages of dimensionality reduction are described in sections 2.2 and 2.3 respectively. The regularization required for the second stage is explained in 2.5 and the information theoretic performance measure in 3.1. Experiments on three public datasets are described in section 4, demonstrating the efficacy of the proposed approach.

2. Proposed Method

The proposed method **LF** uses a feature extraction technique followed by supervised and then unsupervised dimensionality reduction stages. In common with most other investigations, the starting point is a fixed-size sub-image containing an observation of a pedestrian.

2.1. Feature Extraction

The color descriptor $\mathbf{u}_i$ is extracted for the observation indexed with i, working in the HSV color space. It is the concatenation of a parametric $\mathbf{c}$ and non-parametric $\mathbf{h}$ representations of colors in each of the m densely sampled (overlapping) tiles defined within the observation (see Fig. 1). For each tile j , $\mathbf{c}_j$ is the first three moments calculated separately on each of the three components in the color space. Similarly, $\mathbf{h}_j$ is the normalized 8-bin histogram calculated separately for each component of the color space. In addition, the color descriptor $\mathbf{v}_i$ is defined identically except that it uses the YUV color space as the input. Both of these vectors have $33m$ dimensions: in the experiments, $m = 341$, implying an initial dimensionality of 11,253. Below, the dimensionality reduction framework is described.

2.2. Unsupervised Dimensionality Reduction

The first stage employs an unsupervised dimensionality reduction technique to estimate a low dimensional embedding space from the high dimensional feature space. Let $x_i \in \{i = 1, 2, ...N\}$ be the index of each individual available in the dataset and $\mathbf{u}$, the descriptor matrix of all individuals in HSV color space. Similarly, $\mathbf{v}$ is the color descriptor matrix in YUV color space. The use of color descriptors defined in more than one color space is useful for estimating a reliable embedding space in the sibsequent stages. It is common for the descriptor matrix $\mathbf{u}$ to be high dimensional and also the accumulation of descriptors from a dense grid is likely to introduce noise. For a given color space, the matrix $\mathbf{u}$ is de-noised, and its dimensions reduced, by using Principal Component Analysis (PCA) [20].For simplicity,

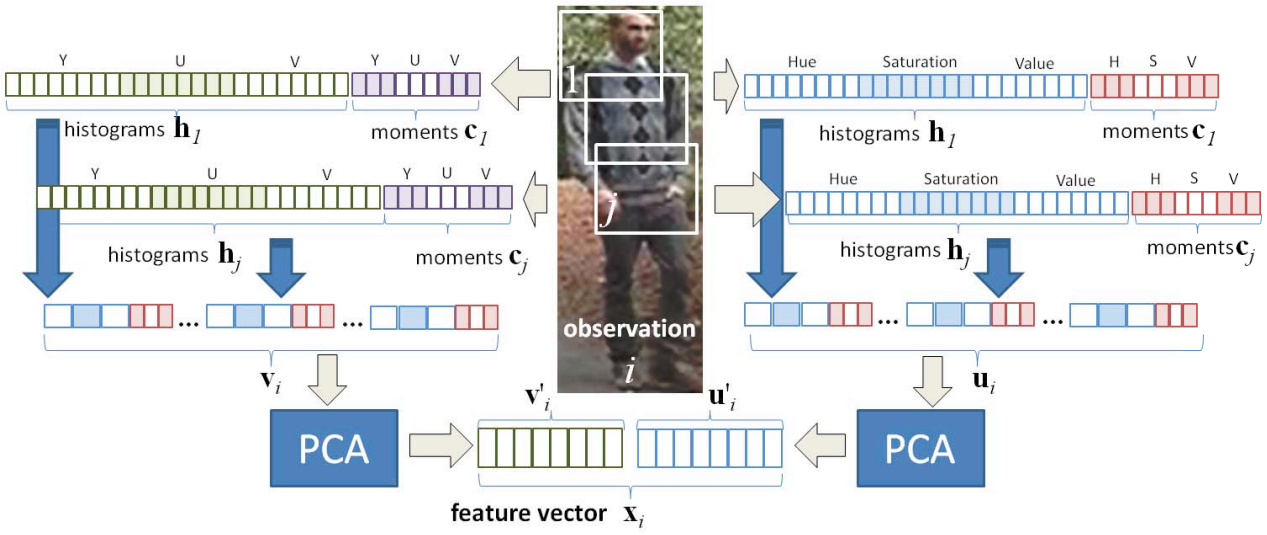

Figure 1. **Proposed feature extraction and first processing stage**

the data projected into the low dimensional manifold estimated by PCA is written $\mathbf{u}'_i = D_u \mathbf{u_i}$, where D_u is the embedding transformation matrix corresponding to the Eigenvectors derived from PCA. Similarly, for the YUV channel, $\mathbf{v}'_i = D_v \mathbf{v_i}$. This differs from the conventional approach of using PCA globally, to the whole feature space [8]. It is hypothesised that separate estimation and use of D_v and D_u retains information more effectively. The overall output $\mathbf{x}_i$ from the first stage is the concatenation of the two sets of principal components: $\mathbf{x}_i = \{\mathbf{u}'_\mathbf{i}|\mathbf{v}'_\mathbf{i}\}$.

2.3. Supervised Dimensionality Reduction

The second stage uses a supervised dimensionality reduction method to estimate a lower dimensional embedding space into which $\mathbf{x}_i$ may be transformed. Supervised techniques which preserve the locality relations of the data perform relatively well compared to unsupervised techniques. In this work, Local Fisher Discriminant Analysis (LFDA) [29] is employed to learn a distance metric between a set of descriptor pairs corresponding to pairs of observations of a set of individuals. The LFDA transformation can be estimated using generalized Eigenvalues and the objective that between-class separability is maximized and multi-class modality is preserved. It combines the supervised aspect of Fisher Discriminant Analysis (FDA) [11] while preserving the local neighborhood structure preserving nature (LPP) [17]. We briefly review LFDA below.

2.4. Local Fisher Discriminant Analysis

A descriptor matrix is defined, $X \equiv (\mathbf{x}_1|\mathbf{x}_2|...|\mathbf{x}_n)$, consisting of n samples of C different classes with labels $\mathbf{x}_i \in \{1, \ldots, C\}$ and feature vectors $\mathbf{x}_i \in \mathbb{R}^d$. We can then define a transformation matrix T such that

$$\mathbf{z}_i = T' \mathbf{x}_i \tag{3}$$

where $\mathbf{z}_i \in \mathbb{R}^m$ $(1 \leq m \leq d)$ is a lower dimensional space of dimensionality m. In Fisher Discriminant Analysis (FDA), the scatter matrices between classes and within a class are defined as

$$S^{(b)} = \sum_{i=1}^{N} n_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)' \tag{4}$$

$$S^{(w)} = \sum_{i=1}^{N} \sum_{j:x_j=i} \left(\mathbf{x}_j - \mu_i\right)\left(\mathbf{x}_j - \mu_i\right)' \tag{5}$$

respectively. Here, μ_i is the mean of data belonging to class i and μ is the mean of all the data. Then, an Eigenvalue problem can be formulated using $S^{(w)}$ and $S^{(b)}$ as $S^{(b)}\varphi = \lambda S^{(w)}\varphi$, where $\{\varphi_i\}_{i=1}^d$ are the eigenvectors associated to the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$. Even though FDA is a supervised process, for multi-modal distributions the the performance tends to be poor. One possible reason for this is that it processes the data as a whole, and so local variations in the transformation cannot be accommodated.

The Locality-Preserving Projection (LPP) uses an $n \times n$ affinity matrix A, describing the affinity between various samples within the data. Typically, the affinity value between two samples separated by a small Euclidean distance will be higher than two samples separated by a larger distance. Let D be a diagonal matrix: each element sums all affinity values over the columns of A, such that $D_{ii} =$

$\sum_{j=1}^{n} A_{ij}$. Subject to the constraint $T'XDX'T = I$, the transformation matrix for LPP is given by

$$T_{LPP} = (\varphi_{d-m+1}|\varphi_{d-m+2}|...|\varphi_d) \tag{6}$$

where $\{\varphi_i\}$ are the d eigenvectors associated with the d eigenvalues $\{\gamma_i\}$ of the following eigenvalue problem:

$$XLX'\psi = \gamma XDX'\psi \tag{7}$$

where $L = D - A$. However, this (LPP) preservation of local affinities is indiscriminate.

LFDA combines the above with a Fischer Discriminat Analysis: both the within-class scatter matrix S^W and between class scatter matrix S^B are weighted by the affinity matrix A of data. The Affinity matrix is calculated by using a local scaling method [33], i.e. choosing the n-th nearest neighbor and assigning individual scaling factors for samples of the same class; for all LF experiments reported in this paper, n is set to 6. With the use of affinity matrix, the contribution made by far apart in-class pairs of samples is almost negligible in the calculation of S^W and S^B:

$$S^W = \frac{1}{2}\sum_{i,j=1}^{n} A_{i,j}^{w}\left(\mathbf{x}_i - \mathbf{x}_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)' \tag{8}$$

$$S^B = \frac{1}{2}\sum_{i,j=1}^{n} A_{i,j}^{b}\left(\mathbf{x}_i - x_j\right)\left(\mathbf{x}_i - \mathbf{x}_j\right)' \tag{9}$$

where

$$A_{i,j}^{w} = \begin{cases} A_{i,j}/n_c & \text{if } y_i = y_j = c \\ 0 & \text{if } y_i \neq y_j \end{cases} \tag{10}$$

$$A_{i,j}^{b} = \begin{cases} A_{i,j}\left(\frac{1}{n} - \frac{1}{n_c}\right) & \text{if } y_i = y_j = c \\ \frac{1}{n} & \text{if } y_i \neq y_j \end{cases} \tag{11}$$

Here, n_c is the number of samples in class c and n is the total. The transformation T_{lfda} can then be defined as

$$T_{lfda} = \arg\max_{T} r\left(\left(T'S^WT\right)^{-1}T'S^bT\right) \tag{12}$$

where $T \in \mathbb{R}^d \times \mathbb{R}^m$. Similar to FDA, the estimation of T_{lfda} is achieved by representing the above as a generalized Eigenvalue problem, $S^B\varphi = \lambda S^W\varphi$, where $\{\varphi_i\}$ and $\{\lambda_i\}$ are the eigenvectors and eigenvalues of this system. Using equation 3, the final projection into the embedding space characterized by LFDA is written as

$$\mathbf{z}_i = T'_{lfda}\mathbf{x}_i \tag{13}$$

In this space, the Euclidean distance is used to measure similarity between two samples i and j:

$$D(i,j) = |\mathbf{z}_i - \mathbf{z}_j| \tag{14}$$

2.5. Regularization in LFDA

The size of the transformation matrix in LFDA is $d \times m$, where d is the dimensionality after stage 1 and m is the dimensionality of embedding space. If the number of classes in the data is C, then the value of m will be $C-1$ or less, a constraint caused by S^B. When d is large compared to the number of samples n, then S^W can be singular, making the Eigenvalue system impossible to solve. This is a drawback of LFDA and to overcome this problem, a small multiple of the identity matrix can be added as follows:

$$\hat{S}^W = (1-\beta)S^W + \beta.\frac{Tr\left(S^W\right)}{n}.I \tag{15}$$

where β is the regularization parameter, $0 \leq \beta \leq 1$ and by default it is set to 0.5. The within-class scatter matrix S^W defined in Eqn.9 is replaced with the above, and the corresponding Eigensystem is solved. The resulting eigenvectors constitute the final transformation matrix, T_{lfda}, for projecting X into the space in which similarities can be measured.

3. Performance Evaluation

The standard methodology to evaluate person re-identification algorithms is adopted here, with one additional contribution. This methodology assumes separate training and test sets, both containing exactly two observations of a number of different subjects. Every subject is a distinct individual and no individual is contained in both training and test sets. The test procedure is to select a 'probe' observation and compare against a 'gallery' of observations including the single remaining 'correct' observation of this probe subject; the remainder of the gallary are distractors. Each re-identification method is required to rank the gallery observations, in order of their likelihood of representing the same individual as the probe observation. The rank of the correct observation is recorded and aggregated over the entire test set to generate a Match Characteristic $M(r)$, the propbability that the rank r choice will be correct. This is accumulated into the familiar CMC curve. These are calculated by repeatedly resampling the partition between training and test sets, to improve confidence in any observed differences between re-identification algorithms.

3.1. The Proportion of Uncertainty Removed

A useful performance indicator can be derived as follows. In these particular experimental conditions, in which a probe observation is compared against a gallery set of S observations, let us assume that each member of the set has an equal prior probability of being correct. Therefore, the initial uncertainty of the state is simply $\log(S)$. After a measurement (by any chosen method), the members of the gallery set are ranked by order of their similarity to the probe measurement. Assuming a stationary and ergodic

data source, the posterior probability that the member at rank r is the correct match, is equal to the match characteristic $M(r)$. The expected uncertainty of this state is therefore $-\sum_{r=1}^{S} M(r)\log(M(r))$. The difference between this and $\log(S)$ is the information added by the identity estimate.

Furthermore, it is useful to normlise this entropy reduction by the original entropy: this 'Proportion of Uncertainty Removed' (PUR) is invariant to the test data set size S, and the base of the logarithm used for the entropy:

$$\text{PUR} \equiv \frac{\log(S) - \sum_{r=1}^{S} M(r)\log(M(r)}{\log(S))} \qquad (16)$$

Thus, the PUR is the normalised entropy reduction between a randomized rank and any given method's output rank. Also it accommodates information from across the entire CMC, rather than at arbitrary values of r.

4. Experimental Results and Discussion

The evaluation of the proposed method is carried out on **VIPER** [13], **3DPES** [3] and **CAVIAR** [9] datasets. Each has specific characteristics, as explained below.

4.1. VIPER

VIPER is a hand-generated dataset containing 632 pedestrian image pairs taken from arbitrary viewpoints under varying illumination conditions. The data was collected in an academic setting over the course of several months. Each image is scaled to 128x48 pixels. The dataset is halved

Figure 2. **VIPER dataset** - Pairs of consecutive images belong to a person in 2 cameras

into testing and training in a random fashion over multiple runs in the experimental set up. In each run, the training images are divided into 8x8 blocks with 50% overlap from which 8 bin histograms and 3 moments are estimated for each of the color channels in HSV and YUV color spaces. For a given color space and a block, the resulting histograms and moments within that block are concatenated as a single vector of 33 dimensions. In the same color space, for each image, the concatenation of all color descriptors will result in a 11,253-dimensional vector.

Such vectors are generated for each of the training images after which PCA is performed on the mean subtracted

RANK	1	10	25	50	PUR
LF	24.18	67.12	85.10	94.12	42.35
eLDFV	22.34	64.04	81.97	88.92	40.56
KISSME	19.81	62.56	80.99	91.93	38.58
LMNN-R	18.28	55.49	74.43	87.65	33.76

Table 1. **Performance comparison on VIPER**. PUR represents the proportion of uncertainty removed (see Sec 3.1) and columns [1,10,25,50] represent the recognition percentage scored by each of the methods at given rank index.

data after which first 20 and 80 principal vectors are chosen for HSV and YUV color spaces respectively. The training data is then projected into each of the respective spaces and the resulting transformed data is concatenated in the two color spaces to form a 100 dimensional subspace for Stage 2. Regularized LFDA is applied on this PCA-projected data to further estimate a 50-dimensional manifold. Color descriptors for test data are extracted in a similar manner in training and they are then projected into PCA and LFDA manifolds (in the same order) using the transformation matrices estimated in training phase. The number of principal components in Stage 1 for each color space and LFDA sub space dimensionality in **LF** are fixed to the above mentioned values for all the reported experimental results.

The method is evaluated and compared with three state-of-the-art methods KISSME [23], LMNN-R [8] and eLDFV [26]. For all four methods, the mean CMC and PUR are computed from100 CMCs generated random re-partition of the test and training set. The performance is also reported with in the range of first 50 ranks in table 1 and the PUR value introduced in 3.1. The VIPER experiment shows that at rank 1, LF was able to achieve 1.68% improvement over eLDFV, 4.37% improvement over KISSME and 5.9% over LMNN-R. The PUR values also indicate that LF has performed significantly better than the other two methods. However, the authors of eLDFV [26] also report the performance of sLDFV, a variant of eLDFV when combined with a metric learning method, which outperforms LF by 2.38% at rank 1.

4.2. 3DPES

3DPES dataset consists images of 191 individuals captured on multiple occasions along their trajectory through an academic campus, from 8 different surveillance cameras.

Data was collected during various times of the day, resulting in strong variations of lighting conditions. Significantly, of the dataset is that it contains a high viewing angle for the cameras which is typical of surveillance cameras in outdoor scenes, *e.g.* town centers. Once again, LF is evaluated and compared with KISSME [23], and LMNN-R [8].

The dataset is partitioned into training and test sets in a random fashion with each containing 95 persons. The re-

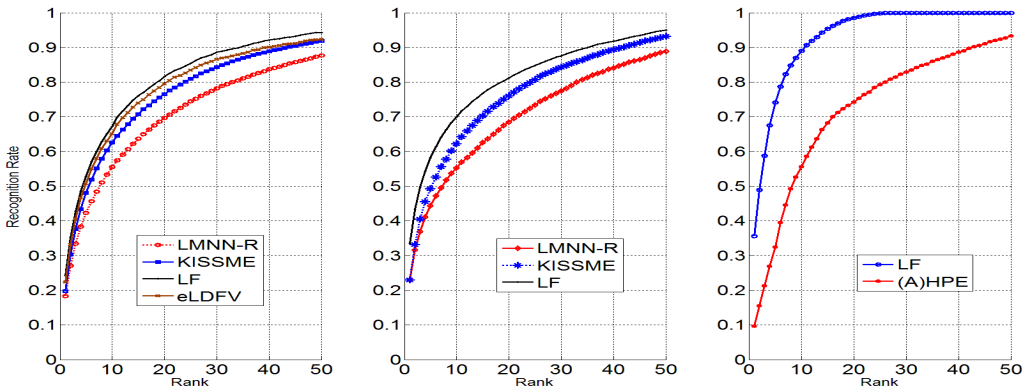

Figure 4. **CMC performance for VIPER, DPES and CAVAIAR datasets**

Figure 3. **3DPES dataset**: Pairs of consecutive images above belong to a person in 2 cameras

RANK	1	10	25	50	PUR
LF	33.43	69.98	84.80	95.07	34.85
KISSME	22.94	62.21	80.74	93.21	25.49
LMNN-R	23.03	55.23	73.44	88.92	21.11

Table 2. **Performance comparison on 3DPES**. PUR represents the proportion of uncertainty removed (see Sec 3.1) and columns [1,10,25,50] represent the recognition percentage scored by each of the methods at given rank index.

sults for this experiment are reported as CMC in 4, recognition percentage at a given rank and PUR values in table 2. LF outperforms the other two methods with significant margins, especially in the lower rank indices $\{1, 10, 25\}$.

4.3. CAVIAR

CAVIAR is another person re-identification dataset containing images of 72 individuals captured from 2 cameras in a shopping center scenario. The main complexity of this dataset is the presence of intra- and inter-camera lighting variation; also images of individuals were extracted from a significantly smaller source image resolution. Moreover, images from the second camera are dominated by background lighting making the dataset very challenging when compared to the existing datasets in person re-identification literature. The complete set of images were resized to to a resolution of 128x48 in experiments.

Figure 5. **CAVIAR Re-identification dataset**: Pairs of consecutive images above belong to a person in 2 cameras. These images are typical of indoor legacy surveillance systems, *e.g.* airports

RANK	1	10	25	50	PUR
LF	36.19	88.56	99.89	100	50.02
(A)HPE	9.70	55.60	79.30	93.30	20.52

Table 3. **Performance comparison on CAVIAR**

x (A)HPE [6], the only method which has been tested on this dataset. It was unclear as to how the authors of (A)HPE have carried out their experiment for this dataset but given that their method does not involve training, it is assumed that all 50 persons in one camera were matched against their observations in the other camera. The experiment conducted for LF makes use of the *multi-shot* nature of the dataset for which 5 images per person were used during training phase. The dataset is split into training and test sets with each set containing 36 persons and 5 images per person. In the test set, the features extracted from all observations belonging to a person are averaged to give one feature vector. The reason for doing is two fold. One is to provide many training examples so that multi-modality is introduced within the feature set of the training data. Two, the average of all features is likely to be an estimate of the centroid for all samples and hence is a good representation for making on-on-one criterion for each person during matching stage.

The performance of LF is evaluated as CMC for 36 individuals and to make a fair comparison between the two methods, CMC reported in [6] and CMC generated for LF

are both normalized on a scale of 1 to 100 ranks. Similar to the results reported for other datasets in this work, recognition percentage for a given rank index and PUR values are reported in 4.3. From the results it is clear that LF was able to achieve good recognition performance compared to (A)HPE and PUR value is as high as 50% for LF. This can be attributed to the multi modal learning ability of LF and also the relatively smaller test set size.

4.4. PCA Encoding

Experiments have been conducted on VIPEr dataset to support the claim in 2.2 by performing PCA on the combined features generated in HSV and YUV colour spaces.

The recognition performance was better for the case of separately encoding each of the colour spaces when compared to the joint encoding, for example, rank 1 result is 24.02% (separate encoding) vs 21.07% (joint encoding). This can be attributed to the fact the colour space conversion between HSV and YUV is non-linear in nature and separately encoding features collected in each colour space with PCA retains the most informative principal components.

4.5. Feature Selection

The use of the parametric and non-parametric features in the method proposed was analysed by the comparison of recognition performance for each of the feature set.

Histogram features are robust compared to colour moments, and they achieve about 7% improvement in recognition performance. The non-parametric nature of histograms does not make any assumptions of the underlying data and hence are very robust. This robustness is simply not enough when compared 6 to the joint feature constructed by adding the colour moments, the parametric representation of data. The joint features outperform histogram features with an increased recognition performance of 6% at rank 1 and 13% for the case of colour moments alone.

4.6. Colour Space Analysis

The features generated from YUV colour space provide better recognition performance compared to the HSV colour space. This is reflected by the number of principal components chosen during PCA projection step for each colour space: 80 of the principal components contribute to the YUV colour space where only 20 components from HSV colour space are used. In this case, the features extracted from both the colour spaces 6 improve the recognition performance by 2% at rank 1 for YUV and 13% for HSV colour space respectively. It is also interesting to note the recognition performance without the use of the stage 1 process of feature data projection into PCA subspace. The application of LFDA alone on accumulated parametric and non-parametric features from both the colour spaces provides a recognition percentage of 17% at rank 1. This observation suggests that PCA encoding is crucial to remove the noise that may reduce the richness of the LFDA learnt subspace.

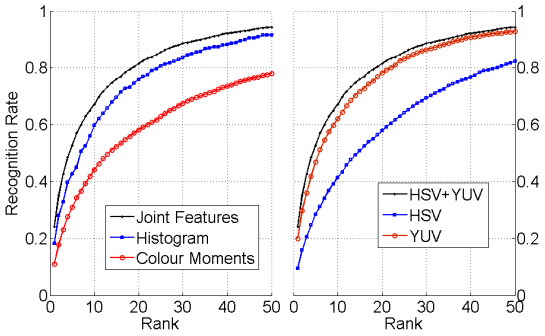

Figure 6. **Inclusion of feature types and colour spaces.**

5. Conclusion

This work presented a novel approach for person re-identification by combining parametric (color moments) and non-parametric (histograms) representation of colors as feature. The proposed method is a two stage training based low manifold learning framework using unsupervised (PCA) and supervised (LFDA) dimensionality reduction methods. A regularized form of the supervised dimensionality reduction was also proposed for LFDA technique used in second stage of the framework. Additionally, for performance evaluation purposes; the percentage of uncertainty removed (PUR) was presented as a dataset size invariant measure for recognition problems. Experiments were conducted to evaluate and compare the proposed method on three different person re-identification datasets with better performance compared to state of the art methods.

6. Acknowledgement

This work was partly funded by EU project ADDPRIV, grant FP7/2007-2013 No. 261653.

References

[1] K.-E. Aziz, D. Merad, and B. Fertil. Person re-identification using appearance classification. In *ICIAR (2)*, pages 170–179, 2011.

[2] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using haar-based and dcd-based signature. In *Proc. Seventh IEEE Int Advanced Video and Signal Based Surveillance (AVSS) Conf*, pages 1–8, 2010.

[3] D. Baltieri, R. Vezzani, and R. Cucchiara. 3dpes: 3d people dataset for surveillance and forensics. In *Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects*, pages 59–64, Scottsdale, Arizona, USA, Nov. 2011.

[4] M. Bauml and R. Stiefelhagen. Evaluation of local features for person re-identification in image sequences. In *Proc. 8th IEEE Int Advanced Video and Signal-Based Surveillance (AVSS) Conf*, pages 291–296, 2011.

[5] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino. Multiple-shot person re-identification by hpe signature. In *ICPR*, pages 1413–1416, 2010.

[6] L. Bazzani, M. Cristani, A. Perina, and V. Murino. Multiple-shot person re-identification by chromatic and epitomic analyses. *Pattern Recognition Letters*, 33(7):898–903, 2012.

[7] A. Bedagkar-Gala and S. K. Shah. Multiple person re-identification using part based spatio-temporal color appearance model. In *Proc. IEEE Int Computer Vision Workshops (ICCV Workshops) Conf*, pages 1721–1728, 2011.

[8] M. Dikmen, E. Akbas, T. Huang, and N. Ahuja. Pedestrian recognition with a learned metric. *Computer Vision–ACCV 2010*, pages 501–512, 2011.

[9] M. S. L. B. Dong Seon Cheng, Marco Cristani and V. Murino. Custom pictorial structures for re-identification. In *British Machine Vision Conference (BMVC)*, page 68.168.11, 2011. http://dx.doi.org/10.5244/C.25.68.

[10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, pages 2360–2367, 2010.

[11] R. A. FISHER. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[12] N. Gheissari, T. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1528–1535. IEEE, 2006.

[13] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 09/2007 2007.

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

[15] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *ICDSC*, pages 1–6, 2008.

[16] M. T. Harandi, C. Sanderson, A. Wiliem, and B. C. Lovell. Kernel analysis over riemannian manifolds for visual recognition of actions, pedestrians and textures. In *Proc. IEEE Workshop Applications of Computer Vision (WACV)*, pages 433–439, 2012.

[17] X. He and P. Niyogi. Locality preserving projections. In *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[18] M. Hirzer, C. Beleznai, P. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. *Image Analysis*, pages 91–102, 2011.

[19] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 26–33. IEEE, 2005.

[20] I. Jolliffe. *Principal Component Analysis*. John Wiley Sons, Ltd, 2005.

[21] K. Jungling and M. Arens. View-invariant person re-identification with an implicit shape model. In *Proc. 8th IEEE Int Advanced Video and Signal-Based Surveillance (AVSS) Conf*, pages 197–202, 2011.

[22] K. Jungling, C. Bodensteiner, and M. Arens. Person re-identification in multi-camera networks. In *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 55–61, 2011.

[23] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 2288–2295, 2012.

[24] T. Kozakaya, S. Ito, and S. Kubota. Random ensemble metrics for object recognition. In *ICCV'11*, pages 1959–1966, 2011.

[25] C. Liu, G. Wang, X. Lin, and L. Li. Person re-identification by spatial pyramid color representation and local region matching. *IEICE Transactions*, 95-D(8):2154–2157, 2012.

[26] B. Ma, Y. Su, and F. Jurie. Local Descriptors Encoded by Fisher Vectors for Person Re-identification. In *12th European Conference on Computer Vision (ECCV) Workshops*, pages 413–422, Italy, 2012.

[27] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 2666–2672, 2012.

[28] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, pages 1–11, 2010.

[29] M. Sugiyama. Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of 23rd International Conference on Machine Learning*, pages 905–912. ACM, 2006.

[30] M. J. Swain and D. H. Ballard. Indexing via color histograms. In *ICCV'90*, pages 390–393, 1990.

[31] Y. Wu, M. Mukunoki, T. Funatomi, M. Minoh, and S. Lao. Optimizing mean reciprocal rank for person re-identification. In *Proc. 8th IEEE Int Advanced Video and Signal-Based Surveillance (AVSS) Conf*, pages 408–413, 2011.

[32] J. Yang, Z. Shi, and P. A. Vela. Person reidentification by kernel pca based appearance learning. In *CRV'11*, pages 227–233, 2011.

[33] L. Zelnik-manor and P. Perona. Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems 17*, pages 1601–1608. MIT Press, 2004.

[34] Y. Zhang and S. Li. Gabor-lbp based region covariance descriptor for person re-identification. In *Proc. Sixth Int Image and Graphics (ICIG) Conf*, pages 368–371, 2011.

[35] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pages 649–656, 2011.

[36] W.-S. Zheng, S. Gong, and T. Xiang. Transfer re-identification: From person to set-based verification. In *CVPR*, pages 2650–2657, 2012.