

Person Re-identification Using Clustering Ensemble Prototypes

Aparajita Nanda and Pankaj K Sa

Department of Computer Science and Engineering
National Institute of Technology Rourkela, India

Abstract. This paper presents an appearance-based model to deal with the person re-identification problem. Usually in a crowded scene, it is observed that, the appearances of most people are similar with regard to the combination of attire. In such situation it is a difficult task to distinguish an individual from a group of alike looking individuals and yields an ambiguity in recognition for re-identification. The proper organization of the individuals based on the appearance characteristics leads to recognize the target individual by comparing with a particular group of similar looking individuals. To reconstruct a group of individual according to their appearance is a crucial task for person re-identification. In this work we focus on unsupervised based clustering ensemble approach for discovering prototypes where each prototype represents similar set of gallery image instances. The formation of each prototype depends upon the appearance characteristics of gallery instances. The estimation of k-NN classifier is employed to specify a prototype to a given probe image. The similarity measure computation is performed between the probe and a subset of gallery images, that shares the same prototype with the probe and thus reduces the number of comparisons. Re-identification performance on benchmark datasets are presented using cumulative matching characteristic (CMC) curves.

1 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 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 appearance based person re-identification concerns 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].



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.

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 are 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) [5] 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]. The prototype strategy for re-identification problem is introduced in [6] by defining a set of prototypes for each body part and representing them with dissimilarity vectors. A matching strategy is followed between the probe and gallery images with respect to the dissimilarity. However in case of crowded environment where the number individuals appearance increases, the prototypes representation of body parts for each individual is considered as a tedious process for re-identification. In addition the prototype representation and dissimilarity measures remains unsuccessful for individuals wearing same combination of attire. Hence a set of prototypes can be created where each prototype defines a set of image instances, that corresponds to local appearance characteristics shared by different individuals. Most of the existing prototype based approaches [7], [8] 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. This motivates the formation of a consensus based prototypes which can be incurred from the ensemble of prototypes representation. The set of consensus prototypes can also termed as clustering ensemble based prototypes that describes the promising representation of features shared by the gallery instances. We depict insights into the optimal prototypes formation of the set of untagged images with given feature sets. The formulation of optimal prototypes representation is assumed as set of consensus prototypes and considered as the best representation of features with respect to commonalities of feature importance of gallery images. The given probe image is classified to a prototype based on its appearance. The similarity measure is computed between the probe and a particular set of gallery images that shares the same prototype.

There exists 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, k-NN classifier assigns a label to each probe image. Similarity measure is computed with a subset of gallery images, that shares the same label with the given probe and hence the number of comparisons are reduced. The resulting scores are listed according to the most similar signature of instances ordered by increasing distance measure. Experimental evaluation of the cluster ensemble based approach is performed on two benchmark datasets.

The paper is structured as follows. In Section 2 problem is formulated. Section 3 describes the detail steps of re-identification using cluster ensemble based approach. Experimental evaluation is described in Section 4 followed by conclusion in Section 5.

2 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 $\{I^{pr}\}$, a prototype P_i is assigned and the matching scores are computed for gallery images $\{I_i^g\}_{i=1}^{n'}$ where $n' \subset n$, the subset instances and the probe shares the same prototype. The gallery and probe images are taken from two different cameras.

3 Cluster Ensemble Based Re-identification Approach

This section depicts the detail of cluster ensemble based re-identification approach which includes the 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.

3.1 Feature Representation

In case of feature representation, different set of components are extracted from sub parts of an image. Each of the subparts are likely to include sets of local features such as color, texture, interest points or visually discrete features which are able to distinguish an individual from others. This feature sets are considered as the signature of the individual. The principle behind this type of subparts representation is to gain robustness against partial occlusion, pose variations and to roughly captures the body parts. 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 $Feature(I_i^g) = \{y_1, \dots, y_d\}^T \in R^d$ 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 ($m \geq 1$).

$$\{I_i^g\} = \{I_{i,1}^g, \dots, I_{i,m}^g\} \quad (1)$$

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 [9] and 13 Schmid filters [10] are applied on the luminance channel. The feature vector of each gallery image is integrated to represent the feature space. There exists no single feature, that can be believed to work universally for all instances of images. So the combination of different types of features lead to more discriminant feature space 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 images 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 2 shows

the matching rate with respect to different types of color and texture features. From Figure 2 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.

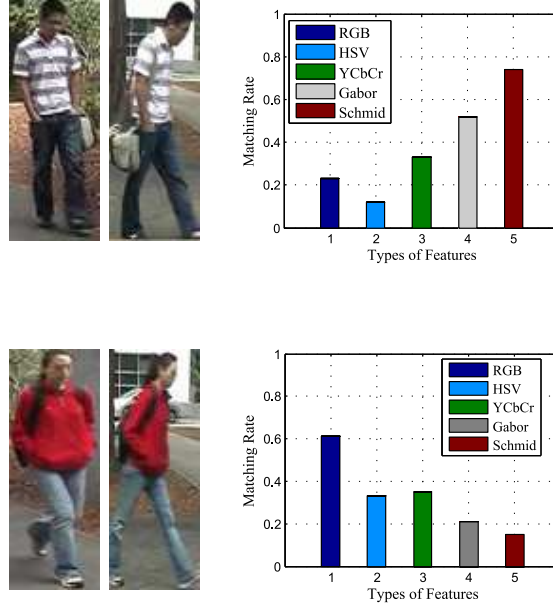


Fig. 2: Matching rate of probe and gallery image on the basis of different color and texture features. RGB, HSV and YCbCr color models are taken for color feature and Gabor and schmid filters are considered for texture feature.

3.2 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 set of untagged images with given features 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 3. The motivation for prototype formation signifies to distinguish the individuals with similar attire in a crowded environment. Generally a set of prototypes $\{P_i\}_{i=1}^K$, is assumed as low-dimensional manifold clusters [11] that group images $\{I_i^g\}_{i=1}^n$ with similar appearance based features. In order to formulate the prototypes we construct an ensemble of T prototypes labels with T different random initialization of K mean clustering algorithm on the

feature space Y . Each prototypes label λ^t where $\lambda^t = \bigcup_{a=1}^K P_a$ and $P_i \cap P_j = \phi$, is obtained from each random initialization of K mean algorithm where K defines the number of partitions of the input image samples $\{I_i^g\}_{i=1}^n$ with respect to their 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 [12]. The objective of the clustering ensemble task is to search for a combination of multiple prototypes label that provides improved overall prototypes of the given untagged gallery image. Cluster based similarity partitioning algorithm (CSPA) [13] is one of the cluster ensemble technique that can be used for prototypes formation.



Fig. 3: 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.

Cluster based Similarity Partitioning Clustering algorithm results a prototype label λ^t for the feature space and it signifies a relationship between sample images in the same cluster and can thus be used to establish a measure of pairwise similarity. In order to compute the ensemble prototype formation, for each prototype label λ^t , 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.

$$S_{ij} = \frac{1}{T} \sum_{t=1}^T I(\lambda_i^t, \lambda_j^t) \quad (2)$$

where λ_i^t represents the prototype to which i th sample belongs in prototype label λ^t .

$$I(\lambda_i^t, \lambda_j^t) = \begin{cases} 1 & (i, j) \in C_a(\lambda^t) \\ 0 & otherwise \end{cases} \quad (3)$$

The entry-wise average of T such matrices representing the T 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.

Algorithm 1: Formation of cluster ensemble based prototypes

Input : Ensemble of prototype labels $\{\lambda^t\}_{t=1}^T$ where $\lambda^t = \bigcup_{i=1}^K P_i$

begin

for $t = 1$ **to** T **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 $\{P_i\}_{i=1}^K$

3.3 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 image. The objective of using k-NN classifier is to assign each untagged probe image I_i^{Pr} to a prototype (class label). So for a given probe image I_i^{Pr} 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 computational overhead.

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 [14] with $\{Y\}$ as inputs and treating the associated prototype labels $\{P_i\}$ as classification outputs. For a given probe image $\{I^{Pr}\}$, we classify it using the learned k-NN classification strategy to obtain its prototype label (class label). Then similarity measure of probe image $\{I^{Pr}\}$ against gallery images $\{I_i^g\}$ of the corresponding prototypes are computed.

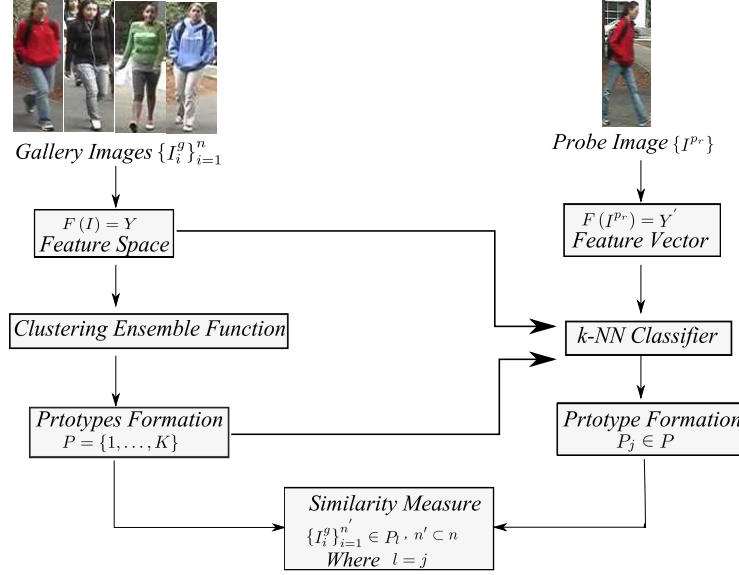


Fig. 4: Overview of cluster ensemble based approach for re-identification

3.4 Similarity Measure

Given a probe image I^{pr} 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 $D(\cdot, \cdot)$.

$$x^* = \arg \min_{I_i^g} D(I_i^g, I^{pr}) \quad (4)$$

where $D(I_i^g, I^{pr})$ is defined as a function of a similarity measure between sets of feature vectors for sequence of parts. It is the final matching score between a probe I^{pr} and gallery images I_i^g calculated as the average distance between the image partitions.

$$D(I_i^g, I^{pr}) = f(\text{dist}(I_{i,1}^g, I_1^{pr}), \dots, \text{dist}(I_{i,m}^g, I_m^{pr})) \quad (5)$$

The measure of $\text{dist}(\cdot, \cdot)$ is the Hausdorff distance d_H [15],[6]. 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 $\text{dist}()$ is defined as the similarity measure of m pairs of parts. For example, $D()$ can be determined as the additive combination of m distances. The $\text{dist}()$ is turned as the k th Hausdorff distance dist_H [16]. The distance dist_H is assumed as k th ranked distance which is the minimum distance between all pair of elements from two sets Q, S .

$$\text{dist}_H(Q, S) = \max\{h_k(Q, S), h_k(S, Q)\} \quad (6)$$

$$h_k(Q, S) = kth \min_{q \in Q, s \in S} (\|q - s\|) \quad (7)$$

$\|\cdot\|$ denotes the distance metric between the elements of the set. The choice of parameter k is useful for altering robustness to partial occlusions. The value of parameter depends upon the image sub partitions. In our case the images are partitioned into six sub parts and the mixture of color and texture based features are extracted from each part. The partition based distance measure strategy helps to attain robustness to pose variations and experimentally validated in [6]. 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. The similarity measure is computed between probe $\{I^{Pr}\}$ and $\{I^g\}$.

4 Experimental Evaluation

Datasets: Experiments were evaluated on two challenging datasets VIPeR [17], i-LIDS [18] used in former works. 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. The i-LIDS dataset contains 476 images of 119 different individuals, captured from disjoint cameras. For i-LIDS the experimental set up of [4] are followed where 208 image pairs from two different camera views are considered. 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.

Feature representation: Each image was partitioned into six horizontal strips of equal size. Similar to [4], [5], [19] 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.

Comparison Methods: Two different existing ensemble based approaches were demonstrated, that includes the state of art ensemble based Rank SVM [4] and Adaboost based ensemble localized feature [17]. All the three methods were compared using the same image feature set representation with same parameters. The experiments were carried on with 5 random trials and the results were presented by averaging over the trials. The experimented were evaluated

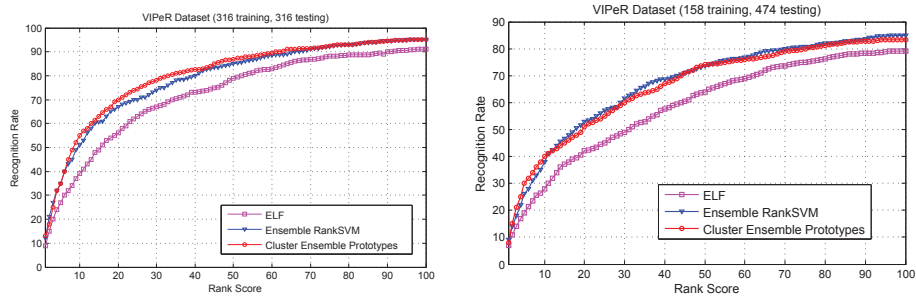


Fig. 5: CMC curves for VIPeR Dataset

for 50% and 75% of testing sets of VIPeR and i-LIDS datasets. The impacts of cluster ensemble based approach were evaluated on both the datasets with the above settings of training and testing. The affects of recognition rates on the formation of cluster ensemble based prototypes and cluster based prototypes were also demonstrated. Table 1 represents the comparison of the proposed approach to current approaches on VIPeR dataset and achieves a higher recognition rate in top ranks.

Implementation and Results: In our experiments we follow the same experimental set up of [4]. The number of prototypes depends upon the appearance of individuals in datasets. We manually assume different value of K prototypes for each dataset based on the appearances. For all experiments we fixed $k = 15$ for the k -NN classifiers. Figure 5 presents the CMC curves of VIPeR dataset with 50% and 25% of the data used as gallery set images while Figure 6 show the CMC curves for i-LIDS with 50% and 75% of testing sets respectively.

Ranks	r=1	10	20	50	100
Cluster Ensemble	13	53	67	86	95
ERSVM [4]	12	51	67	85	94
ELF [17]	12	43	60	81	93
PRDC [5]	16	54	70	87	97
LMNN [20]	18	59	75	91	97
MCM [8]	10	32	42	60	72
PS [21]	22	57	71	87	NA

Table 1: The comparison of recognition rates in % with existing approaches on the VIPeR dataset

Cluster ensemble based prototypes vs. ensemble rank SVM vs. ELF
 From figure 5 it is clearly seen that, by employing the cluster ensemble technique to person re-identification, helps to gain improvement over the ensemble

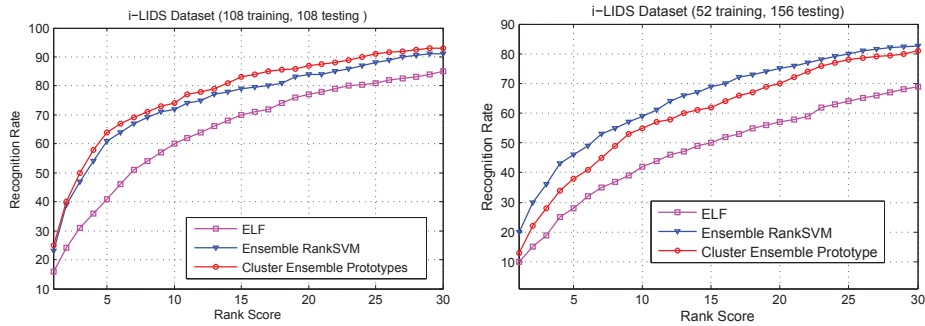


Fig. 6: CMC curves for i-LIDS Dataset

based rank SVM [4] and ensemble localize feature ELF [17]. The performances of distance based learning strategy [17] and the relative ranking scheme degrade, when the individuals with same combination of attire appear in the scene. i. e. images where two individuals wearing combination of white shirt, blue skirt and blue shirt, white skirt, it is difficult to measure the dissimilarity by distance learning or by relative ranking of distance measure between these two images. In contrast the cluster ensemble based approach forms the prototypes from the feature set that best represents the individuals according to their appearance of attire. In case of i-LIDS dataset with 50% of training and testing sets (Figure 6), the recognition rates of the cluster ensemble approach seems better than the existing [4], [17] approaches. However in Figure 5 and figure 6 for 25% training set, the overall recognition rate decreases due to the less number of training instances, because for a small training data, few number of samples are included for each prototype which affects the recognition rate.

Figure 7 depicts the efficiency of cluster ensemble based prototypes approach over cluster based prototypes. The experimentation were carried out on 50% and 25% of testing set for VIPeR and i-LIDS datasets. We followed the same feature set and similarity measure for cluster based prototypes. The CMC curves presented in figure 7 is the recognition rate of 5 random trials of K prototypes. For each random trial K-mean algorithm was employed for prototypes formation and the results were presented by averaging over the trials. From figure 7 it is observed that, for both of the datasets the cluster ensemble strategy always gives commendable results than the cluster based prototypes.

5 Conclusion

The proposed ensemble based framework for the person re-identification performs well under various challenging conditions. Formation of ensemble based prototypes are 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

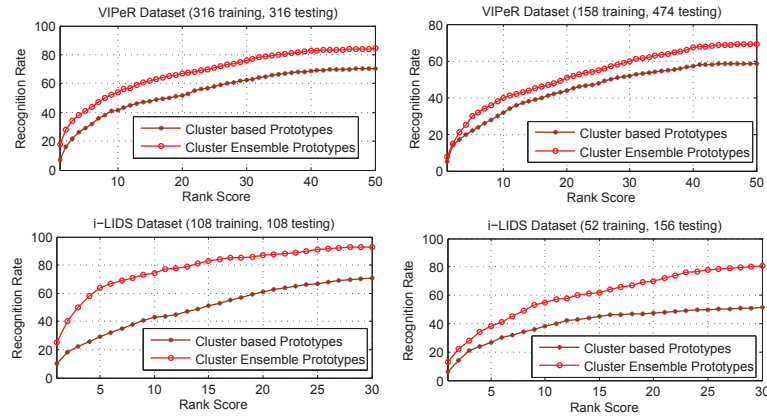


Fig. 7: Recognition rates for VIPeR and i-LIDS datasets by employing cluster based prototypes and cluster ensemble prototypes. The ensemble prototypes are more effective than normal cluster based prototypes. The results are presented by averaging over the five trials.

same prototype reduces the number of comparisons with gallery images. The proposed approach shows a significant improvement over the existing techniques for re-identification.

The ensemble based prototypes is to be worked out for gaining improvement over recognition rates. The prototype based re-identification will be tested over several other datasets for various challenging condition.

References

1. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: *Appearance-based person re-identification in camera networks: problem overview and current approaches*. Journal of Ambient Intelligence and Humanized Computing **2** (2011) 127–151
2. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: *Person re-identification by symmetry-driven accumulation of local features*. IEEE Conference on Computer Vision and Pattern Recognition (2010)
3. Gray, D., Brennan, S., Tao, H.: *Evaluating appearance models for recognition, reacquisition and tracking*. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance **3** (2007)
4. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: *Person re-identification by support vector ranking*. British Machine Vision Conference **2** (2010)
5. Zheng, W.S., Gong, S., Xiang, T.: *Re-identification by relative distance comparison*. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013)
6. Satta, R., Fumera, G., Roli, F.: *Fast person re-identification based on dissimilarity representations*. Pattern Recognition Letters **33** (2012)

7. Satta, R., Fumera, G., Roli, F., Cristani, M., Murino, V.: *A multiple component matching framework for person re-identification*. Image Analysis and Processing-ICIAP (2011)
8. Satta, R., Fumera, G., Roli, F.: *Exploiting dissimilarity representations for person re-identification*. Similarity-Based Pattern Recognition (SIMBAD) (2011)
9. Fogel, I., Sagi, D.: *Gabor filters as texture discriminator*. Biological Cybernetics **61** (1989)
10. Schmid, C.: *Constructing models for content-based image retrieval*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2** (2001)
11. Loy, C.C., Liu, C., Gong, S.: *Person re-identification by manifold ranking*. IEEE International Conference on Image Processing **20** (2013)
12. Topchy, A., Jain, A.K., Punch, W.: *Clustering ensembles: Models of consensus and weak partitions*. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005)
13. Strehl, A., Ghosh, J.: *Cluster ensembles-a knowledge reuse framework for combining multiple partitions*. The Journal of Machine Learning Research **3** (2003)
14. Joachims, T.: *Text categorization with support vector machines: Learning with many relevant features*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1998)
15. Edgar, G.A.: *Measure, topology, and fractal geometry*. Springer-Verlag (2008)
16. Wang, J., Zucker, J.D.: *Solving multiple-instance problem: A lazy learning approach*. (2000)
17. Gray, D., Tao, H.: *Viewpoint invariant pedestrian recognition with an ensemble of localized features*. Computer Vision-ECCV (2008)
18. Zheng, W.S., Gong, S., Xiang, T.: *Associating groups of people*. British Machine Vision Conference (2009)
19. M. Hirzer, P. Roth, M.K.O., Bischof, H.: *Relaxed pairwise learned metric for person reidentification*. Computer Vision-ECCV (2012)
20. Weinberger, K.Q., Saul, L.K.: *Distance metric learning for large margin nearest neighbor classification*. The Journal of Machine Learning Research **10** (2009) 207–244
21. Martin, H., M, R.P., Martin, K., Horst, B.: *Relaxed pairwise learned metric for person re-identification*. Computer Vision-ECCV (2012) 780–793