Day or Night Activity Recognition From Video Using Fuzzy Clustering Techniques

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

***Abstract*—We present an approach for activity state recognition implemented on data collected from various sensors—standard web cameras under normal illumination, web cameras using infrared lighting, and the inexpensive Microsoft Kinect camera system. Sensors such as the Kinect ensure that activity segmentation is possible during the daytime as well as night. This is especially useful for activity monitoring of older adults since falls are more prevalent at night than during the day. This paper is an application of fuzzy set techniques to a new domain. The approach described herein is capable of accurately detecting several different activity states related to fall detection and fall risk assessment including sitting, being upright, and being on the floor to ensure that elderly residents get the help they need quickly in case of emergencies and ultimately to help prevent such emergencies.**

***Index Terms*—Activity labeling, depth images, fuzzy clustering, image moments, infrared images.**

I. Introduction

FALL detection and fall risk assessment are major goals of our research [1] as we continue to conduct experiments at TigerPlace, which is an "aging in place" facility for the elderly. The aim of our research is to build a system to discretely monitor the activity of older adults in their apartments, while addressing their privacy concerns. We also seek to identify diagnostic measures that are predictors of fall risk, which would then fulfill the long-term goal of our project, which is to generate alerts that notify caregivers of changes in a resident's condition so that they can intervene and prevent or delay adverse health-related events [1].

Among previous work related to activity analysis, sit-to-stand was analyzed by Allin and Mihailidis [2] using 2-D and 3-D image descriptors from silhouettes and centroid locations from three different camera views. They computed the following features: distance of the torso from the feet and the angle created by the torso, head, and feet as well as the raw position of the feet; and they used a decision tree to identify the activities of sitting and standing. The ground truth used was obtained by hand labeling the transition data from the video sequence for the two individuals tested.

Manuscript received October 25, 2012; revised February 11, 2013; accepted April 12, 2013. Date of publication April 30, 2013; date of current version May 29, 2014. This work was supported by the U.S. National Science Foundation under Grant IIS-0703692 Grant CNS-0931607.

The authors are with the University of Missouri, Columbia, MO 65211 USA (e-mail: tsbycd@mizzou.edu; kellerj@missouri.edu; skubicm@missouri.edu; erikstone@mail.mizzou.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2013.2260756

Berrada *et al.* [3] used simple image subtraction techniques to extract the pixels indicating motion and then computed the mean and standard deviation along the horizontal and vertical axes to identify sit, attempt to sit, stand, and walk in a given sequence. Checking for motion in the frames in the last 2 min before the stand frame gave the start frame for the beginning of the sit-to-stand activity.

Oikonomopoulos *et al.* [4] used visual operators based on optical flow techniques and B-splines for activity recognition of running, jumping, walking, and other activities. The final classifier used was the relevance vector machine, which is supervised in nature, thus requiring that training data be labeled. In another approach for activity segmentation, Stauffer and Grimson [5] proposed clustering the RGB values of pixels to detect background changes, but the activities were identified using a large database with prototypes of all the activities. This made the segmentation more supervised in nature.

Goffredo *et al.* [6] analyzed sit-to-stand by finding the points of flexion in the shoulder, knee, and hip region using the Gauss–Laguerre transform and subsequently tracking these natural markers to obtain the trajectories of these points. The angles obtained at the points of flexion were compared to analyze the sit-to-stand activity.

While the algorithms mentioned previously were based on vision sensors under normal illumination, McMurdo and Gaskell [7] and Girardi *et al.* [8] conducted several experiments that indicate the severe fall risk of older adults in low lighting conditions. This created a potential problem since nocturnal activities are an important aspect of an independent lifestyle. This, in turn, creates a need for surveillance techniques that can be implemented in the absence of light or under negligible lighting conditions. Maadi and Maldague [9] conducted a study that involves dynamic infrared (IR) sensors. They first implemented background subtraction, then classified different objects, and finally, tracked these objects. The tracker employed iterative systems of location predication (for the next frame) and correction based on the location of detected objects in the current frame. To compensate for global motion, Strehl and Aggarwal [10] used a multiresolution scheme based on the affine motion model to detect independent moving objects using forward looking IR cameras.

Recently, Microsoft introduced the Kinect sensors that incorporate an IR projector and camera, thus enabling it to function as a vision-based sensor under all lighting conditions. This is useful in gathering activity information during all times of the day. Among recent work using Kinect sensors, Sung *et al.* [11] used the sensor to construct an indoor (e.g., office, kitchen, bedroom,

bathroom, and living room) activity dataset for 12 categories of activity detection. In addition to RGBD images (RGB and depth images), the database also created skeleton motion data by defining each pose as a rigid skeleton with 15 joints. They trained a two-layered Markov model for classification. Along similar lines, Ni *et al.* [12] used the RGB information in conjunction with the depth information from Kinect sensors to detect activities such as entering the room, exiting the room, eating, going to bed, etc. Spatiotemporal interest points were considered using features such as Histogram of Oriented Gradients and Histograms of Optic Flows. Classification of the activities was done using K-nearest neighbors and support vector machines (SVMs).

In our system, background subtraction techniques using a mixture of Gaussian models with color and texture features are used on the raw image data to separate the foreground from the background, and the resulting silhouettes are then taken as input to the automatic activity segmentation system. Since our goal is to build an automated video surveillance system to continuously monitor elderly persons as they perform their day-to-day activities, we maintain their privacy by using silhouettes instead of raw images for further analysis. It has been shown previously that silhouettes address the privacy concerns of elderly persons participating in our studies and increase their willingness to accept video monitoring systems in their households [13]. From these silhouettes, image moments are extracted, which are then clustered using fuzzy clustering techniques to produce fuzzy labels in the basic activity categories. Clustering for training sequences followed by human labeling of the cluster centers forms our strategy of activity state recognition. The real-time analysis is done by a fuzzy nearest prototype classifier. As we have done for 3-D processing [14], the resulting activity state memberships can be fed into a higher level rule-based system for fall detection or fall risk assessment.

Clustering is in itself a very fuzzy concept [15]. Depending on the clustering algorithm implemented, the criterion function to be optimized changes, and the nature and shape of the clusters vary. Hence, a key concept in clustering, fuzzy or otherwise, is that there can be multiple viable solutions depending on the dataset and clustering technique used. This issue will be revisited later in this paper. There have been several approaches used to recognize activities using fuzzy clustering. D'Urso and Massari [38] used a multicamera system to create a database of posture vectors of different people. They then compared the fuzzy distances of the test sequence with the mean posture vectors from the database to identify the person. In [39], the researchers tried to distinguish between eating, drinking, and apraxia (eating disorder) by creating a 3-D volume prototype of the different activities using a single camera (time is the third dimension in this case). The test vector was then classified into one of the activities on the basis of the nearest fuzzy distance from the previously created prototypes. While [38] required a multicamera system, [39] was used for specific activities with the requirement that the person was sitting at a fixed distance from the camera. Our algorithm relaxes that constraint and utilizes a single vision sensor with the requirement that the entire silhouette of the person be present in the sensor's field of view.

Our previous work [16]–[18] describes the use of fuzzy clustering techniques in identifying sit-to-stand frames using image moments on visible light data. The work described in here builds on that previous research by extending the results obtained from standard web camera images to cameras with IR lens and finally, using the Kinect sensors. We emphasize that this paper does not extend the state of the art in fuzzy set theory; rather, it describes a novel application of fuzzy clustering and fuzzy prototype classification to video streams from several imaging sensors in a real environment. We are not trying to capture all the activities of daily living, but instead our work focuses on labeling specific activities such as upright, sit, transition, on the bed, or on the floor. These activities are either directly or indirectly related to fall detection and to fall risk assessment activities such as sit-to-stand, stand, or the "timed up and go" test used by physical therapists [41]. Moreover, we have placed the sensors only in the living room area in the apartments in order to address privacy concerns of the residents.

Our group has worked on recognizing indoor activities using several approaches. Zhou *et al.* [33] used a single fisheye Unibrain camera with 180° field of view to measure activity levels using location specific information. The regions of the apartment visible to the camera were manually labeled as the kitchen area, the living area, the bathroom, etc. Then, the location of the identified foreground determined the assumed activity of the person. For example, if the person was in the kitchen area of the perspective image, depending on the activity level detected and location, he could be either cooking or at the dining table. The dining table region was also manually segmented. This approach is more suited to obtaining a rough estimate of the overall level of activity than it is to really identifying specific activities. In another approach, Anderson *et al.* [14], [34] implemented a hierarchical fuzzy rule-based system for activity recognition using a two-camera system to analyze the activity in voxel or 3-D space. This required a rigorous calibration process where locations of common points of interest had to be manually labeled and their world coordinates had to be measured for ground truth. It also relied on the assumption that the locations of both the cameras did not change with respect to each other. Any slight movement of the cameras severely affected the voxel persons, since the intersections changed drastically. The camera system also needed sufficient lighting and cannot "see" in the dark. The processing speed for the Kinect system was approximately 15 frames/s and that for the two-camera system was approximately 8 frames/s, which also requires a graphical processing unit (GPU) (GeForce GTS 450) to increase the processing speed (camera system only). The fuzzy rule-based system required prior training and was built using considerable input from clinicians. Our algorithm uses a single vision sensor (web camera or depth sensor) and only requires the prototype labeling of the initial image sequence. No rigorous training process is needed. It can perform just as well in low lighting conditions, which is an extremely important factor in the case of fall detection as pointed out earlier.

The remainder of this paper is organized as follows. Silhouette extraction and a description of the moments used for clustering are presented in Section II. Section III describes

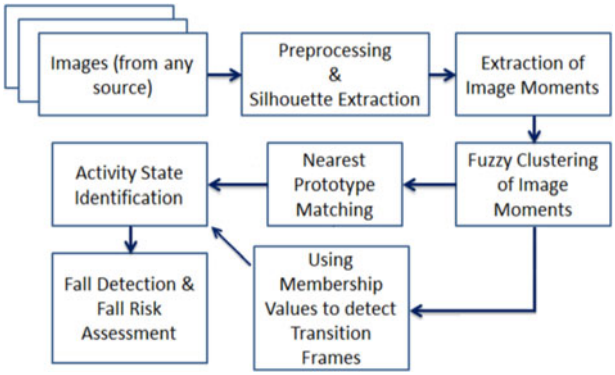

Fig. 1. Block diagram of algorithm.

the fuzzy clustering techniques used for activity analysis and Section IV describes the experimental setup and results using the standard web cameras under normal illumination. Our work using the web cameras with IR filters is described in Section V and the Kinect sensor analysis is described in Section VI. Conclusions and recommendations for future work are presented in Section VII.

II. Methodology

This section describes the silhouette extraction process, the image features employed, and the fuzzy clustering technique that we used in our work. For all three of the image sensors tested in this study (standard web camera under visible lighting, web cameras with IR illumination, and the Kinect sensors), we implemented the same three Zernike image descriptors described in Section II-A and clustered these moments using the Gustafson–Kessel (GK) clustering approach described in Section II-B.

A. Silhouette Extraction and Image Moments

Silhouette extraction is a background change detection technique whose accuracy depends on how well the background is modeled. The background subtraction method implemented in our work uses color and texture features, creating background models using a mixture of Gaussians, and employs shadow removal for greater accuracy. Binary morphological operations [19] are used to fill up holes and remove noise from the extracted silhouettes.

After obtaining the silhouettes from the image sequence, the next step in the algorithm is extracting image moments, as shown in the block diagram in Fig. 1. Image moments are applicable in a wide range of applications such as pattern recognition and image encoding. One of the most widely used is the set of Hu Moments [20]. These are a set of seven central moments taken around the weighted image center. A second set of moment invariants are due to Zernike and are described in [21].

The Zernike polynomials in polar coordinates [20] are given as follows:

$$V_{mn}(r,\theta) = R_{mn}(r) * \exp(jn\theta). \quad (1)$$

The orthogonal radial polynomial is defined by

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s F(m,n,s,r) \quad (2)$$

where

$$F(m,n,s,r) = \frac{(m-s)!}{s!\left(\frac{m+|n|}{2}-s\right)!\left(\frac{m-|n|}{2}-s\right)!} r^{m-2s}. \quad (3)$$

For a discrete image, if P_{xy} is the current pixel intensity (0 or 1 for binary images), the Zernike moments are given by

$$A_{mn} = \frac{m+1}{\pi} \sum_x \sum_y P_{xy} * V_{mn}(x,y). \quad (4)$$

Three of the moments were used in this experiment using (4) with order $m = 2$, 3, and 4 and angular dependence $n = 0$, 1 and 2, respectively. Preliminary experiments on the image moments were conducted to decide which image descriptors would most fit our purpose. We compared the results obtained from Hu moments and those from the Zernike moments and found that the Hu moments are scale and rotation invariant, which make them extremely robust and applicable in different scenarios. However, they are nonorthogonal in nature, i.e., their basis functions are correlated, making the information captured redundant. In contrast, the Zernike orthogonal moments comprise image moments with higher performance in terms of noise resilience, information redundancy, and reconstruction capability. The results of comparing GK clustering on Hu moments with those with the Zernike moments are described in [16]. From those experiments, we focus on the Zernike moments for this work.

The clustering algorithms used for the identification of activity states are explained in the next part of this section.

B. Fuzzy Clustering—Gustafson–Kessel Clustering

Fuzzy clustering techniques are used to partition data on the basis of their closeness or similarity using fuzzy methods. As opposed to the hard clustering, each element can belong to a certain cluster with varying degrees of membership.

The GK [22] fuzzy clustering technique was implemented on the Zernike image moments described previously. This is a very popular clustering algorithm with applications in several fields such as image processing, pattern recognition, system identification, and classification [23]. One of the reasons we chose the GK clustering technique is that it is well suited for the ellipsoidal clusters produced by the Zernike moments. This algorithm is an extension of the fuzzy c-means algorithm in which each cluster has its own unique covariance matrix which makes it more robust and more applicable to various datasets which contain ellipsoidal clusters of different orientations and sizes [24]. The basic clustering approach is well known and is summarized here for completeness.

Algorithm:

1. Fix $c =$ number of clusters & initialize the iteration counter $t = 1$.
2. Initialize membership matrix U for all the data points and for each of the clusters. (The initialization is explained further in this section.)

3. *Do*
4. Compute the cluster centers using

$$\mu_j(t) = \frac{\sum_{i=1}^{N} u_{ij}{}^{q}(t-1) * x_i}{\sum_{i=1}^{N} u_{ij}{}^{q}(t-1)}. \quad (5)$$

5. Compute the covariance matrices for each of the clusters as in

$$\Sigma_j(t) = \frac{\sum_{i=1}^{N} u_{ij}{}^{q}(t-1) * (x_i - \mu_j(t)) * (x_i - \mu_j(t))^T}{\sum_{i=1}^{N} u_{ij}{}^{q}(t-1)}. \quad (6)$$

6. Update the partition matrix:

$$u_{ik}(t) = \frac{1}{\sum_{j=1}^{c} (\frac{D_{ik}}{D_{jk}})^{2/(m-1)}} \quad (7)$$

using the Mahalanobis distance, D_{ik}, which is given by

$$D_{ik}^2 = (x_k - \mu_i(t))^T * \left[|\Sigma_j(t)|^{\frac{1}{l}} * \Sigma_j(t)^{-1} \right] * (x_k - \mu_i(t))$$

where l is the length of feature vector x.
7. Increment the iteration counter t.
8. Until $\| \mu(t) - \mu(t-1) \| < \epsilon$ or $t > t_{\max}$, where ϵ is the maximum permissible error and $t_{\max}$ is the maximum number of iterations specified.

Here, $\mu(t)$ is the vector of all centers, and the distance norm employed to determine convergence is the standard Euclidean distance. An important point to be noted is that it is essential to initialize the membership values to random values but with the means equal to $1/c$ (where c is the number of clusters) and standard deviation equal to 1 so that the algorithm converges at a much faster rate. Standardization is essential because it ensures that equal importance is given to each of the moments used. Otherwise, the algorithm would focus on the moments with the highest range.

It is worth noting that since we constrain the determinant of the covariance matrix to be 1, we impose restrictions on the size of the clusters, and as a consequence, the identified ellipsoidal clusters have to be of similar size [24], [25].

III. Experimental Setup and Results—Web Cameras

A. Experimental Setup

As mentioned earlier in Section II-A, silhouettes from raw RGB image data were extracted by segmenting the human body from the background with the camera at a fixed location. Before silhouette extraction can occur, an accurate background model was acquired. The background is defined as any nonhuman static object. After the background model is initialized, regions in subsequent images with significantly different characteristics from the background are considered foreground objects. Areas classified as background are also used to update the background model. Fused texture and color features are used for background subtraction [26]. The results of the silhouette extraction are shown in Fig. 2.

Fig. 2. (a) RGB image of the side profile of a person sitting on the chair. (b) Silhouette extracted after extracting the foreground from the image.

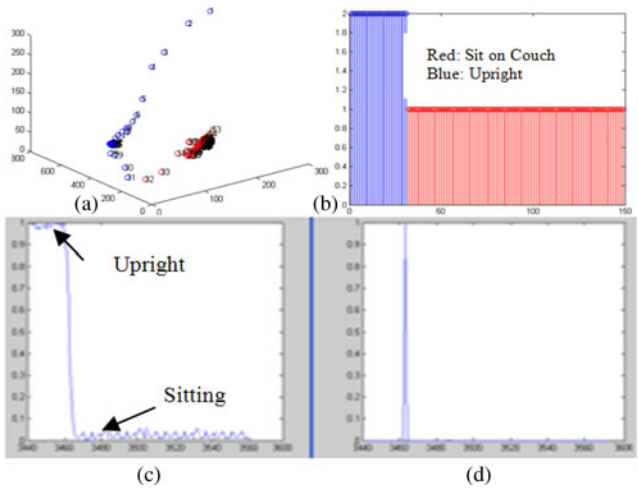

Fig. 3. (a) GK Clustering results on an upright-to-sit sequence (X-, Y-, Z-axes represented by the Zernike image moment values). (b) Hardened fuzzy membership values (X-axis represents the image frame number and Y-axis represents the membership values: 1—upright, 2—Sit). (c) Membership values of a test sequence by frame number for the upright activity. (d) Transition frames based on thresholding the membership values. The trigger shown indicates the identified transition region.

B. Classifying the Transition Frames

Fig. 3(a) shows the results of clustering the Zernike moments from an upright-to-sit sequence, and Fig. 3(b) shows the hardened membership values. The two clusters are color coded with red indicating the sit activity and blue indicating upright. While the clustering technique separates the two activity frames, the labeling itself is implemented by the semisupervised approach described in Section III-C. The transition membership values of the sit-to-stand activity are highlighted in Fig. 3(c). From Fig. 3(c), we can see that the membership (in the "upright" cluster) is initially high and falls to almost zero.

For the frames indicating "transition" motion (sit-to-upright or upright-to-sit or upright-to-floor or floor-to-upright), the membership is intermediate (values 0.2 to 0.8). Subsequently, intermediate memberships are used to identify transitions in activity sequences. This can be seen in Fig. 3(d), which indicates the location of the transition from upright-to-sit in this particular sequence.

C. Prototype Matching

After using the GK, as with other fuzzy clustering algorithms, we obtain a predefined number of clusters. However, without any *a priori* information, there is no way to link a cluster with a specific activity. While we could use the fact that most data runs would begin with a person walking into the room, thus making the initial frames belong to the "upright" cluster, we wanted to make the algorithm more robust and independent of such *a priori* assumptions. To accomplish this, we followed a semisupervised approach, wherein the prototypes of some training data runs were used to identify the activity cluster of the current data. We used the nearest prototype matching for the sequences above and compared the distance between the Zernike moment image vectors of each new image frame with the labeled vectors of previous sequences. The average distance between the vectors over all the sequences of a given activity, such as "on the floor," was averaged and the activity of the new frame was identified.

D. Experimental Results

Experiments were conducted using video sequences collected with visible light, IR illumination in the dark, and depth imagery from the Kinect sensor. The identical approach, which has been described previously, was used successfully with each of these different modalities. The experimental results from the visible web camera experiments are described as follows.

1) Laboratory Settings: We established an image sequence database at a resolution of 640×480 with two fixed fisheye lens Unibrain cameras with range of viewing angle equal to 180°.

We first tested on sit-to-stand runs from seven people aged 18–88. The five participants were three healthy young adults and two elderly adults over the age of 80. The chair used had a standard seat height (approximately 46 cm) as suggested in the Berg Balance Scale test [27]. The video sequences were captured at a rate of 5 frames/s.

To explore the results for a range of sit-to-stand styles, different types of sit-to-stand motions were acted out, including a slouched sit-to-stand, which is common as elderly people start bending forward with age, a sideways slouch (both left and right) to depict a patient with paralysis, and sit-to-stand with legs away from body to portray a patient suffering from a knee injury. Two physical therapists were included in the participant group. They demonstrated the abnormal sit-to-stand motions, which show how paralysis, old age, and knee injuries affect a person's ability to get up from a chair. Each of these motions was repeated multiple times by each of the five healthy young subjects. The two healthy elderly participants were asked to repeat their usual sit-to-stands five times each. In all, 70 runs were taken with 30 of them being the normal healthy runs and the remaining 40 the elderly or abnormal sit-to-stands mentioned previously.

Fig. 4 shows frame examples with the original raw images, the corresponding silhouettes, and the classification results for frames indicating sitting, transition (sit-to-stand), and upright activities. The silhouettes are color coded according to the classified activity. Classification was done by thresholding the membership functions. If the membership value is greater than 0.9, then the frame is assigned to the class that the membership value indicates. If the membership value is between 0.1 and 0.9, then the silhouette is identified as a transition frame. Using this thresholding technique, there are some errors in the transition frames identification, for example, a single frame identified as being upright between a succession of transition frames on both sides, but these are easily rectified using a simple averaging filter on the classification results.

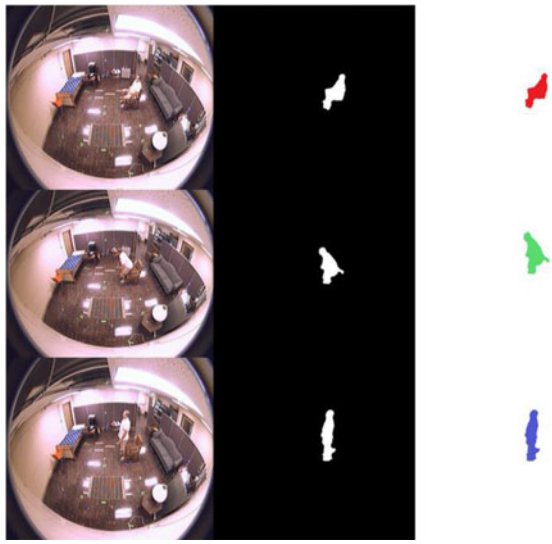

Fig. 4. Segmented activities of a sit-to-stand sequence indicating sitting, transition, and upright with the silhouettes and color-coded silhouettes according to the identified activity. Membership values are thresholded from the GK clustering results using the Zernike Moments as features and then classified after prototype matching.

The results were compared against the Vicon Nexus Motion Capture system that was used as ground truth; 94% accuracy for activity classification from 21 sit-to-stand sequences performed by seven participants was obtained using the GK clustering on the Zernike moments. The sequences contained approximately 300 frames each. The cluster labeling was done using the prototype matching technique described in Section IV-D. Further analysis and discussion has been described in our previous work [16].

A point to highlight here is the strength of this algorithm in being able to connect the results of soft classification of one image sequence and using its results for the successive ones.

2) TigerPlace: Since the goal is to employ our methods in an unstructured environment with older adults, we tested our algorithms at the senior housing facility called TigerPlace. This IRB-approved study included older adults from the ages of 83–97 with multiple health issues such as orthopedic disorders and heart problems as well as some requiring the assistance of a cane to walk. Ten residents from TigerPlace participated in our research and we have tested the algorithm on these ten sequences. A "repairman" was involved in each scenario with the older adult. There is interaction between the two individuals in the scenario. This scenario comprised activities which the subjects normally perform on a daily basis such as walking around, stretching their arm to reach for something, sitting on a chair

TABLE I
CONFUSION MATRIX OF THE GK ALGORITHM FOR ZERNIKE MOMENTS WITH RESPECT TO THE MANUALLY SEGMENTED GROUND TRUTH FOR THE ACTIVITIES OF SIT, TRANSITION, AND UPRIGHT FROM THE STANDARD WEB CAMERA IMAGE SEQUENCES IN THE APARTMENTS AT TIGERPLACE

GT \ GK	Sit	Transition	Upright
Sit	540	6	4
Transition	21	129	22
Upright	3	9	651

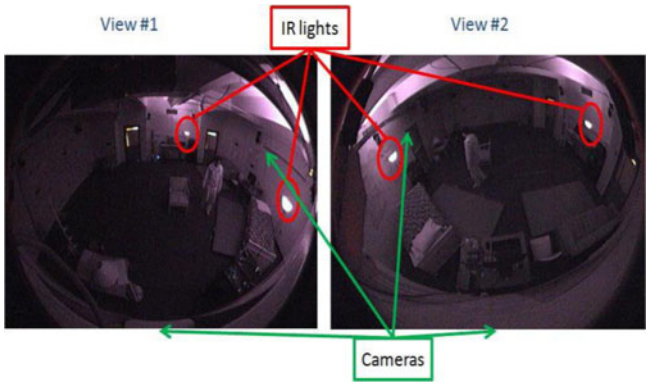

Fig. 5. Two sample images of our two camera views and their positions; red ellipses indicate the positions of the IR lamps, emitting IR radiation which cannot be seen by our eyes.

and getting up, greeting the repairman, and stepping over an object on the floor, which are activities that would occur daily in the lives of the residents at TigerPlace. For the purpose of our experiment, we have focused on the "walking and sit" part of the sequence. The chair in the settings is approximately at right angle to one of the cameras and is facing the other camera in the scenario. The results are shown in Table I. Each of the image sequences was recorded at 5 frames/s and the duration was approximately 150 frames.

In these tests, 94% accuracy was obtained for the sit frames, 90% for transition frames (sit-to-stand or stand-to-sit), and 96% for upright frames. These results were obtained after using a temporal filter with window size 5.

IV. EXPERIMENTAL SETUP AND RESULTS—WEB CAMERAS WITH INFRARED LIGHT

Once the algorithm was established using normal illumination, our next step was analyzing activities taking place at night. This section discusses our preliminary work using the standard Unibrain web cameras with IR filters.

A. Web Cameras Using Infrared Illumination

We established an image sequence database at a resolution of 640×480 with two fixed fisheye lens Unibrain cameras with viewing angle of 180°. With IR lights ON, cameras can see what our eyes cannot. Fig. 5 shows two sample images of these two

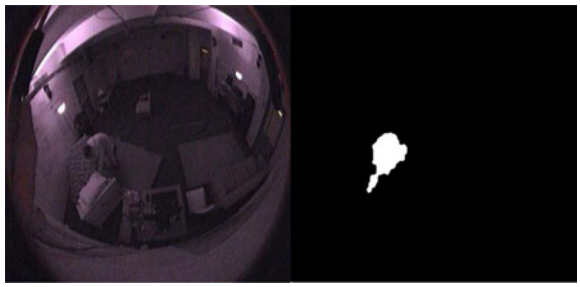

Fig. 6. Silhouette quality obtained using IR emitters in a dark room. This shows a stark difference with the silhouette obtained in Fig. 2.

camera views and their positions. As seen in later parts of this section, we can recognize activities even with the degradation in silhouette quality. Four students were asked to practice several activities under low light conditions in laboratory settings. Note that in the visible spectrum, these images are completely black. IR lighting was used with a wavelength of 850 nm; in total, there were 216 individual IR LEDs distributed between the two lamps with a total power draw of approximately 20 W. The camera lens was a fisheye type with a 180° horizontal field of view, and a 131° vertical field of view when used on a $1/4''$ imager. There is no IR filter on the camera lens. We considered the possible activities practiced at night and included them in our data collection: walking, standing with hand motion, standing without hand motion, sitting down and standing up, sitting on a sofa, going to bed, and getting up from bed. Our data collection also contained the following situations on the bed: sleeping (lying on bed), being sleepless (flipping with some movements, i.e., with a few toss and turns), sitting on the bed, and transitioning from sitting to lying on the bed. In addition, four abnormal activities (falling) were included: walking in the room and falling to the ground due to loss of balance, slipping when trying to get up from a chair, falling when trying to get up from a bed, and falling out of the bed when sleeping. The frame rate is 3 frames/s. We collected more than a half hour of frames (5400 frames) for each person. The camera positions and the laboratory setup are shown in Fig. 5.

Fig. 6 shows the silhouette extracted from a dark room illuminated by IR emitters. We can see that there is a significant difference in the quality of the silhouette in Fig. 6 and that obtained under normal illumination in Fig. 2.

B. Experimental Results

Preliminary experiments were conducted to establish the input parameters and best features to use for these data. Several participants performed different activities. As described in Section II, silhouettes were extracted from the raw image sequences, and the moment features were computed.

The GK clustering technique requires the number of clusters to be specified as an input parameter. In preliminary experiments shown in [16], we demonstrated that clustering the Zernike moments using the GK algorithm with the number of clusters initialized to the number of activities yielded the best results. Since single camera images are used here, the activities of

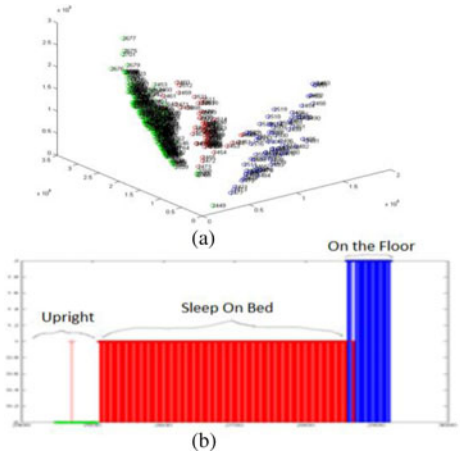

Fig. 7. Test results on an IR video sequence with three activities—upright, asleep on bed, and on the ground. GK on Zernike Moments and clustering results into (a) three clusters and (b) hardened membership results of Zernike Moments by frame number.

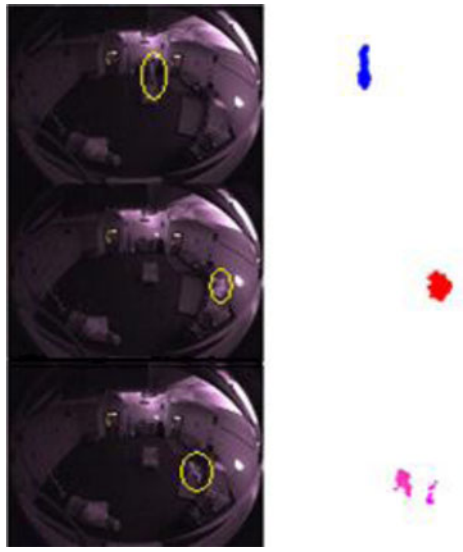

Fig. 8. Segmented activities of a three-activity sequence indicating upright, on the bed, and on the floor activity states with the color-coded silhouettes according to the identified activity. Membership values are thresholded from the GK clustering results using the Zernike Moments as features and then classified after prototype matching.

walking and standing cannot be differentiated in general; thus, they are grouped together as "upright" frames for the purpose of activity recognition.

Fig. 7(a) shows the clustering results of one data sequence using an input of three clusters. Fig. 7(b) shows the clustering results with the X-axis that indicates the frame number in the sequence and the Y-axis that indicates the cluster number after hardening the membership matrix. These results have been color coded for display purposes and black-colored points indicate the areas where the points are densely clustered. Note that while it is evident that the two clusters obtained represent the "sit" and "on the floor" activities, without any prior information, we were unable to identify which cluster indicates which activity. The solution to this is explained in Sections III-B and III-C.

The scenario involved a participant who performs several actions in an unlit room. The three activities performed in the scenario below were nighttime activities of moving around in the room (upright), sleeping on the bed, and then falling onto the floor. As can be seen in Fig. 7(b), the activities were well separated after fuzzy clustering and partition hardening. However, in this scenario, the activity "fall" is equivalent to "on the floor" since no other parameter has been taken into consideration which could differentiate between the two activities.

The detection of the transition frames as well as identifying the activity state using prototype matching is the same as that described in Sections II-B and C. Using prototype matching, the activity states were identified and the sample results are shown in Fig. 8. In the raw images, the person has been circled in yellow to distinguish the person from the background for visualization. We can see that the silhouettes are fairly noisy and their shapes are quite different from those obtained from the standard illuminated data. However, the results show strong clustering of the image moments obtained from the same activity states. This makes activity analysis possible even in the dark using fuzzy clustering. Fig. 8 shows some of the color-labeled image frames with blue indicating an upright frame, red signifying on the bed activity label, and pink representing on the floor activity label.

V. Experimental Setup and Results—Microsoft Kinect

In 2010, Microsoft released a new inexpensive device called the Kinect, to allow controller free game play on their Xbox system. The device uses a pattern of actively emitted IR light to produce a depth image. Since the sensor uses depth information, it is invariant to illumination changes and works extremely well under negligible illumination for indoor environments. The depth data returned from the device (at 30 frames/s) is an 11-bit 640×480 image. The precision of the returned depth values is dependent on the distance. The details are given in [28].

A. Overview of Kinect Sensors

In order to extract the foreground from the background, a standard background subtraction algorithm was employed, different from the one used for visible and IR illuminated scenes. A set of training background images without any person present was collected before each experiment, and the minimum and maximum values of each pixel of the background frames were stored. In the subsequent frames, the pixel values were compared against the minimum and maximum values, and if the value was outside the range, it was considered a foreground pixel. A smoothing operation was used to remove noise in the image frames using anisotropic diffusion [29]. Fig. 9 shows samples of Kinect depth data taken of a person walking, about to sit, and sitting, respectively. The figures below them show the extracted foregrounds using the technique described previously. In the depth images, the person is circled in red to differentiate the silhouette from the background. It is worth noting that parts of the person tend to get absorbed in nearby objects as the person gets further away from the Kinect sensor. In such cases, the object (chair in Fig. 9) and the foreground have similar depth information, which can slightly affect the extracted foreground data. In addition, note

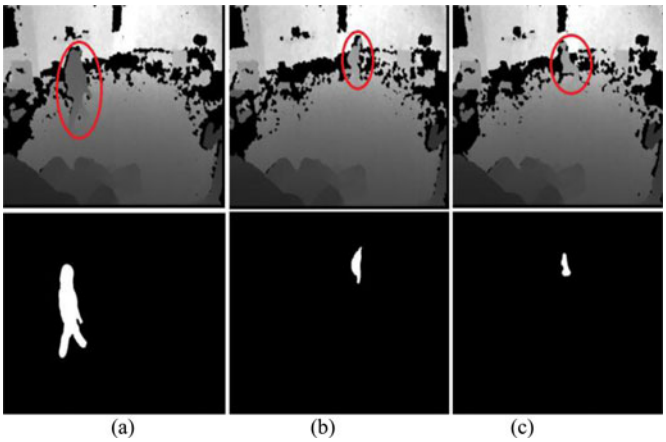

Fig. 9. Sample images of the Kinect depth images of a person (a) walking, (b) about to sit, and (c) sitting and the extracted foreground of each below.

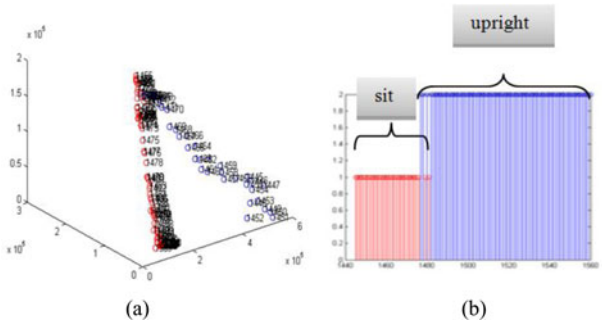

Fig. 10. Test results on a depth video sequence with two activities—sit and upright. GK on Zernike Moments and clustering results into (a) two clusters and (b) hardened membership results of Zernike Moments by frame number.

that for the depth map, brighter pixels mean larger depth values. Some black regions correspond to depth measurement errors due to surface reflections.

B. Experimental Results

1) Laboratory Settings: Preliminary experiments were conducted using four participants with each participant walking back and forth in the lab and performing several sit-to-stands. Using the Zernike moments from the silhouette of each frame, fuzzy clustering using the GK algorithm was implemented with the number of clusters defined as equal to the number of activity states within the data run. The result of clustering a part of a sequence using the Kinect is shown in Fig. 10

The membership results were obtained by hardening the values obtained from clustering the image moments. As can be seen from the results, there are a few frames between the sit and upright regions which have misclassified the results. However, these are rectified by identifying the transition frames as those having membership values between 0.9 and 0.1. The process of obtaining the transition frames as well as prototype matching is the same as described in Sections II-B and II-C.

2) TigerPlace: As a means for testing the algorithms implemented in our research laboratory, as well as continuously assessing gait parameters of older adults in their homes, Kinect cameras have been placed in ten apartments at TigerPlace. We

TABLE II
CONFUSION MATRIX OF THE GK ALGORITHM FOR ZERNIKE MOMENTS WITH RESPECT TO THE MANUALLY SEGMENTED GROUND TRUTH FOR THE ACTIVITIES OF SIT, TRANSITION, AND UPRIGHT FROM THE DEPTH IMAGE SEQUENCES IN THE APARTMENTS AT TIGERPLACE

GT \ GK	Sit	Transition	Upright
Sit	432	15	6
Transition	10	111	16
Upright	5	32	1109

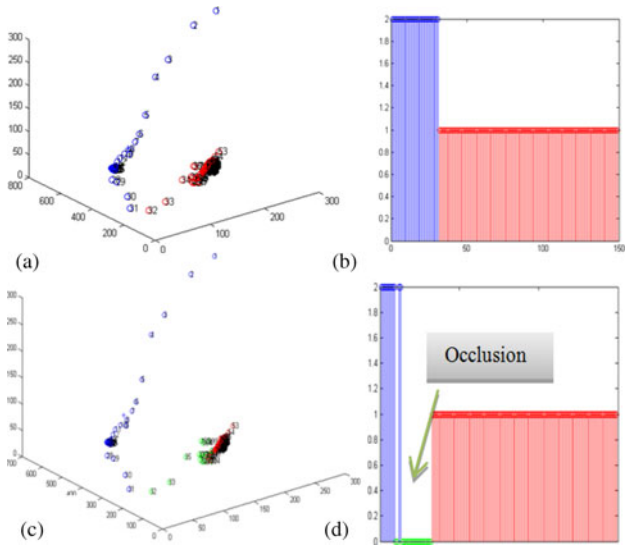

Fig. 11. Test results on a depth video sequence with two activities—sit and upright. GK on Zernike Moments and clustering results into (a) two clusters and (c) three clusters and their hardened membership results of Zernike Moments by frame number (b) and (d). The occluded region is marked in (d).

analyzed sequences taken from nine residents inside their own apartments. For ground truth, we manually segmented the sequences into the three activity states—sit, upright, and transition (sit-to-stand or stand-to-sit). As in Fig. 1, we manually labeled the activity states of the first sequences and used the nearest neighbor approach to classify the remaining sequences (described in Section II-C).

The results are shown in Table II. The total activity classification is ∼95% with the transition frame classification 81.4%, upright frames 96.7%, and sit frames 95%.

In eight of the nine sequences examined here, the number of image frames identified with any of the three states mentioned earlier matched that of the ground truth successfully. However, one of the sequences accounted for approximately 30% of the misclassified transition frames.

The reason for such a high percentage of misclassifications was due to the presence of occlusion in the image sequence, in this case the presence of a chair, which blocked the lower part of the body from camera view. In fact, the occluded frames form a separate cluster by themselves. This is further seen in Fig. 11. Fig. 11 shows the results of clustering the occluded depth

TABLE III
COMPARISON OF THE GK ALGORITHM FOR ZERNIKE MOMENTS WITH RESPECT TO THE MANUALLY SEGMENTED GROUND TRUTH FOR THE ACTIVITIES OF SIT, TRANSITION, AND UPRIGHT FROM THE DEPTH IMAGE SEQUENCES IN THE APARTMENTS AT TIGERPLACE FOR DIFFERENT MACHINE LEARNING TECHNIQUES

% Correct	NN	SVM	HCM	FCM	GK
1	74.1	75	56.3	79.9	96.65
2	68.8	69.4	70.4	79.3	94.4
3	63.1	66	69.3	81	95.7
4	71.3	68.2	50.5	82.2	93.9
5	69.2	76.6	75.4	92.1	97.4
6	66.4	71.2	81	85.5	96.3
7	69.3	70.4	56.2	88.1	94.8
8	77.3	75.5	55.8	61.2	89.7
Average	**69.94**	**71.54**	**64.36**	**81.16**	**94.86**

sequence using GK clustering on the Zernike image moments with two clusters [see Fig. 11(a) and (b)] and three clusters [see Fig. 11(c) and (d)], respectively. As can be seen in Figs. 11(c) and (d), the green cluster points represent the occluded part of the sequence. Currently, we are working on techniques to address the occlusion problem.

C. Comparison With Other Algorithms

In our previous work [16], we compared the results obtained from clustering the Zernike descriptors with the Hu moments and found the former to perform better for our application. For the purpose of comparing our technique with other established supervised and unsupervised techniques, we use the same data described in Table II and test the results obtained from the GK algorithm with those obtained from a multilayer perceptron neural network with back propagation algorithm (NN), an SVM, hard c means clustering (HCM), and fuzzy c means (FCM) with Euclidean distance measure on the Zernike image moments. NNs have been used for activity recognition in several studies. In [35], Foroughi *et al.* used features such as projection of histograms and head position to detect activities like walk, bend, and falls by training a neural network. In [36], the authors used motion sensors placed at different locations of the room and used the neural network to estimate the location of the person to determine his activity. In [37], Xue *et al.* tried to distinguish between different types of gaits like walking with a volleyball, walking with a heavy coat, and normal walking using features such as gait energy image. They also compared the results obtained from using NN and SVM as classifier and found that they achieved 74% classification with NN and 78–91% correct classification for the different gaits using SVM.

The NN used had one hidden layer with ten neurons, which appeared sufficient for the dataset under consideration. The SVM implemented for comparison used the linear kernel. For the two supervised techniques as well as the HCM, the ground truth was divided into three classes for upright, transition, and sit. Of the nine sequences, one was used for training/prototype labeling, and the others were used for testing for all techniques. The results are displayed in Table III.

As can be seen, none of the algorithms perform as well as GK clustering. Overall, both the supervised learning methods yielded unsatisfactory results compared with the FCM and GK techniques, although the SVM appeared to have slightly better results. The HCM performed the worst, especially for the transition frames. This is not unexpected due to the crisp nature of the HCM. The FCM performed the best of the competitors, which was as expected since it is most similar to the GK technique. However, the results appear unstable, especially when the number of frames is not equal in the two clusters. The best results were obtained for the case with almost equal cluster sizes, but still significantly below the output of our approach.

VI. IMPLEMENTATION DETAILS

The total processing time to generate silhouettes from the video streams, including background modeling using mixture of Gaussians, with the web camera system, is currently at 8 and 15 frames/s using the Microsoft Kinect sensors [28]. Much of the speedup in this crucial portion of the web camera system is accomplished using GPUs. Hence, the bulk of processing works in real time. Calculating Zernike moments from binary silhouettes is simple and has the approximate computational complexity $O(np^3)$, where p is the order of the Zernike moment [31]. While this could be a cause for concern for higher order calculation of the Zernike moments, it is negligible for our algorithm since we use a maximum order of 4 (see Section II-A). The clustering for building the activity models is done offline and although the size of the membership matrix depends on the data size as well as the number of clusters, the memory requirements are negligible in this terabyte era. Furthermore, for a sequence of 200 frames—the current size of the training sequences, computation time is approximately 15 s, i.e., roughly 13 frames/s. Since our classifier is based on the fuzzy nearest prototype approach, the computation for that is negligible. Thus, our approach to activity state recognition is extremely efficient with a total computational rate of 8 frames/s for the camera system and 15 frames/s for the Kinect system using a quad core processor, since the frame rate depends on the slowest module in each case.

One potential cause for concern is the data storage. For our applications, privacy concerns are crucial. Hence, we only store the processed silhouettes from the apartments at TigerPlace using the camera system. For the Kinect system, we store the silhouettes as well as the depth images; our University Internal Review Board has approved the storage of the depth information since they do not compromise the privacy of our participants. Even with binary silhouettes and depth imagery that can be very efficiently compressed, storing this information requires considerable disk space over an extended period of time, roughly 11 GB per week per resident. This is accommodated by very large secure server drives and a carefully constructed data management plan.

VII. CONCLUSION AND FUTURE WORK

We have presented a successful and yet simple technique for detecting activity frames using fuzzy clustering methods. A soft

classifier was constructed from the clustering results and activity classification results using the Zernike Moments were obtained. Our previous work [16] using the fuzzy clustering technique was compared against the Vicon Motion Capture system in the laboratory and was shown to be accurate in activity recognition. This technique has shown the ability to link the results of one sequence to the soft classification of others, showing its strength and success over a range of image sensors and in real living quarters.

With the use of improved background update techniques, better quality of silhouettes can be obtained and more information can be extracted from real-time environments. Our application has shown its ability to determine different activities under different environments, both controlled as well as unstructured, and it runs in real-time mode. It has been able to distinguish between an on-ground event as well as an upright event (see Figs. 7 and 8). We also compared the results from testing our technique on the depth sequences collected at the TigerPlace apartments with the residents with other classifiers and found the GK clustering to work best for classifying activities specifically related to fall detection and fall risk assessment. An important application of our activity recognition algorithm is the detection of falls in older adults. As mentioned earlier, research indicates a higher risk of falling at night among the older population [30]. Once our algorithm detects an on-ground event, further analysis can be done to analyze whether a fall has occurred, similar to the work presented in [14]. While identifying the occurrence of a fall may not prevent it from taking place, speedy help can be provided, which can potentially save the lives of the residents or at least ensure that immediate treatment is provided. Our application has also been able to detect sitting (see Figs. 7, 8, 10, and 11) which is important for sit-to-stand analysis as a measure of fall risk assessment. Its ability to segment out upright state is also extremely useful for assessing walk-related information to obtain gait parameters. Once our algorithm detects an upright sequence, further processing can be done to determine if the person is standing or walking and if he/she is walking, then the stride parameters such as walking speed, step time, and step length can be computed [28], [32] for further fall risk assessment. Have we now solved the real-time fall detection and fall risk assessment problem? Not yet, but activity state recognition and analysis is a vital component of an overall eldercare monitoring system, which monitors older adults in their apartments using motion sensors, bed sensors, as well as video sensors for early illness detection [1]. We are also working on sensor fusion algorithms using acoustic sensors to gain more activity-related information in the apartments, especially at regions outside the field of view of the sensors [40].

Future work includes creating an online version of the fuzzy clustering algorithms described previously and validating the results using fuzzy validity measures. This can speed up the activity state identification process and lead to faster fall detection. Apart from that, future work will also include finding ways to detect occlusion in image silhouettes. Experiments are currently being conducted to test the algorithm under different scenarios with different activities to test the performance of the algorithm for different locations of the Kinect sensor. This will make the algorithm useful for automatic activity recognition in unstructured settings with any lighting conditions, which will help prevent unmonitored physical injuries from taking place.

REFERENCES

[1] M. Skubic, G. Alexander, M. Popescu, M. Rantz, and J. M. Keller, "A smart home application to eldercare: Current status and lessons learned," *Technol. Health Care*, vol. 17, no. 3, pp. 183–201, 2009.

[2] S. Allin and A. Mihailidis, "Automated sit-to-stand detection and analysis," presented at the AAAI 2008 Fall Symp. AI Eldercare: New Solut. Old Probl., Arlington, VA, USA, 2008.

[3] D. Berrada, M. Romero, G. Abowd, M. Blount, and J. Davis, "Automatic administration of the Get up and go test," presented at the 1st ACM SIGMOBILE Int. Workshop Syst. Network. Supp. Healthcare Assist. Living Environments, San Juan, Puerto Rico, Jun. 2007.

[4] A. Oikonomopoulos, M. Pantic, and I. Patra, "Sparse B spline polynomial descriptors for human activity recognition," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1814–1825, 2009.

[5] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[6] M. Goffredo, M. Schmid, S. Conforto, M. Carli, A. Neri, and T. D'Alessio, "Markerless human motion analysis in Gauss-Laguerre transform domain: An application to sit-to-stand in young and elderly people," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 207–216, Mar. 2009.

[7] M. E. McMurdo and A. Gaskell, "Dark adaptation and falls in the elderly," *Gerontology*, vol. 37, pp. 221–224, 1991.

[8] M. Girardi, H. R. Konrad, M. Amin, and L. F. Hughes, "Predicting fall risks in an elderly population: Computer dynamic posturography versus Electronystagmography test results," *Laryngoscope*, vol. 111, pp. 1528–1532, 2001.

[9] A. El Maadi and X. Maldague, "Outdoor infrared video surveillance: A novel dynamic technique for the subtraction of a changing background of IR images," *Infrared Phys. Technol.*, vol. 49, no. 3, pp. 261–265, 2007.

[10] A. Strehl and J. K. Aggarwal, "MODEEP: A motion-based object detection and pose estimation method for airborne FLIR sequences," *Mach. Vis. Appl.*, vol. 11, no. 6, pp. 267–276, 2000.

[11] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human activity detection from RGBD images," presented at the AAAI workshop Pattern, Activity Intent Recognit., San Francisco, CA, USA, 2011.

[12] B. Ni, G. Wang, and P. Moulin, "RGBD-HuDaAct: A color-depth video database for human daily activity recognition," in *Proc. 1st IEEE Workshop Consum. Depth Camer. Comput. Vision, Conjunct. ICCV*, 2011, pp. 1147–1153.

[13] G. Demiris, D. Parker, J. Giger, M. Skubic, and M. Rantz, "Older adults' privacy considerations for vision based recognition methods of eldercare applications," *Technol. Health Care*, vol. 17, no. 1, pp. 41–48, 2009.

[14] D. Anderson, R. Luke, M. Skubic, J. M. Keller, M. Rantz, and M. Aud, "Evaluation of a video based fall recognition system for elders using voxel space," in *Proc., Int. Conf. Int. Soc. Gerontechnol.*, Pisa, Italy, Jun. 4–7, 2008, pp. 77–82.

[15] J. Keller and I. Sledge, "A cluster by any other name," in *Proc. IEEE Proc. Fuzzy Inf. Process. Soc.*, 2007, pp. 427–432.

[16] T. Banerjee, J. M. Keller, M. Skubic, and C. C. Abbott, "Sit-to-stand detection using fuzzy clustering techniques," in *Proc. IEEE Int. Conf. Fuzzy Syst./World Conf. Comput. Intell.*, 2010, pp. 1–8.

[17] T. Banerjee, J. M. Keller, Z. Zhou, M. Skubic, and E. Stone, "Activity segmentation of infrared images using fuzzy clustering techniques," presented at the World Conf. Soft Comput., San Francisco, CA, USA, May 2011.

[18] T. Banerjee, "Activity segmentation with special emphasis on sit-to-stand analysis," Master's thesis, Dept. Electr. Comput. Eng., Univ. Missouri, Columbia, MO, USA, 2010.

[19] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice–Hall, 1989.

[20] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Inf. Theory*, vol. IT-8, pp. 179–187, 1962.

[21] A. B. Bhatia and E. Wolf, "On the circle polynomials of Zernike and related orthogonal sets," *Math. Proc. Cambridge Philosoph. Soc.*, vol. 50, pp. 40–48, 1954.

[22] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE Annu. Conf. Decision Control*, San Diego, CA, USA, 1979, pp. 761–766.

[23] M. J. Lesot and R. Kruse, "Gustafson-Kessel-like clustering algorithm based on typicality degrees," presented at the Int. Conf. Inf. Process. Manag. Uncertainty, Paris, France, 2006.
[24] R. Babuska, P. J. van der Veen, and U. Kaymak, "Improved covariance estimation for Gustafson-Kessel clustering," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2002, pp. 1081–1085.
[25] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego, CA, USA: Academic, 1999.
[26] R. H. Luke, D. Anderson, J. M. Keller, and M. Skubic, "Human segmentation from video in indoor environments using fused color and texture features," Elect. Comput. Eng. Dept., Univ. Missouri, Columbia, MO, USA, Tech. Rep., 2008.
[27] K. Berg, S. Wood-Dauphinee, J. I. Williams, and B. Maki, "Measuring balance in the elderly: Validation of an instrument," *Can. J. Pub. Health*, vol. 2, pp. S7–S11, 1992.
[28] E. Stone and M. Skubic, "Evaluation of an inexpensive depth camera for passive in-home fall risk assessment," presented at the Int. Conf. Pervas. Comput. Technol. Healthcare Workshop, Dublin, Ireland, 2011.
[29] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," in *Proc. IEEE Workshop Comput. Vis.*, 1987, pp. 16–22.
[30] E. Tanner, "Assessing home safety in homebound older adults: Home safety instrument development," *Geriatric Nurs.*, vol. 24, pp. 250–254, 256, 2003.
[31] C. W. Chong, P. Raveendran, and R. Mukundan, "A comparative analysis of algorithms for fast computation of Zernike moments," *Pattern Recognit.*, vol. 36, pp. 731–742, 2003.
[32] F. Wang, E. Stone, W. Dai, T. Banerjee, J. Giger, J. Krampe, M. Rantz, and M. Skubic, "Testing an in-home gait assessment tool for older adults," in *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Minneapolis, MN, USA, Sep. 2–6, 2009, pp. 6147–6150.
[33] Z. Zhou, X. Chen, Y. C. Chung, Z. He, T. X. Han, and J. M. Keller, "Activity analysis, summarization, and visualization for indoor human activity monitoring," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1489–1498, Nov. 2008.
[34] 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," *Comput. Vis. Image Understand.*, vol. 113, pp. 80–89, 2009.
[35] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in *Proc. 11th Int. Conf. Comput. Inf. Technol.*, Khulna, Bangladesh, 2008, pp. 219–224.
[36] H. Fang and H. Lei, "BP neural network for human activity recognition in smart home," in *Proc. IEEE Int. Conf. Comput. Sci. Serv. Syst.*, 2012, pp. 1034–1037.
[37] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin, "Infrared gait recognition based on wavelet transform and support vector machine," *Pattern Recognit.*, vol. 43, no. 8, pp. 2904–2910, 2010.
[38] P. D'Urso and R. Massari, "Fuzzy clustering of human activity patterns," *Fuzzy Sets Syst.*, vol. 215, pp. 29–54, 2013.
[39] A. Iosifidis, A. Tefas, and I. Pitas, "Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 25–30, 2012, pp. 2201–2204.
[40] Y. Li, T. Banerjee, M. Popescu, and M. Skubic, "Improvement of acoustic fall detection using Kinect depth sensing," in *Proc. 35th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Osaka, Japan, Jul. 3–7, 2013, to be published.
[41] W. Dite and V. A. Temple, "Development of a clinical measure of turning for older adults," *Amer. J. Phys. Med. Rehabil.*, vol. 81, no. 11, pp. 857–866, 2002.

Tanvi Banerjee (S'10) received the B.S. degree in electronics and telecommunications engineering from Pune University, Pune, India, in 2007 and the M.S. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2010, where she is currently working toward the Ph.D. degree in electrical and computer engineering.

She is a Research Assistant with the Center for Eldercare and Rehabilitation Technology, University of Missouri. Her research interests include computer vision, and machine learning, and fuzzy logic.

James M. Keller (F'00) received the Ph.D. degree in mathematics in 1978.

He holds the University of Missouri Curators' Professorship with the Electrical and Computer Engineering and Computer Science Departments on the Columbia campus. He is also the R. L. Tatum Professor with the College of Engineering. His research interests center on computational intelligence: fuzzy set theory and fuzzy logic, neural networks, and evolutionary computation with a focus on problems in computer vision, pattern recognition, and information fusion including bioinformatics, spatial reasoning in robotics, geospatial intelligence, sensor and information analysis in technology for eldercare, and landmine detection. His industrial and government funding sources include the Electronics and Space Corporation, Union Electric, Geo-Centers, National Science Foundation, the Administration on Aging, The National Institutes of Health, the National Aeronautics and Space Administration/Johnson Space Center, the Air Force Office of Scientific Research, the Army Research Office, the Office of Naval Research, the National Geospatial Intelligence Agency, the Leonard Wood Institute, and the Army Night Vision and Electronic Sensors Directorate. He has co-authored more than 400 technical publications.

Dr. Keller is a Fellow of the International Fuzzy Systems Association, and a past President of the North American Fuzzy Information Processing Society. He received the 2007 Fuzzy Systems Pioneer Award and the 2010 Meritorious Service Award from the IEEE Computational Intelligence Society (CIS). He finished a full six-year term as the Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, followed by serving as the Vice President for Publications of the IEEE CIS from 2005 to 2008 and since then has served as an elected CIS Adcom member. He is the IEEE TAB Transactions Chair and a member of the IEEE Publication Review and Advisory Committee. Among many conference duties over the years, he was the General Chair of the 1991 North American Fuzzy Information Processing Society Workshop and the 2003 IEEE International Conference on Fuzzy Systems.

Marjorie Skubic (SM'13) received the Ph.D. degree in computer science from Texas A&M University, College Station, TX, USA, in 1997, where she specialized in distributed telerobotics and robot programming by demonstration.

She is currently a Professor with the Electrical and Computer Engineering Department, University of Missouri, Columbia, MO, USA, with a joint appointment in Computer Science. In addition to her academic experience, she has spent 14 years working in industry on real-time applications such as data acquisition and automation. Her current research interests include sensory perception, computational intelligence, spatial referencing interfaces, human–robot interaction, and sensor networks for eldercare. In 2006, she established the Center for Eldercare and Rehabilitation Technology, University of Missouri, and serves as the Center Director for this interdisciplinary team.

Erik Stone (S'11) received the B.S. degree in electrical and computer engineering and the M.S. degree in computer engineering from the University of Missouri, Columbia, MO, USA, in 2006 and 2009, respectively, where he is currently working toward the Ph.D. degree in electrical and computer engineering.

He is a Research Assistant with the Center for Eldercare and Rehabilitation Technology, University of Missouri. His research interests include computational intelligence, computer vision, and pattern recognition.