Activity Segmentation of Infrared Images Using Fuzzy Clustering Techniques

Tanvi Banerjee, *Student Member IEEE*, James M. Keller, *Fellow IEEE*, Zhongna Zhou, *Student Member IEEE,* Marjorie Skubic, *Member IEEE &* Erik Stone, *Student Member IEEE.*

***Abstract*—Every year, many older adults are at risk for falling, especially in the dark. Infrared lighting provides a non-intrusive lighting in the dark and our research shows a technique of segmenting human activities using fuzzy clustering of image moments even in the dark. While our research is still in the preliminary stages, it shows promise of being able to detect several different activities and in the future might prevent several falls from taking place.**

I. Introduction

FALL risk assessment has been a goal of our research [1] as we continue to conduct experiments at Tiger Place, an "aging in place" facility for the elderly. McMurdo et al. [17] & Girardi et al. [18] conducted several experiments under indicating the severe fall risk of older adults under insufficient lighting conditions. However, nocturnal activities are an important aspect of an independent lifestyle. This creates a need for surveillance techniques which can be implemented in the absence of light. Maadi and Maldague [20] conducted a study which first subtracts the background, then classifies objects, and finally tracks the objects. The tracker employs 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 [1] used a multi- resolution scheme based on the affine motion model for detecting independent moving objects using forward looking infrared (FLIR) cameras.

In our system, background subtraction techniques using Mixture of Gaussian models with 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 [14]. From these silhouettes, image moments are extracted, which are then clustered to produce fuzzy labels in the basic activity categories.

Clustering is in itself a very fuzzy concept [2]. 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 no clustering result is right or wrong. Depending on the data set, there can be several possible results, each of which is correct. This issue will be revisited later in this paper.

Oikonomopoulos et al. [3] used visual operators based on optical flow techniques and B splines for activity recognition of running, jumping, walking and other activities. However, the final classifier used was the Relevance Vector Machine which is supervised in nature, thus requiring labeled training data. In another approach for activity segmentation, Stauffer et al. [4] proposed clustering the RGB values of pixels to detect background changes, but the activities were identified using a huge data base with prototypes of all the activities which essentially made the segmentation more supervised in nature.

While clustering was employed in some of the above mentioned techniques and silhouettes were extracted in others, none of them use the combination to segment activities. Our previous work [5] describes our work in using fuzzy clustering techniques in identifying sit-to-stand frames using image moments on visible light data and has inspired the work described in this paper.

The remainder of this paper is organized as follows. Silhouette extraction and a description of the moments used for clustering is presented in Section II. Section III describes the fuzzy clustering techniques used for activity analysis, and Section IV describes the preliminary results for selection of the image moments and number of clusters used to initialize the algorithm. The experiments conducted and their results are described in Section V. Finally, the conclusion and future work are presented in Section VI.

II. Silhouette Extraction and Image Moments

Silhouette extraction is a background change detection technique whose accuracy depends on how well the

Manuscript received January 31, 2010. This work was supported by the U.S. National Science Foundation under Grant IIS-0703692.
T. Banerjee is with the University of Missouri, Columbia, MO 65211 USA (e-mail: tsbycd@mizzou.edu).
J. M. Keller, M. Skubic, Zhongna Zhou and Erik Stone are with the University of Missouri, Columbia, MO 65211 USA (e-mails: kellerj@missouri.edu, skubicm@missouri.edu , zz3kb@mail.missouri.edu, erikstone@mail.mizzou.edu)

background is modeled. The background subtraction method implemented in our work uses color and texture features and employs shadow removal for greater accuracy. Finally, binary morphological operations are used to fill up holes and remove noise from the extracted silhouettes. The technique is explained in detail in [6].

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 Figure 1. Image moments are applicable in a wide range of applications such as pattern recognition and image encoding. One of the most important and popular set of moments is the set of Hu Moments [7]. These are a set of seven central moments taken around the weighted image center. In particular, the first three Hu Moments are more robust than the other Hu Moments in the presence of noise and were used in this analysis.

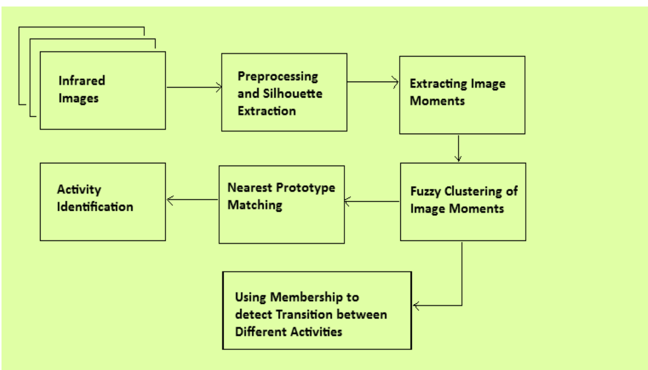


Fig. 1. Block Diagram of Algorithm

One of the most widely used set of image moments are the Hu moments, which are scale and rotation invariant which make them extremely robust and applicable in different scenarios. However, they are non-orthogonal 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 Zernike polynomials in polar coordinates [8] are given as:

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

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:

$$\mathrm{A}_{\mathrm{mn}} = \frac{\mathrm{m}+1}{\pi} \sum_{\mathrm{x}} \sum_{\mathrm{y}} \mathrm{P}_{\mathrm{xy}} * \mathrm{V}_{\mathrm{mn}}(\mathrm{x},\mathrm{y}), \quad (4)$$

Three of the moments were used in this experiment using equation (4) with order, m=2, 3, and 4 and angular dependence, n=0, 1 and 2 respectively. These were selected after implementing Principal Component Analysis to see which moments were most suitable for this application. The clustering algorithms used in the experiments are explained in the next section.

III. FUZZY 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 means, here each element can belong to a certain cluster with varying degrees of membership. The Gustafson Kessel [16] fuzzy clustering technique was implemented on the image moments described above.

Gustafson Kessel clustering technique:

The Gustafson Kessel (GK) Algorithm is an extension of the Fuzzy C-Means algorithm in which each cluster has its own unique covariance matrix. This makes the algorithm more robust and more applicable to various data sets which contain ellipsoidal clusters of different orientations and sizes [13]. The basic clustering approach we use is well known and has been 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 equation (5).

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

5. Compute the covariance matrices for each of the clusters as in equation (6).

$$\Sigma_{\mathrm{j}}(\mathrm{t}) = \frac{\sum_{i=1}^{N} {\mathrm{u}_{\mathrm{ij}}}^{\mathrm{q}}(\mathrm{t}-1) * (x_i - \mu_j(t)) * (x_i - \mu_j(t))^{\mathrm{T}}}{\sum_{i=1}^{N} {u_{ij}}^q(t-1)}, \quad (6)$$

6. Update the partition matrix:

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

using the Mahalanobis distance, D_{ik}, given by:

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

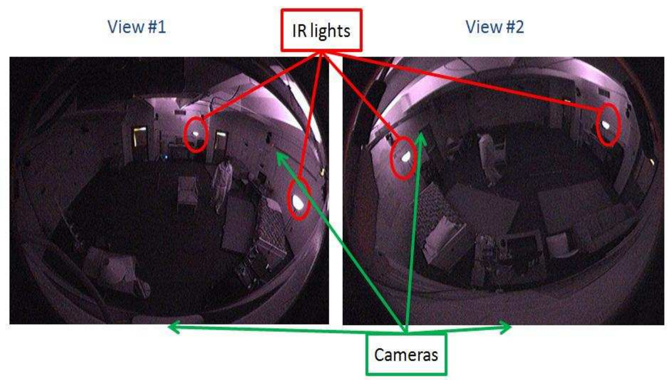
where l is the length of feature vector x.

7. Increment the iteration counter t.
8. *Until* || μ (t)- μ (t-1) || <Є or t > tmax where Є is the maximum permissible error and tmax is the maximum number of iterations specified.

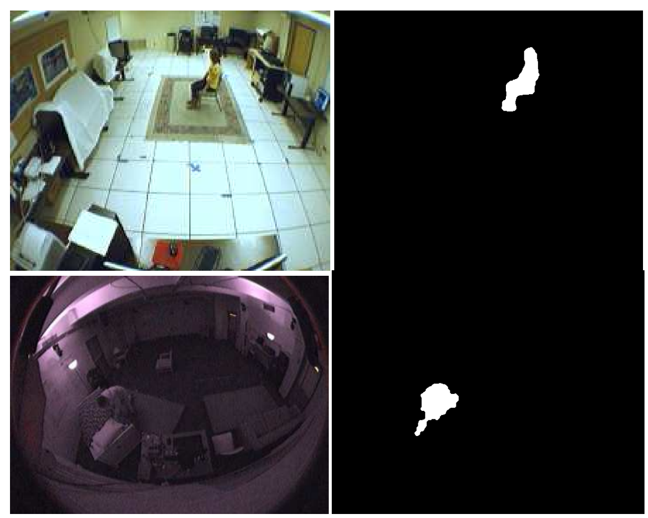
Here, μ(t) is the vector of all centers and the distance norm employed for determining convergence is the standard Euclidean distance measure. An important point to note is that it is essential to initialize the membership values to random values but with the mean equal to 0.5 and standard deviation equal to one so that the algorithm converges at a much faster rate. Another importance of standardization is the fact that it ensures that equal importance is given to each of the moments used or else the algorithm would weigh on the moments whose range is the highest.

IV. EXPERIMENTAL SETUP

We have established an image sequence database at resolution of 640*480 with two fixed fisheye Unibrain cameras which has a viewing angle of 180 degree. With IR lights on, cameras can see what our eyes cannot see. Figure 2(a) shows two sample images of these two camera views and their positions, while Figure 2(b) depicts the differences in silhouette extraction between a visible scene and the type we consider here (dark room illuminated by IR emitters). As seen in section V, we can still recognize activities even with the degradation in silhouette quality. Four students were asked to practice several activities under low light condition in the Computational Intelligence laboratory at the University of Missouri. Note that in the visible spectrum, these images would be completely black. Infrared lighting was used with the following specifications. The wavelength of the IR emitters used is 850nm; in total there are 216 individual infrared LEDs distributed between the two lamps and the total power draw is approximately 20 Watts. The camera lens is fisheye type with a 180 degree horizontal field of view, and 131 degree vertical field of view when used on 1/4" imager. There is no IR filter on the camera lens. We considered the possible activities practiced at night and included them into our data collection: walking, standing with hand motion, standing without hand motion, sitting down and standing up, sitting on a sofa, going to bed, getting up from bed. Our data collection also contained the following situations on the bed: sleeping (lying on bed); being sleepless (flip with some movements); sitting on bed; from sitting to laying on the bed; in addition, four abnormal activities (falling) were included: walking in the room and falling onto ground due to loss of balance; slipping off when trying to get up from a chair; falling when trying to get up from a bed; falling out of the bed when sleeping. The frame rate is 3 frames / second. We have collected more than half hours of frames for each person, i.e. over 3*1800=5400 frames for each person.



(a)



(b)

Fig.2 (a) Two sample images of our two camera views and their positions, red eclipses indicate the positions of the IR lamps, emitting IR lights which can't be seen by our eyes. (b) Difference in silhouette quality between a lighted room (top) and using IR emitters in a dark room (bottom).

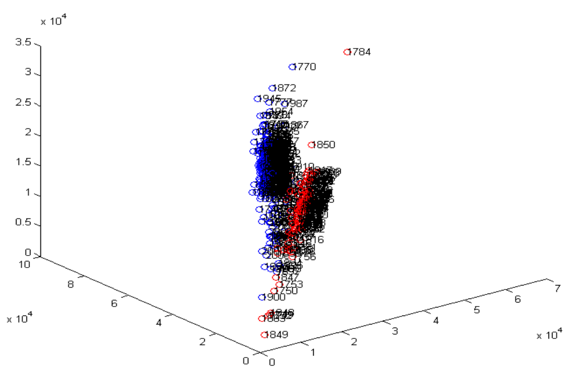
V. PRELIMINARY EXPERIMENTS

A set of preliminary experiments was conducted to establish the input parameters and best features to use for this domain. Several participants performed different activities. As described in Section II, silhouettes were extracted from the raw image sequences, and the moment features were computed.

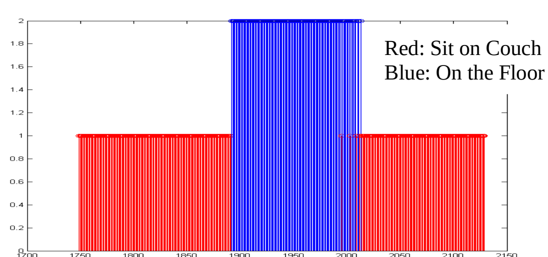
The GK clustering technique requires the number of clusters to be specified as an input parameter. In preliminary experiments shown in [5], we showed that clustering the Zernike moments using the GK algorithm with the number of clusters initialized to the number of activities yielded the best results. Since single camera images are used here, the activities of walking and standing cannot be differentiated in general; thus, they are grouped together as "upright" frames for the purpose of activity recognition.

Figure 3 (a) shows the clustering results of one data sequence using an input of 2 clusters. Figure 3 (b) shows the clustering results with the X axis indicating the frame number in the sequence and the Y axis indicating the cluster number after hardening the membership matrix. The results have been color coded for display purposes. In both the figures, the points colored red represent frames of a person sitting on the couch and the blue colored points represent the image frames indicating the person on the floor. Black colored points indicate the areas where the points are densely clustered.

A point to note here is that while it is evident that the two clusters obtained represent the "sit" and "on the floor" activities, without any prior information, we are unable to identify which cluster indicates which activity. The solution to this is explained in Section IV.B.



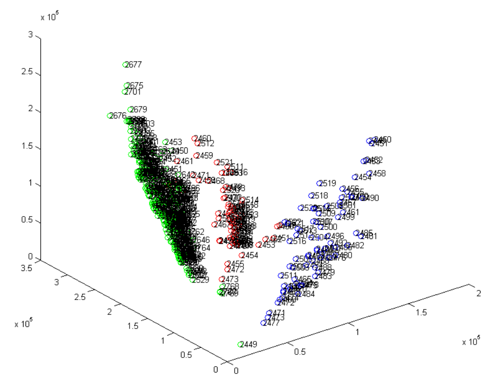
(a)



(b)

Fig. 3. Test results on an Infrared video sequence with two activities - sit on couch and being on the ground. GK on Zernike Moments and clustering results into (a) 2 clusters and (b) membership results of Zernike Moments by frame number.

Figure 4 (a) shows the clustering results of one data sequence using an input of 3 clusters. Figure 4 (b) shows the clustering results with the X axis indicating the frame number in the sequence and the Y axis indicating the cluster number after hardening the membership matrix. Similar to the previous example, these results have been color coded for display purposes. The scenario involved a participant performing several actions in an unlit room. These activities were a part of a planned scenario to enact the possible movements that could occur in the night such as moving around in the room, sleeping in the bed, tossing and turning in the bed and then falling onto the floor. As can be seen in Figure 4, the activities were well separated after clustering. 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.



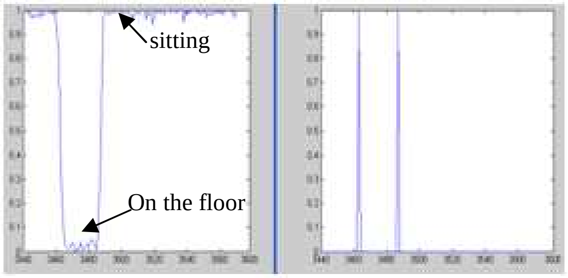
(a)

(b)

Fig. 4. Test results on an Infrared video sequence with three activities – upright, sleep on bed and on the ground. GK on Zernike Moments and clustering results into (a) 2 clusters and (b) membership results of Zernike Moments by frame number.

A. *Classifying the Transition Frames*

Figure 5 shows the membership values of one cluster by frame number for a test sequence. This particular sequence is a part of the video whose results are shown in Figure 3 and indicates a person sitting, then falling, and then resuming his sitting position on the couch. From the figure 5 (a), we can see that the membership (in the "sit" cluster) is initially high, and then it falls to almost zero, and then it again rises.



(a) (b)

Fig. 5. (a) Membership values of a test sequence by frame number for the sitting activity (b) Transition frames based on thresholding the membership values. The first and third transitions are couch-to-fall motions which could indicate a potential fall.

For the frames indicating "transition" motion (sit-to-on floor or on floor-to-sit), the membership is intermediate, approximately in the range of 0.1 to 0.9 in each of the clusters.

B. *Prototype Matching:*

After using the Gustafson Kessel or any other fuzzy clustering algorithm, we can obtain a pre-defined set of clusters. However, without any *a priori* information, there is no way to figure out which cluster indicates which 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* information. To accomplish this, we followed a semi supervised approach wherein the prototypes of the previous data runs are used to identify the activity cluster of the current data. As a means to achieve that, we used the nearest prototype matching for the sequences above. Since we are still in the preliminary phase, we just compared the distance between the Zernike moment image vectors of each new image frame to the labeled vectors of previous sequences. Then 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 identified.

Figure 6 shows frame examples with the original raw images, the corresponding silhouettes and the classification results for frames indicating upright, sleep on couch and on the floor activities. The silhouettes are color coded according to the classified activity. Classification is done by thresholding the membership functions. If the membership value is greater than 0.9, then the frame is assigned to this class. There are some errors in the transition frames identification but these are easily rectified using a simple averaging filter on the classification results.

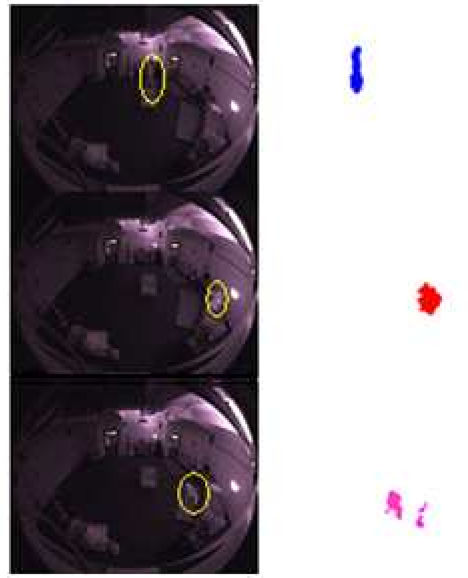


Fig. 6. Segmented activities of a 3 activity sequence: Frames 480 (upright), 586 (on the bed), and 900 (on the floor) with the color-coded silhouettes according to the identified activity. Membership values are thresholded from the GK clustering results using the Zernike Moments as features and then classified after prototype matching.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a successful and yet simple technique of detecting activity frames using fuzzy clustering methods. A classifier was constructed from the clustering results, and the classification results using the Zernike Moments were obtained. Our previous work [5] using the fuzzy clustering technique was compared against the Vicon Motion Capture system and shown to be pretty accurate in activity recognition.

Preliminary work using infrared lighting has been described in this paper. Experiments are currently being conducted to test the algorithm under different settings with different activities on several video sequences. This will make the algorithm more robust and useful for automatic activity recognition in unstructured settings which, in the future, will help prevent unmonitored physical injuries from taking place.

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] J. Keller, I. Sledge, "A Cluster By Any Other Name", *IEEE Proc., NAFIPS*, 2007.

[3] A. Oikonomopoulos, M. Pantic, and I. Patra,"Sparse B spline polynomial descriptors for human activity recognition," *Image and Vision Computing*, 27(12), 2009.

[4] C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 747–757, Aug. 2000.

[5] Banerjee T, Keller JM, Skubic M & Abbott CC, "Sit-To-Stand Detection Using Fuzzy Clustering Techniques," *IEEE International*

Conference on Fuzzy Systems (FUZZ-IEEE) at the World Conference on Computational Intelligence (WCCI), Spain, 2010.

[6] D. Anderson, et al., "Modeling Human Activity From Voxel Person Using Fuzzy Logic", *IEEE Trans. Fuzzy Systems*, 2007.

[7] M.K. Hu, "Visual pattern recognition by moment invariants." *IRE Trans. on Information Theory, IT-8:*pp. 179-187, 1962.

[8] A. B. Bhatia and E. Wolf. "On the circle polynomials of Zernike and related orthogonal sets." *Proc. Cambridge Philosophical Society*, 50:pp. 40-48, 1954.

[9] I. Gath and G. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 773–781, July 1989.

[10] X.L. Xie, G. Beni, "A Validity Measure for Fuzzy Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841-847, Aug. 1991.

[11] M. Goffredo, M. Schmid, S. Conforto, M. Carli, A. Neri, T. D'Alessio, "Markerless Human Motion Analysis in Gauss-Laguerre Transform Domain: An Application to Sit-To-Stand in Young and Elderly People." *Information Technology in Biomedicine, IEEE Transactions on*, 13:2: 207-216, 2009.

[12] D. Berrada , M. Romero , G. Abowd , M. Blount , J. Davis, "Automatic administration of the Get Up and Go Test," *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, June 11-11, 2007, San Juan, Puerto Rico.

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

[14] G. Demiris, D. Parker, J.Giger, M. Skubic, M. Rantz, "Older adults' privacy considerations for vision based recognition methods of eldercare applications," *Technology and Health Care* 2009; 17(1):41-48.

[15] K. Berg, S. Wood-Dauphinee, J.I. Williams , B.Maki," *Measuring balance in the elderly: Validation of an instrument"*. Can. J. Pub. Health 2:S7-11, 1992.

[16] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," *in Proc. IEEE CDC*, San Diego, CA, 1979, pp. 761–766.

[17] McMurdo MET, Gaskell A, "Dark adaptation and falls in the elderly." *Gerontology* 1991; 37:221–4.

[18] Girardi M, Konrad HR, Amin M, Hughes LF., "Predicting fall risks in an elderly population: computer dynamic posturography versus Electronystagmography test results*."* *Laryngoscope* 2001; 111:1528-32.

[19] Delaney, D. K., T. G. Grubb, & Garcelon, D. K., "An infrared video camera system for monitoring diurnal and nocturnal raptors." *Journal of Raptor Research 32: 290–296, 1998.*

[20] 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 Physics & Technology*, 2007. 49(3): p. 261-265.