Online Context-based Person Re-identification and Biometric-based Action Recognition for Service Robots

Ning An[1,2], Shi-Ying Sun[1,2], Xiao-Guang Zhao[1], Zeng-Guang Hou[1]

1. State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
E-mail: ning.an@ia.ac.cn

2. University of Chinese Academy of Sciences, Beijing, 100049, China

Abstract: In this paper, we address the problem of person re-identification and action recognition for service robots, which undergoes lack of training dataset for model learning, reduction of feature set discriminative power in changing scenarios, and high complexity of the algorithm computation. An online context-based person re-identification algorithm is proposed, which learns the person model online without pre-collect dataset and adjusts the weight of features according to the context information. An online biometric-based action recognition algorithm is proposed, actions are recognized by simply matching the skeleton vectors extracted from five linkage mechanisms of human body. The proposed algorithms are evaluated on a service robot system, extensive experimental results show that they performs efficiently and effectively in various real-life scenarios.

Key Words: person re-identification, action recognition, online processing, human-robot interaction, service robot

1 INTRODUCTION

Service robot is one of the most important research field in robotics. The application scenarios of service robots are abundant such as auxiliary transportation, security surveillance, environmental cleanup and health monitoring. Since service robots have to interact with human, the key components of service robots are person re-identification and action recognition. The person re-identification component enables the robot to confirm the person it needs to serve, the action recognition component enables the robot to understand what is the requirement of the person. In this paper, we focus on the problem of person re-identification and action recognition for service robots.

Significant progresses have been achieved in the computer vision domain, however, person re-identification and action recognition are still impractical in robotics on account of the high computation complexity of algorithm and limited description power of 2D images. Recently, RGB-D sensors, e.g. Microsoft Kinect have been employed in computer vision tasks include person re-identification and action recognition and shown great performance improvement [1]. The additional depth cue provides the opportunity to reduce the algorithm complexity and enhance the description power of the features, which makes it possible to put the person re-identification and action recognition algorithms into use in robotics.

Despite the advantages mentioned above, even for RGB-D sensors there are still issues occur in service robot usage person re-identification and action recognition algorithm. For person re-identification, an dataset of the specified persons is needed for model offline learning. However, the dataset is hard to obtain for service robots since the robot has to serve the target person immediately when receives the command, there is no time for the robot to collect samples and learn model of the target person. In addition, the features of person are defined ahead of re-identification, it's hard for the pre-defined features to keep high description power along the service progress with various scenario changes, e.g. illumination variation, which may cause re-identification failure because weak features will make different persons look like the same one. For action recognition, the dataset collection problem still exists. What's more, action recognition algorithm are always time consuming because of the complex model design and unsuited for real-time computation. In contrast, the service robot system should react to the person activity right away. For instance, the health monitoring service robot should give alarm signal once danger condition occurs to obtain timely rescue.

In this work, we propose an online context-based person re-identification algorithm (OCPR) and an online biometric-based action recognition algorithm (OBAR) for service robots to address the above issues. The online context-based person re-identification algorithm learns the model of the target person online, dataset of the person is collected online during the service progress instead of offline before the progress begin. Both 2D appearance features and 3D shape features are employed for target modeling. The weights of each feature are adjusted online according to the context of the scenario to dynamically increase the importance of the higher discriminative features and decrease the importance of lower ones, by which the feature set performs robustly to scenario changes. The on-

This work is supported by the National Natural Science Foundation of China under Grant 61271432, 61673378, 61421004.

line biometric-based action recognition algorithm models the activities using a 3D skeleton vector set, only one snapshot of the activity is needed for model learning instead of offline dataset collection, which can be collected easily during service progress. Both static and dynamic activities can be efficiently and effectively recognized via simple vector set matching.

The main contributions of this paper are:

(i) An online context-based person re-identification algorithm is proposed, in which the model of person is learnt online without offline dataset and the feature set is dynamically adjusted to keep robust discriminative power against scenario changes.

(ii) An online biometric-based action recognition algorithm is proposed, efficient and effective action recognition is achieved by simple snapshot learning and 3D skeleton vector set matching.

(iii) The proposed algorithms are evaluated on a real service robot system, their efficiency and effectiveness are demonstrated in various real-life scenarios.

The remainder of this paper is organized as follows. In Section 2, we review researches related to our work. Section 3 introduces the proposed online context-based person re-identification algorithm. Section 4 introduces the proposed online biometric-based action recognition algorithm. Section 5 presents the experimental evaluation results of the proposed algorithms and conclusions are given in Section 6.

2 RELATED WORKS

In this section, we review the person re-identification and action recognition methods closely related to our work.

2.1 Person Re-identification

Person re-identification can be divided into two components: *pedestrian description* and *distance metric learning*. For pedestrian description, Gray *et al.* [2] use AdaBoost algorithm to learn the best person representation from a defined feature space with different kinds of simple features instead of designing a specific feature by hand. Farenzena *et al.* [3] propose three features that model complementary aspects of person appearance: the chromatic information, the structural information through uniformly colored regions, and the natural information of recurrent patches to achieve robustness to viewpoint, pose and illumination variations. Pedagadi *et al.* [4] reduce the high dimensional of color features by using the unsupervised PCA and Local Fisher Discriminant Analysis. Zhao *et al.* [5] learn view-invariant and discriminative mid-level features instead of existing handcrafted low-level features for pedestrian description via a cross-view training strategy. Yang *et al.* [6] propose a novel salient color names based color descriptor, over which color distributions in different color spaces are obtained and fused to generate the final description feature. Matsukawa *et al.* [7] employ hierarchical Gaussian distribution to describe a local patch, both the mean and the covariance information of pixel features can be modeled properly.

For distance metric learning, Xiong *et al.* [8] propose four methods to learn a discriminative subspaces: regularized Pairwise Constrained Component Analysis, kernel Local Fisher Discriminant Analysis, Marginal Fisher Analysis and a ranking ensemble voting scheme. Chen *et al.* [9] use explicit polynomial kernel feature map and a mixture of similarity functions to maximize the difference between matched and unmatched images for a same person. Liao *et al.* [10] propose a logistic metric learning approach with the PSD constraint to smooth the solution of the metric and an asymmetric sample weighting strategy to address the positive and negative sample pairs unbalanced problem. Chen *et al.* [11] learn a similarity function with multiple subregion similarity measurements. Polynomial feature map is employed to describe each subregion matching. Hirzer *et al.* [12] reduce the computational cost of person matching by relaxing the original hard constraints, a pairwise metric learning method to make use of the structure of the data.

2.2 Action Recognition

Action Recognition can be classified into three categories: skeleton-based, depth map-based, and hybrid methods. Skeleton-based methods model the action of human body by using the 3D position of skeleton joints. Xia *et al.* [13] propose histograms of 3D joint locations to represent human postures. Posture visual words are generated by reprojecting histograms using linear discriminant analysis, and modeled by discrete hidden Markov models. Yang *et al.* [14] combine static posture, motion, and offset action information to generate action recognition features, Naive-Bayes-Nearest-Neighbor classifier is employed for multi-class action classification. Devanne *et al.* [15] formulate the action recognition into the problem of shape and trajectories similarity computation in a Riemannian manifold, and employ k-nearest neighbors on the manifold for classification. Depth map-based methods directly extract spatio-temporal action features from depth frames.

Li *et al.* [16] model the dynamics of actions using an action graph, salient postures correspond to the nodes are characterized by a bag of 3D points. A projection based sampling scheme is proposed to sample the bag of 3D points from the depth maps. Yang *et al.* [17] represent an action video using Histograms of Oriented Gradients extracted from the depth motion maps, which is generated by projecting depth maps onto three orthogonal planes and accumulate global activities through entire video sequences.

Hybrid methods employ information from both the skeleton joints and depth maps. Wang *et al.* [18] learn an actionlet ensemble model to represent each action. A local occupancy pattern is proposed, and the pattern is associated with each 3D joint. Ohn-Bar *et al.* [19] characterize actions using pairwise affinities between view-invariant joint angles features and histogram of oriented gradients extracted on depth maps.

Table 1: Features for online context-based person re-identification

Feature	Feature Symbol	Feature Description	Feature Type	Distance Metric
$\mathbf{f}_1$	$\mathbf{H}_u$	histogram of upper body	color	correlation distance
$\mathbf{f}_2$	$\mathbf{H}_l$	histogram of lower body	color	correlation distance
$\mathbf{f}_3$	$\mathbf{T}$	template of body	texture and edge	correlation distance
$\mathbf{f}_4$	$\mathbf{L}_b$	length of body	physical size	Gaussian distribution
$\mathbf{f}_5$	$\mathbf{L}_w$	length of wingspan	physical size	Gaussian distribution

3 ONLINE CONTEXT-BASED PERSON RE-IDENTIFICATION

The task of person re-identification is to identify the specified person among the detected persons in an image or video. Model of the target person should be trained before re-identification process. For service robots, it is hard to collect the training dataset beforehand like traditional methods since the target person is random assigned. Our online context-based person re-identification algorithm (OCPR) learns and updates the target model online with continuous video frames collected during the robot operation progress. Specifically, for a person-following auxiliary transportation robot, the model of the target person is learned and updated using the data captured by the robot's camera while the robot moving followed the target.

To accurately describe the person, we employ five features ranging from 2D appearance feature to 3D shape feature in our OCPR algorithm. Details of the feature descriptors are shown in Table 1. The $\mathbf{H}_u$, $\mathbf{H}_l$ and $\mathbf{T}$ are 2D appearance features: $\mathbf{H}_u$ and $\mathbf{H}_l$ describe the color information of the upper human body and lower human body respectively, $\mathbf{T}$ describes the texture and edge information of the human body. $\mathbf{L}_b$ and $\mathbf{L}_w$ are 3D shape features: $\mathbf{L}_b$ describes the physical height of the human body and $\mathbf{L}_w$ describes the physical wingspan length of the human body. For an query person q, the feature set is denoted by $Q = \{\mathbf{f}_1^q, \ldots, \mathbf{f}_5^q\}$. The model of the specified target person is denoted by $M = \{\mathbf{f}_1^m, \ldots, \mathbf{f}_5^m\}$. In feature matching, the similarity between two features is calculated as:

$$s(\mathbf{f}_i^q, \mathbf{f}_i^m) = \begin{cases} \frac{(\mathbf{f}_i^q - \overline{\mathbf{f}_i^q})(\mathbf{f}_i^m - \overline{\mathbf{f}_i^m})^T}{\|(\mathbf{f}_i^q - \overline{\mathbf{f}_i^q})\|_2 \|(\mathbf{f}_i^m - \overline{\mathbf{f}_i^m})\|_2}, & i \in \{1,2,3\} \\ \exp(-\frac{(\mathbf{f}_i^q - \mathbf{f}_i^m)^2}{2\sigma_i{}^2}). & i \in \{4,5\} \end{cases} \tag{1}$$

For $\mathbf{H}_u$, $\mathbf{H}_l$ and $\mathbf{T}$, correlation distance is employed as the appearance similarity metric. For $\mathbf{H}_u$ and $\mathbf{H}_l$, we assume the physical size measurement of human body belongs to a Gaussian distribution. The re-identification confidence of the query person is calculated as:

$$c(Q, M) = \sum_{i=1}^{5} w_i s(\mathbf{f}_i^q, \mathbf{f}_i^m), \tag{2}$$

where w_i denotes the weight of feature $\mathbf{f}_i$. If the $c(Q, M)$ exceeds the threshold η and it is the max confidence in the scene, the query person is re-identified as the target.

Consider the scenario changes during robot operation, the discriminative power of each feature varies in different scenarios. Specifically, in a scenario where persons wear uniforms, the color and texture properties of different persons are nearly the same and the 2D appearance features will lose discriminative power, however, the height and wingspan length may be distinctly different between persons and the discriminative power of 3D shape features will maintain. On the contrary, in a scenario where persons have similar stature, the height and wingspan length are nearly the same but the color and texture properties may have strikingly difference, the discriminative power of 2D appearance features are stronger than 3D shape features. Therefore, our OCPR algorithm adjusts the weight of each feature online:

$$w_i = \frac{1 - s(\mathbf{f}_i^o, \mathbf{f}_i^d)}{\sum_{i=1}^{5} 1 - s(\mathbf{f}_i^o, \mathbf{f}_i^d)}, \tag{3}$$

where $\mathbf{f}^o$ denotes the feature of the target person, $\mathbf{f}^d$ denotes the feature of the distractor in the context, which is defined as non-target person with maximum similarity appears in the camera view. For a feature, the lower similarity between target and distractor it achieves, the higher discriminative power and weight it gets.

The model M is updated online during robot operation:

$$\mathbf{f}_i^m = (1 - \gamma)\mathbf{f}_i^{m'} + \gamma \mathbf{f}_i^o, \tag{4}$$

where $\mathbf{f}^o$ denotes the feature of the target person, $\mathbf{f}^{m'}$ denotes the model before update and $\mathbf{f}^m$ denotes the model after update.

4 ONLINE BIOMETRIC-BASED ACTION RECOGNITION

The online biometric-based action recognition algorithm (OBAR) use biometric information to model the human activities, which is extracted from the skeleton joints shown in Figure 1. Consider the connection properties of the human body, we use 3D skeleton vectors to respectively describe the linkage mechanism of the four limbs and torso of the human body. For each linkage mechanism, two skeleton vectors are extracted on both sides of the joint, relative space positions of the linkage mechanism are accurately described by the corresponding skeleton vectors. Ten skeleton vectors extracted from five linkage mechanisms of human body, start and end point of each vector are shown in Table 2.

For static pose recognition, OBAR use the skeleton vector set mentioned above to model the human pose, which is denoted by $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_{10}\}$. The state of whether a query pose $V^q = \{\mathbf{v}_1^q, \ldots, \mathbf{v}_{10}^q\}$ belongs to a specific pose $V^s = \{\mathbf{v}_1^s, \ldots, \mathbf{v}_{10}^s\}$ is calculated as:

$$r(V^q, V^s) = \prod_{j=1}^{10} k(\mathbf{v}_j^q, \mathbf{v}_j^s), \tag{5}$$

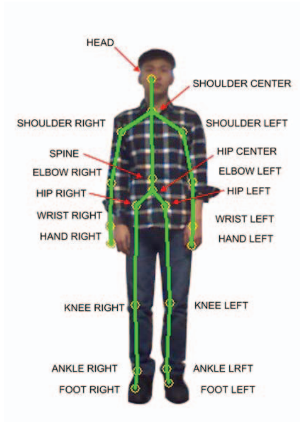

Figure 1: Skeleton joints of human body.

Figure 2: Service robot for experimental evaluation.

Table 2: Skeleton vectors for online biometric-based action recognition

Vector	Start Point	End Point
$\mathbf{v}_1$	head	shoulder center
$\mathbf{v}_2$	shoulder center	hip center
$\mathbf{v}_3$	shoulder right	elbow right
$\mathbf{v}_4$	elbow right	wrist right
$\mathbf{v}_5$	shoulder left	elbow left
$\mathbf{v}_6$	elbow left	wrist left
$\mathbf{v}_7$	hip right	knee right
$\mathbf{v}_8$	knee right	ankle right
$\mathbf{v}_9$	hip left	knee left
$\mathbf{v}_{10}$	knee left	ankle left

where $k(\mathbf{v}_j^q, \mathbf{v}_j^s)$ estimates the similarity between each skeleton vector using threshold θ_j:

$$k(\mathbf{v}_j^q, \mathbf{v}_j^s) = \begin{cases} 1, & \|\mathbf{v}_j^q - \mathbf{v}_j^s\|_2 < \theta_j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

If the distance between the query skeleton vector and the specific skeleton vector is within limit, they are regarded as the same pose. The threshold θ_j is self-adapted according to the physical size to make OBAR robust to different persons, e.g. θ_j is relatively high for an adult with larger physical size and relatively low for a child with smaller physical size. In practice, the skeleton vectors are easy to obtain and the computation complexities of Eq.(5) and Eq.(6) are low for real-time processing.

For dynamic action recognition, OBAR recognizes most dangerous actions in health monitoring case: tumble and faint. The head joint is selected for tumble and faint recognition since if we regard the human body as an inverted pendulum system, head is the farthest point to the system joint hence can reflect the motion of the system with largest margin. At time t, the state of human can be estimated as:

$$\text{state} = \begin{cases} \text{tumble}, & g(\mathbf{h}_y^{t-\Delta t_1}) < \lambda_t \wedge |g(\mathbf{h}_y^t)| < \lambda_c \\ \text{faint}, & |g(\mathbf{h}^{t-\Delta t_2})| < \lambda_c \wedge |g(\mathbf{h}^t)| < \lambda_c \\ \text{normal}, & \text{otherwise}, \end{cases} \quad (7)$$

where $\mathbf{h}^t = (\mathbf{h}_x^t, \mathbf{h}_y^t, \mathbf{h}_z^t)$ denotes the 3D position of head joint at time t. $g(\mathbf{h}^t)$ denotes the gradient (velocity) of head joint at time t, $\lambda_t < 0$ and $\lambda_c < 0$ denote thresholds. If the height of head joint first declines sharply and then keep steady, the tumble action is detected; if the 3D position of head joint keep steady in a long period, faint action is detected. For a health motoring service robot, human can be rescued in time with the immediate dangerous dynamic actions recognition.

5 EXPERIMENTS

We evaluate the proposed algorithms on a service robot operating in real-life scenarios. First, we evaluate the proposed person re-identification method with videos captured by the robot on various scenarios. Then the proposed action recognition method is validated with both static and dynamic actions.

5.1 Experimental Setups

The service robot is equipped with a Microsoft Kinect sensor and a laptop with an Intel I5-2400 3.10 GHz CPU, the drive unit is realized by using a Pioneer-3DX mobile robot. Frame rate of the proposed algorithms is $10fps$, which is practical for real time processing.

For online context-based person re-identification algorithm, HSV color space is employed for color histogram computation; gray patch of the target is extracted for template computation; length of body is computed by accumulating the distance of skeleton joints from head to food, and length of wingspan is computed by accumulating the distance of skeleton joints from hand right to hand left. The standard deviations σ_4 and σ_5 of Gaussian distribution are set to 0.3, the model updating rate γ is set to 0.2.

For online biometric-based action recognition algorithm, the time interval Δt_1 for tumble recognition is set to $2s$ and the time interval Δt_2 for faint recognition is set to $2s$. Thresholds λ_t and λ_c are set to -1 and 0.1 respectively.

5.2 Evaluation of the Proposed Person Re-identification Algorithm

To evaluate the proposed online context-based person re-identification algorithm (OCPR), thousands of frames are collected in various scenarios such as hall and corridor (shown in Figure 4) by the robot during its serving process.

To demonstrate the effectiveness of the proposed OCPR algorithm, we implement an additional algorithm without weight adjustment and model updating mechanisms of OCPR. The success rate curve is employed for performance measurement, whose horizontal axis indicates the threshold of re-identification varying from 0 to 1. The closer the curve approaches to the top right point $(1.0, 1.0)$, the better the algorithm performs.

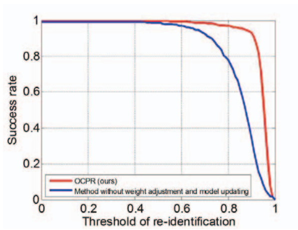

Figure 3: Success rate curve of the proposed OCPR algorithm. Weight adjustment and model updating mechanisms make OCPR robust to scenario changes.

As shown in Figure 3, OCPR performs much better than the algorithm without weight adjustment and model updating. It is because in a scenario with changes, discriminative power of features varies accompany with changes of the sensing environment, meanwhile, the pre-trained model is easy to lose effectiveness due to the appearance variation of the target. The context-based weight adjustment mechanism of OCPR enable its feature set keep high discriminative power in different scenarios, the online updating method makes OCPR update the target model in time to adapt to appearance changes. Figure 4 shows the qualitative evaluation of the proposed algorithm, OCPR performs consistently in various scenarios.

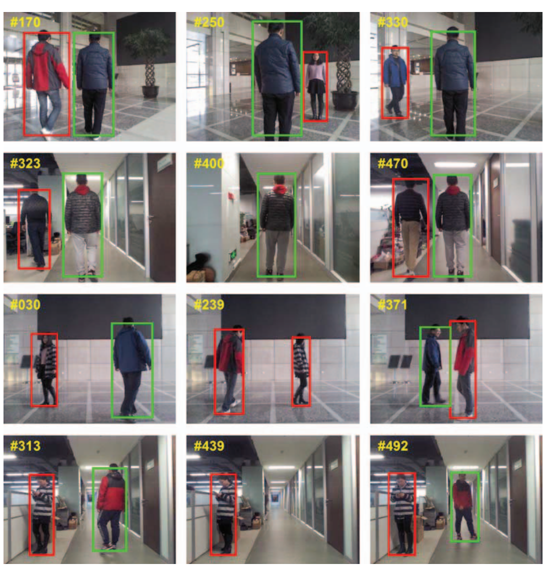

Figure 4: Qualitative evaluation of the proposed OCPR algorithm. Each row indicates one robot working process in real-life scenarios (auxiliary transportation in *row 1, 2* and health monitoring in *row 3, 4*). The target person under service is marked by green bounding box.

5.3 Evaluation of the Proposed Action Recognition Algorithm

To evaluate the proposed online biometric-based action recognition algorithm (OBAR), we measure its performance of static action recognition and dynamic action recognition respectively. Three persons with different physical size are selected for comprehensive evaluation. For static action recognition, we use the normally left arm raising action to test the algorithm, evaluaton results are shown in Figure 5. OBAR algorithm recognizes the static action accurately because of the effectively designed biometric skeleton vectors and effectively matching method.

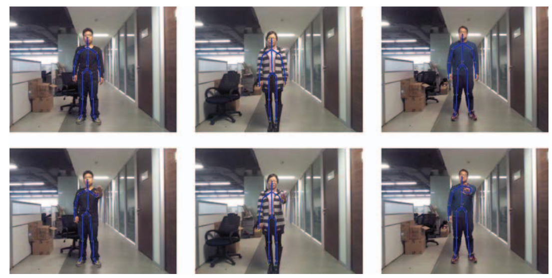

Figure 5: Static action recognition (left arm raising) of the proposed OBAR algorithm on three persons. The first row indicates the un-recognized state and second row indicates the recognized state.

For dynamic action recognition, tumble action and faint action are selected to evaluate the proposed algorithm, which are two most dangerous actions in health monitoring scenario. Evaluation results are shown in Figure 6 and Figure 7. OBAR recognizes the dynamic actions immediately since it models the substantive characteristics of the actions and matches the actions with a simple but effective manner. Low computation complexity is realized for real-time applications.

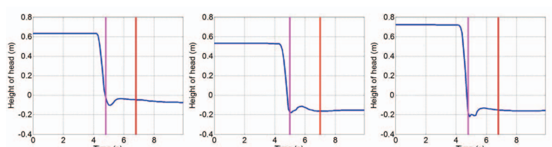

Figure 6: Dynamic tumble recognition of the proposed OBAR algorithm on three persons. The blue curve indicates the height variation of head. The purple and red lines indicate the start and validation time nodes of OBAR.

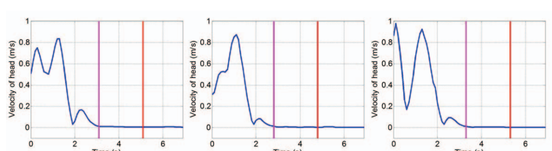

Figure 7: Dynamic faint recognition of the proposed OBAR algorithm on three persons. The blue curve indicates the velocity variation of head. The purple and red lines indicate the start and validation time nodes of OBAR.

6 CONCLUSIONS

In this paper, we propose an online context-based person re-identification algorithm (OCPR) and an online biometric-based action recognition algorithm (OBAR) for service robots. OCPR online learns the model of target person and adjusts the weights of 2D appearance and 3D shape features according to the context information, the robot can re-identification the target person without the difficultly-acquired dataset and the employed feature set can keep discriminative power in various scenarios. OBAR recognizes actions by matching the biometric properties of the five human body linkage mechanisms, the low computation complexity enables it for real-time applications. We evaluate the proposed OCPR and OBAR algorithms on a service robot system, extensive experimental results demonstrate their efficiency, effectiveness and robustness in various real-life scenarios.

REFERENCES

[1] J. Han, L. Shao, D. Xu and S. Jamie, Enhanced computer vision with microsoft kinect sensor: A review, IEEE Transactions on Cybernetics, Vol.43, No.5, 1318-1334, 2013.

[2] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, Proceedings of the European Conference on Computer Vision (ECCV), 262-275, 2008.

[3] M. Farenzena, L. Bazzani and A. Perina *et al.*, Person re-identification by symmetry-driven accumulation of local features, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2360-2367, 2010.

[4] S. Pedagadi, J. Orwell and S. Velastin *et al.*, Local fisher discriminant analysis for pedestrian re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3318-3325, 2013.

[5] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 144-151, 2014.

[6] Y. Yang, J. Yang, J. Yan and S. Liao *et al.*, Salient color names for person re-identification, Proceedings of the European Conference on Computer Vision (ECCV), 536-551, 2014.

[7] T. Matsukawa, T. Okabe and E. Suzuki *et al.*, Hierarchical gaussian descriptor for person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1363-1372, 2016.

[8] F. Xiong, M. Gou, O. Camps and M. Sznaier *et al.*, Person re-identification using kernel-based metric learning methods, Proceedings of the European Conference on Computer Vision (ECCV), 1–16, 2014.

[9] D. Chen, Z. Yuan and G. Hua *et al.*, Similarity learning on an explicit polynomial kernel feature map for person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1565-1573, 2015.

[10] S. Liao and S. Z. Li, Efficient psd constrained asymmetric metric learning for person re-identification, Proceedings of the IEEE International Conference on Computer Vision (ICCV), 3685-3693, 2015.

[11] D. Chen, Z. Yuan and B. Chen *et al.*, Similarity learning with spatial constraints for person re-identification, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1268-1277, 2016.

[12] M. Hirzer, P. M. Roth and M. Kostinger *et al.*, Relaxed pairwise learned metric for person re-identification, Proceedings of the European Conference on Computer Vision (ECCV), 780-793, 2012.

[13] L. Xia, C. C. Chen, J. K. Aggarwal, View invariant human action recognition using histograms of 3d joints, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 20-27, 2012.

[14] X. Yang, Y. L. Tian, Eigenjoints-based action recognition using naive-bayes-nearest-neighbor, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 14-19, 2012.

[15] M. Devanne, H. Wannous and S. Berretti *et al.*, 3-D human action recognition by shape analysis of motion trajectories on Riemannian manifold, IEEE transactions on cybernetics, Vol.45, No.7, 1340-1352, 2015.

[16] W. Li, Z. Zhang and Z. Liu, Action recognition based on a bag of 3d points, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 9-14, 2010.

[17] X. Yang, C. Zhang, Y. L. Tian, Recognizing actions using depth motion maps-based histograms of oriented gradients, Proceedings of the ACM international conference on Multimedia, 1057-1060, 2012.

[18] J. Wang, Z. Liu, Y. Wu and J. Yuan, Mining actionlet ensemble for action recognition with depth cameras, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1290-1297, 2012.

[19] E. Ohn-Bar and M. Trivedi, Joint angles similarities and HOG2 for action recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 465-470, 2013.