Review on Human Re-identification with Multiple Cameras

K.B. Low, U.U. Sheikh
Faculty of Electrical Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
caelynlow@gmail.com

***Abstract*—Human re-identification is the core task in most surveillance systems and it is aimed at matching human pairs from different non-overlapping cameras. There are several challenging issues that need to be overcome to achieve re-identification, such as overcoming the variations in viewpoint, pose, image resolution, illumination and occlusion. In this study, we review existing works in human re-identification task. Advantages and limitations of recent works are discussed. At the end, this paper suggests some future research directions for human re-identification.**

***Index Terms*—Human Re-Identification; Non-Overlapping Views; Multiple Cameras.**

I. INTRODUCTION

Human re-identification is an important task in closed-circuit television (CCTV) surveillance due to its many applications, such as for human tracking and video retrieval. The goal of human re-identification is to determine whether a pair of humans from different cameras with non-overlapping views has the same or different identity. Employing human operators to observe human of interest over multiple cameras is not a good option because human operators are prone to exhaustion. Besides, when human operators change, the new operator will need to spend time to observe again and this is time consuming and may result in missing out suspicious target during that period of change. Therefore, a computer vision system is more suitable to assist human operators in identifying human over a set of cameras with disjoint views.

A human that disappears from one camera view over a period of time needs to be matched in another view of camera at a different location. However, human image taken by different cameras will have different angle of view, illumination, background, occlusion and human pose. A different angle of camera view will cause some important features observed in one camera to disappear in another camera. The difference in illumination will cause the same human in different cameras to have different appearance, while different humans may have similar appearance. Partial occlusion will cause some human body parts to disappear in the camera views. All of these challenging issues may cause variations to human appearance. Examples of images facing these issues are shown in Figure 1.

Many existing works were proposed to overcome these challenging issues [1 – 41].

II. HUMAN RE-IDENTIFICATION

Human re-identification methods can be divided into biometric-based methods and appearance-based methods. Examples of biometric-based methods are iris recognition, fingerprint recognition, face recognition and gait recognition. For biometric-based methods, such as iris and fingerprint recognition, special sensors are required to acquire the biometric information. While high resolution images are required for face and gait recognition. In addition, human cooperation is highly needed to complete the re-identification task for biometric-based methods. Biometric-based methods can only achieve good performance in re-identification task, when certain conditions and constraints are fulfilled. In wide area surveillance systems, biometric-based methods are not suitable for human re-identification task due to the difficulty in acquiring biometric information and obtaining cooperation from the target.

Appearance-based methods are less constrained than biometric-based methods; therefore, they are more suitable for human re-identification in wide area surveillance systems. For human re-identification system, most of the works are based on non-overlapping views of multiple cameras because it is impossible to cover wide area by using multiple overlapping cameras. By using non-overlapping view of cameras, cost and computational expense can be reduced. Camera setup with overlapping views and non-overlapping views are shown in Figure 2. In this paper, appearance-based methods are reviewed in detail.

Figure 1: Some of the challenging issues in human re-identification problem: variations in (a) pose, (b) illumination and (c) partial occlusion [4].

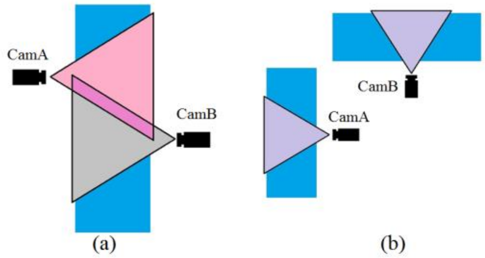

Figure 2: Typical camera setup for: (a) overlapping views and (b) non-overlapping views.

III. APPEARANCE-BASED METHODS

Appearance-based methods for human re-identification can be divided into two categories; single-shot methods and multiple-shot methods, which depend on the number of images that are used to create the human representation. In single-shot methods, only a single image is used to compute the feature vectors, while in multiple-shot methods, a group of images are used to compute the feature vectors to represent the human. Although multiple-shot methods provide more information than single-shot methods, they require more complex algorithms and are more compute intensive. Single-shot methods are useful when no tracking information is provided.

Each category can further be divided into two groups, which are direct-based approach and learning-based approach. Direct-based approach focuses on designing a robust and stable feature representation of human under various conditions. After feature extraction, simple matching is used to determine the similarity between the images.

Learning-based approach uses training data to search for matching strategies, which aim to minimise the distance between similar classes while maximising the distance between dissimilar classes. Besides these two categories, deep learning is now being applied in human re-identification task because it incorporates feature extraction and classification task into one integrated framework.

A. Single-shot Methods

To cope with common challenging issues, researchers prefer to integrate several types of features, such as colour, texture and shape in their works [1–13]. Feature extraction based on colour and edge features were applied to form an appearance descriptor in [9] and correlation based similarity measure was used for matching. A parts-based method was proposed in [42], where they divided the human image into three horizontal partitions and extract signature based on colour histograms in HSV colour space for each part. In [5], shape and appearance context model was proposed by Wang *et al*. They segmented the image of human into regions and registered their colour spatial relationship into co-occurrence matrix. However, their proposed method only worked well when the human images from different cameras were captured from similar viewpoints. Yang *et al.* [10] proposed an appearance model, which was constructed through kernel estimation. Colour Rank Feature and path-length were used in their work. Kullback-Leibler distance was used to compute similarity. Gallagher *et al*. [11] utilised clothing features and facial features for recognising people's task. Appearance of detected human was represented by a set of histograms. Nearest Neighbor classifier was used to find the probable label for detected human. Their proposed method requires high quality image because facial features were involved.

Part-based models, which are used to handle pose variation was proposed in [12]. They detected five human body parts using Histogram of Oriented Gradient (HoG) and then computed covariance regions. Their proposed method only performed well when pose estimators work accurately. Cai *et al*. proposed a patch-based approach in [13] which extracted the patches around the edges by using Canny edge detection algorithm. After obtaining the patches, dominant colour and frequency within the patch were extracted from each patch. The drawback of their algorithm is that it is computationally expensive.

The work in [4] proposed a learning-based method using an ensemble of localised features (ELF) to deal with viewpoint varying issues. In their proposed method, texture and colour features were extracted. To select a subset of optimal features, AdaBoost algorithm was used. Their work is highly dependent on image quality. In [14], pairwise dissimilarity profiles were learned to distinguish a pair of humans; however, the proposed method cannot re-identify a new human unless the new human is added into the training set and re-computed again. Partial Least Squares (PLS) reduction was proposed in [6] by Schwartz *et al*. to transform features into a low dimensional discriminant latent space. Colour, texture and edge features were extracted in their work, and PLS was used to reduce the dimensionality (Figure 3). The drawback of their proposed method is that the need to re-compute if a new novel person is added to the training set.

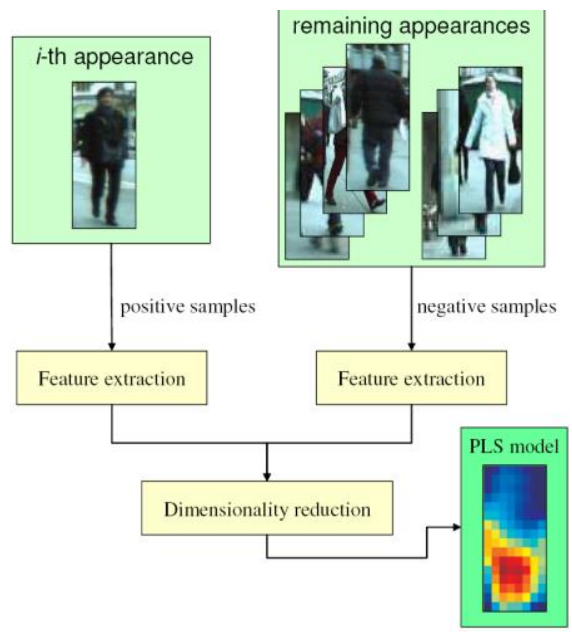

Figure 3: Illustration of the PLS method in [43].

In [15], to obtain the most discriminative features, Haar-like features were extracted from whole body and dominant colour descriptors were extracted from the upper and lower body parts. In [16], semi-supervised method was proposed where multiple features were fused at classification output level by using multi-feature learning (MFL) strategy. Colour and texture features were extracted, such as Black-Value-Tint (BVT) histograms, Hue Saturation Value (HSV), Lightness color-opponent (Lab) histogram, Local Binary Pattern (LBP) histogram and histograms of the Maximum Responses filter banks (MR8).

Existing metric learning methods applied in re-identification task are, such as KISSME [17], RankSVM [1], Probablistic Relative Distance Comparison (PRDC) [3][18] and LMNN-R [19]. Prosser *et al*. [1] formulated re-identification problem as ranking problem by introducing Ensemble RankSVM to learn a subspace where the potential true match gets the highest rank. Zheng *et al*. [18] proposed Probabilistic Relative Distance Comparison (PRDC) to improve the computation of RankSVM; however, their training process is still computationally expensive.

Dikmen *et al*. [19] proposed a metric learning framework by applying a large margin nearest to a neighbour (LMNN-R) to minimise the distance for true matches and maximise the distance for false matches. However, their distance learning based method requires a large labelled human pair, which is not practical.

Table 1 shows the taxonomy of existing methods for single-shot appearance based re-identification. A summary of various features for single-shot methods is shown in Table 2.

B. *Multi-shot Methods*

Bird *et al*. [46] divided the human image into ten horizontal stripes to deal with pose variation problem. They extracted Median Hue, Saturation and Lightness (HSL) colour from these ten stripes. Linear Discriminant Analysis (LDA) was applied on the features in order to reduce its dimensionality. Their proposed method is not reliable for low resolution videos, such as in most surveillance systems. In [26], Gheissari *et al*. proposed a novel spatiotemporal segmentation using ten consecutive frames of human to generate spatiotemporal graph. Appearance of human was represented by colour and edge histograms. The proposed approach is only applicable when the human appearance does not change (similar) in different cameras. This limitation makes their approach impractical for real-world applications. Besides, their proposed method is weak on pose differences and occlusion.

Table 1
Single-shot methods

Direct based		Learning based	
Bak *et al.*	[12]	Bak *et al*.	[15]
Cai *et al.*	[13]	Dikmen *et al*.	[19]
Cheng *et al*.	[7]	Figueira *et al*.	[16]
Farenzena *et al*.	[2]	Hirzer *et al.*	[20]
Gallagher *et al.*	[11]	Hirzer *et al.*	[25]
Park *et al.*	[42]	Ijiri *et al*.	[21]
Wang *et al.*	[5]	Kostinger *et al*.	[17]
Yang *et al.*	[10]	Layne *et al.*	[22]
		Lin *et al*.	[14]
		Loy *et al*.	[8]
		Prosser *et al*.	[1]
		Satta *et al*.	[24]
		Schwartz *et al*.	[6]
		Sivic *et al*.	[44]

Table 2
Summary of various features used in single-shot methods

	Approach	Features
[1]	RankSVM	Color (RGB, HS, YCbCr), texture (Gabor, Schmid)
	Tuning of this algorithm is computationally expensive.	
[2]	SDALF	Color (wHSV, MSCR), texture (RHSP)
	Processing time is higher than other direct-based methods.	
[3]	PRDC	Color (RGB, HS, YCbCr), texture (Schmid, Gabor)
	Training process is still computationally expensive.	
[4]	ELF	Color (RGB, HSV, YCbCr), texture, mean values
	Highly dependent on image quality.	
[5]	Shape context	Shape, color (RGB), texture, HOG
	Only applicable for similar viewpoints.	
[6]	PLS	Color (RGB), texture(Co-occurrence matrices), edge (HOG)
	Cannot re-identify a new human.	
[7]	CPS	Color (wHSV, MSCR)
[8]	Manifold ranking	Color (RGB, HSV, YCbCr), texture (Gabor, Schmid)
[9]	Stochastic models	Color, edge
[10]	Kernel density estimation	Color (RGB), path-length
[11]	Clothing co-segmentation	Clothing, facial
	High quality image is required because facial feature is involved.	
[12]	SCR	Color (RGB), gradient, position
	Their approach performed well when pose estimators work accurately. Their approach was tested in indoor environment only.	
[13]	Patch-based approach	Dominant color, frequency within the patch
	Their approach was computationally expensive.	
[14]	Pairwise dissimilarity profiles	Normalized color, color Rank
	Cannot re-identify a new human.	
[15]	AdaBoost	Color (Dominant color descriptor), Haar-like features
[16]	MFL	Color (BVT, HSV, Lab), texture (LBP,MR8)
[17]	KISSME	SIFT
[19]	LMNN-R	Color (RGV, HSV)
	Large labelled human pairs are required for their approach.	
[44]	pictorial structure model	Color (RGB)
	High quality images are required. Not suitable for surveillance system.	
[20]	Impostor learning	Color (HSV, Lab), texture (LBP)
[22]	Attributes	Color (RGB, HSV, YCbCr), texture (Gabor, Schmid)
	High resolution images are required.	
[23]	Saliency	color (LaB), SIFT
	Computationally expensive learning-based method.	
[18]	PRDC	Color(RGB, YCbCr, HSV), Texture (Schmid, Gabor)
	Training process is still computationally expensive.	
[25]	RPLM	Color (HSV, Lab), Texture mean color, LBP
[45]	Group context	Color (RGB), texture, HOG
	High resolution images are required. Their approach was tested in indoor environment only.	

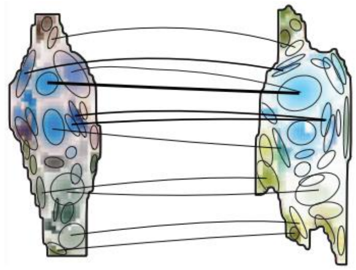

Figure 4: Illustration of the EMD blocks matching [47].

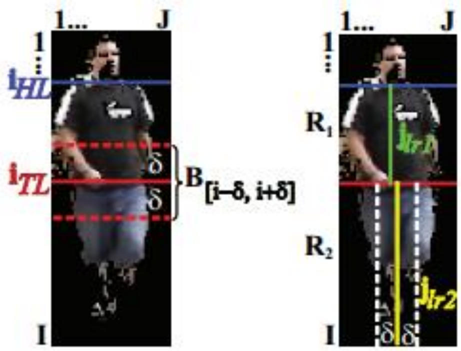

Figure 5: Illustration of symmetry and asymmetry based partitions [2].

In [27], Speeded Up Robust Features (SURF) interest points were collected in multiple frames and stored in a KD-tree. To measure the similarity of interest point descriptors, Sum of Absolute Differences (SAD) was applied. Weighted Region Matching (WRM) method was proposed by Oreifej *et al*. in [47]. In their work, they divided image into patches based on their appearance consistencies. Earth Movers Distance (EMD) was used to measure the similarity between patches, and this is shown in Figure 4. Their work was applied on low-resolution aerial images. Their proposed method is prone to mismatched patches.

Chromatic content, which is HSV histogram, Maximally Stable Colour Regions (MSCR) and Recurrent Highly Structured Patches (RHSP) have been proposed in [2]. These features were extracted based on their proposed symmetry and asymmetry axes, which are shown in Figure 5. Their Symmetry-Driven Accumulation of Local Features (SDALF) approach is very time-consuming when compared to other existing hand-crafted methods.

Global features consisting of HSV histogram and a set of local features were extracted from highly informative patches using epitome analysis in [28]. Histogram Plus Epitome (HPE) was proposed to fuse global features and local features. Pictorial Structures (PS) was proposed by Cheng *et al*. [7] to estimate human body configuration. Multiple images of human were used to compute features, which are based on different body parts. Their proposed method can also be applied in single shot methods to localise body parts to deal with pose variation problem. However, the pose estimator must work accurately to perform re-identification task.

In learning-based methods, Nakajima *et al*. [48] proposed the extraction of a set of local and global features to recognise full body people by using multi-class SVMs. However, when a new person appears, they need to re-train their learning-based approach.

A non-surveillance application of person re-identification was proposed in [44] to find occurrences of person in a sequence of photographs. Their proposed method is highly dependent on image quality, so it is not suitable for cases where the captured image is from far distance and in low resolution. In [29], Bazzani *et al*. proposed the histogram plus epitome feature, which incorporates both global and local statistical descriptors for person re-identification.

Table 3 shows the taxonomy of existing appearance based on multiple-shot methods for re-identification. A summary of the various features in multiple-shot approaches is shown in Table 4.

IV. Deep Learning Approaches

In recent years, deep learning has been used in performing re-identification task because deep learning methods integrate feature extraction and classification into a single framework. Convolution Neural Network is one of the most popular deep learning methods that can be used because it can automatically extract optimal features from raw input image instead of using hand-crafted features.

In [37], Li *et al*. proposed deep filter pairing neural network (FPNN), which is the first work that applied deep learning on person re-identification task. Their proposed architecture consists of multiple layers composed of convolution layer, max pooling layer, patch matching layer, maxout pooling layer, fully connected layer and softmax layer, as shown in Figure 6.

Another deep learning method was proposed in [38], which applied Siamese Convolution Neural Network (SCNN) architecture in human re-identification task is as shown in Figure 7. They divided the human image into three equally overlapping parts: Each of the part was fed into SCNN architecture. At the end, they obtained three similarity metrics from three overlapping parts by using cosine function. Finally, they fused the three similarity metrics by sum rule to get the final similarity value.

In [39], Zhang *et al*. introduced a linear support vector machine (linear SVM) in their deep convolutional neural network (deep CNN). In order to compute the similarity of pair of images, they adopted a margin-based loss.

Table 3
Multiple-shot methods

Direct based		Learning based	
Bak *et al.*	[36]	Bak *et al.*	[35]
Bazzani *et al.*	[28]	Bak *et al.*	[49]
Cheng *et al.*	[7]	Bazzani *et al.*	[29]
Farenzena *et al.*	[2]	Ma *et al.*	[32]
Gheissari *et al.*	[26]	Sivic *et al.*	[44]
Hamdoun *et al.*	[27]	Truong *et al.*	[30]
Ma *et al.*	[34]	Truong *et al.*	[31]
Oreifej *et al.*	[47]	Yang *et al.*	[50]
Salvagnini *et al.*	[51]		
Shishir *et al.*	[33]		

Table 4
Summary of various features used in multiple-shot methods

	Approach	Features
[2]	SDALF	Color (wHSV, MSCR), texture (RHSP)
	Processing time is higher than other direct-based method.	
[7]	CPS	Color (HSV, MSCR)
	Performance highly dependent on pose estimator.	
[44]	Pictorial structure model	Color (RGB)
	High quality images are required. Not suitable for surveillance system.	
[46]	Blob	Color (HSL)
	Not reliable at low resolutions.	
[26]	Spatiotemporal model	Color (HSV, LAB), edges
	Their approach is only useful in situations where sequences of frames are provided. Only applicable for similar viewpoints. Weak on pose differences and occlusion.	
[27]	SURF	Interest-points
[47]	PageRank	Color (HSV), HOG
	Easily affected by mismatched patches.	
[28]	HPE	Color (HSV), Epitomic Analysis, local epitome
[48]	SVM	Color (RGB), shape
	Cannot re-identify new person.	
[30]	Graph	Color (RGB, color/path-length, spatiograms)
	Their approach re-identifies people in similar poses walking inside a train.	
[31]	Spectral analysis, SVM	Color (RGB), position
[49]	Correlation space Re-ID	Position, color (RGB), gradient
[32]	Fisher vector	Position, color (HSV), texture (Gabor)
	Computationally expensive.	
[33]	Multi-feature model	Color (ACM, RMC), face
	Highly dependent on image quality, not suitable for low resolution image.	
[50]	SBDR	Color (RGB, YCbCr, HSV), texture (Schmid, Gabor)
[34]	BiCov	Color (wHSV, MSCR, gBiCov), texture(Gabor)
	The approach assumes uniformity of texture in the background.	
[35]	Boosted Re-ID	Position, color (RGB), gradient
[36]	MRCG	Color (RGB), shape, MRCG
	Their proposed method only tested in indoor environment.	
[51]	Max-var	Color (HS), texture (LBP)

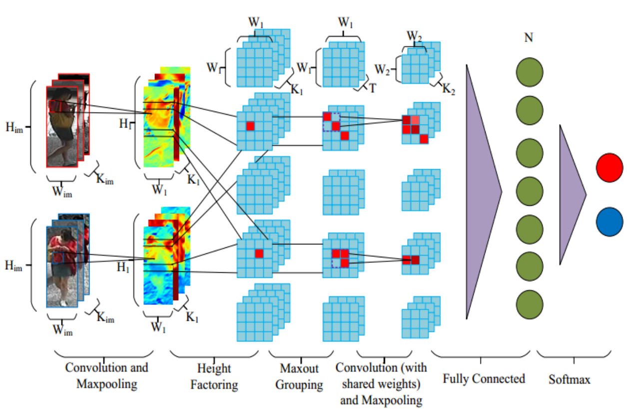

Figure 6: Flowchart of the FPNN proposed in [37].

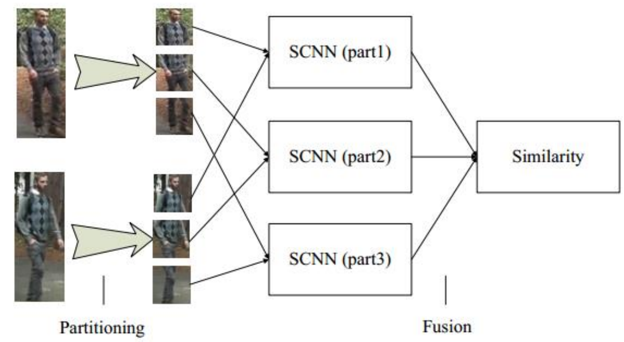

Figure 7: Flowchart of Siamese Convolution Neural Network [38].

In the proposed CNN in [40], they designed a cross-input neighbourhood difference layer to compute the local relationships between images from different views. They also incorporated a patch summary layer. Recently, Wu *et al*. [41] proposed a CNN architecture by adding a layer of computing neighbourhood range differences. Their network is deep with small (3x3) convolution filters. Deep learning approaches that are mentioned above have shown good performance in dealing with human re-identification task with reported Rank-1 accuracy ranges between 28% and 35% for VIPeR dataset.

V. Summary and Future Work

Appearance based methods are the preferred way of performing human re-identification task in surveillance systems because they are less constrained than biometric based methods. Appearance based methods can be divided into single-shot and multiple shot approaches, which depend on the number of images used per person. In each approach, there are two different ways to deal with re-identification task: direct or through learning. For direct-based methods, handcrafted features are designed to be distinctive and stable under challenging conditions. For learning-based methods, training data are used to select features, and they are dependent on the training data and cardinality.

Recently, deep learning methods, such as the Convolution Neural Network (CNN) were used for re-identification task. Deep learning approach is anticipated to work well on human re-identification due to being independent of handcrafted features. However, substantial experimental work needs to be done to measure its effectiveness in the presence of challenges such as low resolution, noisy image and partial occlusion.

Acknowledgment

The authors would like to express their gratitude to Universiti Teknologi Malaysia (UTM) for supporting this research work under GUP grant number Q.J130000.2523.11H73.

References

[1] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person Re-Identification by Support Vector Ranking," Bmvc, vol. 1, no. 3, p. 21.1-21.11, 2010.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 2360–2367, 2010.

[3] W.-S. Zheng, S. Gong, and T. Xiang, "Re-identification by Relative Distance Comparison," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 653–668, 2013.

[4] D. Gray and H. Tao, "Viewpoint Invariant Pedestrain Recognition with an Ensemble of Localized Features," Eur. Conf. Comput. Vis., pp. 262–275, 2008.

[5] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and Appearance Context Modeling," Int. Conf. Comput. Vis., pp. 1–8, 2007.

[6] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," Brazilian Symp. Comput. Graph. Image Process., pp. 322–329, 2009.

[7] L. B. and V. M. Dong Seon Cheng, Marco Cristani, Michele Stoppa, "Custom Pictorial Structures for Re-identification," Bmvc, p. 68.1--68.11, 2011.

[8] C. C. Loy, C. Liu, and S. Gong, "Person Re-identification by Manifold Ranking," IEEE Int. Conf. Image Process., 2013.

[9] Kang, I. Cohen, and G. Medioni, "Object reacquisition using invariant appearance model," Proc. - Int. Conf. Pattern Recognit., vol. 4, pp. 759–762, 2004.

[10] Y. Yu, D. Harwood, K. Yoon, and L. S. Davis, "Human appearance modeling for matching across video sequences," Mach. Vis. Appl., vol. 18, no. 3–4, pp. 139–149, 2007.

[11] A. C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," 26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, 2008.

[12] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person Re-identification Using Spatial Covariance Regions of Human Body Parts," Int. Conf. Adv. Video Signal Based Surveill., pp. 435–440, 2010.

[13] Y. Cai, K. Huang, and T. Tan, "Human appearance matching across multiple non-overlapping cameras," Icpr, pp. 1–4, 2008.

[14] Lin, Z. Lin, L. S. Davis, and L. S. Davis, "Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance," Adv. Vis. Comput., pp. 23–34, 2008.

[15] S. Bk, E. Corvee, F. Bremond, M. Thonnat, S. Bąk, E. Corvee, F. Brémond, and M. Thonnat, "Person Re-identification Using Haar-based and DCD-based Signature," 2010 7th IEEE Int. Conf. Adv. Video Signal Based Surveill., pp. 1–8, 2010.

[16] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," Avss, 2013.

[17] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large Scale Metric Learning from Equivalence Constraints," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., no. March, pp. 2288–2295, 2012.

[18] W. S. Zheng, S. Gong, and T. Xiang, "Person Re-identification by Probabilistic Relative Distance Comparison," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 653–668, 2013.

[19] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," Proc. Asian Conf. Comput. Vision,2010, pp. 501–512, 2011.

[20] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," Proc. - 2012 IEEE 9th Int. Conf. Adv. Video Signal-Based Surveillance, AVSS 2012, pp. 203–208, 2012.

[21] Y. Ijiri and S. Lao, "Human Re-Identification Through Distance Metric Learning Based on Jensen-Shannon Kernel," Visapp, no. c, pp. 603–612, 2012.

[22] R. Layne, T. Hospedales, and S. Gong, "Towards Person Identification and Re-identification with Attributes," ECCV Work. 2012, vol. 7583, pp. 402–412, 2012.

[23] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3586–3593, 2013.

[24] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino, "A multiple component matching framework for person re-identification," Int. Conf. Image Anal. Process., pp. 2360–2367, 2011.

[25] M. Hirzer, P. M. Roth, M. Koestinger, and H. Bischof, "Relaxed Pairwise Learned Metric for Person Re-identification," Comput. Vis. - Eccv 2012, Pt Vi, vol. 7577, pp. 780–793, 2012.

[26] N. Gheissari, T. B. Sebastian, P. H. Tu, J. Rittscher, and R. Hartley, "Person reidentification using spatiotemporal appearance," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2, pp. 1528–1535, 2006.

[27] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person Re-identification In Multi-camera System By Signature Based On Interest Point Descriptors Collected On Short Video Sequences," Robotics, pp. 0–5, 2008.

[28] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-shot person re-identification by HPE signature," Proc. - Int. Conf. Pattern Recognit., pp. 1413–1416, 2010.

[29] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot Person Re-identification by Chromatic and Epitomic Analyses," Pattern Recognit. Lett., vol. 33, no. 7, pp. 898–903, 2012.

[30] D. N. T. Gong, C. Achard, L. Khoudour, and L. Douadi, "Video Sequences Association for People Re-identification across Multiple Non-overlapping Cameras," Image Anal. Process. - Iciap 2009, Proc., vol. 5716, pp. 179–189, 2009.

[31] D. N. Truong Cong, L. Khoudour, C. Achard, C. Meurie, and O. Lezoray, "People re-identification by spectral classification of silhouettes," Signal Processing, vol. 90, no. 8, pp. 2362–2374, 2010.

[32] F. Ma, B., Su, Y. and Jurie, "Local Descriptors Encoded by Fisher Vectors for Person Re-identification," Comput. Vis. – ECCV 2012, pp. 413–422, 2012.

[33] A. Bedagkar-Gala and S. K. Shah, "Part-based spatio-temporal model for multi-person re-identification," Pattern Recognit. Lett., vol. 33, no. 14, pp. 1908–1915, 2012.

[34] B. Ma, Y. Su, and F. Jurie, "BiCov: a novel image representation for person re-identification and face verification," Procedings Br. Mach. Vis. Conf. 2012, p. 57.1-57.11, 2012.

[35] S. Bak, E. Corvée, F. Brémond, M. Thonnat, and I. S. Antipolis, "Boosted Human Re-identcation using Riemannian Manifolds," Image Vis. Comput., pp. 443–452, 2012.

[36] S. Bąk, E. Corvee, F. Brémond, and M. Thonnat, "Multiple-shot human Re-identification by Mean Riemannian Covariance Grid," 2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011, pp. 179–184, 2011.

[37] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 152–159, 2014.

[38] D. Yi, Z. Lei, and S. Z. Li, "Deep Metric Learning for Practical Person Re-Identification," Icpr, vol. 11, no. 4, pp. 1–11, 2014.

[39] G. Zhang, J. Kato, Y. Wang, and K. Mase, "People re-identification using deep convolutional neural network," VISAPP 2014 - Proc. 9th Int. Conf. Comput. Vis. Theory Appl., vol. 3, pp. 216–223, 2014.

[40] E. Ahmed, M. Jones, and T. K. Marks, "An Improved Deep Learning Architecture for Person Re-Identification," IEEE Conf. Comput. Vis. Pattern Recognit., 2015.

[41] L. Wu, C. Shen, and A. van den Hengel, "PersonNet: Person Re-identification with Deep Convolutional Neural Networks," CoRR abs/1601.07255, 2016.

[42] U. Park, A. K. Jain, I. Kitahara, K. Kogure, and N. Hagita, "ViSE: Visual Search Engine Using Multiple Networked Cameras," 18th Int. Conf. Pattern Recognit., vol. 3, pp. 1204–1207, 2006.

[43] G. L. Prado, W. R. Schwartz, and H. Pedrini, "Person re-identification using partial least squares appearance modeling," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8259 LNCS, no. PART 2, pp. 382‑390, 2013.

[44] J. Sivic, C. L. Zitnick, and R. Szeliski, "Finding people in repeated shots of the same scene," Procedings Br. Mach. Vis. Conf. 2006, p. 93.1-93.10, 2006.

[45] J. Sivic, C. L. Zitnick, and R. Szeliski, "Finding people in repeated shots of the same scene," Procedings Br. Mach. Vis. Conf. 2006, p. 93.1-93.10, 2006.

[46] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs, "Detection of loitering individuals in public transportation areas," IEEE Trans. Intell. Transp. Syst., vol. 6, no. 2, pp. 167–177, 2005.

[47] O. Oreifej, R. Mehran, and M. Shah, "Human Identity Recognition In Aerial Images," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 709–716, 2010.

[48] T. P. Chikahito nakajima, massimiliano Pontil, Bernd Heisele, "Full-Body Person Recognition System," Pattern Recognit., vol. 34, 2004.

[49] S. Bąk, G. Charpiat, E. Corvée, F. Brémond, and M. Thonnat, "Learning to Match Appearances by Correlations in a Covariance Metric Space," Proc. 12th Eur.Conf. Comput., pp. 806–820, 2012.

[50] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao, "Set Based Discriminative Ranking for Recognition," Proc. 12th Eur. Conf. Comput. Vis, pp. 497–510, 2012.

[51] P. Salvagnini, L. Bazzani, M. Cristani, and V. Murino, "Person Re-identification with a PTZ Camera: An Introductory Study," ICIP, 2013.