Covariance Based Person Re-identification Using Spectral Matching

Aparajita Nanda
Department of CSE
National Institute of Technology
Rourkela, India, 769008
Email:512CS102@nitrkl.ac.in

Pankaj K. Sa
Department of CSE
National Institute of Technology
Rourkela, India, 769008
Email:pankajksa@nitrkl.ac.in

Banshidhar Majhi
Department of CSE
National Institute of Technology
Rourkela, India, 769008
Email:bmajhi@nitrkl.ac.in

***Abstract*—This paper presents an approach for re-identification based on appearance. The person re-identification is recently introduced and yet an unsolved problem in computer vision. Re-identification refers to identify an individual who has already been observed by different cameras. The appearance of an individual in different cameras looks unlike due to illumination variations and arbitrary pose alternations. The identity of an individual is represented by a distinct signature model that should invariant to illumination, pose variation and occlusions. This paper focuses on the formation of distinct signature models based on mean covariance patch. A patch homogeneity is proposed which handles the clutter in the image of a specific individual. The signature model of each individual needs to find its corresponding signature model over the network. The idea of spectral matching is used for the computation of matching between the models signature. The matching signature models are ranked according to matching scores. The performance of our approaches is evaluated on ETHZ and VIPeR data sets and the results are shown in cumulative matching characteristics.**

I. Introduction

Today the necessity of video surveillance is expanded over a broad scope of fields such as from small indoor surveillance to wide area monitoring. For making the surveillance system more automated and intelligent, camera networks are being deployed which leads to the extraction of high level information from a large bulk of recorded video. In such surveillance system the identity of an individual with respect to the position and time is the most significant matter for detection as well as tracking. Moreover, tracking over a camera network means maintaining the identity of an individual from one camera to other. Re-identification can be defined as the ability to recognize a previously observed individual, who leaves one camera view and enter the view of another camera or re-enter in the same place after a period of time. Alternatively, it may be considered as the process of associating the identity of one individual with others. The identity of an individual depends upon the accessibility of information cue in the system. Biometric cues such as iris, gait and face can be considered for identification, however, due to low resolution, arbitrary changes of the poses and occlusions makes the biometric cues inefficient for identification [6]. Figure 1 illustrates the difficulties of considering biometric cue for identification. Therefore, the appearance of the whole body of an individual can be seen as an alternative cue for re-identification with the premise that the individual's attire remains unchanged during the surveillance. The use of appearance based cue can also referred as appearance based re-identification. It requires forming a signature model that represents the identity label of an individual that must be distinctive and invariant to pose changes, illumination variations and occlusions. These distinctive signature models, search for their reliable matching signature pairs over the network within a large set of individual extracted from different cameras.

978-1-4799-5364-6/14/$31.00 ©2014 IEEE

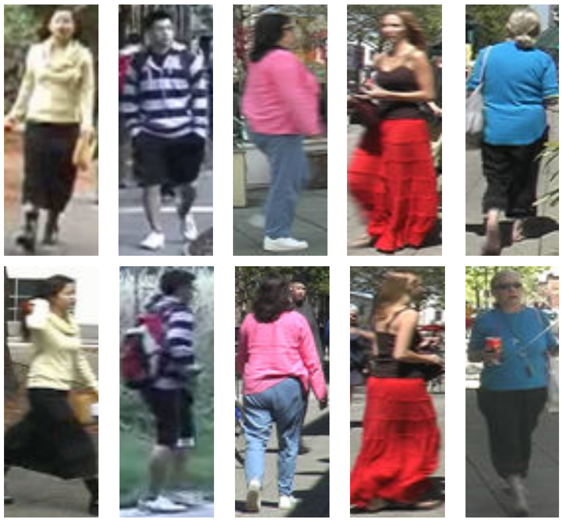

Fig. 1. Sample images from two cameras with different view point

This paper describes the appearance based re-identification that forms a discriminative signature model for each individual appears in the surveillance video followed by matching strategy. The region based mean covariance patches are computed for the formation of distinct signatures that provides invariance to pose changes and illumination variation and occlusions. These patches are the local image instances that are extracted from body part subdivision. The patch homogeneity is introduced, that computed over the covariance patches for avoiding the clutter in the background. The idea of spectral matching is followed for finding the matching pairs of signature models over the networks.

The paper is structured as follows. Section II describes the related work on appearance-based person re-identification, In section III problem is defined and Section IV describes the proposed approach for re-identification. Section V shows an experimental evaluation of proposed approach in terms of matching score followed by conclusion in section VI.

II. Related Work

Appearance based re-identification relies on forming a signature that represents the identity of an individual. This can be achieved either by extracting the appearance based distinct information from a single image (single shot) or by integrating the information extracted from multiple images of an individual (multi shot). In event of single shot the whole body part of an individual is segmented into multiple parts and features are extracted from each part to forge a signature model for the individual. Afterwards it is accompanied by the concept of template matching. In case of multi shot the sequence of images is considered for training to form signatures and trying out a probe image of an individual for computing the similarity.

In case of single shot, in [1] an object is represented by sequence of parts. It follows the multiple component matching (MCM) approach and it is basically modeled as a recognition problem which can be resolved by matching approach. Further MCM extended to [2] where it endures for a dissimilarity based framework named as multiple component dissimilarity (MCD). In [3] the MCD based re-identification is further improved with respect to its computational speed. It reports about the trade off between accuracy and matching time and a model is developed for its evaluation. The proposed approach of [4] works for both single shot as well as multi shot. The symmetry and asymmetry axis are found out which helps to separate the image into multiple parts. Weighted color histograms, maximally stable color regions (MSCR) and structured patches are considered as local features extracted from each image. [5] focuses on computation of human signature. The person re-identification approach uses the covariance matrices as the feature descriptor. The spatial pyramid matching strategy is applied for matching of corresponding regions between images. Regarding multi shot, [6] proposed a human signature model for multi shot case where the mean Riemannian covariance (MRC) patches are extracted which keeps the information regarding temporal changes. Patch reliability and boosting scheme are also developed for enhancing the discriminating nature of MRC patches. A new similarity measure is also introduced to measure the discriminating power between the MRC patches. Person re-identification problem is considered as a ranking problem in [7], [8]. In [7] an ensemble RankSVM is developed to overcome the scalability issue of SVM based ranking for person re-identification. Normalized graph Laplacian and unnormalized graph Laplacian are used in [8]. The matching gallery images are assigned a positive label and ranking score is calculated along the base of the two ranking algorithms. In [9] the re-identification problem is modeled as a relative distance learning problem. This model maximizes the likelihood of a couple of true matches with relatively smaller distance than the wrong match pair. In [10] an approach for re-identification based on supervised salience learning is proposed that develops an unsupervised framework for extracting the distinctive feature, without identifying label for the persons during training. The human salience or distinct features are learned and integrated with patch matching.

The approaches for re-identification require to handle the issues such as pose, viewpoint and illumination variations. But a single feature descriptor alone is not able to handle all these issues. A discriminative feature needs to be designed by integrating multiple features into a single descriptor. The accuracy and computing time of re-identification approach depends on the matching strategy of the set of probe and gallery images.

III. Problem Definition

Let $G_{ij} = \left\{g_i^j\right\}$ be the signature model where $i = \{1, \ldots, N\}$ denotes the identity label of N persons and $j > 1$ denote the number of cameras. The signature models of one camera can be assumed as gallery set whereas the probe set is the set of signature model from other camera. For a given probe signature model g_i^j, the corresponding signature is searched in gallery set signature model. Moreover the signature of probe and gallery are compared by computing a matching score. The highest matching score is considered as the true match.

IV. Proposed appearance based re-identification

This section depicts the appearance based re-identification that involves the formation of a discriminative signature model for each individual followed by a matching approach. Covariance patch extraction, computation of optimum mean covariances are described in detail. The reliable patches are obtained through the patch homogeneity. For each signature model and its corresponding pairs are determined by spectral matching.

The color distribution of an object depends upon the illumination which varies for different lighting conditions. So an object appears in two different camera looks dissimilar with respect to color. In order to compensate the color variations, image histogram equalization is applied to each of the color channels (RGB) to obtain a nearly invariant image. This represents invariation in intensity component only. For extracting the color component equalized image is mapped to HSV color space, as HSV separates the image intensity and the color information. In addition mapping of RGB to HSV color space provides robustness to illumination changes.

A. Covariance Patch Extraction

Multiple feature descriptor provides more detailed information about an object and more robust to invariants. Features like color, edge, texture, intensity, gradient etc. can be combined together into a single descriptor. The covariance matrix is one that fuses the multiple features and also find a global optimum descriptor. These are scale and rotation invariant as well as robust to illumination variations, since this methods had successfully been applied to object tracking [11] and [6]. Every region of an image can be described by a covariance patch descriptor regardless of its dimensionality.

Let I be an image and Z be a $W \times H \times d$ dimensional feature vector extracted from I.

$$Z(x,y) = \Phi(I,x,y) \tag{1}$$

Where the function ϕ can be any mapping function that is typified by a feature vector as follows

$$\left[x, y, H\left(x,y\right), S\left(x,y\right), V\left(x,y\right), G^i(x,y), \nabla^i(x,y), NFP(u,v)\right]^T \tag{2}$$

Where x, y denotes the pixel locations. $H(x,y)$, $S(x,y)$, $V(x,y)$ are the color feature of HSV channels. $G^i(x,y)$, $\nabla^i(x,y)$ and $i \in [H,S,V]$ corresponds to gradient magnitude and orientation of each channel to extract the change in intensity and color of an image. Gradient magnitude and orientation are computed by convolving the image with a bank of log-Gabor filters [12]. $NFP(u,v)$ represents the normalized Fourier coefficient of an image. It is an approach to texture description [13]. NFP is obtained as

$$NFP\left(u,v\right) = \frac{\left|FP\left(u,v\right)\right|}{\sqrt{\sum_{(u\neq 0)\wedge(v\neq 0)} \left|FP\left(u,v\right)\right|^2}} \tag{3}$$

where

$$FP\left(u,v\right) = T\left(P\left(x,y\right)\right)$$

$FP\left(u,v\right)$ is the Fourier transform of pixel data $P\left(x,y\right)$ respectively. The advantage of using Fourier transform is that it posses shift invariance i.e. the transform of a bit of large or uniform segment will remain same. The magnitude transformed data is considered and is normalized by the sum of the squared values of each magnitude component. So that the obtained normalized Fourier coefficient is invariant to linear shifts in illumination.

For each rectangular region $R \subset Z$, let $\{z_k\}_{k=1\ldots n}$ be the $d-$dimensional feature points within the region R. Each of the regions R is represented with $d \times d$ covariance matrix of the feature points.

$$C = \frac{1}{n-1}\sum_{k=1}^{n}(z_k - \bar{z})(z_k - \bar{z})^T \tag{4}$$

$\bar{z}$ is the mean feature vector. The covariance patch descriptor is a symmetric matrix, where the diagonal elements represent the variance of each feature and the non diagonal elements represents the correlation of features. To heighten the local illumination variation in an image the covariance descriptors are normalized by using Pearson Product Moment Correlation Coefficients as in eq. 5. These normalized matrices are also termed as correlation matrix. Afterwards, every patch matrix is considered to be the normalized covariance patch matrix. The normalization is computed as

$$\overset{*}{C}\left(i,j\right) = \frac{C\left(i,j\right)}{\sigma_i\,\sigma_j} \tag{5}$$

The detected object of interests (human) are resized into a fixed size of $W \times H$. Histogram equalization is used for color normalization of resized images and mapped to HSV color channel for extracting color information. A number of covariance patches of size $\frac{W}{4} \times \frac{W}{4}$ are obtained from respective positions by shifting horizontally and vertically $\frac{W}{8}$ over the image. The position and the spatial correlation between the patches signify the discriminating power of feature set.

Let $\left\{X^k\right\}_{k=1\ldots N}$ be the set of covariance patches extracted from each image of a person appearing in N consecutive frames. $X^k = \{C_i^p\}_{i=1\ldots s}$ represents the set of normalized covariance patches with respect to the position p of the patch. Thus, the signature for a particular person is computed by combining the sets of covariance patches incurred from the sequence of N images. This signature is considered as the unique identification of the human. It not only keeps the information about feature description, but also keeps the essential cues about temporal changes of appearances with respect to the position. For forming signature mean covariance patches are computed.

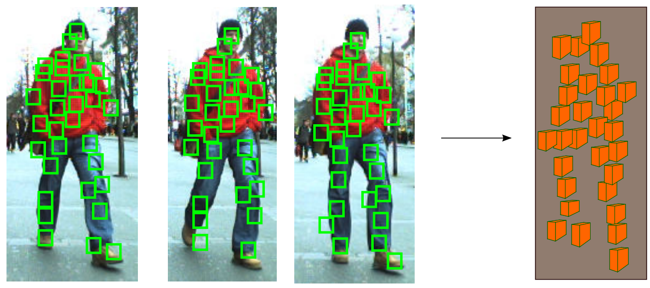

Fig. 2. Multi shot case: Illustration of the covariance patches extracted from the sequence of frames (left) and mean covariance patches formation (right)

B. Mean Covariance Patch Matrices

Let $C_1,\ldots,C_N$ be a set of covariance matrices for a particular position p . The normalized mean covariance matrices signify as a human signature or a descriptor. The covariance matrices are the positive definite and symmetric can be considered as tensors. Such defined tensor space is assumed as a manifold [14]. Manifold is a topological space that is logically similar to Euclidean space. The Karcher or Frchet mean is the set of tensors minimizing the sum of squared distances. For tensor the manifold has a non-positive curvature, so there is a unique mean value m [14], [6]. Figure 2 demonstrates the mean covariance patch formation by combining the extracted patches from the sequence of frames.

$$m = \underset{C\in MF}{\operatorname{argmin}} \sum_{i=1}^{N} \rho^2\left(C, C_i\right) \tag{6}$$

where ρ is the covariance patch matrix distance and MF represents a manifold (tensor space).

C. Covariance matrix distance

The distance measure proposed in [6] is used to measure the dissimilarity of two covariance matrices

$$\rho\left(C_i, C_j\right) = \sqrt{\sum_{q=1}^{d} \ln^2 \lambda_q\left(C_i, C_j\right)} \tag{7}$$

where $\left\{\lambda_q\left(C_i, C_j\right)\right\}_{q=1,\ldots,d}$ are the generalized eigenvalues of C_i and C_j, computed from

$$\lambda_q\, C_i\, X_q - \lambda_q\, C_j\, X_q = 0, q = 1\ldots d$$

and $X_q \neq 0$ are the generalized eigen vectors.

D. Mean optimum

The obtained mean covariance patch space is considered as a manifold [14]. The Quasi Newton method is used to compute the optimum mean of covariance patch. Quasi Newton methods are iterative, involving a series of line search with less computation [15]. In case of multi shot the mean covariance obtained from N images are assumed to be the initial value and a sequence of mean m_t are produced that converge to the optimum mean m^*. The mean computed for next iteration at $t+1$ step as

$$m_{t+1} = \exp_{m_t}\left[\frac{1}{N}\sum_{i=1}^{N}\log_{m_t}(C_i)\right] \tag{8}$$

$\exp_{m_t}$ and $\log_{m_t}$ are denoted as the operators of Reimannian Manifold [14]. Quasi Newton method retains good convergence properties with faster convergence. It arrives at the optimum value in a finite number of steps. In case of single shot the optimum mean is computed on the extracted covariance patches. Quasi Newton method is considered to be the most effective of the unconstrained multivariate search technique as in [15].

E. Patch Homogeneity

Let $m_1, \ldots, m_s$ be a set of normalized mean covariance patches incurred from the respective image square regions. In addition to mean covariance patches the patch homogeneity is defined as

$$H(C_i) = \frac{1}{|C_i|}\sum_{z_i \in C_i} S(z_i, \bar{z}_i) \tag{9}$$

$$S(z_i, \bar{z}_i) = \left[1 - \frac{|z_i - \bar{z}_i|}{D\text{max}}\right] \tag{10}$$

$$D\text{max} = \max_{z_i \in C_i}(|z_i - \bar{z}_i|) \tag{11}$$

where $\bar{z}_i$ is the mean of the feature vector of the normalized covariance patch belongs to a specific region. $S(.)$ is the distance measure of the feature vector from its mean feature vector. $Dmax$ corresponds the maximum distance of any feature vector in the corresponding region to its mean. Homogeneity of the covariance signifies the purity or the strength of the pixels of the region with respect to color, gradient and orientation. It helps to separate the coherent covariances with high information content. The idea is to filter out the most variable features because the feature belong to the background region are considered as the noisiest with less homogeneity value. The coherent covariance patches are selected as the patch descriptors which participate on the matching process.

F. Spectral Matching

This section describes how the human signature models (gallery images) is compared with the given probe image and the similarity between two signatures is determined. For this purpose the notion of spectral matching for re-identification is introduced. The original formulation of spectral matching has been proposed in [16]. Spectral matching is an efficient technique used to resolve the correspondence problem between the probe image and the gallery image. The purpose of using this technique is to establish the pair wise relationship between the feature sets of probe and gallery respectively. It facilitates one to one as well as one of many matching strategies for the feature sets. Spectral correspondence is one of the robust technique that requires less computational complexity even for large datasets. The feature vectors extracted from the probe and gallery are considered as two sets of features. In our case the covariance patch matrices are represented as feature sets. Spectral matching follows the graph matching technique. It makes a weighted undirected graph where the nodes of the graph are the pair of candidate assignments from both feature sets. The edges denote the agreement between two pair of candidate assignment. A positive edge or agreement link is established when two pair of features is correspondence with another pair of assignment with respect to the pairwise relationship. Otherwise there is no link between the pair of assignment. All the positive edge together forms a strongly connected cluster. The cluster formation begins by observing the degree of association of each assignment with the cluster by scrutinizing the eigenvector of the graph corresponding to its largest eigenvalue.

Let us assume P and G represent the two feature sets of probe and gallery image respectively. n_p and n_g are the number of data features in P and G. Now the objective is to construct an undirected weighted graph A from the feature sets. A list L of n candidate assignment is formed, where the candidate assignments denote the corresponding mapping from one feature set to another one. Let (i, i') be a candidate assignment where $i \in P$ and $i' \in G$. For each candidate assignment $k = (i, i')$, an affinity score is computed. The affinity score refers the matching between the $i \in P$ and $i' \in G$. Also for pair of assignment (k, l) where $k = (i, i')$ and $l = (j, j')$ affinity measures are worked out which represents the compatibility between the features (i, j) and (i', j'). In the undirected graph A, the candidate assignment $k = (i, i')$ from L are considered as nodes of the graph, with pairwise score $A(k, l)$ as weights on the edge and $A(k, k)$ as weights at the nodes. For every assignment $k \in L$ and every pair of assignment $k, l \in L$, a $n \times n$ affinity matrix A is formed. $A(k, k)$ represents the individual assignments $k = (i, i')$ from the candidate list L. The computed patch homogeneity value defines the affinity score for individual assignment. The correct assignments are extracted by thresholding the homogeneity value and reform the affinity matrix A. On the other hand $A(k, l)$ represents the pairwise relationship between the correspondence features (i, j) and (i', j'). It draws the compatibility of (i, i') with (j, j'), otherwise $A(k, l)$ is set to zero. The matrix A became a $n \times n$ sparse symmetric matrix where $n = an_p$ and a is the average number of matching candidates for each feature $i \in P$. Each feature $i \in P$ have a different candidate matches (i, i'), $i' \in G$. All matching candidates are considered as positive link edge or cluster of assignments. A cluster Cl of assignments (i, i') is to be chosen that maximizes the inter cluster score R. A cluster is denoted with indicator vector I_v, where $I_v(k) = 1$ if $k \in Cl$ and else zero.

$$R_{score} = \sum_{k,l \in Cl} A(k, l) = I_v^T A I_v \tag{12}$$

The inter cluster score depends upon the number of candidate assignments in the cluster, number of positive linkages and the weights on the link. The cluster with maximum inter cluster score is resolved by eigen vector technique. The optimal solution I_v^* that maximizes the score as

$$I_v^* = \text{argmax}(R_{score}) \tag{13}$$

For a given probe image spectral matching technique is iterated for each feature set of respective gallery images. The

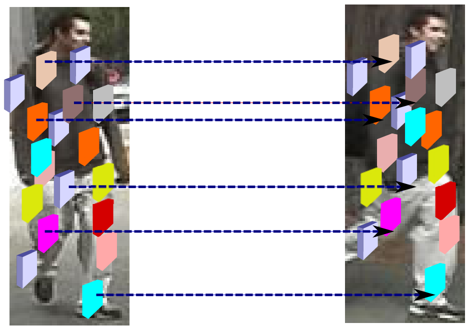

Fig. 3. Spectral matching between probe and gallery signatures. The set of optimum mean covariance patches are denoted as signature. The patches in probe signature find its best corresponding patches in gallery signature. Some of the patches are illustrated for clarity

inter cluster scores are the matching score of all gallery images and are arranged in diminishing order of their scores. The ranks are assigned in increasing order. The lowest ranks are taken as the true match for the given probe.

V. Experiment and Results

This section describes a series of experiments that have been carried out for testing the re-identification in terms of the matching capabilities of the model. We evaluated our approach on two publicly available datasets, the ETHZ dataset [17], [18], and the VIPeR dataset [19]. These datasets are most widely practiced in person re-identification due to its challenging factors in real world scenario such as viewpoint and illumination variations, pose, different background, low resolution and occlusion. The execution is presented in standard Cumulative Matching Characteristic (CMC) curve [19]. The CMC curve represents the rate of correct matches among the top matches. The results are compared with the state of the art methods and likewise with some recent methods.

Features : All the images were resized into a fixed size of 64×192 pixels. 16×16 patch sizes were generated from the image regions and shifted 8 positions horizontally and vertically. From each patch a 12 dimensional feature vector was extracted from each pixel position. Thus, for each image region at 12×12 covariance matrices were formed. In multi shot case, the mean covariance patch was computed at their respective positions and iterated for optimum value. In addition, for single shot the normalized covariance patches were considered and optimized. Patch homogeneity was calculated for each patch and spectral matching technique was applied for obtaining the matching score between probe and gallery pictures.

A. *ETHZ Dataset*

ETHZ dataset comprises of 3 video sequences. The original video sequences are captured from a mobile camera. The images appearing in the video sequence have varied in appearance, illumination and occlusion. The modified ETHZ dataset in [18] is weighed for our experimentation. According to [18], all the images were normalized to 64×32 pixels, where SEQ-1 contains 83 pedestrians with 4857 images, SEQ-2 contains 35 pedestrians with 1936 images and SEQ-3 have 28 pedestrians with 1762 images. However the images of the dataset have low variation with respect to pose.

For each person a set of 10 images were randomly selected from the sequences. These images were used for the signature formation for each person. The formed signatures were counted as the gallery set. For a given probe image, the covariance based features were extracted and compared with the gallery set image signatures by using the spectral matching. Our approach was compared with SDALF [4], and recent methods such as LCP, RCP [6], unsupervised based eSD_knn, eSD_ocsvm as in [8]. The results of eSDC_knn and eSDC_ocsvm presented in fig 4, fig 5, fig 6 are the single shot case of ETHZ. For SEQ-1 our method provides better results than SDALF, eSD_knn, eSD_ocsvm and comparable to LCP and RCP results. On SEQ-2 our method outperforms all other methods. On SEQ-3 our method provides better results than SDALF, RCP, eSDC_knn and eSD_ocsvm. The performance results for SEQ-1, SEQ-2 and SEQ-3 are shown in Fig 4, fig 5 and fig 6 respectively.

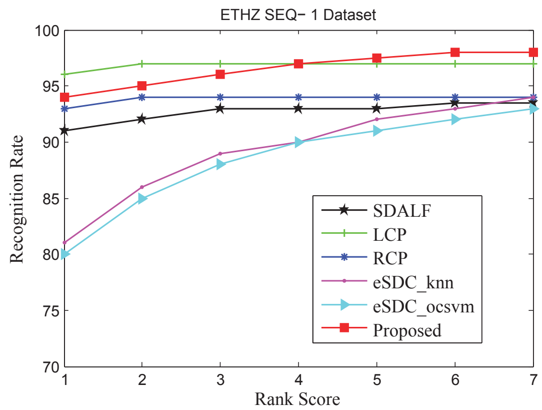

Fig. 4. Recognition results of ETHZ SEQ-1 Dataset

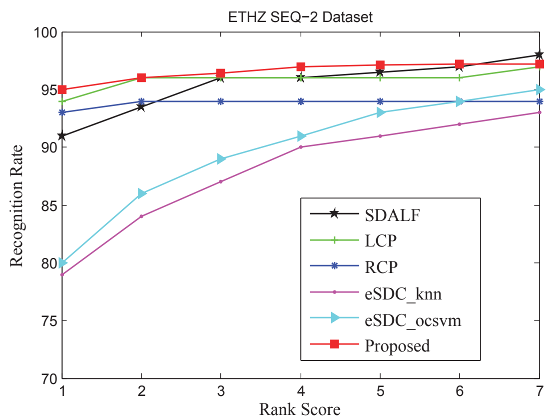

Fig. 5. Recognition results of ETHZ SEQ-2 Dataset

B. *VIPeR Dataset*

The VIPeR [19] is one of the challenging dataset for the person re-identification. It consists of 632 pedestrian pairs taken from two different cameras at arbitrary view points.

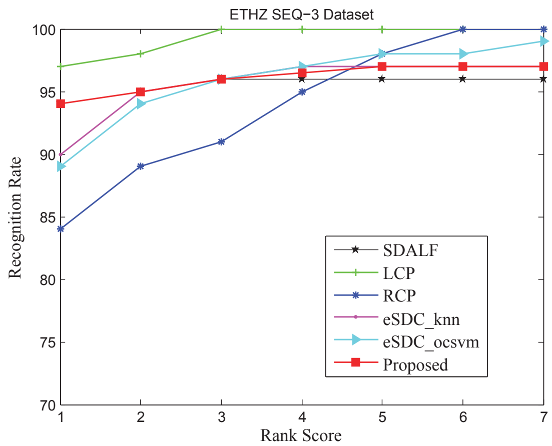

Fig. 6. Recognition results of ETHZ SEQ-3 Dataset

Each pair contains two images of the same person. The image pairs differ with respect to viewpoint, pose and illumination. Images were normalized to 128×48 pixels. Images from one camera were considered as a gallery where as the probe was the images captured from another camera. The highest matching score for each probe with the gallery is considered as the true match and ranking them in order of their matching score. As VIPeR belongs to single shot images, we compared our results with the SDALF [4], ELF [20], and recent ranking approaches reported in [6]. The comparison results are presented in Fig 7. Our method out performs all other considered approach.

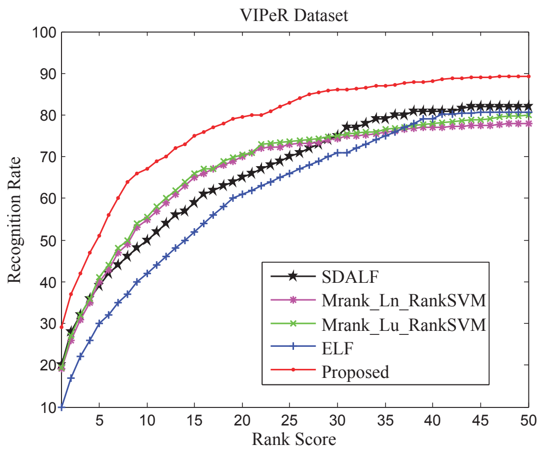

Fig. 7. Recognition results of VIPeR Dataset

VI. Conclusion

In this work we have addressed the person re-identification problem as an appearance based matching problem. The signature model based on covariance patch handles the view point and arbitrary pose changes. The color variations are handled by the color normalization scheme. The computation of patch homogeneity makes the distinction between the foreground and cluttered background patches. The extracted mean covariance patches are matched through the spectral matching which provides sufficient flexibility to the viewpoint changes. The proposed also supports both the multi shot as well as single shot modality. This needs more customization in terms of reducing the time complexity of relevant real time application.

References

[1] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino. A multiple component matching framework for person re-identification. *Image Analysis and Processing-ICIAP*, 2011.

[2] R. Satta, G. Fumera, and F. Roli. Exploiting dissimilarity representations for person re-identification. *Similarity-Based Pattern Recognition (SIMBAD)*, 2011.

[3] R. Satta, G. Fumera, and F. Roli. Fast person re-identification based on dissimilarity representations. *Pattern Recognition Letters*, 33(14), 2012.

[4] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[5] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. *Advanced Video and Signal Based Surveillance (AVSS), Seventh IEEE International Conference*, 2010.

[6] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Boosted human re-identification using riemannian manifolds. *Image and Vision Computing*, 30(6), 2012.

[7] B. Prosser, W. S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. *British Machine Vision Conference*, 2(5), 2010.

[8] C. C. Loy, Chunxiao C. Liu, and S. Gong. Person re-identification by manifold ranking. *IEEE International Conference on Image Processing*, 20, 2013.

[9] W. S. Zheng, S. Gong, and T. Xiang. Re-identification by relative distance comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3), 2013.

[10] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*, 2013.

[11] O. Tuzel, F. Porikli, and P. Meer. Region covariance: a fast descriptor for detection and classification. *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, 2006.

[12] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of The Optical Society of America A*, 4(12), 1987.

[13] M. Nixon and A. Aguado. Feature extraction & image processing. *Academic Press*, 2008.

[14] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journals on Computer Vision*, 2006.

[15] SJ. Wright and J. Nocedal. Numerical optimization. *Springer New York*, 2, 1999.

[16] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. *Proceedings of 10th IEEE International Conference on Computer Vision (ICCV)*, 2005.

[17] A. Ess, B. Leibe, and L. Van Gool. Depth and appearance for mobile scene analysis. *Proceedings of the 11th International Conference on Computer Vision (ICCV)*, 2007.

[18] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. *Proceedings of the 22nd Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, 2009.

[19] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. *PETS*, 2007.

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