Resident Identification Using Kinect Depth Image Data and Fuzzy Clustering Techniques

Tanvi Banerjee, *Student Member, IEEE,* James M. Keller, *Fellow, IEEE,* and Marjorie Skubic, *Member, IEEE*

Abstract—As a part of our passive fall risk assessment research in home environments, we present a method to identify older residents using features extracted from their gait information from a single depth camera. Depth images have been collected continuously for about eight months from several apartments at a senior housing facility. Shape descriptors such as bounding box information and image moments were extracted from silhouettes of the depth images. The features were then clustered using Possibilistic C Means for resident identification. This technology will allow researchers and health professionals to gather more information on the individual residents by filtering out data belonging to non-residents. Gait related information belonging exclusively to the older residents can then be gathered. The data can potentially help detect changes in gait patterns which can be used to analyze fall risk for elderly residents by passively observing them in their home environments.

I. INTRODUCTION

THE occurrence of falls in the elderly is a major public health problem causing many obstacles in their independent lifestyle. Studies show that there is a high risk of falling in the population of older adults of age 65 or older [9]. The estimated incidence of falls for independent people over age 75 is at least 30 percent per year [4].

Passive, continuous, in-home vision-based monitoring systems can greatly facilitate clinicians in diagnosing older adults at higher risk of falling. This creates a need for daily monitoring of activities using vision-based sensors, bringing up privacy concerns. Research has shown that older adults are more amenable to vision-based systems than using wearable sensors in an unstructured home environment, if they use anonymized imaging systems, e.g., silhouettes [10]. However, in order to gather more information pertaining to gait parameters or other activity related information, it is important to separate the residents from the visitors in the apartment settings or, in multi-resident homes, to distinguish among the individual residents. This is especially difficult if only silhouettes can be used. Preliminary analysis has shown that simple height and speed information are not always sufficient to distinguish the residents from each other.

Among other gait-based person identification techniques using vision sensors, one involved computing the silhouette

of a human and unwrapping it by evenly sampling the contour. The distance between each contour point and its center of gravity was computed, termed, "distance profiling." The unwrapped contour was then processed by principal component analysis, and compared against the profiles of the class exemplars using the nearest neighbor approach, testing with a standardized dataset [12]. Other work using template matching techniques included that of Collins et al. [13] who first extracted key frames from a sequence and then measured the similarity between two sequences using the normalized correlation. Among gait recognition work using Hidden Markov Models (HMM), Liu et al. [2] used a population HMM to align gait sequences from different subjects. The shape distances between two silhouettes from the same generic gait stance was computed using linear discriminant analysis. Although this algorithm achieved high identification rates, it is very computationally intensive (on the order of the number of pixels).

Here, we present a study from TigerPlace, a senior housing facility, in which we identify residents using depth images. Data have been continually collected during the residents' day-to-day activities for about eight months. To our knowledge, no other research projects have collected such gait data continuously in the home for this duration. Section II describes the features we use to identify the residents using the Microsoft Kinect system. Section III describes the fuzzy clustering method used for classification, and Section IV describes the data analysis and results. Concluding remarks are in Section V.

II. FEATURE EXTRACTION

Foreground was extracted on the raw depth images from the Kinect using a simple background subtraction algorithm. The background was extracted using training images and the minimum and maximum depth value of each individual pixel was stored. Foreground was identified if the depth value of any pixel was outside the range of minimum and maximum values [1].

Once the 2D silhouettes were extracted from the 3D depth images of a single Kinect, features were extracted from the image sequence. The bounding box of the silhouette was extracted and height, width and width/ height ratio were part of the feature set used for resident identification. The angle made by the top most point of the head and body center with respect to the normal from the body center was also computed. This is shown in Fig. 1(a) as θ.

The distance profile was computed using the technique implemented by Wang et al [12] where the distance from each point in the outline of the silhouette was measured with

respect to the body center and normalized to scale to 1. For our application, we used the difference in the distance profile values from consecutive frames in the image sequences and computed the average of these differences as one of our features. Since the distance profile is related to the shape of the silhouette, the difference profile indicates the change in shape of the silhouette over an image sequence which captures the gait "signature" of an individual. The distance profile of a silhouette is shown in Fig. 1(b).
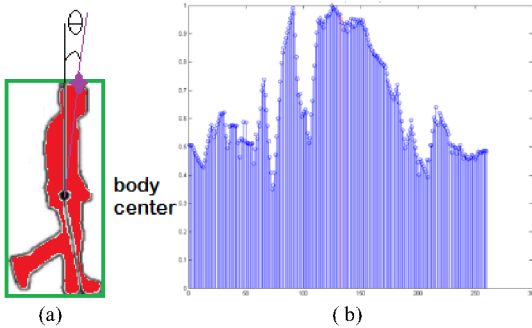

(a)   (b)

Fig. 1(a) Silhouette with outline in blue and bounding box indicated in green and the measured head angle with respect to the normal from the body center (theta). Fig. 1b) shows the distance profile of the silhouette.

In addition, the Zernike image moments were calculated for each depth image silhouette. The particular Zernike image moment set used is described in previous work in [3, 6-7]. Similar to the distance profile, the image descriptors are primarily shape descriptors and as part of our "gait signature" feature set, the difference between the moments from consecutive frames was computed. We call this the "Zernike velocity"; the minimum and maximum values of the magnitude of the difference vectors were incorporated as part of the seven-dimensional feature set. Fig. 2 shows the plot of the Zernike velocity for two residents living in an apartment for a single gait sequence.
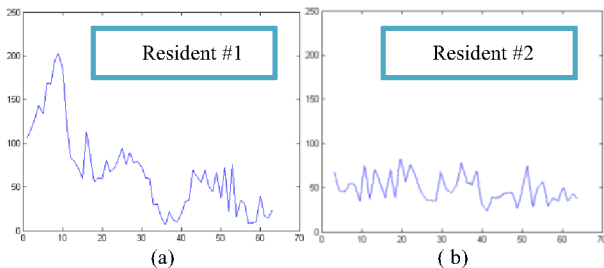

(a)   (b)

Fig. 2 Magnitude of the Zernike velocity of two residents for a single gait sequence with the X axis representing the image frame number and Y axis representing the magnitude of the Zernike difference moments.

As can be seen, Resident 1 has a much higher range of fluctuation in the gait pattern as compared to Resident 2 whose gait is more consistent. Hence, we can see that the shape descriptors can provide useful information regarding the gait patterns in individuals.

III. RESIDENT IDENTIFICATION – POSSIBILISTIC C MEANS CLUSTERING APPROACH

Clustering is an unsupervised learning technique which helps in finding similarities between unlabeled data. Fuzzy clustering has been found to be more advantageous than crisp clustering since any feature vector can be a part of more than one cluster and with varying degrees of membership [14]. An extension of the Fuzzy C Means clustering [11] is the Possibilistic C Means (PCM), which relaxes the constraint that the sum of the membership values of a feature vector should add up to 1 [14]. The algorithm is described below.

Possibilistic C Means clustering technique (PCM):

  The generalized optimization function for clustering N feature vectors into C clusters is described below in equation 1. Here, u_{ij} is the membership of feature vector $\bar{x}_j$ in cluster i and m is the fuzzifier. Equation 2 indicates the additional constraint on the objective function which forces the membership values to be as high as possible to prevent the trivial solution of all the membership values being 0.

$$J_{FCM}(U, A) = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^{m} d_{ij}^{2} \tag{1}$$

$$J_{PCM}(U, A) = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^{m} d_{ij}^{2} + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1 - u_{ij})^{m}$$

where $m > 1$ (2)

The parameter η_i can be chosen as the average intra cluster distance with the formula in Equation 3.

$$\eta_i = k \frac{\sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^{m} d_{ij}^{2}}{\sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^{m}} \tag{3}$$

where k is a constant greater than 0. The update equations for the membership values are shown in equation 4 and the cluster centers are calculated by equation 5.

$$u_{ij} = \frac{1}{1 + \left(d_{ij}^{2} \middle/ \eta_i \right)^{1/(m-1)}} \tag{4}$$

$$a_i(t) = \frac{\sum_{j=1}^{N} u_{ij}^{m} * x_j}{\sum_{j=1}^{N} u_{ij}^{m}}, \tag{5}$$

PCM Algorithm:

1. Fix C = number of clusters & initialize the iteration counter t=1. Set the fuzzifier value m such that m is in the range $(1, \infty)$.
2. Initialize membership matrix U for all the data points and for each of the clusters. Compute η_i using (3)
3. *Do*
4. Compute the cluster centers using equation (5).
5. Update the partition matrix using (4)
6. Increment the iteration counter t.
7. *Until* $\| U(t) - U(t-1) \| < \epsilon$ or $t > $ tmax where ϵ is the minimum permissible error and tmax is the maximum number of iterations specified.

One of the advantages that possibilistic clustering has over its fuzzy counterpart is that the membership values for a particular feature vector do not have to add up to 1. It is one of the few fuzzy clustering algorithms that can input the

number of clusters as 1. This means that the data do not have to belong to any cluster (which is useful in case of outliers), or can also belong to more than one cluster with strong membership values. This makes it extremely useful in our application, where we know the number of residents but not the number of visitors. We expect that each resident's data will cluster, and the visitors will appear as outliers. The details are given in Section IV.

IV. DATA ANALYSIS AND RESULTS

Data are analyzed from three apartments at TigerPlace. Two are two-resident apartments (Apartments A and B) and the third one (Apartment C) is a single resident apartment. These data have been collected by placing a single Kinect sensor in the living room of each apartment. For our analysis, we used some of the data from a single Kinect sensor from each of these apartments. Fig. 3(a) shows the Kinect positioned in the living room of an apartment at TigerPlace and the walk patterns recorded for six weeks using the Kinect sensor.
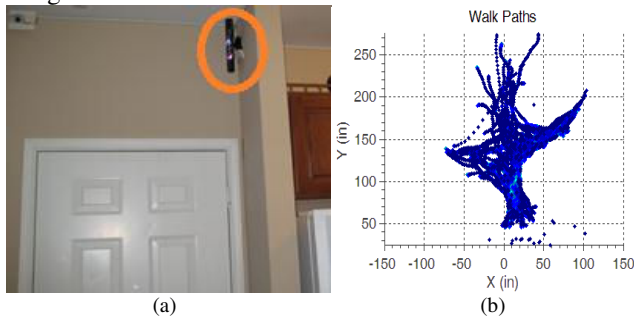

Fig. 3 (a) An apartment at TigerPlace with the Kinect mounted on the wall close to the ceiling (marked with the circle) and (b) the walk patterns of the resident in the living room for six weeks.

A. Two Resident Apartments:

Fig. 4 shows the results of clustering using the Possibilistic C Means with $C = 2$ as described in Section III, using the features explained in Section II. 30 gait sequences were extracted from Apartment A; 15 sequences belonged to each resident. The average duration of each of these gait sequences is about 4 seconds. In this particular apartment, the residents have dissimilar gaits and body shapes. This can be further seen in Fig. 2 where the magnitude of the Zernike velocities shows a significant difference between their gait signatures.
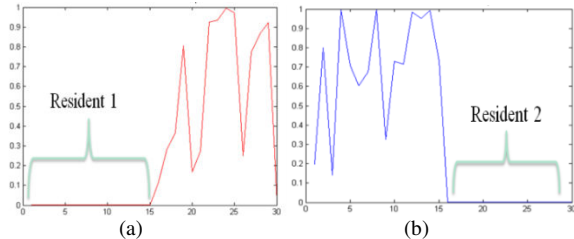

Fig. 4. Results of possibilistic c means clustering using two clusters on data collected from Apartment A with a) showing the membership values in cluster 1 and b) showing the membership in cluster 2. The X axis indicates the gait sequence number and the Y axis indicates the membership value.

As can be seen from Fig. 4, the membership of the data sequences belonging to resident 1 has negligible membership values in cluster 1 (Fig. 4(a)) and higher membership values in cluster 2. Similar results are obtained on the second resident's gait sequences. Fig. 5 shows the results from Apartment B. Note that the gait patterns of the two residents in Apartment B, as well as their body shapes, are more similar than in the previous apartment. However, PCM results from Fig. 4 show that the membership values are significantly different for the two residents.
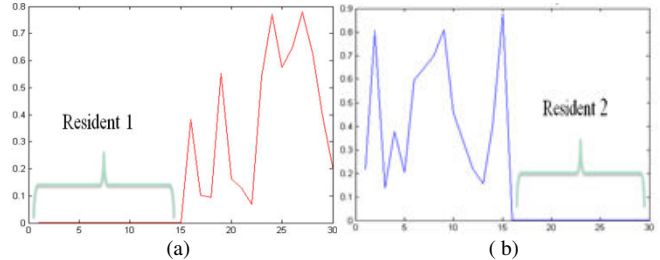

Fig. 5. Results of possibilistic c means clustering using two clusters on data collected from Apartment B with (a) showing the membership values in cluster 1 and (b) showing the membership in cluster 2. The X axis indicates the gait sequence number and the Y axis indicates the membership value.

VAT or Visual Assessment of Tendency is a way of visualizing the number of clusters in high dimensional data [15]. The algorithm utilizes the dissimilarity matrix which is an N x N matrix, where N is the number of data points and the data are reordered so that the similar data points are clustered together. The dark blocks around the diagonal represent the number and size of clusters present in the entire data set. This visualization tool helps answer basic clustering questions such as whether there are any clusters present in the data, and if so, how many and how strongly they are grouped. Fig. 6 shows the VAT results obtained from reordering the seven-dimensional feature vectors of the thirty sequences from the two apartments discussed previously.
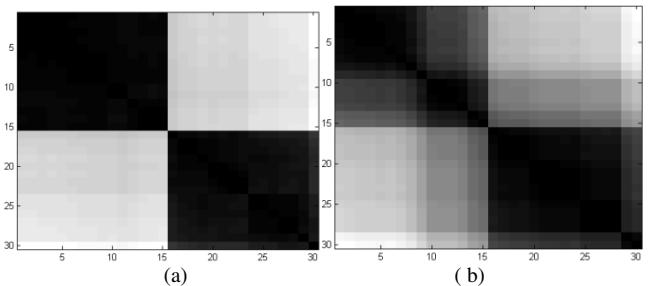

Fig. 6. Results of VAT on data collected from a) Apartment A and b) Apartment B respectively.

As can be seen from Fig. 6 (a), there are two large dark blocks present along the diagonal of the matrix. This indicates the presence of two clusters. Since they are similarly sized, this indicates that the clusters are of similar size. Since the rest of the matrix is fairly light colored, this indicates high separability between the two clusters. Similarly, if we look at Fig. 6 (b), there are two large dark blocks present along the diagonal. Both are similarly sized which indicate that the clusters are similarly sized. However, if we compare the VAT images obtained from apartment A and B, the clusters obtained from Fig. 6 (a) are more separable than those from 6 (b). There are also several sub

clusters present in Fig. 6(b). The gray regions present around the dark blocks show overlap between the two clusters which indicate the lower separability of the two clusters as compared to the VAT image in Fig. 6 (a). This corroborates the fact that the gait signatures obtained from the residents in apartment A are more distinguishable that those obtained from the residents in apartment B.

B. Single Resident Apartments:

Gait sequences extracted from Apartment C were analyzed. Features were extracted (Section II) and the possibilistic c means clustering technique (Section III) with $C = 1$ was computed; 18 gait sequences belonging to the resident and 17 sequences of visitors to the apartment were extracted. The membership results of the clustering with a single cluster are shown below in Fig. 7.
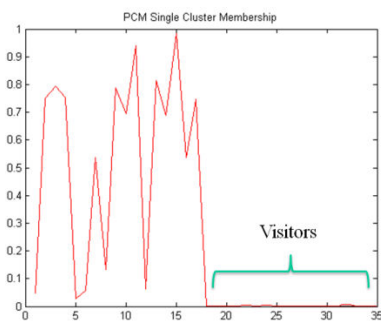

Fig. 7. Results of possibilistic c means clustering using a single cluster on data collected from a single resident apartment.

The advantages of using possibilistic clustering is clear since we want to find out which sequences belong to the resident and which do not as opposed to which sequences belong to cluster 1 and which belong to cluster 2 as computed by other clustering algorithms. The sequences (Fig. 7) belonging to the visitors have negligible membership values as compared to those belonging to the resident. Another observation from Figures 4, 5 and 7 is that some of the sequences, while belonging to the cluster, have a much smaller membership value than the other sequences. These low membership values are most often caused by occlusions present in the image sequences such as the presence of a chair in front of the person blocking the view.

V.   CONCLUSION & FUTURE WORK

We have developed an algorithm to identify gait sequences belonging to elderly residents, to distinguish gait patterns between the residents in a two resident apartment, and to identify visitor gaits from the resident in a single resident apartment. A point to note here is that the data collected here has been taken from the homes of the residents at TigerPlace and is not a standardized data set such as the one described in [12]. The technique described here can not only help identify individuals on the basis of their gait but also help filter gait sequences which have noisier silhouettes and/or occlusions.

Future work includes analyzing data from several apartments at TigerPlace and being able to identify visitors in multi-resident apartments and using the membership results to segment out gait sequences with clean silhouettes and no occlusions present. This can then be used in collaboration with other gait parameter extraction techniques described in [16]. This technique can also be used along with activity detection algorithms to analyze the behavior patterns of individual residents. In the future, this can provide useful information pertaining to the residents' health issues and provide early detection of health changes which could potentially help maintain their independent lifestyles.

REFERENCES

[1] E. Stone, & M. Skubic, "Silhouette Classification Using Pixel and Voxel Features for Improved Elder Monitoring in Dynamic Environments.", *Workshop on Smart Environments to Enhance Health Care,* Seattle, USA, March, 2011.

[2] Z. Liu and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 863–876, 2006.

[3] T. Banerjee, J. M. Keller, M. Skubic, & C. Abbott, "Sit-To-Stand Detection Using Fuzzy Clustering Techniques," *IEEE World Congress on Computational Intelligence (FUZZ-IEEE)*, Barcelona, Spain, July 18-23, 2010.

[4] R. Cucchiara, A. Prati, R. Vezzani, "A multi-camera vision system for fall detection and alarm generation", *Expert Systems,* vol. 24, no. 5, pp. 334–345, Nov. 2007.

[5] R. H. Luke, "Moving Object Segmentation from Video Using Fused Color and Texture Features in Indoor Enviroments", technical report, CIRL, University of Missouri, 2008.

[6] T. Banerjee, "Activity segmentation with special emphasis on sit-to-stand analysis." Master's Thesis, April 2010.

[7] T. Banerjee, J. M. Keller, Z. Zhou, M. Skubic, and E. Stone "Activity Segmentation of Infrared Images Using Fuzzy Clustering Techniques," *Workshop on Soft Computing*, San Francisco, May 23-26, 2011.

[8] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", *Academic Press,* San Diego, CA (1999).

[9] Sattin RW. "Falls among older persons: a public health prospective." *Annu Rev Public Health*., Vol. 13:489–508, 1992.

[10] G. Demiris, D. Oliver, J. Giger, M. Skubic, M. Rantz, "Older adults' privacy considerations for vision based recognition methods of eldercare applications," Technology and Health Care 2009; 17: 41-48.

[11] JC Bezdek, "Pattern Recognition With Fuzzy Objective Function Algorithms," New York, *Plenum Press*, 1981.

[12] L. Wang, T. Tan, H. Ning, W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (12) (2003) 1505–1518.

[13] R.T. Collins, R. Gross, and J. Shi, "Silhouette-Based Human Identification from Body Shape and Gait," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 351-356, 2002.

[14] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.,* vol. 1, no. 2, pp. 98–110, May 1993.

[15] J. C. Bezdek and R. J. Hathaway, "VAT: A tool for visual assessment of (cluster) tendency," in *Proc. Int. Joint Conf. Neural Networks*, Honolulu, HI, 2002, pp. 2225–2230.

[16] F. Wang, E. Stone, W. Dai, T. Banerjee, J. Giger, J. Krampe, M. Rantz & M. Skubic, "Testing an In-Home Gait Assessment Tool for Older Adults," *Proceedings, 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, Minnesota, September 2-6, 2009, pp 6147-6150.