Hand-Crafted System for Person Re-Identification:A Comprehensive Review

Salwa Baabou[1,3], François Bremond[2] , Awatef Ben Fradj[3], Mohamed Amine Farah[3] and Abdennaceur Kachouri[3]

[1]National Engineering School of Gabes,Tunisia
baabousalwa@gmail.com
[2]INRIA Sophia Antipolis-Mediterranee, France
francois.bremond@inria.fr
[3]National Engineering School of Sfax, Electronics and Information Technology Laboratory (LETI),Sfax, B.P.W.3038 , Tunisia
benfradj_awatef@yahoo.fr, med.farah@yahoo.fr, abdennaceur.kachouri@enis.rnu.tn

***Abstract*— In video surveillance, Person Re-Identification(Re-ID) consists in recognizing an individual who has already been observed (hence the term Re-Identification) over a network of cameras. Usually, the person Re-Id system is divided into two stages: i)constructing a person's appearance signature by extracting feature representations which should be robust against pose variations, illumination changes and occlusions and ii)Establishing the correspondence/matching between feature representations of probe and gallery by learning similarity metrics or ranking functions. A gallery is a dataset composed of images of people with known IDs whereas a probe is collected of detected persons with unknown IDs from different cameras. Specifically, the process of person Re-Identification aims essentially at matching individuals across non-overlapping cameras at different instants and locations. However, the matching is challenging due to disparities of human bodies and visual ambiguities across different cameras. This paper provides an overview of hand-crafted system for person Re-identification, including features extraction and metric learning as well as their advantages and drawbacks. The performance of some state-of-the-art person Re-ID methods on the commonly used benchmark datasets is compared and analyzed. It also provides a starting point for researchers who want to conduct novel investigations on this challenging topic.**

Keywords*— *Person Re-Identification, hand-Crafted system

I. INTRODUCTION

Person Re-Identification aims to match individuals appearing across non-overlapping camera networks. This task has attracted more interest over the recent few years. The goal is to return a list of probabilistic matched images ranked by degree of similarity.

In the literature, to tackle the Person Re-Identification, two major directions are taken into consideration which aim to: Extract robust and invariant feature representations for both probe and gallery images and then learn specialized distance metrics to person matching using this representation. This Kind of approaches is called *Appearance-based*.

Appearance-based methods can be divided into two categories. For the first category, color and texture based features are widely used [1,10]. However, these features representations are sensitive to pose and illumination change, which may result in larger intra-person variation (difference between features of the same person) than inter-person variation (difference between features of different persons). Besides, in order to improve the recognition accuracy, low level image features and attribute or shape information have been applied in conjunction with color or texture features. For the second group of methods [2,3], to guarantee more reliable matching, feature transformations or distance metrics are learned. Support Vector Machine (SVM) with ranking [5] and transfer learning [3] have also been proposed to obtain better matching correspondence. The distance metric learning based methods have been proved to be effective in matching person images [6]. However, the drawback of these methods is that the learned model tends to overfit the training data.

This paper is organized as follows. An overview of a person Re-Identification system is described in section 2: Re-ID diagram, the main challenges, principle methods and a description of the hand-crafted system for person Re-ID. The experimental results and performances of the state-of-the-art Re-ID methods which were evaluated on iLIDS, ViPeR, CAVIAR and 3DPeS datasets are shown in section 3. Section 4 concludes the paper.

II. PERSON RE-IDENTIFICATION: AN OVERVIEW

A. System Design

Fig.1 shows the diagram of person Re-Identification system. It starts with automatic person detection. In recent years, most of the existing person re-identification works have ignored this step and assume perfect pedestrian detection. However, perfect detection is impossible in real scenarios and misalignment can seriously affect the person Re-ID performance. Therefore this factor should be carefully studied in future works.

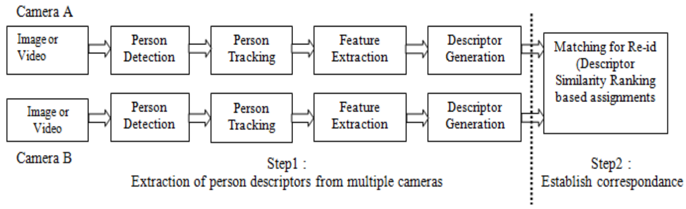

Fig. 1. Re-ID system Diagram.

In order to build a strong visual signature of people appearances, people have to be accurately detected and tracked, so the step of person tracking should be also taken into consideration. However, person detection and multiple person tracking are difficult problems with their own hurdles. Significant amount of work has gone into the problem of person detection over the years as well as Multiple Object Tracking (MOT) within a single camera's FOV which has also been widely researched, but sustained tracking under varying observation environments remains an open problem.

For feature extraction and descriptor generation, the most commonly used features are color, shape, position, texture, and soft-biometry. The adopted feature is determined by different factors. On one side, the signature should be unique and discriminative enough which can lead to the selection of biometry or soft-biometry features. On the other side, camera resolution, computational load and other implementation issues can prevent or limit their usage and more generic features are required.

B. *Main challenges of Person Re-ID*

The main problem in Re-ID resides in the variation in a person's appearance across different cameras.

A typical Re-ID system may have an image (single shot) or a video (multi-shot) as input for feature extraction and signature generation. Thus, the first step in Re-ID is to learn a person's visual signature or model and then compare the two models to get either a match or a non-match.

Extracting a reliable signature depends on the availability of good observations. Besides, faulty trajectory estimation and incorrect detections introduce errors in signature generation and extraction that affect the Re-ID quality. The most obvious and simplest signature of a person is characterized by features like color, texture and shape. However, these features are hardly unique, not descriptive enough and prone to variations.

Color/texture varies due to cross view illumination variations, pose variations, view angle or scale changes in multi-camera settings. To solve this problem, the equalization between cameras is needed. Different camera geometries also make shape descriptors less discriminative.

A subject may be fully or partially occluded by other subjects or carrying items that lead to errors in matching between tracklets. Furthermore, some works in person Re-ID used body-parts methods (such as SDALF, MPMC…) to solve the issue of signature alignment but this problem is still difficult and not efficient as these methods require real detections and many annotations.

We can also cite the low image quality as another problem in person Re-ID where the captured image of a person may suffer from low resolution, noise or blur due to limited imaging quality of surveillance cameras.

All these issues may affect the performance of person Re-ID which is still not robust enough to guarantee high accuracy in practice.

C. *Principle Methods of Person Re-ID*

In general, the existing methods predominating the person Re-ID area can be classified into two major categories: Single-shot and Multi-shot person Re-Identification. The first category only analyzes a single image for each subject assuming no tracking information is available. The second one assumes multiple images are available for each person through tracking. These existing methods are divided into two groups: unsupervised and supervised approaches. Unsupervised methods mainly focus on feature design and feature extraction and do not require manually labeling training samples. However, Supervised methods generally require the assistance of manually labeled training samples which lead to better performance. Most existing works [1, 3, 5, 7, 9] choose training and test samples from the same camera views. Only very recently, people started to study the cases when training and testing samples are from different camera views [10].

Unsupervised methods include: Symmetry-Driven Accumulation of Local Features (SDALF) [1], Biologically inspired features and Covariance descriptors (BiCov) [9], Local Descriptors encoded by Fisher Vectors combined with other features (eLDFV) [9],etc.

Among Supervised methods, we can cite: Ensemble of Localized Features learned with AdaBoost (ELF) [7], distance metric learning for Large Margin Nearest Neighbor classification (LMNN) [2], Information Theoretic Metric Learning (ITML) [19], Pairwise Constrained Component Analysis (PCCA) [11], Large Margin Nearest Neighbor with Rejection (LMNN-R) [3], supervised Local Descriptors encoded by Fisher Vectors (sLDFV) [8],etc.

III. HAND-CRAFTED SYSTEM FOR PERSON RE-ID

A. *Features Extraction and descriptor Generation*

State-of-the-art descriptors have been reviewed from two different viewpoints, namely the kind of body model and the kind of features used to represent a person.

Each body part (or the whole image of the individual, if no body part subdivision model is used) is described using one or more different global, local or patch-based features.

That's why, appearance-based methods rely mainly on designing discriminative features such as viewpoint invariance features, low-dimensional discriminative features[7], combination of local and global features, accumulation of multiple features [1], bio-inspired features, Fisher vector encoded features and attribute-based features [9]. To select the most descriptive features for person Re-ID, the work in [7] used Adaboost to learn effective representations from a set of local features. In [1] the authors showed a strategy based on the localization of perceptual relevant human parts, driven by asymmetry/ symmetry principles to suggest a method called Symmetry Driven Accumulation of Local Features (SDALF), which is robust to background clutter. A model in a covariance metric space is proposed in [12] in order to extract features from different regions of each person which should be matched specifically. The work in [13] presented a discriminative signature from multiple local features and designed a distance measure by exploiting different body parts. A Local Maximal Occurrence (LOMO) feature is proposed[14], in which the horizontal occurrence of local features is used to achieve a constant representation against view changes. Attribute-based features are more robust to image translations comparing with low-level color and texture features. Recently, the authors in[16] gathered a richly annotated dataset for pedestrian attribute.

B. Distance Metric learning

Many distance metric learning and matching process have been proposed for person Re-Identification. These methods aim at learning the best metric between appearance features of the same pedestrian across camera pairs.

Support Vector Machine and boosting [7] have been widely used, which cast the problem into two-class or multi-class classification problem. However, in [5] the person Re-ID scheme is considered as a ranking problem which trained a primal RankSVM ranker and tried to find a linear function to weigh the absolute difference between samples. In [2] the Large-Margin Nearest Neighbor metric (LMNN) is proposed which put a limited area (perimeter) for the matched pairs of subjects whilst punishes the unmatched ones, LMNN is time consuming and suffer from the problem of overfitting. In [3] in order to improve on previous results, a reject option for unfamiliar matches, as an LMNN variant, named LMNN-R was introduced. To achieve encouraging re-identification performance, several distance metric learning methods have also been proposed, such as Relative Distance Comparison (RDC), Pairwise Constrained Component Analysis (PCCA)[11] and Local Fisher Discriminant Analysis (LFDA) [4]. However, these methods are prone to overfitting problems especially in large scale and high dimensional learning scenarios. Information Theoretic Metric Learning (ITML) [19] and Logistic Discriminant Metric Learning (LDML) [6] have as well been applied. A strategy called Keep It Simple and Straightforward Metric (KISSME) has been proposed to learn a distance metric from equivalence constraints. This method is based on the class of Mahalanobis distance functions which generalizes Euclidean distance. KISSME does not rely on complex optimization and computationally expensive iterations. However, because of the intra-class and inter-class variation, the Mahalanobis metric is more suitable for person Re-ID problem. An effective method called Geometric Preserving Local Fisher Discriminant Analysis (GeoPLFDA) was proposed in [19].The method combines LFDA with geometric preserving method which uses a nearest neighbor graph to approximate local manifold. LFDA [4] provides descriptive information by separating differently labeled samples and gathering similarly labeled ones together. Taking the benefit of the complementarily between them, the proposed method achieves considerable advance over state-of-the-art approaches.

Most previous distance metrics are learned by supervised approaches and they are not practical in real-world applications in which the data comes in without any manully labeling efforts. Furthermore, a drawback of the distance metric learning methods is the overfitting problem caused by the small sample size (SSS) problem in person Re-ID where the number of samples per subject is fewer than the dimension of the feature.

There are also some other works in the literature using neural network models to address the person Re-ID problem[18]. A Set Label Model was presented in [11] which applies Neighborhood Component Analysis (NCA) and Deep Belief Network (DBN) on the features of the query and gallery images to improve person Re-Identification performance.

IV. PUBLIC DATASETS AND EVALUATION METRICS

A. Benchmark Datasets

Several frequently used datasets have been adopted for evaluation of Re-ID approaches, namely iLIDS,VIPeR, ETHZ , CAVIAR and 3DPeS. Additional datasets should be mentioned in this context like V-47 ,GRID ,Chokepoint,Terrascope, Person RE-ID(PRID), SAIVT, CUHK02 and Sarc3D. The details of the most commonly used datasets are given in TABLE I.However, the estimation of long period Re-ID requires data to be collected over several days using the same or different set of cameras. None of the available datasets offers such instances of people collected on different days. A recent RGB-D person Re-ID dataset captures depth information that can be used for the evaluation of depth based features for Re-ID.As RGB-D datasets we can cite the BIWI RGBD-ID Dataset and the IAS-Lab RGBD-ID Dataset.

For More details and descriptions of these datasets, we refer to survey[6].

TABLE I. SIZE OF SOME COMMONLY USED DATASETS.

Datasets	#images	#subjects	#cameras	#view	label
iLIDS	479	119	8	2	hand
ViPeR	1264	632	2	2	hand
CAVIAR	1220	72	2	2	hand
3DPeS	1011	192	8	3	hand

B. Evaluation Metrics

The metric known as the Cumulative Matching Characteristic (CMC) curve is the most widely used evaluation metric for performance of person Re-ID.

Since person Re-ID is considered as a ranking problem, this metric is adopted where each element in the gallery is ranked based on its comparison to the probe. The probability that the correct match is ranked the same as or less than a particular value is plotted against the size of the gallery set .The Synthetic Re-ID Rate (SRR) curve is derived from the CMC curve in order to evaluate the performance of the simultaneously matching multiple probe images of the gallery. It gives the probability that any of the given fixed number of matches is correct. The normalized Area Under the CMC curve (nAUC) and Rank 1 recognition rate is also an important performance metric. The nAUC is the probability that the Re-ID system will produce a true match over a false (incorrect) match. However, these metrics are inadequate for evaluating the ability of the system to determine if a probe ID exists in the gallery or not (novelty detection).

C. Performance on the iLIDS dataset

The iLIDS dataset contains 476 images (128x64 pixels) of 119 pedestrians taken from 2 non-overlapping cameras and captured at a busy airport arrival hall. We randomly select the images of 30 and 50 people to form the test set and the rest of images are used for training. Fig.2 proves the CMC results of LMNN, ITML,KISSME, PCCA, LFDA and GeoPLFDA when the rank r is set to 1,5,10 and 20.

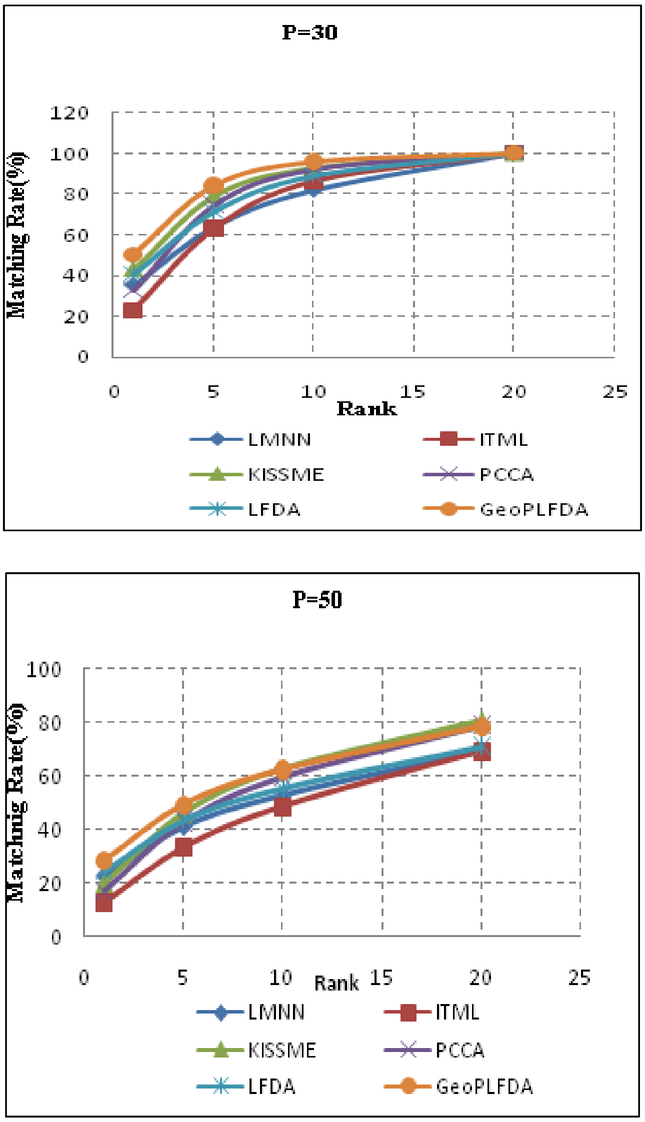

Fig. 2. CMC RESULTS ON ILIDS DATASET. P IS THE SIZE OF THE GALLERY SET.(P=30 AND P=50)

D. Performance on the ViPeR dataset

The ViPeR dataset contains 632 people taken outdoor with 2 images (128x48 pixels) for each pedestrian. The images of 316 and 532 people are selected for the testing set and the rest for the training set. Fig.3 shows the top ranked matching rate(%)on ViPeR dataset when the rank r is set to 1,5,10 and 20.

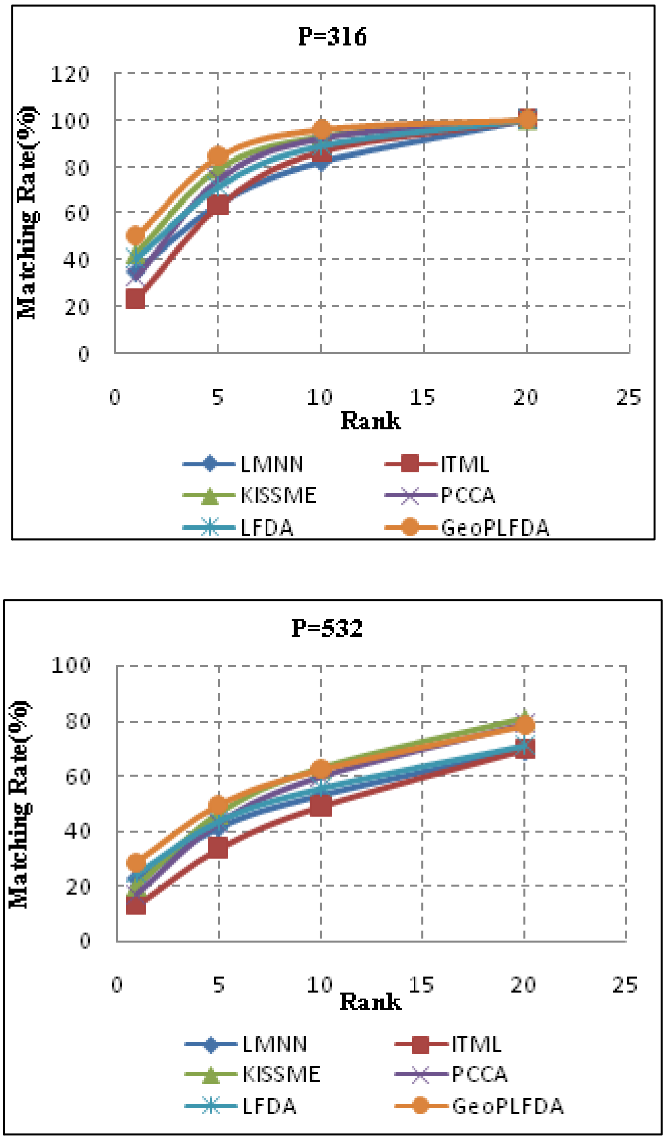

Fig. 3. CMC RESULTS ON VIPER DATASET.(P=316 AND P=532)

E. Performance on the CAVIAR dataset

The dataset contains 1220 images with 10 to 20 images for each person. The minimum and maximum sizes of the images are 17x39 and 72x144, respectively. We selected 18 and 54 people for the test and the remaining images were used for training. Fig.4 indicates the CMC results when the ranks are 1,5,10 and 20.

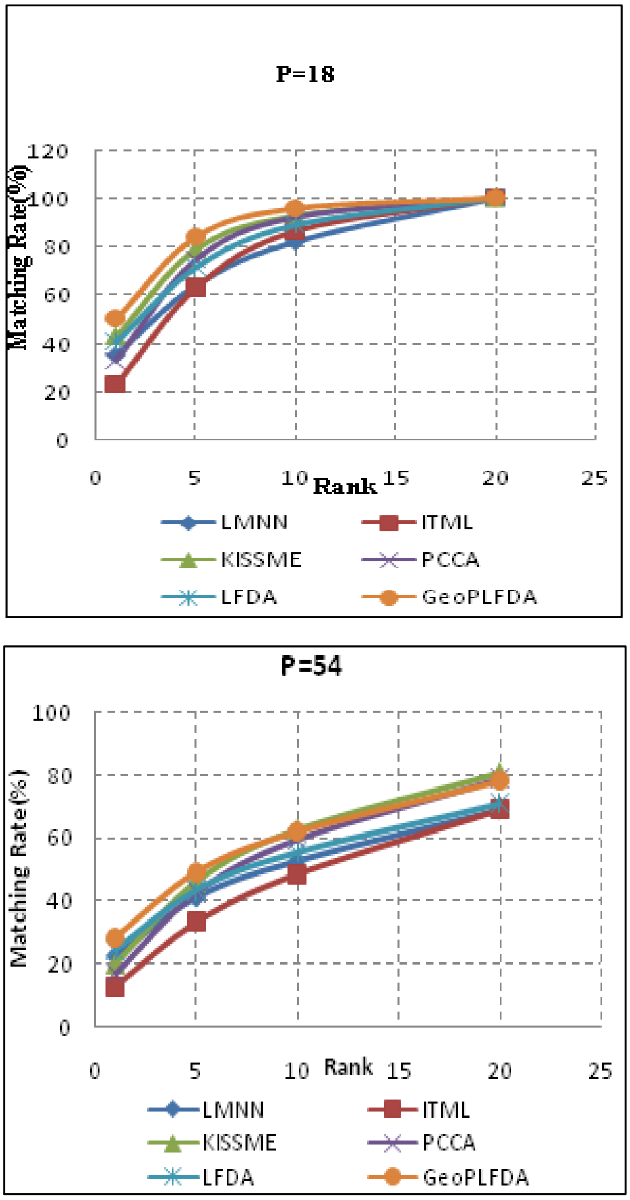

Fig. 4. CMC RESULTS ON CAVIAR DATASET.(P=18 AND P=54)

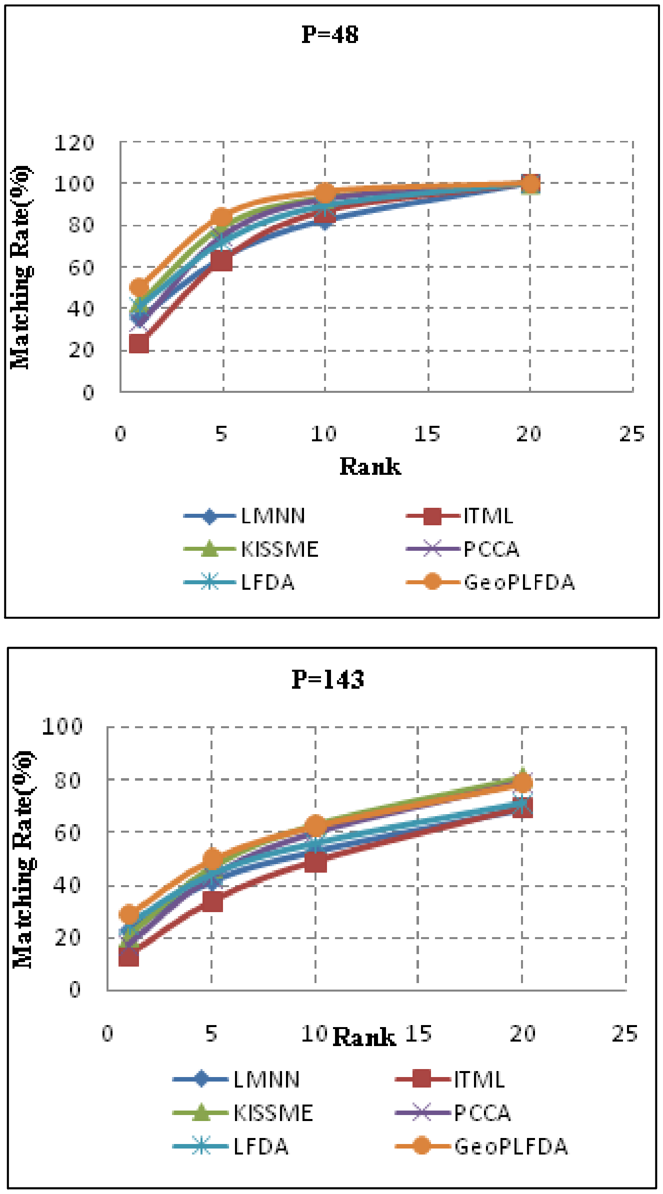

Fig. 5. CMC RESULTS ON 3DPES DATASET.(P=48 AND P=143)

F. Performance on the 3DPeS dataset

The 3DPeS dataset contains 1011 images of 192 persons (each person has 2 to 26 images) captured from 8 outdoor different cameras viewpoints. We selected 48 and 143 people for the test and the rest of images were used for training. Fig.5 shows the CMC results when the ranks are 1,5,10 and 20.

G. Discussion

From the above experiments and among the comparing methods, LMNN, ITML, KISSME and PCCA, we notice that these methods necessitate an iterative optimization scheme which is very expensive on both memory and computation. GeoPLFDA take over from the merits of LFDA and can efficiently find the optimal solution without any iteration, that's why we observe that GeoPLFDA provide the best rank 1 matching rate and high performance compared with other methods .On ViPeR dataset, the results showed 27% and 12,8% when p is set to 316 and 532 respectively. For the 3DPeS dataset, GeoPLFDA attained also the best rank 1 matching rate (50.3% and 32.6% when p is set to 18 and 54, respectively) which proves that this method can measure the distance between probe and gallery images effectively. We can see then that LFDA with geometric preserving projection is more efficient than performing it alone.

Finally, on all the four datasets, when the value of p increases, less data are available for training thus learning becomes more difficult. Therefore the matching becomes harder due to large test set. As a result, the matching rate decreases with the increase of p.

V. CONCLUSION

This paper provided a survey of current approaches for constructing appearance descriptors by extracting feature representations from person detections which are discriminative and robust against illumination change, occlusion and pose variations.

State-of-the-art appearance-based methods as well as matching methods using distance metric learning or ranking classifiers for person Re-Identification were reviewed. We tried to provide a comprehensive review and we also highlight some important issues of person Re-ID that may attract the attention of researches in the future.

In summary, as the process of person Re-Identification is significantly a challenging field due to high intra-class variation and inter-class similarity with vast opportunities for improvements and research, we can say that the integration of discriminative and robust feature learning, detector/tracking optimization, and the emergence of large-scale datasets and deep learning systems will lead to a successful person re-identification system. We hope that this paper will be useful as reference for anyone willing to work on this interesting topic.

REFERENCES

[1] Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. “Person re-identification by symmetry-driven accumulation of local features”. In Computer Vision and Pattern Recognition (CVPR), pp. 2360-2367, June 2010.

[2] Weinberger, Kilian Q., and Lawrence K. Saul. "Distance metric learning for large margin nearest neighbor classification." Journal of Machine Learning Research ,207-244, 10.Feb 2009.

[3] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, “Pedestrian recognition with a learned metric,” in Proc. Asian Conf. Comput. Vis. (ACCV), pp. 501–512, 2011.

[4] Pedagadi, S., Orwell, J., Velastin, S., & Boghossian, B. “Local fisher discriminant analysis for pedestrian re-identification”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,pp. 3318-3325, 2013.

[5] B. Prosser , W.-S. Zheng , S. Gong , T. Xiang , “Person re-identification by support vector ranking”, in: Proceedings of the 2010 British Machine Vision Conference (BMVC), pp. 21.1–21.11, 2010.

[6] A. Bedagkar-Gala and S. K. Shah, “A survey of approaches and trends in person re-identification,” Image and Vision Computing,vol. 32, no. 4, pp. 270–286, 2014.

[7] D. Gray , H. Tao , “Viewpoint invariant pedestrian recognition with an ensemble of localized features”, in: Proceedings of the 2008 European Conference on Computer Vision (ECCV),pp.262–275,2008.

[8] B.Ma, Y.Su, F.Jurie, “Local descriptors encoded by Fisher vectors for person re- identification”, in International Workshop Conjunction with European Conference on Computer Vision, pp.413–422, 2012.

[9] Munaro, M., Fossati, A., Basso, A., Menegatti, E., & Van Gool, L. “One-shot person re-identification with a consumer depth camera”. In Person Re-Identification,pp. 161-181,Springer London,2014.

[10] Liu, C., Gong, S., Loy, C. C., & Lin, X. “Person re-identification: What features are important?”. In European Conference on Computer Vision, pp. 391-401, Springer Berlin Heidelberg, October 2012.

[11] Mignon, A., & Jurie, F. “PCCA: A new approach for distance learning from sparse pairwise constraints”. In Computer Vision and Pattern Recognition (CVPR),IEEE Conference,pp.2666-2672,2012.

[12] S. Bak , G. Charpiat , E. Corvee , F. Bremond , M. Thonnat , “Learning to match appearances by correlations in a covariance metric space”, in: Proceedings of the 2012 European Conference on Computer Vision (ECCV), pp. 806–820, 2012.

[13] N. Martinel , C. Micheloni , “Re-identify people in wide area camera network”, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 31–36, 2012.

[14] S. Liao , Y. Hu , X. Zhu , S.Z. Li , “Person re-identification by local maximal occurrence representation and metric learning”, in: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition,pp. 2197–2206, 2015.

[15] Jia, Jieru, Qiuqi Ruan, and Yi Jin. "Geometric Preserving Local Fisher Discriminant Analysis for person re-identification." Neurocomputing 205 , vol. 205, p. 92-105, 2016.

[16] D. Li, Z. Zhang, X. Chen, H. Ling, and K. Huang, “A richly annotated dataset for pedestrian attribute recognition,”arXiv:1603.07054,2016.