People Identification for Domestic Non-overlapping RGB-D Camera Networks

Boris Takač

Andreu Català Technical Research Centre for Dependency Care and Autonomous Living - CETpD Universitat Politécnica de Catalunya - BarcelonaTech Vilanova i la Geltrú, Spain e-mail: boris.takac@estudiant.upc.edu Matthias Rauterberg Wei Chen Industrial Design Department Eindhoven University of Technology Eindhoven, The Netherlands

Abstract—The ability to identify the specific person in a home camera network is very relevant for healthcare applications where humans need to be observed daily in their living environment. The appearance based people identification in a domestic environment has many similarities with the problem of re-identification in public surveillance systems, but there are also some additional beneficial and constraining factors (e.g., less people, non-pedestrian behaviour, unusual camera viewpoints). In this paper, we are considering the problem of people identification in a small home RGB-D camera network, for which we have developed a method based on appearance learning and classification using a combination of SVM and the Naive Bayes classifier. The method is evaluated using the prototype of a real-time multiple camera system on a 16 people dataset.

I. INTRODUCTION

One popular application of camera networks is in the field of ambient intelligence for continuous elderly domestic healthcare monitoring. The main purpose of camera networks is to provide localization data for the specific person (i.e. patient), whose health state can then be inferred from the observation of long-term changes in her movement patterns or from her immediate interaction with the environment. To successfully complete its mission the installed camera network needs to be able to track and identify the patient even when there are other people living in the house.

The availability of affordable color and depth (RGB-D) cameras, i.e. Microsoft Kinect, opened the possibility for their use in small domestic camera networks. RGB-D cameras are able to overcome illumination caused drawbacks of color only cameras, providing robust background subtraction which simplifies multiple people tracking. Assuming that the people tracking for one RGB-D camera is adequately solved, the problem of having functional patient localization in a home healthcare system is reduced to the problem of the identification of inhabitants, or more specifically the identification of the patient.

A. Identification with RGB-D camera

Up to date, there have been several systems and algorithms developed specifically for the identification of persons using RGB-D cameras. Basso et al. [1] presented a tracking approach for multiple people in which an online classifier based on Adaboost was used for learning identities of people. Randomized parallelepipeds inside a 3D color (RGB) histogram space were used as weak classifiers for boosting. Since the algorithm is geared towards the use for tracking in mobile robots, the approach was not tested in the conditions matching those of domestic camera networks in which multiple static displaced cameras are used.

Similar online boosting technique, using three types of RGB-D features and the confidence maximization search in 3D space, was used to build people models for tracking by Lueber et al. [2]. The system evaluation was done on the data from a populated indoor environment using the setup of three Kinect sensors with a joint field of view. The evaluation of this algorithm for a true distributed non-overlapping camera configuration is not available.

Barbosa et al. [3] presented the set of 3D soft-biometric cues, such as skeleton-based and surface-based distances calculated on the 3D point cloud, which they used to build identity signatures. The 3D soft-biometric cues could be the best approach for home network, since the biometric signature of household members needs to be taken only once. However, at the current moment the approach does not seem to be suitable for a range of different camera viewpoints, low resolution and unconstrained poses that can be expected in a multitude of a different home network configurations, since many of the proposed anthropometric measures are hard to extract under non-perfect real-world conditions.

Satta et al. [4] developed a fully functional prototype of a real-time multiple RGB-D camera re-identification system which uses the fast re-identification method based on their own dissimilarity representation descriptors. We consider this method to be appropriate for the application in a domestic camera network. However, the method is primarily oriented towards the typical surveillance re-identification scenario, with a large number of cameras and persons and presumed presence of the human operator for the system. To see how the method could be adapted for our application, we should discuss the advantages and the limitations of a small home camera network compared to a public surveillance network.

B. (Re)identification in a home camera network

The usual approach to the re-identification for public surveillance is to build a database of descriptors generated for each person and apply some distance measure between them. The database of descriptors is known as the gallery, while the distance measure is known as the matching score. In the reidentification for public surveillance, there are two possible frameworks [5]. The first framework works with the complete gallery, which means that the appearance descriptors of all the persons that need to be identified are available when the reidentification needs to be done. In this case the re-identification is solved as the ranking problem, by searching for the minimal matching score against all other descriptors in the gallery. The second framework is the one in which the complete gallery is not available, so that new persons are added to the gallery by the human operator and the re-identification problem is solved by setting the similarity threshold value on the matching score.

The re-identification in a home camera system, which has to be automatically executed during long-term daily monitoring, involves components of both frameworks. It needs to improve on the functionality of the second framework in terms of providing automatic detection of new person appearances, and it has to be able to assign available appearance when the system is operating with the complete gallery of all household members.

Once there exists a short-term re-identification approach for multiple cameras that is able to support both previously described functionalities, the advance towards absolute longterm person identification can be made by using the contextual knowledge of the living habits of inhabitants or by adding additional sensors in the environment/on the persons. For example, the primary initialization of the gallery could be supervised by contextual information, such as the time of a day and the room where the specific appearance first showed up. Or, if the person wears an inertial sensor, the association between the learned appearance model and the person identity could be reinforced using the matching of movement data between inertial and video sensor modalities.

At the moment, our final goal of automatic association between the appearance and person identity in a vision-based healthcare monitoring system is predicated by the need to firstly obtain a robust re-identification framework capable of automatically deciding whether track appearances are a part of the existing household gallery or not. Instead of solving the re-identification problem using ranking and threshold on the similarity matching score of a single probe image, in this paper we propose the use of appearance learning (similar as was done in [3] or [2]) to find a more dynamic, multi-view appearance model of each individual in the home gallery. During re-identification, the binary classification is first applied to discriminate between all the learned appearance models of gallery members and unknown people outside the gallery. Then, if a new appearance (probe) is classified as a part of the gallery, the subsequent identification between known individuals in the gallery is performed by using a simple prediction on a multi-class classifier trained with the known gallery members.

In the following section, we present the appearance learning approach that uses novel feature descriptors with a combination of the Support Vector Machine (SVM) and the Naive Bayes (NB) classifier. The description of our multi-camera system and the interaction between the tracking and the person appearance extraction are explained in Section 3. The experiment data gathering, classifier training and the evaluation of results are given in Section 4.

II. APPEARANCE LEARNING WITH DISSIMILARITY REPRESENTATION

In this section we present features, classifiers and algorithms used for re-identification with learned appearances.

A. Dissimilarity representation and MCDimpl descriptor

The Multiple Component Matching (MCM) framework proposed by Satta et al. [6] provides an unifying view of appearance-based person re-identification methods, by embedding the common concepts of multiple instance representation (e.g. patches, strips, interest points) and body part subdivision. An extension of MCM framework called the Multiple Component Dissimilarity (MCD) framework [7] adopts very compact representation of individuals, while still trying to keep the discriminative capability and robustness of the original identification method. Under the MCD framework, the descriptor \mathbf{I}_p^D for the image of a person subdivided into M body parts, is obtained as the concatenated vector of M dissimilarity vectors, where each dissimilarity vector represents dissimilarity between each body part and a set of bag of components for that part, which are called prototypes. In this paper, we use a specific implementation of MCD descriptor, called MCDimpl [8]. The MCDimpl descriptor subdivides the body into torso and legs and uses components patches randomly extracted from each body part, represented by their HSV colour histogram. A great density of information, along with a fixed small size, make the MCDimpl descriptor a good candidate for use in statistical classifiers.

B. One-class SVM and Naive Bayes classifier cascade

We have emphasized that the ability to automatically confirm or reject a track as a part of the existing set of household members is an important property. The problem of detecting an unknown identity can not be solved by any supervised binary classification method, since the training data for unknown identities can not be collected. Problem can be posed as a novelty detection case, for which a suitable solution can be found in the application of the one-class SVM (OCSVM) introduced by Scholkopf et al. [9].

In the OCSVM, the kernel function is used to map feature vectors in a higher dimensional space, and find a hyperplane that separates the trained class from the origin. Given the training vectors $\mathbf{x}_i = \mathbf{I}_n^D$, $\mathbf{x}_i \in \mathbb{R}^n$, $i \in [l]$, the model is



Fig. 1. Identification using two classifier cascade.

estimated as follows:

$$\min_{\mathbf{w},b,\xi,\rho} \frac{1}{2} \mathbf{w}^T \mathbf{w} - \rho + \frac{1}{\nu l} \sum_{i=1}^{l} \xi_i$$
(1)
subject to $\mathbf{w}^T \phi(\mathbf{x}_i) > \rho - \xi_i, \xi_i \ge 0,$

where $\rho/||\mathbf{w}||$ specifies the distance from the decision hyperplane to the origin, and ξ_i are introduced slack variables. The trade-off parameter $\nu \in (0, 1)$ corresponds to an expected fraction of outliers within the feature vectors. As is the usual case in other SVMs, a kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ is needed to form a decision function. One of the most common kernel functions used in experiments is the Gaussian Radial Basis Function (RBF) $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-(\gamma ||\mathbf{x}_i - \mathbf{x}_i||^2)}$, with the parameter γ which sets the kernel width. For the multi-class classification of the persons inside the household gallery, the Naive Bayes classifier is used. We use it because it is simple, non-parametric and there are implementations which enable on-line training.

The identification algorithm based on the combination of the two classifiers is shown in Fig. 1. Each descriptor of the unidentified track is first classified by the OCSVM. In the case of a positive result, the NB classification on the same descriptor is invoked. The outputs of both classifications are stored in their respective buffers. When the sufficient number *num_dsc* of SVM classifications for the track has been reached, the contents of both buffers are used for the final identification class decision according to SVM and NB decision functions (written on Fig. 1). If the SVM decision function confirms that the current track appearance is a part of the gallery, the class given by NB decision function is forwarded to the system output. Otherwise, the appearance is declared to be unknown.

III. PEOPLE TRACKING AND APPEARANCE FEATURE VECTOR EXTRACTION

The localization system for which we are developing people identification capabilities is the pervasive home monitoring



Fig. 2. System architecture overview.

system for context based detection of symptoms in Parkinson's disease, with its concept presented in [10]. The system comprises out of a network of Kinect sensors expanded with a wearable inertial sensor worn by the patient.

The software architecture of the system is enabled by the Robotic Operating System (ROS) middle-ware [11]. For each Kinect there is an independent process node able to communicate with all other nodes via ROS communication infrastructure. All vision nodes provide their track information to a process node labelled as *ID Manager*, which is dedicated to identity assignment. Fig. 2 shows the system overview.

A. Person tracking

We implemented a tracking system with a generalized and robust tracking model based on the 3D point cloud of persons in the foreground of the scene. The vision tracking subsystem and its position estimation properties were described with more detail in [12]. Basic position tracking of persons is done only on the 2D floor plane, but regarding the appearance feature vector extraction it is important to note that the system has the ability to change representations between the floor plane (where person tracking in real world coordinates is executed) and camera image plane (where structural information about person appearance is given) through transformations on 3D point clouds.

B. Context based extraction of appearance feature vector

For building a dissimilarity representation of a person, we use a two part model of a human (torso and legs) similarly to how it was done in [8]. We base the body part extraction on knowing fixed ratios of human proportions towards the standing height. This approach has contextual dependency, since the model is taken only for the people which are having the standing posture. In a home environment, where frequent posture changes between standing, sitting and lying are to be expected, this approach ensures that appearances will be taken under consistent and similar conditions. The necessary contextual information about the standing posture is obtained by tracking person's height and applying pre-set height thresholds.



Fig. 3. Body part images extraction. a) color image, b) tracked 3D point cloud with bounding cube c) re-projection of points inside the bounding cube on image plane forming person mask and bounding rectangle d) upper and lower body part mask based on metric anthropomorphic ratios.

Starting with the 2D bounding rectangle (a = 0.6m) of the person track, the appearance extraction algorithm recovers the bounding cube for the 3D point cloud of all the points in the column above the tracked area (Fig. 3-b). Using the known transformation matrix between the 3D world coordinate space and 2D pixel coordinates of the Kinect camera image plane, it is possible to re-project back to the image plane the points only inside the bounding cube of the person. Since the pixels of the re-projected image all belong to the tracked person, by applying a threshold operator it is straightforward to obtain the person mask and the bounding rectangle (Fig. 3-c). The masks for torso and legs are then obtained by applying body region limits on the existing mask of a person. Body region limits are given as follows, taking as the basis the anthropomorphic ratio towards height in meters:

$$hs = \frac{H}{lpha}, \ \alpha = 7.5$$
 (2)

$$ts = H * \beta * \frac{2}{3}, \ \beta = 0.46$$
 (3)

$$ls = H * \delta, \ \delta = 0.5 \tag{4}$$

where *H* is standing height of a person, *hs* is head size, *ts* is torso size, *ls* is lower legs size, and the coefficients α , β and δ are based on values given in [13]. By applying 2/3 coefficient only the upper two-thirds of the torso are taken into account, leaving unmasked space between the upper torso and the legs (Fig. 3-d). This ensures that only color consistent parts will be taken by part masks, nullifying a potential influence of the change of perspective. In the end, body part images extracted in this manner are given to the original implementation of *MCDimpl* to produce person's appearance feature vector.

IV. EXPERIMENT

The purpose of the experiment was to evaluate the feasibility and statistical properties of the proposed person identification approach based on the OCSVM and NB classifier cascade for learning and detection of the people living inside a home. We used the prototype of localization system consisting out of two Kinect sensors and a PC server (Intel Core i7 @ 2.3 Ghz; 8 GB RAM; Linux) on which all the processing nodes were run.



Fig. 4. Examples of FOVs of each of the cameras and eight entry events into those FOVs. There are 7 different person views (1 frontal, 2 right profile, 3 left back overhead, 4 frontal with door, 5 semi-frontal with obstacle, 6 left profile, 8 right back overhead), with one view (7 right profile) repeated twice in order to enable a continuous walking path during experiment.

A. Dataset creation

We constructed a new dataset of 16 persons. For each person in a dataset there are two recorded videos. One used for the appearance learning, and the other used for the testing. A learning data video for each participant consists of the participant walking for approximatively 20 seconds in a random pattern in front of each Kinect camera. A testing data video for each participant consists of the participant walking the predefined route which alternates between the scenes in front of the two cameras. The walk along the testing route takes about a minute and a half, and it was designed to maximize the number of possible ways a person can enter camera FOV. For the given camera configuration set in a laboratory environment, there were seven possible views of a person when entering into camera FOVs. These views are shown in Fig. 4. In route planning, we tried to introduce possible obstacles. One of the camera FOV entries (Fig. 4-4) is performed by opening the door which were previously shut, forcing the system to extract the appearance while dealing with an occlusion. The other similar situation is entering the FOV behind the obstacle (Fig. 4-5), when only a part of the body is visible.

Collected videos were used to build a training database and a testing database of person appearances. To obtain the training database, learning video for each person was replayed with the system put into the learning mode in which tracking nodes continuously (every 0.5 seconds) extracted one appearance (an image and two masks) and sent them to *ID Manager* node, which recorded those appearances in the SQL database associated with that person's class label. During the testing database construction, testing videos were replayed with the system set into the normal operation mode in which appearances were extracted only when a new track was detected. For each new track detection $num_dsc = 7$ appearances per track were taken, with the extraction frames being at least 0.1 s spaced in time.



Fig. 5. Examples of 16 person appearances. Top row, left to right, appearances 1-8. Bottom row, left to right, appearances 9-16.

On-line re-identification in *ID Manager* was turned off and feature vectors were recorded in the SQL database.

B. Classifier model training

For this experiment, the size of the target household gallery was set to C = 3. The NB and the OCSVM classifiers needed to be trained for an each separate instance of the household gallery. To get the statistical data about the classifier cascade performance, pairs of both classifiers were trained 16 times, each time with a different triplet of persons inside the gallery. The sets of three people were chosen at random, with a constraint that the same person could not be represented more than three times. Obtained galleries are presented in the first column of Table I, where numbers in the triplets are related to the appearances in Fig. 5.

The prototype gallery for obtaining dissimilarity descriptors using *MCDimpl* was pre-built using the 1264 pedestrian images of the VIPeR dataset [14], the same as in [4]. Given the triplet of persons forming a household gallery, their stored appearances were retrieved from the training database and turned into the dissimilarity representation forming the training data set $\mathcal{X} = \{X_{r_1}, X_{r_2}, X_{r_3}\} = \{x_i, ..., x_l\}$, where $r_1, r_2, r_3 \in \{1, ..., 16\}, r_1 \neq r_2 \neq r_3$ are class indexes for the persons in the dataset, and $l = N_{r1} + N_{r2} + N_{r3}$ is the total number of the feature vectors used. The number of collected training descriptors for each person varied between 25 and 50, due to the random walking patterns in learning videos.

OpenCV Machine Learning library was used for training of both classifiers. The training of the Naive Bayes classifier is non-parametric, requiring only feature vectors and labels. On the other hand, the performance of the OCSVM classifier with the RBF kernel is dependent upon the used value of the hyperparameters ν and γ during the training. Since using the usual cross-validation method in order to optimize the hyperparameters of OCSVM is not possible [15], we had to find another approach.

We decided to use 642 different appearances taken from the VIPeR dataset as the validation data on which performance

TABLE I ACCURACY OF NAIVE BAYES CLASSIFIER

Gallery models	Correct	Tested	Accuracy [%]
8,12,13	24	24	100.0
4,6,15	22	23	95.7
4,8,12	23	23	100.0
4,9,14	20	23	87.0
9,10,11	19	22	86.4
6,10,16	14	22	63.6
5,8,11	23	23	100.0
3,5,10	22	23	95.7
7,9,13	19	23	82.6
1,7,13	18	20	90.0
2,5,14	22	24	91.7
3,6,7	19	23	82.6
2,15,16	20	23	87.0
1,14,16	19	20	95.0
1,13,15	20	21	95.2
2,12,16	20	23	87.0
TOTAL	324	360	90.0 ± 9.2

metrics is calculated to be used as the guidance in the grid search optimization process. Since the VIPeR dataset is diverse and does not contain the exact appearances from the household gallery on which the OCSVM was trained, the intuition is that the fraction κ of the VIPeR images classified by OCSVM as belonging to the household gallery should tend towards zero as the OCSVM is improving in the rejection of the unknown appearances. Still, in order to avoid over-training, we can consider that there is a possibility of having a few appearances in the VIPeR set that are similar to the ones in the household gallery, which means that there should exist some small fraction κ of VIPeR images classified as positive by OCSVM. To investigate this behaviour, the OCSVM classifier was trained for each of 16 galleries in Table I with the target κ_{thresh} values of 0.5%, 1%, 2% and 4%.

C. Results

Statistic results for the accuracy of Naive Bayes classification are given in Table I. Each row contains the test result for one NB classifier trained on the data of a randomly chosen household gallery triplet. The column "Tested" contains the sum of the entries into camera FOV taken by the persons in the triplet during testing trials for which the sufficient number $num_dsc = 7$ of the track appearance images was collected by the underlying tracker, while the column "Correct" shows how many of those entries were correctly classified. In the final NB decision function the nb_detect_rate threshold set to 75% was used. According to the aggregated testing results for all 16 trained NB classifiers, the achieved average accuracy of the classification is 90.0 \pm 9.2%.

Table II demonstrates the influence of the OCSVM classification on the identification system output. The dependent variable is the newly introduced performance metric κ used for OCSVM hyperspace parameter optimization. The OCSVM classifiers for each person triplet from Table I were re-trained four times, each time with a different value of κ_{thresh} . Each row of Table II shows average results obtained by testing the

TABLE II CLASSIFICATION RESULTS FOR ONE-CLASS SVM AND COMBINED CLASSIFICATION SYSTEM OUTPUT

κ_{thresh}	Sens.SVM[%]	Spec.SVM[%]	Sens.Sys[%]	Acc.Sys[%]
0.5%	71 ± 24	87 ± 10	64 ± 21	83 ± 7
1%	78 ± 17	82 ± 11	71 ± 19	80 ± 8
2%	81 ± 16	75 ± 12	74 ± 19	75 ± 8
4%	85 ± 10	64 ± 15	76 ± 13	66 ± 11

classifier cascade with 16 different OCSVM models trained with the same value of κ_{thresh} . To test the OCSVM and the overall system, the same set of valid FOV entries from the test database was used as when testing the NB classifier. In OCSVM decision function the *svm_detect_rate* acceptance threshold was set to 70%.

The average specificity and the sensitivity (along with the standard deviation) of the OCSVM classifier are given in the 2^{nd} and the 3^{rd} column. The best balance of positive and negative SVM classifications, with almost 80% sensitivity and specificity, can be be expected for κ_{thresh} value around 1%. The last two columns of Table II express the average sensitivity and accuracy of the classifier combination. The true positives for a given person inside a gallery are influenced by the misdetections of the NB classifier, resulting in the lower sensitivity output for the classifier cascade. Since the NB classifier only acts when there is a positive SVM classification, it has no influence on the true negatives that influence system specificity. The consequence is that the system specificity is the same as the specificity of the OCSVM classifier alone, so there is no need to show it separately in Table II.

V. CONCLUSION

In this paper we have presented a method for identification by appearance learning in a small domestic camera network using a combination of two classifiers. The advantage of the method is the ability to automatically discern between known and unknown appearances. In appearance learning, we have employed a novel dissimilarity representation descriptor, and we have shown how to control the acquisition of those descriptors in the learning phase by using available contextual information from the trackers. Furthermore, we have shown how to train the OCSVM classifier for the given problem using totally independent dataset for the training parameter optimisation. Our method was tested on the 16 people dataset collected with the prototype of a real-time localization system in a laboratory environment.

The experimental results have shown that the best balance of the specificity and the sensitivity of the classifier combination is achieved with around 75% for both (Table II; 3^{rd} row). This results should be improved if the system is to be applied in an uncontrolled environment. In the experiment, we presumed that only three people live in a house. It would be interesting to see how other possible household sizes influence the system output. For our initial problem of the patient identification in a home, the ability to train the OCSVM classifier opens interesting possibilities, since given some additional contextual information the system can focus only on learning the appearance of the patient and discriminate against everyone else. Combining wearable inertial sensor data with video for improved patient identification is the line of work that we would like to explore more in the future.

ACKNOWLEDGMENT

This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the Education, Audiovisual & Culture Executive Agency under the FPA n° 2010-0012. The authors would also like to thank to dr. Riccardo Satta for his help and provision of the original code for MCD framework descriptors.

REFERENCES

- [1] F. Basso, M. Munaro, S. Michieletto, E. Pagello, and E. Menegatti, "Fast and robust multi-people tracking from RGB-D data for a mobile robot," in *Intelligent Autonomous Systems 12*, ser. AISC. Springer Berlin Heidelberg, Jan. 2013, no. 193, pp. 265–276.
- [2] M. Luber, L. Spinello, and K. O. Arras, "People tracking in rgb-d data with on-line boosted target models," in *In IEEE/RSJ Int. Conf. on*, 2011.
- [3] I. B. Barbosa, M. Cristani, A. D. Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D sensors," in *Computer Vision ECCV* 2012. Workshops and Demonstrations, ser. LCNS. Springer Berlin Heidelberg, Jan. 2012, no. 7583, pp. 433–442.
- [4] R. Satta, F. Pala, G. Fumera, and F. Roli, "Real-time appearancebased person re-identification over multiple KinectTM cameras," in 8th International Conference on Computer Vision Theory and Applications (VISAPP 2013), Barcelona, Spain, 2013.
- [5] D. Chen, S. C. S. Pittsburgh, and H. Wactlar, "People identification across ambient camera networks." [Online]. Available: http://www.informedia.cs.cmu.edu/documents/PeoIDCamNet.pdf
- [6] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino, "A multiple component matching framework for person re-identification," in *Proceedings of the 16th international conference on Image analysis* and processing - Volume Part II, ser. ICIAP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 140–149.
- [7] R. Satta, G. Fumera, and F. Roli, "Exploiting dissimilarity representations for person re-identification," in *Proceedings of the First international conference on Similarity-based pattern recognition*, ser. SIMBAD'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 275–289.
- [8] R. Satta and G. Fumera, "Fast person re-identification based on dissimilarity representations," *Pattern Recogn. Lett.*, vol. 33, no. 14, pp. 1838–1848, Oct. 2012.
- [9] B. Schlkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," 2000.
- [10] B. Takač, A. Català, D. Rodríguez, W. Chen, and M. Rauterberg, "Ambient sensor system for freezing of gait detection by spatial context analysis," in *Proceedings of the 4th international conference on Ambient Assisted Living and Home Care*, ser. IWAAL'12. Berlin, Heidelberg: Springer-Verlag, 2012, p. 232239.
- [11] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng, "ROS: an open-source robot operating system," 2009.
- [12] B. Takač, A. Català, D. Rodríguez Martín, N. van der Aa, W. Chen, and M. Rauterberg, "Position and orientation tracking in a ubiquitous monitoring system for parkinson's disease patients with freezing of gait symptom," *JMIR Mhealth Uhealth*, vol. 1, no. 2, pp. 1–21, 2013.
- [13] ISO TC 159/SC 3 N 416, "Basic human body measurements for technological design part 3: Worldwide and regional design ranges for use in ISO product standards," Tech. Rep.
- [14] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *In IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, Rio de Janeiro*, 2007.
- [15] H. Lukashevich, S. Nowak, and P. Dunker, "Using one-class SVM outliers detection for verification of collaboratively tagged image training sets," in *IEEE International Conference on Multimedia and Expo*, 2009. *ICME* 2009, 2009, pp. 682–685.