Proceedings of the 2016 IEEE
International Conference on Robotics and Biomimetics
Qingdao, China, December 3-7, 2016

Human Recognition for Following Robots with a Kinect Sensor

Shiying Sun, Ning An, Xiaoguang Zhao and Min Tan

Abstract— **In this paper, a human recognition method based on soft biometrics is proposed for the human following robot. Two soft biometric traits (clothes color and body size) are calculated as features of the human. First, the human region detected by the Kinect is segmented to obtain the torso and leg parts of the body. Then the weighted HSV histograms of the body parts are calculated to describe the clothes color information. Body height, arm length and shoulder width values are measured to represent the body size information. Finally, human recognition is implemented by evaluating the similarity among the objects. The effectiveness and robustness of the proposed method is verified by the experiments. And the robot system with the recognition method can recognize and follow a human target reliably.**

I. INTRODUCTION

Human detection and recognition are basic and important tasks for service robots. The RGB-D camera Kinect [1] have provided a human skeleton tracking method which can recognize up to six users simultaneously. And each of the detected human is assigned a TrackingID. With the tracker of Kinect, human detection and tracking could be easily implemented. For the human following robot, the 3D positions of the tracked humans can be used as the input of the robot controller. However, the tracker of Kinect merely solves the local tracking problem, which means if the human is lost from view during tracking, the TrackingID of the same human changes and the target tracking fails. Therefore, the ability to recognize and re-identify the target is of great importance for the human following robot.

In recent years, many relative works are proposed to solve the human recognition and re-identification problems. Some of the methods are based on modeling the color or shape of the objects. In [2], Corvee *et al.* proposed a new appearance model which applied spatial pyramid scheme to capture the correlation between human parts in order to obtain a discriminative signature. Gray and Tao [3] introduced an ensemble of localized features method in which they used AdaBoost to find the best representation. Ma *et al.* [4] transformed the local descriptors into the Fisher Vector and then extracted the global features for object recognition. In [5] and [6], a multi-shot appearance model was proposed to condense a set of human samples into a highly informative signature. Moreover, some commonly used features, such as SURF [7], SIFT [8], HOG [9], Haar [10] and LBP [11], are adopted in human recognition domain. The recognition problem has also been regarded as a discriminate model learning problem. A metric learning framework was used in [12] to obtain a robust metric for large margin nearest neighbor classification with rejection. The method in [13] reformulated the problem as a ranking problem and developed a novel Ensemble RankSVM to rank probe images. In [14], the appearance matching problem was formulated as the task of learning a model with most descriptive features. Camera transfer approaches have also been proposed to learn metrics from pairs of samples from different cameras [15] [16]. However, these methods are either not robust to the variation of pose and appearance or too complex to be implemented on the mobile robot platform with constrained resource.

Soft biometrics are personal characteristics that are easily measurable in a distance without any user cooperation [26]. Soft biometrics commonly used for human recognition include gait [17], body weight [18], height [19], body size, facial features [20], clothes color and so on. Southwell *et al.* [22] proposed a human recognition approach using the color information of the shirts. Usually, two or more soft biometrics are used to improve the recognition accuracy. Denman *et al.* [23] proposed three part height and color soft biometric models to represent a person. In [24], face and body information were fused for person recognition. 3D soft biometric cues including skeleton-based features and surface based features based on RGB-D sensors are proposed in [25]. Dantcheva *et al.* [21] proposed an idea of Bag of Soft Biometrics (BoSB) that was inspired by the principle of Bag of Words with combining several traits. And a method combining body weight and fat measurements was introduced in [26]. Considering the following robot with a Kinect, the faces of human targets are not always captured and the gait based methods are computationally too expensive. On the contrary, color of clothes is a very remarkable characteristic. Besides, with the Kinect SDK, the joints information can be used to calculate the body size information directly. Therefore, this paper utilizes both clothes color and body sizes for human recognition.

In this paper, a human recognition method is proposed based on the soft biometrics traits. This method is applied on a human following robot system to recognize the enrolled target after the target is lost. Clothes color and body size values are obtained based on information gathered from the Kinect and matched to determine if a candidate human is the enrolled one. The proposed method includes four phases: first, the human region detected by the Kinect is segmented

This work was supported by the National Natural Science Foundation of China (61271432, 61673378, 61421004, 61633020) and State Grid Corporation of China Technology Projects (52053015000G), and in part by the Beijing Natural Science Foundation (4161002).

Shiying Sun, Ning An, Xiaoguang Zhao and Min Tan are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with University of Chinese Academy of Sciences, Beijing 100049, China.

Email: {sunshiying2013, ning.an, xiaoguang.zhao, min.tan}@ia.ac.cn

to obtain the regions of shirts and trousers from the image; second, a weighted HSV histogram is calculated to represent the color feature of clothes; then body height, arm length and shoulder width values are measured to represent body size information; finally, similarities of color and body size are fused using a weighting approach and the weighted similarity between the human target and candidate is evaluated for human recognition and identification.

The organization structure of the remainder of this paper is as follows, In Section II, the detail of the proposed human recognition method is presented. Section III shows the human following robot system. The experimental results are shown in Section IV. Finally, conclusions are presented in Section V.

II. HUMAN RECOGNITION METHOD BASED ON SOFT BIOMETRICS TRAITS

The core problem of human recognition is how to represent the human object. In this paper, two soft biometrics (clothes color and body size) are used to describe the features of humans. The proposed recognition method involves four steps: human body region segmentation, color feature extraction, body size feature extraction and feature fusion.

A. Human Region Segmentation

A human body can be divided into three portions: head, torso and legs. And the clothes color information is contained in the torso and legs regions. To acquire more discriminative features to recognize a human, we extract the color features of the two parts respectively. Therefore the human region is segmented firstly.

The Kinect can provide the RGB-D image and a skeleton tracker which can be used to detect humans and determine their positions. Coordinates of 3D positions for 20 skeletal joints are obtained in real-time using the depth sensor. The position of joints is shown in Fig. 1 (a). We can crop the human regions by checking the playerID of every pixels in depth image. If a playerID is not equal to zero, it means this pixel belongs to the region of a detected human. Then boundaries are obtained from 3D positions of the joints. Locations of boundaries in RGB image coordinates are determined by transforming the 3D positions of shoulder center and left hip among 20 skeletal joints. The segmentation result is shown in Fig. 1. In the following process, we focus on the torso and leg regions.

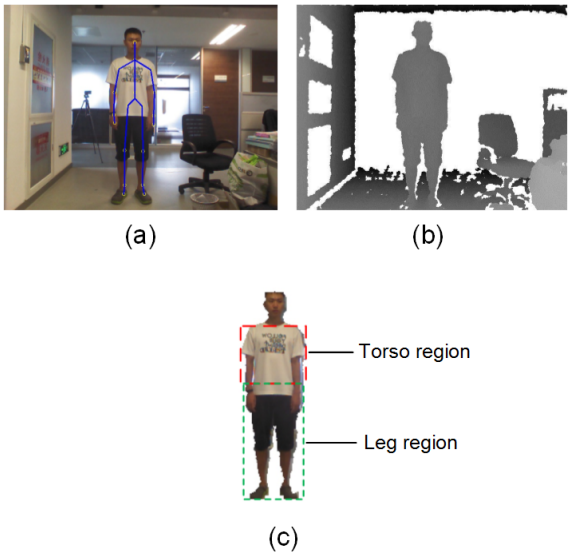

Fig. 1. (a) The original RGB image with the detected skeletons. (b) The depth image. (c) The cropped human body region.

B. Color Feature Extraction

Generally, the appearance of the human is supposed to be invariable in a certain time for the following robot. So the color is the most popular and effective information for human recognition. However, there is no spatial information in traditional color histogram which limits the description ability. In this article, we adopt a weighted color histogram in which the spatial position of a pixel is regarded as its weight. As RGB color space has shortages to distinguish objects [28]. To handle varying lighting conditions by partially decoupling the chromatic and achromatic information, the segmented regions are transformed from RGB color space into an HSV color space representation and only hue and saturation channels are used for histogram computing.

The weighted HSV color histogram in kth region is represented by R^k, and each dimension in the histogram can be expressed as:

$$R^k_{h,s} = \sum_j \sum_i \delta^{h,s}_{i,j} \eta Q_{i,j} \tag{1}$$

where $k \in \{1, 2\}$ denotes the torso and legs sub-regions, $h \in \{1, 2, \cdots, H\}$, H is the number of bins for the hue channel, $s \in \{1, 2, \cdots, S\}$, S is the number of bins for the saturation channel. $\delta^{h,s}_{i,j} = 1$, if a pixel $p_{i,j}$ is within the sub-region k and its hue value is within h and its saturation value is within s; and $\delta^{h,s}_{i,j} = 0$ otherwise. $Q_{i,j}$ is the weight of $p_{i,j}$, and it depends on the distance between the pixel to the gravity center of the sub-region. It means that the pixels near the gravity center will have stronger weight. And η is a constant and is applied to normalize $Q_{i,j}$. Spatial information is integrated into color histogram which will bring more robustness. The dimension of R^k is determined by bins of H and S. To measure the similarity of color information, we compare the histograms using the histogram intersection approach:

$$P_C(I, \hat{I}) = \sum_{k=1}^{2} \omega_k \frac{\sum_h \sum_s \min(R^k_{h,s}, \hat{R}^k_{h,s})}{\sum_h \sum_s R^k_{h,s}} \tag{2}$$

where $P_C(I, \hat{I})$ is the color similarity score between humans I and $\hat{I}$, and ω_k is the weight of the sub-regions.

C. Body Size Feature Extraction

Skeletal joints positions acquired by the Kinect can not only be used to recognize the actions but also to measure the body size value. Body height, shoulder width and arm

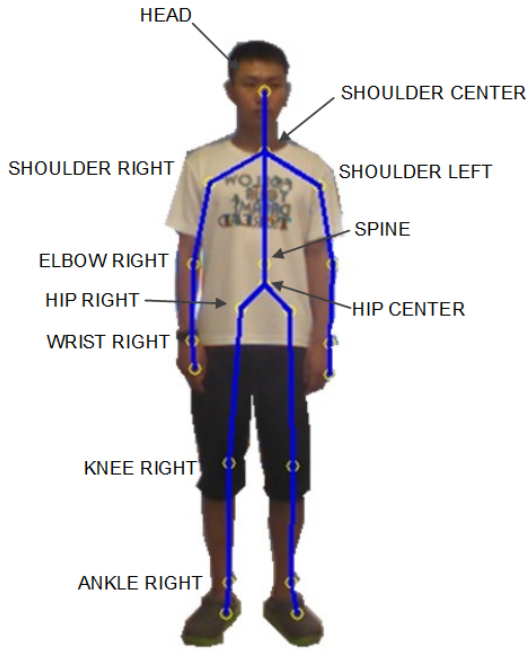

Fig. 2. Joints used for body size computation.

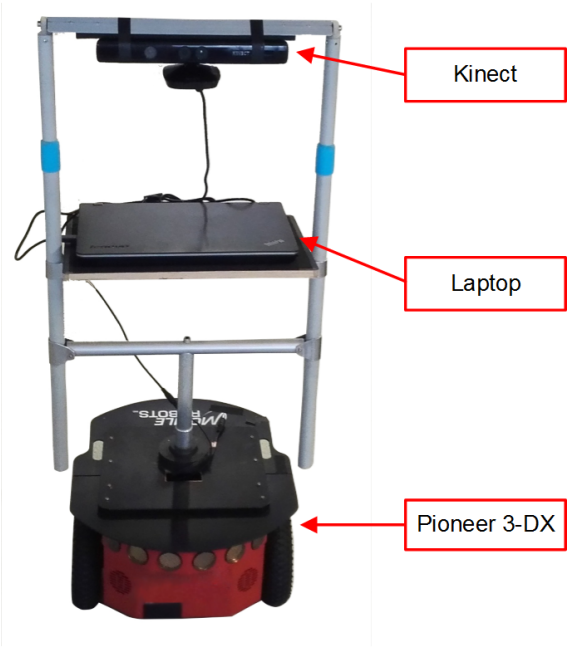

Fig. 3. The human following robot.

length are basic characteristics for humans which are used to describe the body size features in this paper. The characteristics can be easily measured by computing the Euclidean distance between joints:

$$\begin{aligned} L_{BodyHeight} = & D(Head, ShouldCenter) \\ & + D(ShouldCenter, Spine) \\ & + D(Spine, HipCenter) \\ & + D(HipCenter, HipRight) \\ & + D(HipRight, KneeRight) \\ & + D(KneeRight, AngleRight) \end{aligned} \tag{3}$$

$$\begin{aligned} L_{ArmLength} = & D(ShoulderRight, ElbowRight) \\ & + D(ElbowRight, WristRight) \end{aligned} \tag{4}$$

$$L_{ShoulderWidth} = D(ShoulderRight, ShoulderLeft) \tag{5}$$

where $D(\cdot)$ represents the Euclidean distance between two joints. The joints used to compute body size are shown in Fig. 2.

Because of the measurement error, values of the same human change over time. We assume that each value in the body size model follows a Gaussian distribution. Then a set of training data of the same human target is used to calculate the average and variance of each value. For a candidate human target, body size values are computed and the matching probabilities for body height, arm length and shoulder width are defined as:

$$P_i = \frac{1}{\sigma_i \sqrt{2\pi}} exp(\frac{-(L_i - \mu_i)^2}{2\sigma_i^2}) \tag{6}$$

where $i \in \{1, 2, 3\}$ denotes the body height, arm length and shoulder width; μ_i is the mean and σ_i is the standard deviation of each value.

Finally, the similarity between the candidate and the target is expressed as:

$$\begin{aligned} P_B = & \alpha_1 P_{BodyHeight} + \alpha_2 P_{ArmLength} \\ & + \alpha_3 P_{ShoulderWidth} \end{aligned} \tag{7}$$

where α_1, α_2 and α_3 are the weights of the body size values, P_B is the body size similarity.

D. Combination of Color and Body Size

As both the color and body size describe different characteristics, it is reasonable to fuse them for representing a human. Then we adopt a weighted sum method,

$$P = \beta P_C + (1 - \beta) P_B \tag{8}$$

where β is the weight parameter for fusing the two features.

III. THE HUMAN FOLLOWING ROBOT SYSTEM

The proposed method is integrated in the human following system. The robot platform is presented in Fig. 3. It is composed of a Pioneer-3DX and a Kinect camera. The computing unit of the robot is a notebook computer with a Core i5 2.5GHz chipset, and 6GB RAM.

The task of the human following robot is to detect humans in the view and follow the human who gives a specific command. The robot detects and tracks humans with the skeleton tracker of Kinect. 3D positions of the detected humans are the inputs of the motion controller. The hand waving action of a human is regarded as the following command. When the robot recognizes the action, following begins. Meanwhile, color histogram and body size feature

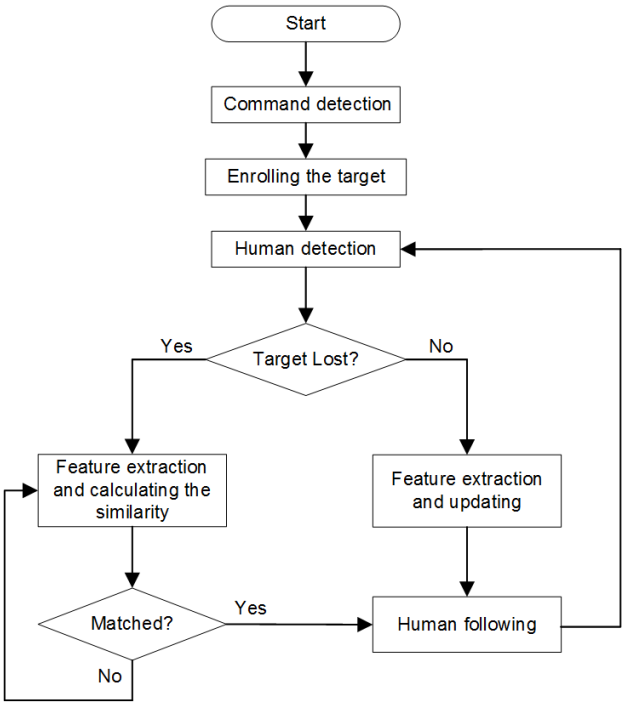

Fig. 4. The flowchart of the human following robot system.

of the target in first 100 frames are used to compute the average histogram model and train the body size model for enrolling. Afterwards, color histogram and body size model are updated to adapt to the appearance variance. The updating method can be expressed as:

$$\hat{R}^k(t) = \lambda \hat{R}^k(t-1) + (1-\lambda) R^k(t) \tag{9}$$

$$\mu_i(t) = \lambda \mu_i(t-1) + (1-\lambda) L_i(t) \tag{10}$$

$$\sigma_i^2(t) = \lambda \sigma_i^2(t-1) + (1-\lambda)[L_i(t) - \hat{\mu}_i(t)]^2 \tag{11}$$

where $\hat{R}^k(t)$ is the value of the average color histogram model at time t, $R^k(t)$ is the histogram computed for the frame at time t; $i \in \{1,2,3\}$ denotes the body height, arm length and shoulder width, $L_i(t)$ is the body size value measured at time t, $\mu_i(t)$ is the mean value at time t, $\sigma_i(t)$ is the standard deviation at time t, and λ is the learning rate. This ensures that new information is integrated gradually when the object appearance changes.

If the human target is lost, the tracker of the Kinect fails. A new ID is assigned when the human target re-enters into the view. Then the system recognize the human target by calculating the color histogram and body size values and matching enrolled human. If the similarity is larger than the threshold θ_s, the robot will continue following the target. The flowchart of the human following robot system is show in Fig. 4.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To test the effectiveness of the proposed human recognition method, experiments are conducted in two aspects. Firstly, we use the dataset collected by the Kinect in our robot to evaluate the performance of the proposed human recognition method. Then the method is implemented in the human following system to show the performance of human re-identification.

A. Evaluation of the Human Recognition Method

The dataset used for evaluation contains 15 different persons and 100 samples for each human. The samples are collected in different scenes and with the human standing in different positions within the view of field of the Kinect. The parameters in the method are chosen empirically as follows: $\beta = 0.65$, $\omega_1 = 0.6$, $\omega_2 = 0.4$, $\alpha_1 = 0.3$, $\alpha_2 = 0.4$, $\alpha_3 = 0.3$, $\lambda = 0.95$. And for the HSV color histogram, $H = 12$ and $S = 8$.

The human recognition method is proposed to re-identify the human target after the human re-enters into the field of view. So we can regard the human recognition problem in the following robot system as the re-identification problem and evaluate the proposed method with CMC (cumulative matching characteristic) [27] curve which is commonly used in object re-identification domain. In the experiment, we select one human as the target randomly, and 50 samples of the human are used to calculate the average histogram and body size models. Then one sample in the rest 50 samples and 99 samples of other humans are selected to form the test sample database whose size is 100. The similarity between the target and the test samples are computed to obtain a ranking. The experiment is conducted for 100 times. We compare the result with the methods using color feature only and body size only. And Fig. 5 shows the experimental result. In the CMC curve, the rank n represents that the recognition result is accepted if the correct match is within the n top matches. The result shows that color only approach significantly outperforms the body size only approach because of the large position errors of joints in some frames. While the proposed method that fuses the color and body size increases the recognition rate significantly. The reason is that clothes color and body size describe the human in two different aspects. The color information is robust to different poses and the body size information is robust to the changes of lighting conditions and environments. The proposed fused method can make full use of the two characteristics which results in the best performance.

Fig. 6 and Fig. 7 show some of the similarity measurement results. It can be seen that the similarity between samples belonging to the same human is high even when the lighting conditions and poses are changed. And for the different human target, similarity value is relatively lower. So it is reasonable to recognize the human target by setting a proper similarity threshold.

B. Experiments of the Human Following System

Finally, we implement the human recognition method in the human following robot system and the system is tested in real-life scenarios. In the human following system, the similarity threshold is set to 0.7. We design two scenarios to test the system. The experimental results of the two scenarios

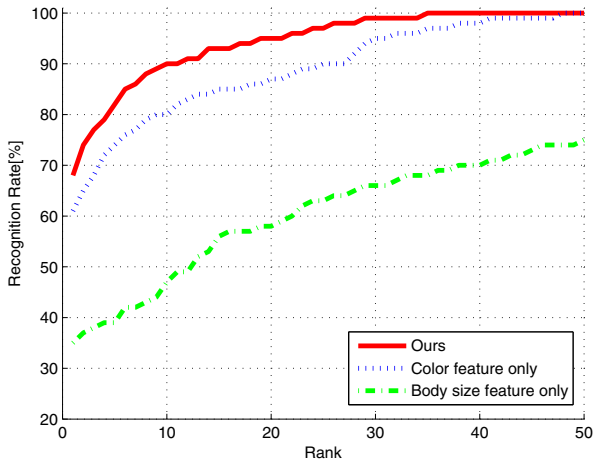

Fig. 5. The CMC curves of different features.

Fig. 6. The first target and some of the similarity results.

Fig. 7. The second target and some of the similarity results.

are depicted in Fig. 8. In the first scenario, as shown in Fig. 8 (a), the robot is static and the human target is asked to walk beyond the view of robot. Then another human steps into the view, the robot detects the human and determines that she is not the target to follow. When the human target re-enters into the robots view, she is recognized by the system and following should be continued. Experimental process of the second scenario is shown in Fig. 8 (b). Firstly, the robot follows a human target. Then another human goes back and forth between the target and the robot frequently and the

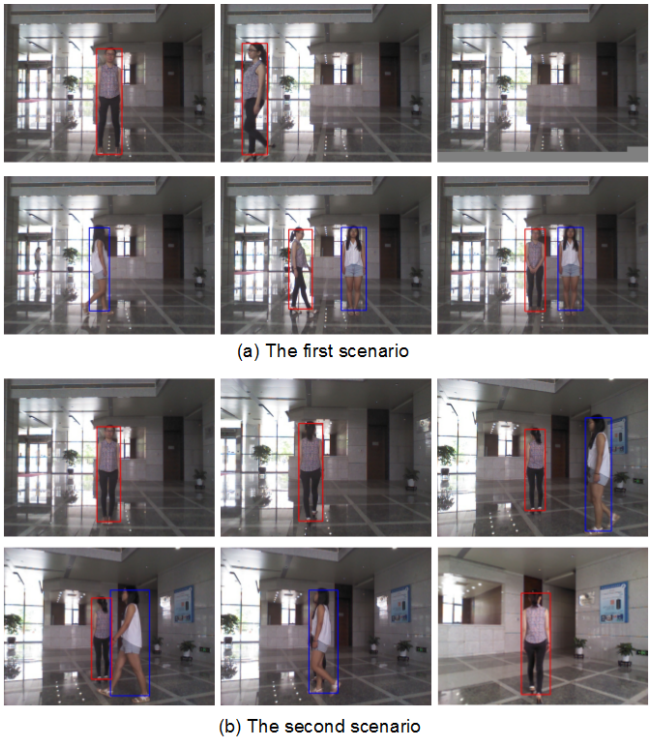

Fig. 8. Snapshots of human following results. The human target is in red bounding box and the other detected humans are in blue bounding box.

target is occluded for many times. Experimental results show the proposed method can recognize the human accurately and the robot system can follow the human correctly although sometimes the target is lost.

V. CONCLUSIONS

In this paper, we proposed a human recognition method based on soft biometrics for the human following robot. Two soft biometric traits (clothes color and body size) are calculated to represent a human. The proposed method involves four steps: human body region segmentation, color feature extraction, body size feature extraction and feature fusion. In this way, the clothes color and body size information are made the best of. Then human recognition is implemented by evaluating the similarity among the targets. Experimental results show that the proposed method is effective and robust to the change of environments and human poses. And the human following system with the proposed method can recognize and follow the human target reliably.

REFERENCES

[1] *Microsoft Kinect SDK* [Online]. Available: http://www.microsoft.com/en-us/kinectforwindows/.

[2] 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 (AVSS), 2010 Seventh IEEE International Conference on*, Conference Proceedings, pp. 435-440.

[3] D. Gray and H. Tao, "Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features," in *European Conference on Computer Vision, 2008*, Conference Proceedings, pp. 262-275.

[4] B. Ma, Y. Su, and F. Jurie, "Local Descriptors Encoded by Fisher Vectors for Person Re-identification," in *Computer Vision-ECCV 2012. Workshops and Demonstrations*, Conference Proceedings, pp. 413-422.

[5] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-Shot Person Re-identification by HPE Signature," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Conference Proceedings, pp. 1413-1416.

[6] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses,"*Pattern Recognition Letters*, vol. 33, no. 7, pp. 898-903, 2012.

[7] 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,"in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, Conference Proceedings, pp. 1-6.

[8] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised Salience Learning for Person Re-identification,"in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, Conference Proceedings, pp. 3586-3593.

[9] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person Re-identification by Descriptive and Discriminative Classification,"in *Image Analysis: 17th Scandinavian Conference, SCIA 2011*, Conference Proceedings, pp. 91-102.

[10] E. Corvee, F. Bremond, and M. Thonnat, "Person Re-identification Using Haar-based and DCD-based Signature," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. IEEE, Conference Proceedings, pp. 1-8.

[11] Y. Zhang and S. Li, "Gabor-LBP Based Region Covariance Descriptor for Person Re-identification,"in *Image and Graphics (ICIG), 2011 Sixth International Conference on*, Conference Proceedings, pp. 368-371.

[12] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian Recognition with a Learned Metric," in *Computer Vision-ACCV 2010: 10th Asian Conference on Computer Vision*, Conference Proceedings, pp. 501-512.

[13] B. Prosser, W. S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person Re-Identification by Support Vector Ranking," in *Proc. British Machine Vision Conf., 2010*, Conference Proceedings.

[14] S. Bak, G. Charpiat, E. Corve, F. Brmond, and M. Thonnat, "Learning to Match Appearances by Correlations in a Covariance Metric Space," in *Computer Vision-ECCV 2012: 12th European Conference on Computer Vision, Florence*, Conference Proceedings, pp. 806-820.

[15] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning Implicit Transfer for Person Re-identification," in *Computer Vision-ECCV 2012. Workshops and Demonstrations*, Conference Proceedings, pp. 381-390.

[16] M. Hirzer, P. M. Roth, M. Kostinger, and H. Bischof, "Relaxed Pairwise Learned Metric for Person Re-identification," in *Computer Vision-ECCV 2012: 12th European Conference on Computer Vision*, Conference Proceedings, pp. 780-793.

[17] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, "3D ellipsoid fitting for multi-view gait recognition," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on.* IEEE, Conference Proceedings, pp. 355-360.

[18] C. Velardo and J. L. Dugelay, "Weight estimation from visual body appearance," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, Conference Proceedings, pp. 1-6.

[19] C. Madden and M. Piccardi, "Height measurement as a session-based biometric for people matching across disjoint camera views." in *Image and Vision Computing New Zealand, 2005*, Conference Proceedings, pp. 282-286.

[20] A. Dantcheva and J. L. Dugelay, "Frontal-to-side face re-identification based on hair, skin and clothes patches," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE, Conference Proceedings, pp. 309-313.

[21] A. Dantcheva, C. Velardo, A. Dangelo, and J. L. Dugelay, "Bag of soft biometrics for person identification," *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 739-777, 2011.

[22] B. J. Southwell and G. Fang, "Human object recognition using colour and depth information from an RGB-D Kinect sensor,"*International Journal of Advanced Robotic Systems*, vol. 10, 2013.

[23] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-Biometrics: Unconstrained Authentication in a Surveillance Environment," in *Digital Image Computing: Techniques and Applications, 2009*, Conference Proceedings, pp. 196-203.

[24] S. Gharghabi and R. Safabakhsh, "Person recognition based on face and body information for domestic service robots," in *Robotics and Mechatronics (ICROM), 2015 3rd RSI International Conference on*, Conference Proceedings, pp. 265-270.

[25] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D Sensors," in *Computer Vision-ECCV 2012. Workshops and Demonstrations*, Conference Proceedings, pp. 433-442.

[26] H. Ailisto, E. Vildjiounaite, M. Lindholm, S.-M. Makela, and J. Peltola, "Soft biometricsłcombining body weight and fat measurements with fingerprint biometrics,"*Pattern Recognition Letters*, vol. 27, no. 5, pp. 325-334, 2006.

[27] R. Vezzani, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey,"*ACM Computing Surveys (CSUR)*, vol. 46, no. 2, pp. 1-37, 2013.

[28] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects,"*Pattern Recognition*, vol. 34, no. 12, pp. 2259-2281, 2001.