# Different Techniques for Person Re-Identification: A Review

**O. S. Reshma[1] and Reshma Sheik[2]**
[1&2]Department of Computer Science and Engineering, TKMCE, Kollam, Kerala, India
E-Mail: osreshma@yahoo.com, rezmasheik@gmail.com

*Abstract-* In investigation, person re-identification is also a tough task of matching persons determined from utterly completely different camera views. It is necessary applications in AI, threat detection, human trailing and activity analysis. Person re-identification is also a tough analysis topic as a result of partial occlusions, low resolution images and massive illumination changes. Also, person determined from utterly completely different camera views has very important variations on poses and viewpoints. This paper summarises the challenges related to the person re-identification jointly discuss varied techniques utilized person re-identification.

*Keywords:* **Gaussian Mixture Model (GMM), Video-Surveillance, Histogram of Oriented Gradients (HOG), Bidirectional Ranking**

## I. INTRODUCTION

In recent years, closed-circuit television plays a vital role publically, and has become a lot of necessary in crime investigation. During this state of affairs, investigators wish to find and track the person in vicinity lined by multiple cameras. Here, manual browsing is time overwhelming and is not economical for the crime investigation. To unravel this downside, person re-identification that is matching persons across totally different camera views, attracts a lot of and a lot of analysis interests.

In video surveillance system, a sequence of video frames is obtained from their source mainly from closed circuit television (CCTV) and processing of these frames help to extract relevant information. Surveillance in public places is mainly for monitoring various locations and people in that locations and observing their behaviours. Nowadays, events like terrorist attacks have occurred more frequently in different public places. So, there is a growing need for video network systems to guarantee the safety of people in those areas.

In addition, intelligent surveillance has proven to be a useful tool for detecting and preventing potentially violent situations in public transport such as airports, train stations or even inside trains and air planes. The growth of computational capabilities in intelligent systems provided more opportunities in video surveillance system. This includes segmentation, object detection, tracking, classification etc. Person re-identification is one of the most important topic in this area.

In 1961, Alvin Plantinga [1] provided one of the first definitions of re-identification: "To re-identify a particular, then, is to identify it as (numerically) the same particular as one encountered on a previous occasion".

According to Paul McFedries [2], re-identification is the process of matching unknown huge amount of data with the individuals who provided the data. In generally, person re-identification can be defined as the process of recognizing individuals over different camera views in various locations under the condition of large illumination variations. In other words, it is the process of finding a person of interest from large number of images and video frames that track the persons in a network of cameras. The important applications of person re-identification are in video surveillance such as human tracking, human retrieval, and activity analysis. Searching a person from a large number of images and videos are time consuming. In this situations person re-identification technique saves a lot of human efforts. At the same time, it is a challenging research topic in computer vision due to large illumination variations, low resolution images, pose variation, background noises and occlusions.

When a person stays within a single camera view, the system has knowledge about his location, position, background and lighting condition. When the person moves from one camera view to another camera view the important question is, how does the system know that person observed from the camera was same as the person observed in another camera. This issue is called re-identification problem. It is the technique of recognising persons separated in time and location. Person re-identification is a complex problem due to the lack of spatial continuity for the information received from different camera observations.

Traditionally, person re-identification can be considered as a matching problem. There is a gallery set and probe set. Gallery set contains huge amount of candidate person images and probe set contains query person images. For each test image or group of test images of an unknown person, the goal of person re-identification is to return a ranked list of individuals from the gallery set as shown in figure 1. It is a key point of many applications, at the same time a most challenging research topic in Pattern Recognition.
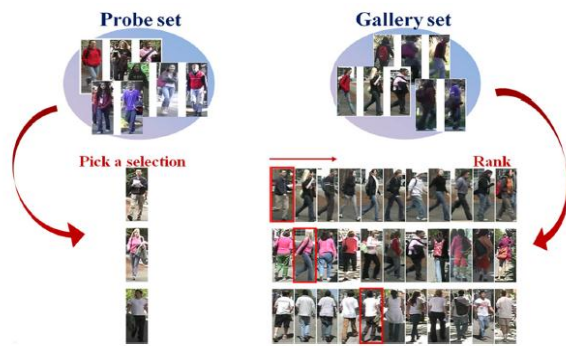
Fig. 1 Person Re-identification

## II. TECHNIQUES USED FOR PERSON RE-IDENTIFICATION

In the past, to resolve person re-identification, single-view and multi-view strategies, as well as single-shot and multi-shot strategies are projected. In single-view strategies information are extracted from a single pose of the person to create an appearance model, while in multi-view strategies an individual's appearance model is created based on information captured from multiple poses. In the case of single-shot strategies information from a single frame is used to create an appearance model, while multi-shot strategies when a person's appearance model is created based on information from multiple frames. Single-shot and multi-shot classification depends on the number of images available for a person. In single-shot case only one individual is available in the probe set and gallery set, while in multi-shot case multiple individuals are available in probe set and gallery set. Single-shot case is harder than multi-shot case because of the insufficient information. Whether the case is single-shot or multi-shot, person re-identification assumes that individuals wear the same cloths under the multi-camera surveillance network.

Doretto *et al.* [3] have studied the person re-identification problems and have proposed different methods to solve it. It discussed about the limitation of holistic appearance models based on histograms. In which two persons may have same histogram based signature even though they are dressed differently. This will happen, if they have same quantity of body surface covered with the same colors, in spite of the distribution of those colors. To overcome this drawback, they have proposed some methods which use appearance context modelling and parts-based modelling. Here, part-based strategies are useful when we want to create a unique signature for a person in a group.

Garcia *et al.* [4] have captured information from multiple poses of a person in order to create their appearance model. So, this is a multi-view and multi-shot case. To find the orientation of the person's trajectory with respect to camera, they use tracked or known tracklets. During tracking, the orientation and feature vector of a person is stored. After that, calculating distance between feature vectors from different orientations in the unknown tracklet and stored feature vectors from known tracklets. If the distances

between the tracklets are small, there is a large probability that they belongs to same individual.

Kuo *et al.* [5] have proposed a single-view and multi-shot approach. They have used multiple frames but not multiple poses of a person. Here, tracklets in the same frames are not belonging to the same person. Hence, features from these tracklets are taken as negative samples. At the same time; features from two different responses of the same track are considered as positive samples. Adaboost, which carries out a binary classification on the features from these positive and negative samples, then given two tracklets, that determines if they belong to the same person or not.

Farenzena *et al.* [6] have proposed a parts-based approach. They have used a STEL generative model for pedestrian segmentation. An asymmetry human partition is used to divide each person into head, torso and legs and symmetry partition used to separate these three parts into left and right subparts each. Finally, features are extracted by using a weighted HSV histograms and high structured patches. Here, result matching distance is the combination of the distance computed on each features. They have also mentioned three ways through which they can perform feature matching. First are single-shot vs single-shot, that is each image in the gallery and probe represent the different individual. Second is multiple-shot vs single-shot, in which each image in the probe set represent a different individual at the same time each individual in the gallery is represented by multiple images. Third are multiple-shot vs multiple-shot, here signatures from multiple images are included in both probe and gallery set.

Liu *et al.* [7] have examined different set of features to determine what features are important to person re-identification. They have used an unsupervised approach for learning a bottom-up feature importance. So, here features captured from different persons are weighted adaptively. They conclude that better is to selectively allocate some weights to features that are specific to certain appearance attributes of an individual. For example, texture will be more important feature for an individual wearing a textured shirt, at the same time it will be irrelevant for an individual wearing plain, textureless shirt.

Zhao *et al.* [8] have used super pixels to segment person's image. After segmentation C-SIFT local features are extracted as visual words then build a TF-IDF vocabulary index tree to speed up people search. They proposed an image retrieval technique for person re-identification. Here, vocabulary trees are building using training images.

Kviatkovsky *et al.* [9] have used an invariant color signature for person re-identification. In which, observations from upper and lower parts are marked with red or blue markers then plot them onto a log chromaticity histogram. From which, they found that for each person, two main modes are generated. Here distribution structure for the same person is preserved, but distribution structure for the different person is different.

Person re-identification is again classified into supervised, unsupervised and methods other than supervised and unsupervised. Unsupervised methods focus on feature extraction and feature design. They do not require any training samples. Supervised methods require training samples and it focus on learning feature representation.

*A. Unsupervised Methods*

Unsupervised methods focus on feature design of given images. Chang *et al.* [10] have proposed a single-shot person re-identification system. The proposed algorithm consists of four stages; pedestrian segmentation, human region partitioning, feature extraction and human feature matching. They introduced an improved random walk algorithm into person re-identification. Here, pedestrian segmentation is done by combining the shape information and color seed into the Random Walk formulation. Then, HSV color histograms and 1-D RGB signal along with texture features are used for person re-identification. Here, local binary pattern (LBP) and scale invariant local ternary pattern (SITP) are used as texture features. The limitation of this framework is from similar appearance of person images.

Zhao *et al.* [11] have done person re-identification by using unsupervised salience learning. In which, combination of LAB color histogram and SIFT features are used as feature. Here, first image is divided into several blocks and from each block LAB color histogram is extracted. Illumination variation and pose variations are handled by SIFT features. So, SIFT feature is complementary to color feature. Finally, they have used adjacency search and k-nearest neighbour algorithm for selecting possible candidates.

Ma *et al.* [12] have proposed a new image representation that measure similarity between two persons without requiring any pre-processing step like background removal and human body partition. They introduced a BiCov descriptor for person re-identification. BiCov representation is the combination of Biologically Inspired Features (BIF) and covariance descriptors which used to compute the similarity of the BIF features at neighboring scales. This method is a two stage process; first stage is the extraction of BIF features using Gabor filters and MAX operator. In the second stage covariance descriptor is applied to compute the similarity between BIF features. This method is also useful to face verification.

Malocal *et al.* [13] proposed a person re-identification system which is based on simple seven dimensional feature representation and Fisher vector method. The combination of the Fisher vectors with seven dimensional descriptors is termed as Local Descriptors encoded by Fisher Vector or LDFV. And this LDFV is used to describe person images. The seven dimensional local features contain the coordinates, the intensity, the first order and second order derivative of the pixel. These features are encoded and aggregated into a global Fisher vector that is the LDFV representation.

One disadvantage of unsupervised methods is that, information represented by their feature designs is not fully utilized. Also, it takes too much time for feature extraction and matching, which reduces its application in real time systems.

Unsupervised methods are also used for refining initial ranking list. Some unsupervised re-ranking techniques are described below:

Wu *et al.* [14] have constructed a scalable face image retrieval system. In which, both global and local features are used for representing a face. In the indexing stage, they have designed new component based local features by utilising special properties of faces. Then a novel identity based quantization schema is used for quantizing these local features into visual words. Then discriminative features for each face are encoded by hamming signature. To refine candidate person images, they have constructed a reference images set. In the retrieval stage, candidate images are retrieved from the inverted index of visual words. Finally, candidate images are re ranked using a multi-reference distance using hamming distance. However, a new feature extraction process is needed for re-rank the initial ranking list.

Pedronette *et al.* [15] have proposed a new post-processing method that re-ranks the initial ranking list by considering contextual information. In the content based image retrieval there is a query image, the approach aims to retrieve more similar images from a collection. Here, contextual information improving the efficiency of CBIR systems for generating ranking list. For retrieving contextual information they proposed an approach by creating gray scale image representation of distance matrices computed by CBIR descriptors. For the k-nearest neighbours of a query image context image is constructed and analyzed using image processing techniques. Context images are created by using the ranked list and distance measures, which brought computation complexity.

Huang *et al.* [16] have studied about two mechanisms, visual consistency and visual latency. In current web image search engines, search query and images that are closely related to that search query are visually similar. Also, user's eyes can capture salient images than ones in low level vision. Here, results from the search engines are re-ranked based on the visual saliency and consistency. So, re-ranked images consider both vision and content. However, the characteristic of saliency is not suitable for person re-identification.

Zhu *et al.* [17] have proposed a clustering algorithm in re-ranking. Rank-order distance, the core of the algorithm, uses neighbouring information in the dataset for measuring the dissimilarity between pair of faces. Here, faces of the same person will have similar neighbours. First a ranking order list is created for every faces by sorting all the faces in the dataset. Then ranking order is used for calculating the rank

order distance. Finally, a rank-order distance based clustering algorithm is designed based on the new distance, which iteratively groups all faces into small number of groups. Here, only context distance is used for re-ranking which ignores content distance. But content distance is essential in some situation, so ignoring content distance reduces detection rate.

Most of the above re-ranking methods are based on object retrieval. The datasets for general object retrieval task contain several images from one object. And these traditional re-ranking methods compare gallery image with both probe image and its k-nearest neighbours. Our proposed method provides a bi-directional comparison that is, here probe image is compared with gallery image and their k-nearest neighbours at the same time, gallery image is compared with probe image and its k-nearest neighbours.

### B. Supervised Methods

Generally, supervised methods require training samples with identity labels. These methods have lower flexibility for large gallery objects. Du et al. [18] have proposed a new approach called, random ensemble of color features (RECF). In which color features are obtained from six kinds of popular color spaces such as RGB, normalized RGB, HSV, YCbCr, CIE XYZ and CIE Lab. This information is used by a random forest to learn the similarity function between pair of person images.

Gray and Tao [19] have proposed a viewpoint invariant pedestrian recognition by using an ensemble of localized features (ELF), an efficiently designed object representation. In computer vision, one of the most important problem is the recognition of viewpoint invariant pedestrian. It is very difficult to matching two persons with unknown view point and poses. Here, designing a feature space instead of a specific feature to solve the problem and a machine learning algorithm is used for getting the best representation. Also, this approach generates a single similarity function by combining different kinds of single features. They also showed that, how to use Adaboost algorithm to learn an object class specific representation and discriminative recognition model. Both [22] and [23] proved that color histograms are effective for person re-identification.

Bak et al. [20] have solved person re-identification problem by using haar-like features and dominant color descriptor (DCD). Generally, persons observed from one camera view is different from those observed in another camera. Human signatures are used for person re-identification in order to handle difference in illumination, pose and camera parameters. For obtaining most discriminative haar-like feature set for each individual, the AdaBoost schema is used in haar-based approach. Finally, human signature is obtained by combining the set of haar-like features through a cascade. The dominant colors of upper and lower parts are extracted for obtaining DCD signature. Haar-like features

are suitable for handling view point and pose changes, however foreground extraction seems to be a difficult task in person re-identification.

Prosser et al. [21] have presented reformulation of person re-identification problem. This method considered person re-identification as a relative ranking problem. Previous approaches tried to extract discriminative features after that matching is performed based on distance measures. In this approach, team find out most similar persons for a given query image, by focusing on a highest ranked person. So, person re-identification problem is considered as a relative ranking problem rather than an absolute scoring problem. Here, an Ensemble RankSVM is used for resolving the computational scalability limitations of existing RankSVM models.

Schwartz et al. [22] have proposed partial least square (PLS), a feature selection approach, for learning discriminative appearance models. The appearance information is used in the applications like tracking and people recognition. In the case of appearance based models there is a chance for ambiguities among classes, when the number of persons was being considered increases. Also for each appearance limited numbers of training samples are available. Here, ambiguities are reduced by using a rich set of feature descriptors based on color, textures and edges. Partial least square (PLS) is a powerful statistical tool used for weighting the features based on their discriminative power for different appearance. This approach can reduce the background influence, but in the case of variable number of persons it is not flexible.

### C. Other Methods

This category includes those methods other than supervised and unsupervised for handling pose variations, light conditions and occlusions. Wang et al. [23] have used appearance models for computing the similarity between image regions. The local descriptors in an image are computed for constructing an appearance model, and then different strategies are used for aggregating them. In this work discriminative features are extracted for computing the similarity by introducing a multi-layer appearance modelling framework. The local description of the image is computed by the first layer, and second layer computing spatial relations between appearance labels. This work is mainly focus on computational complexity.

Bak et al. [24] have computed human signature through a spatial covariance descriptor. Human signatures should handle the variations in illumination, pose and camera parameters in order to re-identify persons. First human body parts are detected by using Histogram Oriented Gradients (HOG), and then spatial covariance regions are extracted from these body parts. Then a covariance descriptor is used for computing the similarity between corresponding body parts. Finally, spatial pyramid matching is used for computing a new dissimilarity measure between human signatures.

## III. CONCLUSION

In this paper, we briefly discuss the application and challenges of person re-identification. In addition, we also discuss the general methods used for person re-identification. Existing methods are mainly based on single directional matching. Single directional matching is not efficient in the case of variation of illumination and poses. Based on the drawbacks introduce a bi-directional re-ranking method for getting better results. To provide associate insight for the long run analysis direction, we tend to highlight the strategies that are capable of being helpful within the forthcoming analysis.

## REFERENCES

[1] A. Plantinga, "Things and persons", The *Review of Metaphysics*, pp.493–519, 1961.

[2] Wordspy. [Online] Available at http://www.wordspy.com/words/ reidenti_cation.asp.

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

[4] J. Garcia, C. Kambhamettu, A. Gardel, I. Bravo, and J. L. Liazaro, " Features accumulation on a multiple view oriented model for people re-identification", *in Eurographics 2013 Workshop on 3D Object Retrieval*, pp. 105–108, 2013.

[5] C. H. Kuo, C. Huang, and R. Nevatia, "Multi-target tracking by on-line learned discriminative appearance models", *in Computer Vision and Pattern Recognition (CVPR)*, *2010, IEEE Conference*, pp. 685–692, 2010.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino and M.Cristani, "Person re-identification by symmetry-driven accumulation of local features", *In computer vision and pattern recognition(CVPR)*, pp. 2360-2367, 2010.

[7] C. Liu, S. Gong, C. C. Loy, and X. Lin", Person re-identification: what features are important?", *in Computer Vision–ECCV* , pp. 391–401, 2012.

[8] Z. Zhao and C. Liu**,** "Person re-identification by local feature based on super pixel", *in Advances in Multimedia Modeling*, pp.196–205, 2013.

[9] I. Kviatkovsky, A. Adam and E. Rivlin, "Color Invariants for Person Re-Identification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, pp. 1622–1634, 2013.

[10] Y.-C. Chang, C.-K. Chiang and S.-H, Lai, "Single-shot person re-identification based on improved random-walk pedestrian

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

segmentation", In *intelligent signal processing and communication systems (ISPACS)*, 2012.

[12] B. Ma, Y. Su and F. Jurie, "Bicov: A novel image representation for person re-identification and face verification", *British Machive Vision Conference*, Guildford, United Kingdom, p. 11, 2012.

[13] B. Ma, Y. Su and F. Jurie. "Local descriptors encoded by fisher vectors for person re-identification", *In European Conference on computer vision (ECCV)*, pp.413-422, 2012.

[14] Z. Wu, Q. Ke, J. Sun, H. Y. Shum, "Scalable face image retrieval with identity-based quantization and multireference reranking", *IEEE Trans Pattern Anal Machine Intell,* Vol. 33, No. 10, pp. 1991–2001, 2011.

[15] D. Pedronette, C. Guimares, R. da S Torres, "Exploiting contextual spaces for image re-ranking and rank aggregation", *In: Proceedings of the 1st ACM international conference on multimedia retrieval,* 2011.

[16] J. Huang, X. Yang, X. Fang, W. Lin and R. Zhang, "Integrating visual saliency and consistency for re-ranking image search results", *IEEE Trans Multimedia*, Vol. 13, No. 4, pp. 653–66, 2011.

[17] C. Zhu, F. Wen and J. Sun, "A rank-order distance based clustering algorithm for face tagging", *In: IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 481–488, 2011.

[18] Y. Du, H. Ai and S. Lao, "Evaluation of color spaces for person re-identification", *In International Conference on pattern recognition (ICPR)*, pp. 1371–1374, 2012.

[19] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensamble of localized features", *In European Conference on computer vision (ECCV)*, pp. 262-275, 2008.

[20] S. Bak, E. Corvee, F. Bremond and M. Thonnat, " Person re-identification using Haar-based and DCD-based signature", *In Workshop on Activity Monitoring by Multi-Camera Surveillance Systems,* 2010.

[21] B. Prosser, W.-S. Zheng, S. Gong and T. Xiang, "Person re-identification by support vector rankin", *In British Machine Vision Conference (BMVC),* 2010.

[22] W. Schwartz and L. Davis, "Learning discriminative appearance-based models using partial least squares", *In XXII Brazilian symposium on computer graphics and image processing (SIBGRAPI),* 2009.

[23] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, P. Tu, "Shape and appearance context modelling", *In International Conference on computer vision (ICCV),* 2007.

[24] S. Bak, E. Corvee, F. Bremond and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts", In *Advanced Video and Signal-Based Surveillance,* 2010.