Detecting Foreground Disambiguation of Depth Images Using Fuzzy Logic

Tanvi Banerjee, James M. Keller, Marjorie Skubic
Department of Electrical and Computer Engineering
University of Missouri
Columbia, MO 65211
Email: {tsbycd,kellerj,skubicm}@ mizzou.edu

***Abstract*—We present a unique occlusion and foreground overlap detection technique from depth sensor data using a fuzzy rule-based system. Features such as bounding box parameters and skeletonization were extracted from the foreground images and then input to the Fuzzy Inference System. Overlap and occlusion confidence measures were taken for each frame in the image sequence and compared against the extracted ground truth. This technique can help filter out occluded regions in the image sequence which, in an Eldercare environment, can then be used to compute accurate estimates of fall risk parameters such as stride time, stride length, and walking speed on a daily basis in in order to monitor the well-being of older adults in an ambient assisted living facility.**

Keywords—occlusion; fuzzy rules; depth image; activity analysis; machine learning

I. INTRODUCTION

FALL risk measurement in an in-home setting using non-intrusive, non-wearable, passive sensors has been the focus of our ongoing research at TigerPlace, an independent living facility in Columbia, Missouri [1]. Falls are one of the top ten major causes for death among elderly populations [2]. In fall risk assessment, researchers have found a strong correlation between gait variability and institutionalization [3] [4]. This has been further corroborated by researchers from our team who have been working with the residents at TigerPlace. Algorithms have been developed to measure gait parameters such as stride length, stride time, and walking speed using data collected continuously from vision and depth sensors in the apartments at TigerPlace [5][6]. One of the major challenges in this continuous assessment of gait data has been the identification of gait patterns of individual residents. Many of the apartments at TigerPlace have multiple residents and we have developed algorithms to differentiate between individual gait parameters [5][7]. Another major obstacle has been occlusion detection. Since our monitoring systems are collecting data from actual apartments, the environment is constantly changing: objects such as furniture are constantly getting moved around; and multiple persons are present in the field of view when visitors like housekeeping staff enter the apartment.

This work was supported in part by the U.S. National Science Foundation under Grant R01-HS018477.

This dynamic environment creates a strong need to be able to detect object-person overlap, multiple person overlap as well as to detect occlusion so that the relevant sequences can be either filtered out so that they are not used to evaluate gait parameters or, in some cases, be further processed to gain more information about the environment. Among previous work on occlusion, Masoud et al developed a real-time system in which pedestrians were tracked as blobs over rectangular patches using Kalman filtering to obtain the tracking parameters [8]. The system allowed for overlap of pedestrians which helped their tracking results but the use of a single vision camera created issues like shadows getting identified as blobs which resulted in erroneous results. In related work, Wu et al. used image descriptors such as edgelets to detect body parts separately to deal with partial occlusion [9]. However, the feature set was in the order of thousands which increased the computation time as well as memory requirements of the detection system. In occlusion work using depth sensors, Shen et al [10] tried to address human occlusion by using a complex two-step skeletonization process to extract interior as well as exterior boundary points from the human body silhouette. However, the processing speed per image frame is about one minute which is too high for real time processing.

In this paper we propose a unique method for quantifying self-occlusion and multiple person occlusions of human silhouettes extracted from depth data using a fuzzy rule based system. By self-occlusion, we mean the obstruction of the silhouette of one person by an object such as a chair or a table which is already a part of the background. To our knowledge, no one has tried to quantify partial self-occlusion and multiple person occlusions in human silhouettes before. In order to differentiate between the two measures in this paper, we use the terms overlap and occlusion, which describe multiple person-object occlusions and self-occlusion, respectively. The idea of using fuzzy logic in activity analysis is not new. Anderson et al [11] created linguistic summarizations of activity analysis, and effectively utilized the rule based system for fall detection. For our system, fifteen features were computed from the foreground data extracted from the Microsoft Kinect sensor's depth images. These were then fed into a Fuzzy Inference System (FIS) to measure the overlap and occlusion confidence. One of the major advantages of using a FIS is its interpretability [12]. Most of the rules generated are intuitive in nature. The rule based system is easy to interpret for researchers from different disciplines, which is very beneficial for a diverse team of clinicians, physical

therapists, social workers, and engineers. The other great advantage is its flexible structure: rules can be added or removed without the need of building another FIS structure from scratch.

The rest of the paper is divided into the following sections. Section II describes the preprocessing involved to extract the foreground before the features are computed. Section III describes the features used as input to the FIS. Section IV describes the rules and the reasoning behind the overlap and occlusion confidence measurement. Section V discusses the experiments and the ground truth extraction. Section VI concludes the paper and discusses the future work.

II. Preprocessing - Silhouette Extraction

Foreground was extracted on the raw depth images from a single Microsoft Kinect sensor using a simple background subtraction algorithm. The minimum and maximum depth value of each individual pixel was obtained from an initial set of background only images. Any depth value outside these limits was identified as a foreground pixel. Some more filtering steps were taken to smooth the foreground images [13]. We implemented the dynamic background update algorithm to account for the constant changes in the environment in real world setting in the apartments at TigerPlace as well as in hospital settings [14]. Once the 2D silhouettes were extracted from the 3D depth images of a single Kinect, features were extracted. The basic block diagram as well as an example of a depth image and the corresponding silhouette is shown in Figure 1.

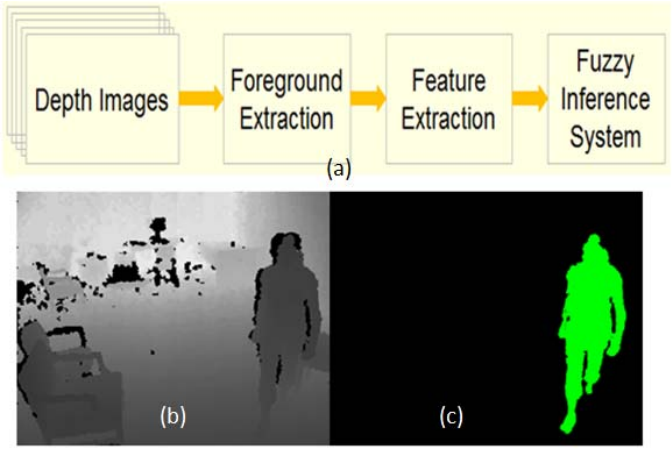

Fig. 1. Basic Block Diagram of the system (a). The second row shows an example of a depth image of a person walking (b) and the corresponding foreground is shown in (c).

III. Inputs and Outputs to the FIS System

Once the foreground for a given image sequence was obtained, features are extracted. For a given sequence, the maximum number of moving objects is computed to check for presence of overlap. For every image frame evaluated, a corresponding image was used as reference. This was the last frame where the maximum number of foreground objects was present. The occlusion confidence is measured separately for each foreground object in the field of view at a given frame. The inputs to the FIS are described next.

A. Bounding box features

These are the parameters of the minimum dimension rectangle that can be fit to the foreground objects. The two features are the width (BBX) and height (BBY) of this bounding box (Fig. 2d). For normalization, each value is divided by the sum of the bounding box values of the reference image. For input to the FIS system, the difference in these values in consecutive frames was used (temporal differencing).

B. Peak-to-peak difference in Vertical and Horizontal Projections

This is the peak to peak difference between the vertical and horizontal projections of the silhouettes. Figure 2 shows an example of the bounding box parameter and the projections. Figure 2(a) shows the single foreground formed when two people merge together. Figure 2(b) and 2(c) show the horizontal and vertical projection respectively. The dip between the peaks is apparent in the vertical projection where the two silhouettes merge. The bin size is different for both the histograms to highlight the dip present in the vertical projection. Figure 2(d) shows the bounding box with the minimum dimensions which fits the entire foreground. For input to the FIS system, the difference in these values in consecutive frames was used (temporal differencing). These will be referred as DHP (difference in horizontal projection) and DVP (difference in vertical projection), respectively.

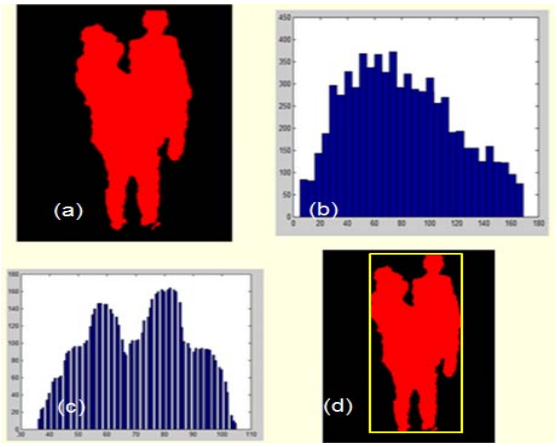

Fig. 2. Features extracted for (a) the foreground detected with (b) and (c) showing the horizontal and vertical projections and (d) showing the bounding box.

C. Area of the Foreground

This gives the number of foreground pixels detected for each object. For normalization, this value is divided by the sum of the area of the blobs from the reference image, hence called ratio area (RA).

D. Number of Points on the Boundary

This parameter is important to determine the confidence of the person exiting/entering the field of view. The number of points on the boundary (NB) is measured by the number of points of the object present on the boundary of field of view. For example, the number of points on boundary for the image

in Figure 2(a) is 33 as we see a part of the head is outside the field of view.

E. Time Separation

This parameter measures the time elapsed or Time Gap (TG) since objects were last separate. The idea is that the longer the time elapsed since the objects were last seen apart, the more likely the overlap.

F. Distance Separation

This parameter measures the distance between the two people/ objects when they were last separate (DBO). The idea is that the larger the distance between the objects when last seen apart, the less likely the overlap.

G. Skeletonization

The skeletonization feature helps identify silhouettes where the entire body is in view (i.e. no occlusion). This feature was added to take into account the possibility that the silhouette could be occluded at the point it is first identified as foreground. Hence, it is important to be able to identify the segments in the gait sequence when the entire silhouette is visible. To do so, we compute the gradient of the distance transformation DT(x, y) of the foreground objects in the process described in [15][16] :

$$|\nabla DT| = \sqrt{\left(\frac{\partial DT}{\partial x}\right)^2 + \left(\frac{\partial DT}{\partial y}\right)^2} \qquad (1)$$

where $|\nabla DT|$ is the magnitude of the gradient response. To reduce the number of boundary points, the end points obtained using (1) were compared with the corners extracted from the Harris corner detector and only the points belonging in both sets were kept [17]. The skeleton points were then linked and thinned iteratively to obtain the final results. Instead of using the skeletal points directly as features and trying to track them over successive frames, we clustered the x and y locations of these points into 5 clusters using the Fuzzy C-Means (FCM). The idea behind five clusters was to have a separate cluster for the head and the four appendages. Distance between these cluster centers from the body centroid was measured. Figure 3 (a) shows the non-occluded output and (b) shows the occluded output. The filled blue points are the cluster centers and the red point is the body centroid. The three largest distance values were highlighted by the yellow points in each case. For a non-occluded silhouette, these would be the cluster points for the head and the two legs. The ratios of the second over the largest distance value, called S21 (Skeletal cluster point 2 divided by skeletal cluster point 1) and third largest distances over the largest values (S31) were input to the FIS. For a non-occluded silhouette, both these ratios would be close to 1 since the three largest values would be approximately equal. The input variable used along with these two values is the location of these three clusters with respect to the centroid. This is to ensure that two of the selected cluster centers are below the centroid (legs) and one is above (head). To do this, we took the three y values of the selected cluster centers, subtract the centroid y value from them and sort them in ascending order. The two larger values were kept for input to FIS (Y1 and Y2). Both of these values need to be positive for the complete silhouette with both legs visible.

In addition, the difference in the two largest distance values (DS1 and DS2) in consecutive frames was used (temporal differencing). This was used in conjunction with the bounding box feature width over height (X/Y) for the current silhouette.

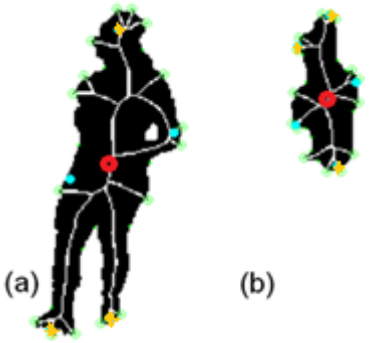

Fig. 3. Results of (a) Skeletonization of non-occluded silhouette and (b) occluded silhouette.

In total, there are 15 inputs to cover the various cases of overlap and occlusion. The outputs of the FIS were the amount of overlap (AO), self-occlusion (OCC) and the confidence of being at the boundary of field of view (BND) i.e. representing a person entering or exiting the scene.

IV. Fuzzy Rules

This section describes the rules which determine the level of occlusion, boundary position as well as overlap. For our current experiments, 37 rules were used. In the interest of computational efficiency, we kept the rule base as small as possible. If we were to implement the complete set of rules, assuming (on average) three fuzzy labels for each input variable, we would have to implement 3^{15} or approximately 14 million rules!

We categorized the rules into four groups. For the first group, the three parameters used are the bounding box parameters (BBX and BBY) and the Time Gap elapsed when last seen separate (TG). For each of these inputs, we use trapezoidal membership functions with parameters [a b c d], where parameters a, d are the "feet" of the trapezoid and b, c are the "shoulders". Six rules have been implemented using these variables (Table 1). BBX and BBY are Low = [-0.3 -0.03 0.04 0.3], Medium = [0.2 0.4 0.5 0.7] and High = [0.7 0.96 1.04 1.36]. TG values were not normalized but bounded between 0 and 200 with membership functions: Low = [3.9 16.9 30.4 65.9], Medium = [65.8 95.2 143 175], and High = [128 175.9 208 272]. The membership functions for the output variables are as given as shown. The AO membership functions take values: Very Low = [-0.3 -0.02 0.03 0.11], Low = [-0.3 -0.02 0.1 0.25], Medium = [0.14 0.4 0.6 0.86], and High = [0.6 0.86 1.04 1.36]. BND membership functions of Low, Medium and High were kept with same parameter values. The linguistic variables are Very Low (VL), Low (L), Medium (M) and High (H). All the antecedents are joined using the AND operator as a connective for the rule generation.

TABLE I. FIRST SET OF FUZZY RULES

		BBX	BBY	TG		AO	OCC
1		L	L			H	VL
2		M	L			H	VL
3	IF	L	M		THEN	H	VL
4		H	H	M		L	L
5		M	M	L		M	VL
6		H	H	H		M	VL

The second set of features decides both OCC and AO output variables. The inputs in this category are the difference projections DHP and DVP described in Section III along with the Time parameter TG. The memberships for DVP and DHP are Low = [-0.3 -0.04 0.04 0.36], Medium = [0.13 0.46 0.54 0.86], High = [0.64 0.86 1 1.36]. The DVP and DHP parameters are symmetric in their response since the overlap and occlusion can take place both horizontally and vertically.

TABLE II. SECOND SET OF FUZZY RULES

		DHP	DVP	TG		AO	OCC
7		H	H	L		M	VL
8		H	M	L		H	VL
9	IF	M	H	L	THEN	H	VL
10		M	M	L		M	VL
11		H	H	H		H	VL

The next set of rules contains the area variable RA, the distance measure DBO, the number of boundary points NB and time gap TG. NB is bounded between 0 and 100 and has membership parameter values of Low = [-32.6 -0.5 3.5 13.1], medium = [7.28 14.4 27 33.7] and High = [22.3 30.8 104 136]. RA has similar membership values as BBX and BBY: Low, Medium and High. DBO is bounded between 0 and 50 and has values of Low = [-18 -2 2 11], Medium = [9 20 27 34] and High = [30 43 52 67].

An example for the linguistic interpretation of a rule (Rule 28) is

IF the Area Ratio (RA) is ***High*** and the Distance before Overlap is ***High*** and the Number of foreground points at the boundary of field of view is ***Low*** and the Time Gap (TG) elapsed is ***Medium***, ***THEN*** the Amount of Overlap is ***Very Low***, the Occlusion is ***Low*** and the confidence of Exit/ Entry is ***Low***.

This rule applies to the case when the ratio of the area of the current foreground to the reference foreground is high, and the no foreground object is close to the boundary of the field of view. The time elapsed since the objects were last seen separate is Medium and the Distance between the objects was High so the confidence of overlap (AO) is very low.

The confidence of the person entering or exiting the scene is low and the corresponding occlusion (OCC) confidence is also low since none of these conditions indicate the presence of occlusion.

TABLE III. THIRD SET OF FUZZY RULES

		RA	DBO	NB	TG		AO	OCC	BND
12		L	L	L	H		H	L	
13		L	M		M		L	M	
14		M	L		M		M	VL	
15		M	M		M		L	L	
16		H	M		M		L		
17			M	M	M		L	H	H
18				H			L	H	VH
19		L		L			M		L
20	IF		M	L	M	THEN	L		
21			M	M	L		L	M	M
22			M	L	L		L	M	L
23		L	L		M		H	VL	
24		M	L		M		H	VL	
25		H	H		M		VL		
26		L	M	L			L		L
27		M	L	L			L		L
28		H	H	L	M		VL	L	L
29		M	H		H		L	L	
30		H	M		H		L	L	

We can see that the TG parameter has been used in most of the rules which makes sense since time plays a crucial role in understanding the gait dynamics specifically about the location. The number of boundary points influences the boundary confidence BND as well as impact on the self-occlusion since if the person is close to the boundary; it is more likely that part of the body is hidden from field of view.

The final set of rules involves the skeletonization parameters described in Section III. Briefly, these are the cluster centers to centroid distance ratios S31 and S21, the temporal difference cluster center distances DS1 and DS2, and the vertical location of the cluster centers with respect to the centroid Y1 and Y2. X/Y is the width to height ratio from the bounding box. DS1 and DS2 are bounded in the interval [-100 100] and have membership functions of Negative (N) = [-172 -108 -51.6 -6.6], Zero (Z) = [-10.8 2.4 9.0 15], and Positive (P) = [4.5 31.5 108 172]. For simplicity, X/Y has the same membership function parameter values as BBX. Y1 and Y2 are also bounded between -100 and 100 with the trapezoidal function parameter values of Negative = [-172 -108 -92 -28], Zero = [-72 -8 8 72], Positive = [28 92 108 172]. Table IV describes the rules using these inputs.

TABLE IV. FINAL SET OF FUZZY RULES

		Y1	Y2	S21	S31	DS1	DS2	X/Y		AO	OCC
31						N		M			M
32							N	M			M
33	IF					P		L	THEN	L	L
34						P		L		L	L
35						Z		L		L	VL
36							Z	L		L	VL
37		P	P	H	H						VL

For these experiments, the rules and membership values were developed heuristically using a small one minute training image sequence. Rules were added in an iterative manner and those which did not affect the error significantly were removed. In the future, we plan to create an automated way to generate the optimized set of rules.

All the rules implement the fuzzy intersection operator which takes the minimum membership values for three input sets IP1, IP2 and IP3 as shown in (2).

$$\mu_{IP1 \cap IP2 \cap IP3} = \min(\mu_{IP1}, \mu_{IP2}, \mu_{IP3}) \quad (2)$$

The last rule, i.e. rule # 37 has been added as a standalone condition to decide the presence of self-occlusion. This specific rule does not have any temporal inputs and it just tries to ascertain that both the legs of the person as well as the head are visible at that particular frame. This is particularly useful if part of the person is partially hidden from view when he / she enters the field of view and we get incorrect initial bounding box information from the occluded silhouettes. Once all the rules are evaluated, the final confidence measures are computed using centroid defuzzification on the aggregate fuzzy responses corresponding to the fuzzy rules described above.

We observed that the bounding box features alone can also cause errors when an object becomes part of foreground and is attached to the person's silhouette, for example when a person is carrying a large object. Our rule based system uses the bounding box features in conjunction with skeletonization features which help rectify the readings by giving us locations in the sequence where the OCC confidence is very low. These are locations which can be used as accurate information related to the foreground. Also, when the AO membership is High or Medium, membership of OCC has been kept Low or Very Low. This does not mean that the two events are mutually exclusive; for example the overlap caused when a person is partially blocked by a chair. However, once a frame is detected as having high or medium overlap, it will be discarded from further activity analysis.

V. EXPERIMENTAL RESULTS

Three fifteen minute videos were recorded using the Microsoft Kinect sensors at a frame rate of approx. 6.5 frames per second for person–object overlap, called Scenario 1. This comprised of a single person moving around the scene, moving a chair and carrying it around, and moving a table around the scene. Another three fifteen minute videos were recorded with two persons walking around in the scene that contained a chair and table to test for multiple person overlap. We named this Scenario 2. In all, approximately 35,100 frames were captured to test our algorithm. The training data was not a part of the test sequences.

For ground truth, the raw Kinect depth images were manually segmented using bounding boxes to measure the actual overlap. For overlap, the quotient of the bounding box area of the current frame over the bounding box area of the two persons were last seen separate was considered. For occlusion, the ratio of bounding box of the current frame over the bounding box of the person when last seen non-occluded was considered. The set of linguistic rules was implemented using Matlab Fuzzy Inference System toolbox. The average processing time for the occlusion and overlap membership calculation is approximately 1 frame/second. This does not include the silhouette processing time which is approximately 15 frames per second.

Figure 4 shows a sample one minute result for overlap of two persons walking around. The ground truth has been color coded as low medium and high as shown in the legend.

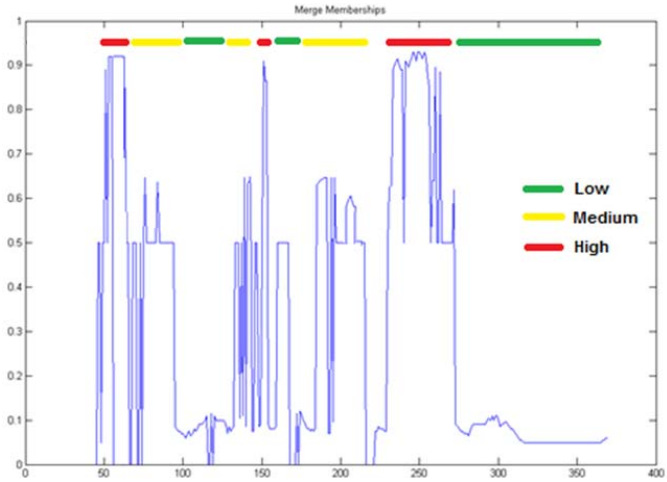

Fig. 4. Results of the confidence measures for amount of overlap (AO) for a two person scene for about one minute duration (400 frames). The x axis represents the frame number and the y axis measures the AO confidence measure between 0 and 1.

These results were further smoothened by using an averaging filter of window size 5. The results are shown in Figures 5 and 6 below. Figure 5 shows the occlusion confidence measures for five sample silhouettes. These have been color coded in different shades of green to indicate the membership degrees in the OCC class. Black and dark green (Frames 5(a) and 5(e)) indicates low confidence measures with values less than 0.1 and between 0.1 and 0.2 respectively,

i.e., non-occluded persons. Figures 5(b) and 5(d) have memberships between 0.4 and 0.6. Figure 5(c) has membership value 0.8.

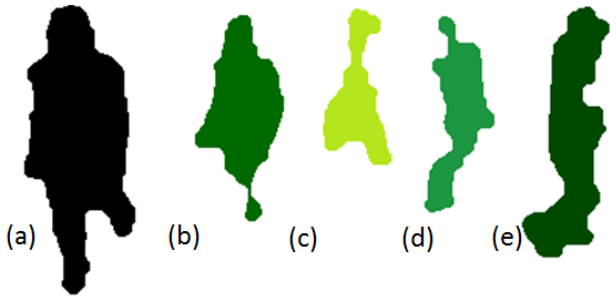

Fig. 5. Results of the confidence measures for occlusion (OCC) for five sample image frames. The darker color indicates lesser occlusion.

Figure 6 shows the confidence measures for AO class for four sample silhouettes. These have been color coded in different shades of red to indicate the degrees of membership. Black (Frame 6(a)) indicates zero confidence measure in overlap class since the two persons are separate. Figure 6(b) has a higher membership of 0.46. Figure 6(c) has membership value 0.72. Figure 6(d) has a slightly lower membership value of 0.55 because rule # 19 reduces the overlap confidence a little since the number of boundary points is low. This could indicate that one of the persons is about to leave the field of view.

Fig. 6. Results of the confidence measures for amount of overlap (AO) for four sample image frames. The darker color indicates lesser overlap.

Table 5 shows the results for the difference between the AO confidence measure and the ground truth for the two scenarios described earlier in this section.

TABLE V. RESULTS FOR OVERLAP CONFIDENCE (AO) FOR THE TWO CATEGORIES

Overlap	Avg diff (abs)	Max diff (abs)	Std
Object-person (1st scenario)	0.09	0.21	0.1
Person-person (2nd scenario)	0.08	0.28	0.15

The largest difference between the computed confidence and the ground truth took place for the two person sequence when the foreground of one person suddenly got absorbed into the background. This occurred when the distance between the person and sensor exceeded the maximum range of the Kinect (4-5m depending on acceptable error rate) which caused high errors in the depth values received by the sensor, thus causing the foreground to blend into the background [18]. In this situation, rules 13 and 26 gave it low overlap and rule #1 gave a high overlap confidence which effectively resulted in medium confidence response. The ground truth computation gave low overlap confidence so the difference was the largest in this case. However, it was interesting to see that if the situation continued, i.e. the silhouette of the second person remained a part of the updated background, rule #13 was triggered which reduced the AO confidence, thus stabilizing the system. This was caused by the increasing time gap from when both the persons were last seen separately, which reduced the confidence in the measured silhouette parameters. In this way, the system adapted to the environment and corrected its initial assumption.

Table 6 describes the occlusion results for the two scenarios. Since there were two persons in the second scenario, the occlusion confidence was computed separately for each person. The results appear more stable than the AO measures.

TABLE VI. RESULTS FOR OCCLUSION (OCC) FOR THE TWO SCENARIOS

Occlusion	Avg diff (abs)	Max diff (abs)	Std
Person (1st scenario)	0.09	0.15	0.11
Person 1 (2nd scenario)	0.08	0.17	0.09
Person 2 (2nd scenario)	0.1	0.17	0.12

Note that the error computed here is caused by the entire system described in Figure 1 including the foreground extraction, the feature computation process, and the fuzzy inference system. This also contains the error in the disparity values obtained from the Kinect sensor which increases exponentially as the distance from the sensor increases [18]. Considering this limitation of the sensor, the results obtained are quite promising.

VI. CONCLUSION

We present a unique method for quantifying overlap and occlusion in this paper using a fuzzy rule based system, focusing on gait related occlusion and overlap in video data collected in laboratory settings. In order to improve the results presented in this paper, more features will be considered; specifically features related to the relative locations of the foreground objects in the field of view of the sensor. Automated tuning of rules and membership functions using evolutionary computation techniques will also be considered. Our next step is to implement the system on the depth data collected in the apartments at TigerPlace. This will help

strengthen our algorithm by testing it in an in-home, dynamic environment.

Future work includes evaluating occlusion in other activities such as lying on the bed or couch, lying on the floor, as well as sitting. This will need different parameters such as shape features to determine the activity and we will also have to consider the dynamic background update parameters as well as the orientation of the person (horizontal in the case of a person lying on the bed or on the floor) to measure the overlap and occlusion parameters. This will be useful for identifying the presence of a person on a chair or lying in a bed even when the foreground is completely blended into the background due to the dynamic update. Occlusion detection can also be used as an inference tool to better understand the room setup. For example, if a person is always partially occluded from field of view at a particular location, this would indicate that there is a large object such as a table or chair in that location which blocks part of the silhouette from view. By inference, this could give more information regarding the clutter and furniture location in the apartment which could generate an alert for safety concerns.

Detection of foreground disambiguation in the form of occlusion and overlap can help filter video segments to gain better information about the gait parameters of the residents at TigerPlace and room setup, providing invaluable information related to fall risk and resident and patient monitoring in homes or hospital rooms.

REFERENCES

[1] M. Skubic, G. Alexander, M. Popescu, M. Rantz and J. Keller, "A Smart Home Application to Eldercare: Current Status and Lessons Learned," *Technology and Health Care*, vol. 17, no. 3, pp. 183-201, 2009.

[2] Y. Gorina, D. Hoyert, H. Lentzner, M. Goulding, "Trends in causes of death among older persons in the United States." *Aging Trends*. (6):1-12, 2005.

[3] B.E. Maki, "Gait changes in older adults: predictors of falls or indicators of fear." *J Am Geriatr Soc*, 45, pp. 313−320, 1997.

[4] J. M. Hausdorff, D. A. Rios, H. K. Edelberg, "Gait variability and fall risk in community-living older adults: a 1-year prospective study*." Arch Phys Med Rehabil* 82:1050–1056, 2001.

[5] E. Stone & M. Skubic, "Capturing Habitual, In-Home Gait Parameter Trends Using an Inexpensive Depth Camera," *Proceedings, 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego CA, August 28-September 1, 2012.

[6] F. Wang, E. Stone, M. Skubic, J. M. Keller & C. Abbott, "Towards a Passive Low-Cost In-Home Gait Assessment System for Older Adults," *IEEE Transactions on Information Technology in Biomedics*, in press.

[7] T. Banerjee, J. M. Keller & M. Skubic, "Resident Identification Using Kinect Depth Image Data and Fuzzy Clustering Techniques," *Proceedings, 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* San Diego CA, August 28-September 1, 2012.

[8] O. Masoud and N. P. Papanikolopoulos, "A novel method for tracking and counting pedestrians in realtime using a single camera," *IEEE Trans. on Vehicular Tech.*, Vol. 50, No. 5, pp.1267-1278, 2001.

[9] B. Wu and R. Nevatia. "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors." *International Journal of Computer Vision*, 75(2):247 – 266, 2007.

[10] W. Shen, S. Xiao, N. Jiang, and W. Liu, "Unsupervised human skeleton extraction from Kinect depth images." *Proceedings, 4th International Conference on Internet Multimedia Computing and Service*, pp.66-69, Wuhn, China, 2012.

[11] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz and M. Aud, "Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic," *Computer Vision and Image Understanding,* Vol. 113, pp. 80-89, 2009.

[12] S. Guillaume, "Designing fuzzy inference systems from data: An Interpretability-oriented review*," IEEE Trans. Fuzzy Syst.,* vol. 9, no. 3, pp. 426–443, Jun. 2001.

[13] 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.

[14] T. Banerjee, M. Rantz, M. Li, M. Popescu, E Stone, and M. Skubic, "Monitoring Hospital Rooms for Safety Using Depth Images" *AAAI in Gerontechnology,* Washington DC, November 2-4, 2012.

[15] J. Ding, Y. Wang, L. Yu, "Extraction of Human Body Skeleton Based on Silhouette Images", *Second International Workshop on Education Technology and Computer Science,* 2010.

[16] L.J. Latecki and R. Lakämper. "Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution." *Computer Vision and Image Understanding (CVIU),* vol. 73, pp. 441-454, 1999.

[17] C. Harris and M. Stephens, "A combined corner and edge detector". *Proceedings of the 4th Alvey Vision Conference.* pp. 147–151, 1988.

[18] K. Khoshelham and S. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors,* vol. 12, 2012.