Non-contextual Long-Term Person Re-ID Under Sudden Illumination Conditions

B. Yogameena[1,2,3](✉), K. Menaka[1,3], and S. Saravana Perumaal[1,2,3]

[1] Department of ECE, Thiagarajar College of Engineering, Madurai, India
{ymece,sspmech}@tce.edu, ece.menaka@gmail.com
[2] Department of ECE, V.S.B. Engineering College, Karur, India
[3] Department of Mechanical Engineering, Thiagarajar College of Engineering, Madurai, India

Abstract. Person Re-ID (Re-ID) has been an emerging topic and there is a need for non-contextual, long term person Re-ID algorithm since most of the crimes occur in public places such as airports, railway stations where the video is recorded for a long duration across arbitrary camera views. The proposed work comprises of two fusion frameworks. First, the people detection is accomplished by fusing Histogram of Oriented Gradients (HOG) and extended Center Symmetric Local Binary Pattern (XCS-LBP) features which overcomes several disadvantages like missing detection and false detection due to sudden illumination, and near/far field of view changes. Next, each body part of the tracked person is learned using Deformable Part Model (DPM) which is robust to different viewpoints. Secondly, the feature level fusion of Gabor (appearance feature) and Skeleton Recurrent Motion Image (SRMI Gait biometric feature) is proposed to overcome the homogeneity issue of distinguishing people when both the texture and color of the people attire are similar. Finally, person Re-ID is achieved by relevance metric learning method with list wise constraints (RMLLCs). The performance measure, Cumulative Matching Curve (CMC) Rate shows the improved matching accuracy compared to other state-of-the-art algorithms with benchmark datasets.

Keywords: Person Re-identification · Long-term · Skeleton recurrent motion image sudden illumination · Gait and appearance features

1 Introduction

Deployment of large networks of cameras is drastically increasing in public places like airports, railway stations and office buildings. Fascinatingly, automated analysis of huge video data can improve the quality of surveillance by processing the video faster. Such automated analysis is more useful for high-level surveillance tasks like suspicious activity detection or undesirable event prediction for timely alerts or behaviors or events of interest. These tasks can only be inferred from a long term analysis of the target person across the camera network. It is inferred from the literature that till now there is a lack of research in non-contextual long term person Re-ID where multiple cameras have been employed [1]. Since most of the crimes occur in public places like airports, railway stations where the video is recorded for a long duration, this paper aims to

R. V. Babu et al. (Eds.): NCVPRIPG 2019, CCIS 1249, pp. 302–311, 2020.
https://doi.org/10.1007/978-981-15-8697-2_28

provide a framework for non-contextual long term person Re-ID. It includes four phases like people detection, tracking, identification and Re-ID across multiple camera views.

2 Related Work

Contextual methods rely on external contextual information which considers camera geometry as well as camera calibration as context of a single static camera for feature extraction [1]. Non-contextual methods rely on dissimilar context information across arbitrary cameras and view. Short term person Re-ID utilizes the video taken only for few hours and the external factors influencing the scenario is less. Nevertheless, the appearance features such as Symmetry Driven Accumulation of Local Features (SDALF) and HOG possibly will fail due to varying illumination, occlusions and noise. Hence, these features are not suitable for long term person Re-ID which utilizes videos recorded in a day, week or a month which is the main focus of the proposed framework. Existing non contextual, long term Re-id methods in literature are categorized into color calibration, descriptor learning and metric learning methods. Color calibration category consists of direct and indirect approaches and claimed Weighted Brightness Transfer Function (WTBF) outperforms than Brightness Transfer Function (BTF) and Cumulative BTF [2]. Recently the authors in [3, 4] have adopted calibration methods to be robust against uncalibrated images and less space constraints. However, background clutter, large pose variations still remains as a major challenge. Descriptor learning type consists of Shape Context [5], Re-id with attributes [6] and Re-id by saliency. The closed set Re-ID problem has been solved by the endeavor of recent state-of-the-art methods [1, 7]. Moreover, various approaches using appearance descriptors for person Re-ID with their merits and demerits have been reviewed [8]. Amongst these appearance descriptors, it is recommended to utilize Gait biometric for long term Re-ID and most of the recent research is concentrated on non-contextual methods [1]. HoG [5] is used for shape and appearance based person detection and they had achieved only 82% of matching accuracy. HOG performs poorly when the background is cluttered with noisy edges and the detection rate decreases further in sudden illumination conditions. An attempt has been done by fusing HoG [9] and LBP features [10] to improve the detection rate since both HOG and LBP individually revealed bigger achievement in person detection applications. Also LBP filters the noisy background. However, the false positives which occurs more frequently in sudden illumination conditions is not addressed. Later, XCS-LBP [11] has been proposed and it is robust to sudden illumination changes. Hence, a framework which considers the non-contextual and long-term person Re-ID issues by adapting Gait as well as the appearance features is needed. Moreover, to obtain the correct match against query is another major challenge. Distance metric learning includes Dynamic Time Warping, Distance Metric, Set Based Discriminative Ranking (SBDR), Mahanolobis Metric Distance Learning [12], and Probabilistic Relative Distance Learning (PRDC) [13]. It is inferred from the literature that distance metric learning methods pay no attention to geometry information as well as sensitive to outliers especially Mahalanobis distance metric learning. More recently in [14, 15], asymmetric metric learning is used to eradicate the issues such as similar appearance, motion and actions in video representations. Even though, methods based on metric learning lack adequate feature representation and efficiency.

Gait Energy Image (GEI) proposed by Liu et al. in [16] requires perfect alignments of the silhouettes for comparison and it is sensitive to segmentation errors too especially in real-world applications. Recently, Chen et al. [17] proposed RMLLC (R) model to solve the person Re-ID problem based on cross view gait recognition. As per the survey of various human detection algorithms, till now there is no algorithm which is robust to detect persons even in cluttered background and sudden illumination changes. The authors well known that the traditional methods are obsolete as compared with recent deep learning methods. However, recent deep learning algorithms need larger dataset for training is obvious. Traditional approaches characterizes a rapid and effective method for processing computer vision algorithms, whereas CNN need lakhs of training samples of known objects for effective training. The benchmark person Re-ID dataset under sudden illumination conditions is very few [18]. The recent datasets Market1501, MARS, Duke MTMC, MSMT17 are larger dataset, lacks challenging conditions like sudden illumination changes. Hence, the traditional approach, the combination of HoG and XCS-LBP descriptor, is proposed and attempted for further validation with deep learning. Re-ID datasets such as ETH and PETS 2009 have sudden illumination variations. However, these datasets are insufficient to implement with deep learning algorithms. Taking this idea into consideration, in this paper, the fusion of appearance and gait biometric (Skeleton Recurrent Motion Image (SRMI) with relevance metric learning for non-contextual long term person Re-ID is proposed. It is detailed in Sect. 3. The discussion is elaborated with results and discussion is conferred in Sect. 4. Finally, the work is summarized in Sect. 5.

3 Proposed Method

First, the person is detected by combining HOG and XCS-LBP feature vectors. Subsequently, tracking of multiple persons is carried out across multiple cameras using Kalman filter [20]. Followed that, the images of tracked persons are automatically cropped and saved. DPM [19] is used to model each body part of a tracked person. Further, the Gabor feature descriptor [21] is extracted from each body part. The fusion of appearance (Gabor feature) and Gait (Skeleton RMI) features are proposed for long term person Re-ID. Finally, the correct match of the given query image against Gallery is achieved by relevance metric learning approach [16]. The overall methodology of the proposed framework is illustrated in Fig. 1.

3.1 Person Detection by Fusing HOG and XCS-LBP and Multiple Person Tracking

The first and the foremost step is to detect multiple persons across multiple views. Hence, the fusion of HOG and XCS-LBP features is proposed since HOG provides person shape information and XCS-LBP is illumination invariant as well as provides texture information. Extension of Centre Symmetric Local Binary Pattern (CS-LBP) operator is performed by comparing the gray values of center symmetric pixel pair which is quite sensitive to quick illumination variations. It is expressed as:

$$XCS-LBP_{P,R}(C) = \sum\nolimits_{i=0}^{(p/2)-1} s(g_1(i,c) + g_2(i,c))2^i \tag{1}$$

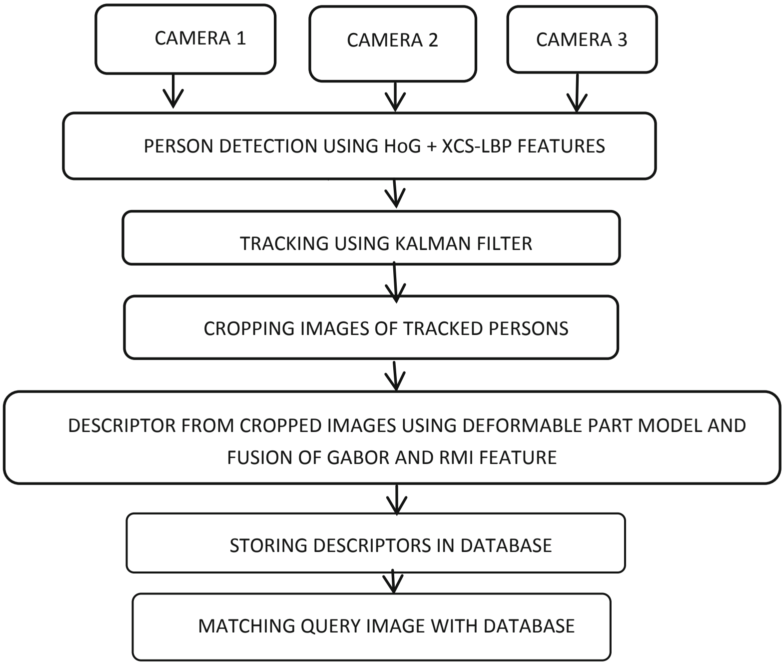

Fig. 1. The proposed methodology for person Re-ID for sudden illumination conditions.

$$XCS-LBP = \mathrm{s}((\mathrm{g_o}-\mathrm{g_4})+\mathrm{g_c}+(\mathrm{g_o}-\mathrm{g_c})(\mathrm{g_4}-\mathrm{g_c}))2^0 + \mathrm{s}((\mathrm{g_1}-\mathrm{g_5})+\mathrm{g_c}+(\mathrm{g_1}-\mathrm{g_c})(\mathrm{g_5}-\mathrm{g_c}))2^1 + \mathrm{s}((\mathrm{g_2}-\mathrm{g_6})+\mathrm{g_c}+(\mathrm{g_2}-\mathrm{g_c})(\mathrm{g_6}-\mathrm{g_c}))2^2 + \mathrm{s}((\mathrm{g_3}-\mathrm{g_7})+\mathrm{g_c}+(\mathrm{g_3}-\mathrm{g_c})(\mathrm{g_7}-\mathrm{g_c}))2^3 + \ldots\ldots..$$

Where $\mathrm{g_c}$ is the center pixel and $\mathrm{g_o}$, $\mathrm{g_1}$....... $\mathrm{g_7}$ are the neighborhood pixels. Subsequent to the detection of multiple persons, Kalman filter is used for tracking the labelled persons. It estimates and up-dates the new observation with minimum prediction (estimation) error [20]. At time k, each target has state and observation,

$$x_k = F_k x_{k-1} + w_k \tag{2}$$

Where $w_k \sim \mathrm{N}(0, Q_k)$ F_k is the state transition model which is applied to the previous state X_{k-1}; W_k is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_k.

$$y_k = H_k x_k + v_k \tag{3}$$

where, H_k is the observation model which maps the true state space into the observed space and V_k is the observation noise which is assumed to be zero mean Gaussian white noise with co-variance R_k.

3.2 Deformable Part Model

Subsequently, the image regions of tracked persons are only considered for further process. DPM [19] is adopted for modeling each part of the person's body. The advantages of DPM comprises the robustness to cluttered background changes since it will not include most of the background in the detection window as well as to partial inter-object occlusions besides less sensitive to different viewpoints and illumination changes [19]. Modeling of person's whole body into different parts provides high accuracy in feature extraction.

3.3 Gabor Feature Extraction

After modeling each person using DPM, Gabor feature is extracted for each part of a person's body [21]. In real scene, vertical change is smaller than horizontal change since viewpoint change of a pedestrian is around vertical axis. Gabor filters are orientation-sensitive that capture texture and edge information on an image. Gabor bank is formed by selecting Gabor filters with orientation equals to 0. Convolution of Gabor kernels with image I form the Gabor feature which is given as follows:

$$g_{u,v}(x, y) = I(x, y) * \varphi_{u,v}(x, y) \tag{4}$$

Where $\varphi_{u,v}(x, y)$ 2-D Gabor kernel, u and v is the orientation and scale of the kernel.

3.4 Skeleton Recurrent Motion Image

A specific feature vector called RMI is used to estimate repetitive motion behavior of moving objects. At pixels, where the motion occurred repetitively, RMI will have high values whereas at pixels with little or no motion occurred, RMI constitutes low values. Using (2) and (3), RMI is computed to determine the areas of moving object's silhouette with repetitive changes.

$$DS_a(x, y, t) = S_a(x, y, t-1) \oplus S_a(x, y, t) \tag{5}$$

$$RMI_a = \sum_{k=0}^{T} DS_a(x, y, t-k) \tag{6}$$

Where S_a is a binary silhouette for object at frame t, DS_a is a binary image which indicates areas of motion for object a between frame t and $t-1$ and $t-k$ is the RMI for object a calculated over T frames. Subsequently, to compute the average recurrence for each block, RMI is partitioned into N equal-sized blocks.

Hence, the fusion of appearance feature (Gabor) and Gait feature (Skeleton RMI) is proposed for long term person Re-ID across multiple cameras.

3.5 Relevance Metric Learning

After that, the framework utilizes person Re-ID algorithm, Relevance Metric Learning method with List wise Constraints (RMLLC) proposed by Chen et al. [17]. This algorithm treats the person Re-ID as an image retrieval task, and measures the similarity of two feature vectors by using their inner product, or, relevance, which is angle between two vectors. To overcome sparse pairwise limitations, similarity score list [17] is predefined for every probe image x_i and gallery set x_j through their inner product and initializing binary numbers for similar and dissimilar pairs.

4 Results and Discussion

Experimentations are carried out using benchmark datasets and the results are evaluated using Matlab 2018a. The PETS2009 dataset comprises multi-sensor sequences. This scenario contains one group who navigate around a stationary group. Here, PETS2009-S3-View 5, View 6, and View 7 datasets (Flow Analysis and Event Recognition) as well as VIPER, ETHZ, i-LIDS, CAVIAR4REID and TRECVid 2008 dataset are used for the evaluation of person Re-ID. The specifications of benchmark datasets in terms of frame rate, scenario with place, people size, and number of cameras employed are depicted in Table 1.

Table 1. Specifications of various benchmark dataset

Dataset	Image/Video	Frame rate	Place	People size	No. of cameras
VIPER	Still images	–	Outdoor surveillance	128×48	1
ETHZ	Video	15	Moving cameras on city street	13×30 to 158×432	1
i-LIDS	Video	25	Collection from different scenarios	21×53 to 176×326	5
Choke	Video	25	Corridor	Normal	1
CAVIAR4REID	Still images	–	Shopping centre	17×39 to 72×144	2
TRECVid 2008	Video	25	Gatwick International Airport –London	21×53 to 176×326	5
PETS2009	Video	7	Outdoor surveillance	26×67 to 57×112	8

Initially, HOG feature is used for person detection in PETS 2009 dataset. But, few shortcomings like when the person is just entering the field of view, the algorithm fails to detect that person. Also, 'false positives' occur in sudden illumination changing conditions. Such failure cases are illustrated in Fig. 2a and Fig. 2c respectively. These shortcomings overcome by the proposed framework and it is depicted in Fig. 2b and Fig. 2d. Subsequently, the appearance features (Gabor texture) are extracted from the DPM of the detected persons. The output of Gabor feature extraction is shown in Fig. 3 (d). The fusion of appearance (Gabor) and Gait (Skeleton RMI) features is done for long term person Re-ID across multiple cameras. It is illustrated in Fig. 3 (a, b, c, and d). Re-ID process is carried out by matching the probe image with the gallery images using RMLLC with rectification (RMLLC (R)) [16] by adopting list wise similarities, which consist of the similarity list of each image with respect to all remaining images.

The appearance features fail in case if a person wears different attire or two persons wear same color attire and it is illustrated in Fig. 4. From Fig. 4 (a) it is inferred that,

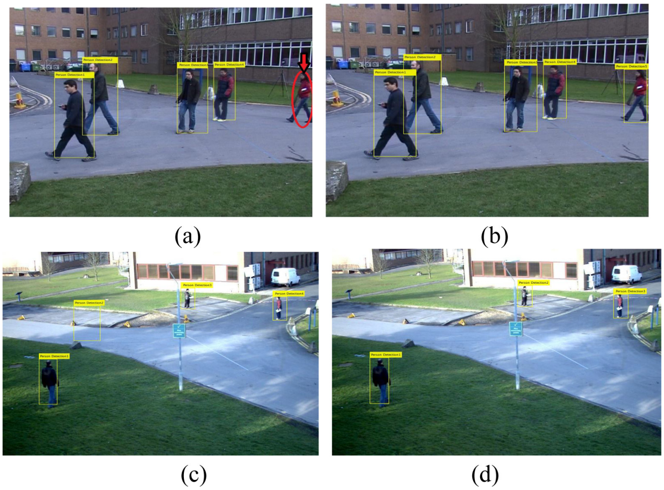

Fig. 2. Detection of multiple person across multiple views by fusion of HOG and XCS-LBP (a) Camera 1, (b) Camera 1 (c), Camera 2 (d) Camera 2.

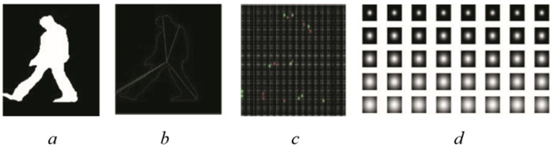

Fig. 3. (a) Silhouette sequence of person 1 in View - 1 (b) Skeleton features (c) Partitioned RMI of skeleton features (d) Gabor filter banks used and the output of Gabor feature extraction

Fig. 4. (a) 'Correct match' in case of person-23 in gallery image and person-23 in query image (Same color clothes) in i-LIDS dataset. (b) Re-ID of person-1 with gallery image (View-1) and person-2 with query image (View-2) is 'Wrong match' in i-LIDS dataset.

'Correct match' is correctly indicated in case of person-23 in gallery image of camera 1 and person-23 in query image of camera 2 are wearing same clothes for i-LIDS dataset using the proposed framework. Figure 4 (b) shows the performance of using Gabor feature under challenging scenarios and it displays 'Wrong match' for i-LIDS dataset. This shows the advantage of the proposed long term Re-ID using Gait features and fusing it with Gabor.

The performance metric widely used for person Re-ID is the Cumulative Matching Characteristics (CMC) rate. This metric is adopted where each image in the database is ranked based on its comparison to the query image. Various existing algorithms such as SDALF, Re-ID by Saliency and the proposed work are evaluated on benchmark datasets, given in Table 1 and the results obtained is depicted in Fig. 5. It is inferred that proposed framework with rank 20 provides better accuracy as compared with the state-of-the-art algorithms. The CMC rate obtained for the person Re-ID using GEI, Gabor, RMI, GEI + Gabor, SDALF, context-based methods and the proposed fusion framework of Gabor + Skeleton RMI on i-LIDS is illustrated in Table 2. The CMC rate obtained for person Re-ID compared with non-contextual state-of-the-art methods is depicted in Table 3. It is inferred that, the proposed framework significantly provides better performance than the state-of-the-art methods.

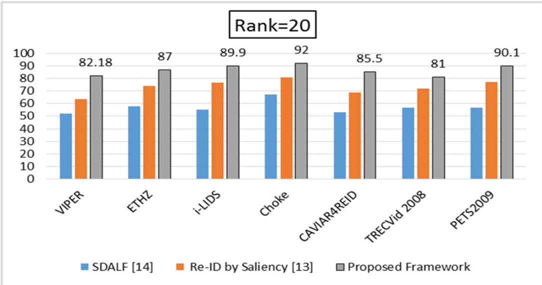

Fig. 5. CMC rate for the proposed person Re-ID framework and the state-of-the-art non contextual methods.

Table 2. The CMC rate obtained for person Re-ID using GEI, Gabor, RMI, GEI + Gabor and proposed RMI + Gabor for i-LIDS

Rank	r = 1	r = 5	r = 10	r = 15	r = 20
(2009) Context-based method [9]	16.15	39.71	48.12	54.52	60.13
(2013) SDALF	17.43	41.44	50.00	56.72	63.35
(2012) RMI	16.30	35.90	52.50	62.30	70.29
(2015) GEI + Gabor [17]	22.11	51.32	64.37	75.42	82.18
(2015) Gabor [17]	19.32	46.31	62.27	73.15	80.45
(2015) GEI [17]	13.27	31.52	49.15	58.55	67.18

Apart from this validation, the ablation study has been carried out with HOG, XCS-LBP, DPM, Gabor and Skeleton RMI. The entire proposed system is experimented with only HOG for person detection. Unfortunately, missing detection and false detection rate is high in sudden illumination conditions and near/far field of views and it is illustrated in Fig. 2a and Fig. 2c respectively. It has been overcome by the proposed fusion framework and it is illustrated in 2b and Fig. 2d. This experimentation validates the fusion of XCS-LBP and HOG as well as Gabor and skeleton RMI features for person detection and Re-ID.

Table 3. The CMC rate, Training Time (TT (ms)) and Testing Time (TS (ms)) for the proposed person Re-ID framework and the state-of-the-art non contextual methods for i-LIDS

Passive Approach	Features	Rank in % (Accuracy)				Computational Time	
		1	5	10	20	TT	TS
(2011) CPS [5]	Color	21.84	X	57.21	X	70.12	0.010
(2013) SDALF [14]	Color, Texture	X	X	X	50-60	80.18	0.012
(2012) Re-ID Attributes [10]	Color, Texture	15-32	43-53	36-63	X	90.03	0.017
(2013) Re-ID by Saliency [13]	Color, Texture	X	X	62.37	76.36	82.15	0.029
Proposed framework	Color, Texture, HOG	25.62	54.55	69.10	89.90	68.02	07.34

5 Conclusion

Non-contextual long term person Re-ID framework has been proposed to re-identify the person across multiple camera views under sudden illumination conditions. The combination of HOG and XCS-LBP improved the accuracy of person detection under sudden illumination conditions and at the same time it reduces the false positives. Subsequently DPM has been used on the detected persons. Afterwards, fusion of Appearance (Gabor) and Gait (Skeleton-RMI) features is proposed to overcome the homogeneity issue of distinguishing people even the texture and color of their attire are similar. When considering long term person Re-ID, possibility of wearing similar color and textured clothes by different persons is high. Hence, fusion of appearance and Gait features has been proposed. Even though deep learning booms, the limited dataset under the challenging conditions like sudden illumination encourages to h with traditional features. The future work aims to create such dataset in high dense crowd and to use deep learning to overcome severe occlusions.

References

1. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. Image Vis. Comput. **32**(4), 270–286 (2014)
2. Datta, A., Brown, L.M., Feris, R., Pankanti, S.: Appearance modeling for person re-identification using Weighted Brightness Transfer Functions. In: International Conference on Pattern Recognition, Tsukuba, Japan, pp. 2367–2370 (2012)
3. Chen, L., Chen, H., Li, S., et al.: Person re-identification by color distribution fields. J. Chi. Comput. Syst. **38**(6), 1404–1408 (2017)
4. Yang, M., Wan, W., Hou, L., et al.: Person re-identification using human salience based on multi-feature fusion. In: International Conference on Smart and Sustainable City and Big Data, Shanghai, China, pp. 1–5 (2016)

5. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.H.: Shape and appearance context modeling. In: International Conference on Computer Vision, Rio de Janiero, pp. 1–8 (2007)
6. Layne, R., Hospedales, T.M., Gong, S.: Towards person identification and re-identification with attributes. In: European conference on Computer Vision, Florence, Italy, pp. 402–412 (2012)
7. Vezzani, R., Balteiri, D., Cucchiara, R.: People reidentification in surveillance and forensics: a survey. ACM Comput. Surv. **46**(2), 29 (2013)
8. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. In: International Conference on Computer Vision and Pattern Recognition, CoRR abs/1307.5748 (2013)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, pp. 886–893 (2005)
10. Ma, Y., Chen, X., Chen, G.: Pedestrian detection and tracking using HOG and oriented-LBP features. In: International Conference on Network and Parallel Computing, Changsha, China, pp. 176–184 (2011)
11. Silva, C., Bouwmans, T., Frelicot, C.: An eXtended center-symmetric local binary pattern for background modeling and subtraction in videos. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Berlin, Germany, pp. 1–8 (2015)
12. Yang, L., Jin, R.: Distance metric learning: a comprehensive survey. Technical report, Michigan State University (2006)
13. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: International Conference on Computer Vision and Pattern Recognition, Washington, DC, pp. 649–656 (2011)
14. Yu, H.X., Wu, A., Zheng, W.S.: Symmetry-driven accumulation of local features for human characterization and re-identification. Comput. Vis. Image Underst. **117**(2), 130–134 (2013)
15. Chen, J., Wang, Y., Qin, J., et al.: Fast person re-identification via cross-camera semantic binary transformation. In: IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, pp. 3873–3882 (2017)
16. Liu, Z., Zhang, Z., Wu, Q., Wang, Y.: Enhancing person re-identification by integrating gait biometric. Neurocomputing **168**, 1144–1156 (2015)
17. Chen, J., Zhang, Z., Wang, Y.: Relevance metric learning for person re-identification by exploiting listwise similarities. IEEE Trans. Image Process. **24**(12), 1657–1662 (2015)
18. Li, W., Zhao, R., Xiao, T., et al.: Deepreid: deep filter pairing neural network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 152–159 (2014)
19. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2010)
20. Li, X., Wang, K., Wang, W., Li, Y.: A multiple object tracking method using Kalman filter. In: International Conference on Information and Automation, Harbin, pp. 1862–1866 (2010)
21. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Process. **11**(4), 467–476 (2002)