Person Re-identification Using Prototype Formation

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

***Abstract*—Person re-identification is to match the appearance based images of an individual who has already been captured by different camera perspectives. This paper presents an appearance based model for person re-identification. It consists of the formation of prototypes followed by a matching strategy. The prototypes are discovered from the visual appearances among the individuals with similar characteristics. When a probe comes along, it necessitates to be classified through prototype assignment. To determine the correct matching of a given probe, the similarity computation is performed between the probe and a subset of gallery images, that shares the same prototype with the probe. Thus the strategy of finding the correspondence within a subset of gallery instances reduces the computational overhead. This model is useful in scenarios where individuals appear with similar attire. The performance measure of the proposed method is evaluated on benchmark data sets and presented using cumulative matching characteristic (CMC) curves.**

I. Introduction

Associating individuals across different cameras in a wide coverage space at different instances of time is known as person re-identification. It is a vital task to facilitate cross-camera tracking of people and understanding their global behavior in a wider context. The temporal transition between cameras varies significantly from individual to individual with a great deal of uncertainty. These uncertainty results in images with arbitrary change in pose, variation of illumination, occlusions etc. Figure 1 shows some sample images of individuals who captured from two different cameras. It can be seen that there is a significant change in pose and illumination as well. It also demonstrates the difficulty in segmenting the biometric traits like face and iris. This clearly disapproves the use of such traits as prospective candidate for identification. Hence these issues are addressed on a model, that must rely on appearance based features alone. The objective of appearance based person re-identification deals with the establishment of visual correspondence between instances of same individual at different locations and times. Appearance based person re-identification is also considered as non-trivial problem due to visual ambiguities and illumination changes, unknown viewpoint and pose variations, and inter-object occlusions [1].

The state-of-art person re-identification methods have majorly focused on two strategies: (i) formulating discriminative feature representations of individuals which are invariant to viewing angle and illuminations [2], [3], [4] and (ii) applying learning methods that is capable of making fine distinctions by optimizing the parameters of re-identification model [5]. RankSVM method in [4] aims to find a linear function to weigh the absolute difference of samples through optimization given pairwise relevance constraints. The Probabilistic Relative Distance Comparison (PRDC) [6] shows the probability of a pair of true match having a smaller distance is maximized than that of a wrong matched pair. The requirement of labeled gallery images to discover gallery specific feature importance are described in [3]. In [7] prototype strategy is introduced in re-identification problem. Prototypes are defined as set of instances, that correspond to local appearance characteristics shared by different individuals. Most of the existing prototype based approaches [8], [9] do follow the simple clustering technique for the prototype formation, whereas these prototypes are not offering a promising representation of features, because for each random initialization of clustering algorithm yields dissimilar prototype labels of representation. Hence the prototypes representation depend upon the selection of random points in clustering algorithm. The better qualitative prototypes representation signifies the qualitative features representation with regard to the commonalities. Thus it motivates the formation of prototypes that describes the promising representation of features shared by the gallery instances.

Fig. 1: Samples from VIPeR Dataset with pose and illumination variations. Top row depicts images of seven different individuals from one angle. Bottom row shows images of same individuals from another angle.

There exist some commonality in terms of visual features among the instances of gallery representing different individuals. In this work such common features are exploited to form prototypes representing similar instances in the untagged gallery set images. Considering the prototypes as class labels

of gallery images, k-NN classifier assigns a class label to each probe image. Similarity measure is computed with a subset of gallery images, that shares the same class label with the given probe and hence the number of comparisons between probe and gallery instances are reduced. The resulting scores are listed according to the most similar signature of instances ordered by decreasing similarity measure. Experimental evaluation of the qualitative prototype based approach is performed on three benchmark datasets.

The paper is structured as follows. In Section II problem is formulated. Section III describes the detail steps of re-identification using prototype formation based approach. Experimental evaluation is described in Section IV followed by conclusion in Section V.

II. Problem Formulation

Let $\{Y_i^g\}_{i=1}^n$ be the feature space representing n feature vectors of set of gallery images $\{I_i^g\}_{i=1}^n$. The feature vector of each gallery instance is assumed as signature of the instances. The $\{I_i^g\}_{i=1}^n$ are assigned to $\{P_i\}_{i=1}^K$ prototypes based on the features. For a given probe the objective is to find its corresponding signature in gallery. So for each probe $\{I^{p_r}\}$, a prototype P_i is assigned and the matching scores are computed for gallery images $\{I_i^g\}_{i=1}^t$ where $t \subset n$, the subset instances and the probe shares the same prototype. The gallery and probe images are taken from two different cameras.

III. Proposed Prototype based re-identification approach

This section depicts the detail of prototype based re-identification approach which includes feature space representation, prototypes formation, classification and similarity measure. The color and texture features are extracted from each image of gallery and represented as feature space. Prototypes are discovered from the feature space based on the appearance characteristics of the gallery instances. Assuming the prototypes as the true label, the feature space is trained by using k-NN classifier. Distance based similarity measure is computed between the subset of gallery images and the given probe image.

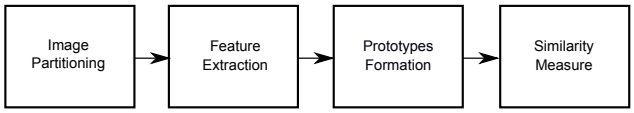

Fig. 2: Sequential steps for Person Re-identification

A. Feature Extraction

In case of feature extraction, different set of components are extracted from sub parts of an image. The principle behind this type of representation is to gain robustness to partial occlusion and pose variations [8]. Formally, let $\{I_i^g\}_{i=1}^n$ be the given input of the n untagged gallery set images, where only one image is available for each individual. A d-dimensional feature vector, that is $Feature\,(I_i^g) = \{y_{1,\ldots,}y_d\}^T \in R^d$ is extracted from each image instance. Thus $Y = \{Y_i^g\}_{i=1}^n$ represents the feature space of gallery images. Each image of gallery is denoted as an ordered sequence of m parts where $(m \geq 1)$.

$$\{I_i^g\} = \left\{I_{i,1}^g, \ldots, I_{i,m}^g\right\} \tag{1}$$

Each part $I_{i,m}^g$ is represented with a set of $d^{'}$ dimensional feature vector $f_{i,m}^{d^{'}}$, $d^{'} \subset d$ and $f_{i,m}^g \in Y$. Where Y denotes the feature space. The feature vectors of all parts are assumed to be represented with same dimensions. In order to roughly capture the head, torso and leg part,the image is partitioned into six equal sized horizontal strips as in [4]. From each strip color features are extracted based on the mixture of color models such as RGB, HS and YCbCr and for texture features 8 Gabor filters [10] and 13 Schmid filters [11] are applied on the luminance channel. The feature vector of each gallery image is integrated to represent the feature space. Considering different types of features lead more discriminate feature space. Because a single feature is not enough for the formation of distinct feature that stands for all image instances. i. e. for individual wearing colorful and bright clothes, the color features yields higher precedence whereas for an individual with high textured clothes, texture features tend to more influencing. To illustrate this, two different image of same individual are considered and the matching rate is computed with regard to different color models, texture features. The matching rates for each feature are determined separately through the average of Euclidean distance measure. Figure 3 shows the matching rate with respect to different types of color and texture features. From Figure 3 it is observed that, a single feature alone is not able to well perform for all image instances where as the combination of features provide more detailed information.

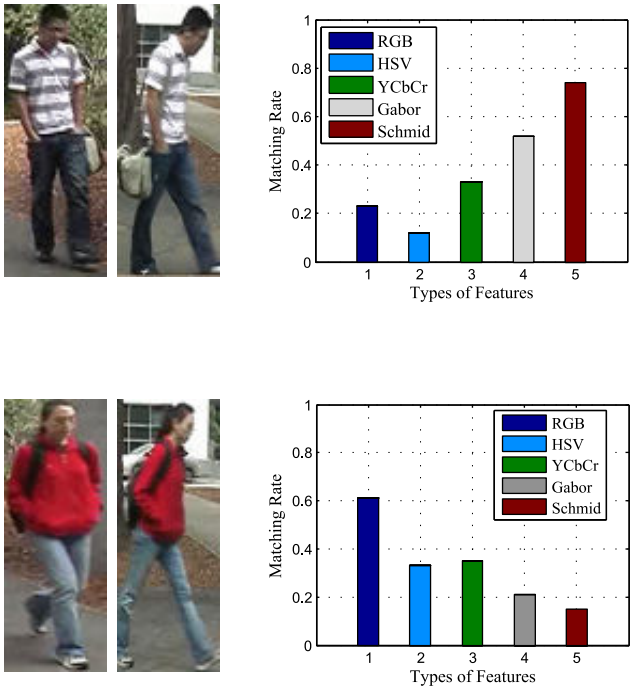

Fig. 3: Matching rate of probe and gallery image on the basis of different color and texture features. RGB, HSV and YCbCr color models are taken for color feature and Gabor and schmid filters are considered for texture feature.

B. Cluster Ensemble Based Prototypes Formation

The set of feature vectors is denoted as feature space where each element represents an image instance. The aim of prototype formation is to cluster a given feature set of untagged images into several prototypes representation. Each prototype represents images with similar visual appearance based features such as colors, textures and shapes with colorful shirts, blue jeans, dark jackets or back pack as in Figure 4. The motivation for prototype formation signifies to distinguish the individuals with similar attire in a crowded environment. A set of prototypes $\{ P_i\}_{i=1}^{K}$, is assumed as low-dimensional manifold clusters [12] that group images $\{I_i^g\}_{i=1}^{n}$ with similar appearance based features. We treat the prototype formation problem as a clustering ensemble problem. Cluster ensemble methods have emerged as powerful tools for improving the robustness as well as the accuracy of clusters [13]. The objective of the clustering ensemble task is to search for a combination of multiple prototype labels that provides improved overall clustering of the given untagged gallery image. Cluster based similarity partitioning algorithm (CSPA) [14] is one of the cluster ensemble technique that can be used for prototype formation.

Fig. 4: Example of prototype formation on few images of gallery set of VIPeR Dataset. P_1 and P_2 denote the prototypes. P_1 and P_2 represent the images with similar appearances. Based on the feature of the probe image, it only compares with the images belonging to prototype P_1. The green bounding box signifies as true match.

1) Cluster based Similarity Partitioning: In order to formulate the prototypes we construct an ensemble of $N_{cluster}$ prototype labels using several runs of K-mean clustering algorithm on the feature space Y. Each prototype label λ^{cl} where $\lambda^{cl} = \bigcup_{a=1}^{K} P_a$ and $P_i \cap P_j = \phi$, is obtained from each run of K-mean and defines K partitions of the input image samples $\{I_i^g\}_{i=1}^{n}$ with respect to their features.

A prototype signifies a relationship between sample images in the same prototype and can thus be used to establish a measure of pairwise similarity. For each prototype label, a co-association matrix is computed. Co-association matrix is a symmetric binary square matrix of size $n \times n$, n being the number of image samples to be classified. The similarity between two sample images is 1 if they are in the same prototype and 0 otherwise.

$$S_{ij} = \frac{1}{N_{cluster}} \sum_{cl=1}^{N_{cluster}} I\left(\lambda_i^{cl}, \lambda_j^{cl}\right) \tag{2}$$

where λ_i^{cl} represents the prototype to which ith sample belongs in prototype label λ^{cl}.

$$I\left(\lambda_i^{cl}, \lambda_j^{cl}\right) = \begin{cases} 1 & (i,j) \in P_a\left(\lambda^{cl}\right) \\ 0 & otherwise \end{cases} \tag{3}$$

The entry-wise average of $N_{cluster}$ such matrices representing the $N_{cluster}$ sets of groupings yields an overall co-association matrix that is used to re-cluster the sample images, yielding a combined prototype label. The overall similarity matrix is considered as an undirected graph where vertex represents an object and edge weight represents similarities. Given the co-association matrix, a normalized cut algorithm is employed to partition the weighted graph into K clusters. These K clusters are considered as K prototypes. Thus, each untagged probe image is assigned to a prototype P_i. The K value is manually decided by observing the datasets or can be estimated automatically using alternative methods.

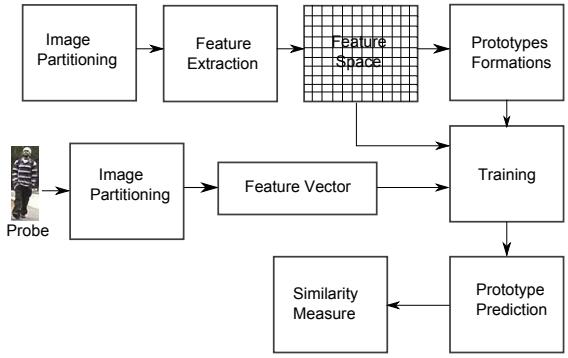

Fig. 5: Overview of re-identification using prototype formation

Algorithm 1: Ensemble based prototypes formation

Input : Ensemble of prototype labels $\left\{\lambda^{cl}\right\}_{cl=1}^{N_{cluster}}$ where $\lambda^{cl} = \bigcup_{i=1}^{K} P_i$

begin
  for $cl = 1$ **to** $N_{cluster}$ **do**
    Compute the co-association matrix by Eq.3 ;
  end
  Compute the average co-association matrix by Eq.2;
  A normalized cut algorithm is employed to partition the matrix into K prototypes;
end

Output: Cluster ensemble prototypes $\{P_i\}_{i=1}^{K}$

C. Training Through k-NN Classifier

The K prototypes characterized by different appearance characteristics and that are assumed to be the efficient representation of images with similar appearance based features. Moreover each prototype P_i has its own appearance based feature importance which is learned by the k- nearest neighbor. The prototypes that are obtained from the cluster ensemble approach are considered as the class label for the feature set of gallery images. The objective of using k-NN classifier is to assign each untagged probe image $I_i^{p_r}$ to a prototype (class

label). So for a given probe image $I_i^{p_r}$ is need to be compared only with a set of gallery images that belongs to the same prototype with the probe image. Thus instead of comparing the probe with all feature vectors of gallery set image, it only compares with the subset of image feature vectors of the prototype that it belongs to and reduces the number of comparisons.

Based on the above intuition, we compute the importance of robust prototype assignment of probe according to its ability in discriminating different set of feature vector of image samples. Specifically, we train a k-NN classifier [15] with $\{Y\}$ as inputs and treating the associated prototype labels $\{P_i\}$ as classification outputs. For a given probe image $\{I^{p_r}\}$, we classify it using the learned k-NN classification strategy to obtain its prototype label (class label). Then similarity measure of probe image $\{I^{p_r}\}$ against gallery images $\{I_i^g\}$ of the corresponding prototypes are computed.

D. Similarity Measure

Given a probe image I^{p_r} is represented as sequence of parts with feature vectors as well, the task is to find the most similar feature vector $x^* \in Y^{'}$, where $Y^{'} \subset Y$, according to similarity measure $D(\cdot,\cdot)$.

$$x^* = \arg\min_{I_i^g} D\left(I_i^g, I^{p_r}\right) \tag{4}$$

where $D\left(I_i^g, I^{p_r}\right)$ is defined as a function of a similarity measure between sets of feature vectors.

$$D\left(I_i^g, I^{p_r}\right) = f\left(dist\left(I_{i,1}^g, I_1^{p_r}\right), \ldots, dist\left(I_{i,m}^g, I_m^{p_r}\right)\right) \tag{5}$$

The measure of $dist(\cdot,\cdot)$ is the Hausdorff distance d_H [16]. Given two set Q and S, d_H is defined as the distance among the minimum distances between all pairs of elements from Q and S. The dist() is defined as the similarity measure of m pairs of parts. For example, $D(\cdot)$ can be determined as the additive combination of m distances.

$$d_H(Q,S) = \max\{\ H(Q,S),\ H(S,Q)\} \tag{6}$$

$$H(Q,S) = \min_{q\in Q, s\in S}(\parallel q - s \parallel) \tag{7}$$

$\parallel \cdot \parallel$ denotes the distance metric between the elements of the set. The partition based distance measure helps in attaining robustness to outlying parts that come from partial occlusion. The result of the similarity measure of the probe is given by the list of the most similar feature vector of the gallery images ordered by increasing dissimilarity. The identity of the probe is determined by finding the gallery images that are most similar to the probe using similarity measure.

E. Evaluation Criteria

The recognition rates are evaluated with the Cumulative Matching Characteristic (CMC) curves [17]. The CMC curve represents the expectation of finding the correct match in the top rank matches. In other words, a rank recognition rate shows the percentage of the probes that are correctly recognized from the top matches in the gallery images.

Algorithm 2: Re-identification by prototypes formation

Notations: I^{p_r}: probe image, n' : # gallery images share the same prototype with probe
Input : $\{Y_i^g\}_{i=1}^n$ is feature space representing n feature vectors of set of gallery images $\{I_i^g\}_{i=1}^n$.

begin
- Formation of prototypes $\{P_i\}_{i=1}^K$ by algorithm 1;
- Feature space $\{Y_i^g\}_{i=1}^n$ and prototypes $\{P_i\}_{i=1}^K$ are trained in k-NN classifier ;
- For a given probe I^{p_r}, it is classified to a prototype P_i using k-NN classifier ;
- **for** $j = 1$ **to** n' **do**
  - Distance measure computed between I^{p_r} and I_j^g where I^{p_r} and I_j^g belong to same P_i by Eq.5, Eq. 6, Eq.7 ;
- **end**
- Compute the similarity score of x_q^* by Eq.4 ;

end
Output : Similarity scores for probe image x_q^*

IV. Experimental Evaluation

Datasets: VIPeR [17], ETHZ [18] and QMUL under Ground Re-identification (GRID) [19] are the publicly available person re-identification datasets used for experimental evaluation. The VIPeR dataset consists of 632 pedestrian image pairs taken from two camera views. VIPeR is one of the most promising and challenging dataset with differences in pose, orientation and illumination. It contains only one image for each individual. ETHZ dataset was originally used for human detection and these data sets have been adjusted for re-identification purposes in [18]. The modified dataset consists of three sequences. Experimental evaluation was performed only on Sequence 1 with 83 pedestrians. For this dataset re-identification is performed with the same camera. For Gallery set and probe set images single-shot were considered for each individual with different pose. GRID dataset is captured in a busy underground station, with severe inter-object occlusion and large viewpoint variations. We compared our methods with ELF [17] and ensemble RankSVM [4] and present the results in Figure 6, figure 7 and figure 8 respectively. We have also demonstrated the impact of the formation of clustering ensemble based prototypes and cluster based prototypes on the feature space of the set of gallery image. Table I shows the performance of ensemble based prototypes for recognition.

Feature Extraction: Image was partitioned into six horizontal strips of equal size. Similar to [4], [5], [20] mixture of color (RGB, HSV and YCbCr) and texture features (8 Gabor filters and 13 Schmid filters) were extracted and forming a 2784-dimensional feature vector for each image. Each feature channel was represented with 16 dimensional feature vector. The Gabor filter used had parameters γ, λ, θ and σ^2 that were set to (0.3,0,4,2), (0.3,0,8,2), (0.4,0,4,1), (0.4,0,4,1), (0.3,$\frac{\pi}{2}$,8,2), (0.4, $\frac{\pi}{2}$,2 ,4,1) and (0.4, $\frac{\pi}{2}$, 2 ,8,2) respectively. The Schmid filters used parameters were set to (2,1), (4,1), (4,2), (6,1), (6,2), (6,3), (8,1), (8,2), (8,3), (10,1), (10,2), (10,3) and (10,4) respectively.

Implementation and Results: In our experiments we followed the feature extraction on subparts of the image. In order to the formation of prototypes, the number of prototypes depends upon the apperance characteristics of image instances in the datasets. We manually assumed different values of K prototypes for each dataset. For ETHZ dataset each image was scaled as 84×192 and images of GRID datasets were scaled into a fixed size of 112×272 and the VIPeR images were scaled to 48×128. For all experiments we fixed $k = 15$ for the k-NN classifiers. In our experiment, for each of the dataset 80% and 20% of image instances were considered as training and testing set respectively. In case of VIPeR dataset with $80\%(506)$ of the data used as gallery set images and 20% as probe set images while (400) samples for gallery and (100) samples for probe were used in case of GRID datasets. For ETHZ sequence-1, 80% images were considered as gallery and rest 20% images were probe set.

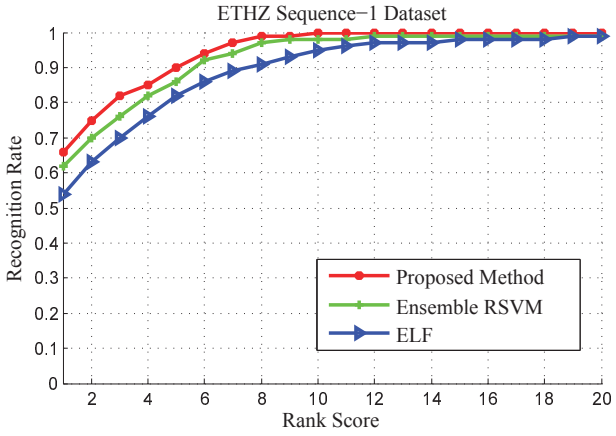

Fig. 6: CMC curves obtained on ETHZ Sequence 1

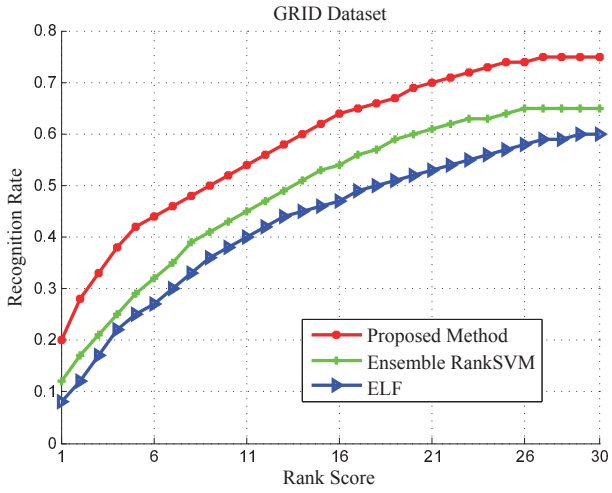

Fig. 7: CMC curves obtained on GRID Dataset

Our proposed method performs better for the top rank scores Figure 6, however after a few ranks curves are crossing.

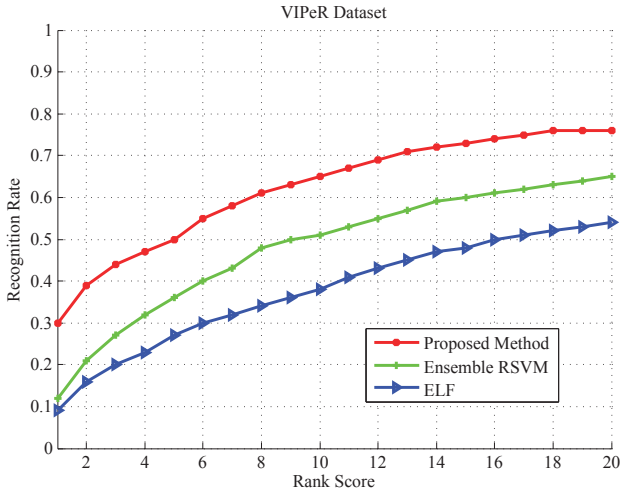

Fig. 8: CMC curves obtained on VIPeR Dataset

TABLE I: Comparison of recognition rate of the cluster based prototype and cluster ensemble based prototype over the datasets.

	ETHZ Sequence 1			VIPeR			GRID		
Ranks (r)	r=1	r=5	r=10	r=1	r=5	r=10	r=1	r=5	r=10
Cluster based Prototype	0.62	0.88	0.94	0.25	0.48	0.61	0.17	0.40	0.49
Cluster ensemble based Prototype	0.66	0.90	1.00	0.30	0.50	0.65	0.20	0.42	0.52

In spite of such challenging prospects as illumination alterations and occlusions, the ETHZ dataset is not challenging enough as it contains images from single cameras. Different camera settings, different color responses, different camera view points are the most intriguing issues for re-identification problem, which is not the case for ETHZ dataset. Hence, we have also evaluated our approach on images from more challenging GRID and VIPeR datasets. Figure7 shows the result of GRID dataset our methods performs better upto a certain rank scores. Results for VIPeR dataset in figure 8, it is worth noting that results are not very high because person images from the datasets are very challenging since they were captured from disjoint cameras views lead to large variations in both view angle and illumination. However our ensemble based proposed method outperforms the existing methods of [17], [4]. We have also demonstrated the importance of prototype formation. For each dataset prototypes are formed using both the cluster and cluster ensemble based approach and the recognition rates are computed.

The Table I illustrates the recognition rate for each data set with regard to the rank scores. It depicts the efficiency of cluster ensemble based prototypes over cluster based prototypes. From I it is observed that the ensemble based cluster improves the recognition results on the average of 4% for considered datasets. The outcomes show that both the prototypes formation approaches complement each other to make improvement on the recognition rate.

V. Conclusion

The proposed ensemble based framework for the person re-identification performs well under various challenging conditions. Formation of prototypes is able to describe individuals with similar appearance as well as improve the reliability and accuracy under crowded environment. The matching strategy of probe image with a certain group of images, where both shares the same prototype reduces the number of comparisons between gallery and probe images. The proposed approach shows a significant improvement over the existing techniques for re-identification.

References

[1] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person re-identification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, pp. 127–151, 2011.

[2] 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.

[3] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition and tracking," *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, vol. 3, 2007.

[4] B. Prosser, W. S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," *British Machine Vision Conference*, vol. 2, no. 5, 2010.

[5] W. S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, 2013.

[6] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person reidentification: What features are important ?" *Workshop on Computer Vision-ECCV*, 2012.

[7] R. Satta, G. Fumera, and F. Roli, "Fast person re-identification based on dissimilarity representations," *Pattern Recognition Letters*, vol. 33, no. 14, 2012.

[8] 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.

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

[10] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, no. 3, 1989.

[11] C. Schmid, "Constructing models for content-based image retrieval," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2001.

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

[13] A. Topchy, A. K. Jain, and W. Punch, "Clustering ensembles: Models of consensus and weak partitions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, 2005.

[14] A. Strehl and J. Ghosh, "Cluster ensembles-a knowledge reuse framework for combining multiple partitions," *The Journal of Machine Learning Research*, vol. 3, 2003.

[15] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998.

[16] G. A. Edgar, "Measure, topology, and fractal geometry," *Springer-Verlag*, 2008.

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

[18] W. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares," *22nd Brazilian Symposium on Computer Graphics and Image Processing*, 2009.

[19] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *IJCV*, vol. 90, no. 1, 2010.

[20] M. K. O. M. Hirzer, P. Roth and H. Bischof, "Relaxed pairwise learned metric for person reidentification," *Computer Vision-ECCV*, 2012.