A Study of Detecting Social Interaction with Sensors in a Nursing Home Environment

Datong Chen, Jie Yang and Howard Wactlar

School of Computer Science
Carnegie Mellon University
{datong, yang+, hdw}@cs.cmu.edu

Abstract. Social interaction plays an important role in our daily lives. It is one of the most important indicators of physical or mental diseases of aging patients. In this paper, we present a Wizard of Oz study on the feasibility of detecting social interaction with sensors in skilled nursing facilities. Our study explores statistical models that can be constructed to monitor and analyze social interactions among aging patients and nurses. We are also interested in identifying sensors that might be most useful in interaction detection; and determining how robustly the detection can be performed with noisy sensors. We simulate a wide range of plausible sensors using human labeling of audio and visual data. Based on these simulated sensors, we build statistical models for both individual sensors and combinations of multiple sensors using various machine learning methods. Comparison experiments are conducted to demonstrate the effectiveness and robustness of the sensors and statistical models for detecting interactions.

1 Introduction

The worldwide population over age 65 is expected to more than double from 357 million in 1990 to 761 million by 2025 [17]. At present, five percent of Americans over age 65 reside in nursing homes, with up to 50 percent of those over the age of 85 likely being placed in a nursing home at some point in their lives [13]. Among these nursing home residents, about 80% of them are believed to suffer from a psychiatric disorder, and 90% of patients with Alzheimer's disease experience behavioral complications leading to increased functional disability, medical morbidity, mortality and premature institutionalization [32]. In many nursing homes, physicians might visit their patients for only a short period of time once a week. Assessment of a patient's progress is based mainly on reports from staff (nurses and nurse assistants). The reports may be incomplete or even biased, due to schedule shift and the fact that each staff person has to take care of many patients. This may result in insufficient observation for monitoring either progressive change or brief and infrequent occurrences of aberrant activity for diagnosing some diseases. For example, dementia is very common among residents in nursing facilities. One obvious characteristic of dementia is a sustained decline in cognitive function and memory [24]. Studies indicate that the elderly with dementia may exhibit measurable agitated behaviors that increase confu-

sion, delusion, and other psychiatric disturbances [27][31]. In the early stages of dementia, these agitated behaviors occur occasionally and only last a very short period of time and are frequently missed by caregivers. Therefore, a long-term observation and care become increasingly important for the elderly with dementia in nursing homes [9]. Although no widely accepted measure exists for dementia care environments [5], quantitative measures of daily activities of these patients can be very useful for dementia assessments.

In this research, we are interested in automatically extracting information from sensors for geriatric care applications within skilled-care facilities. We would develop a system that can automatically extract and classify important antecedents of psychosocial and health outcomes. One such indicator is the frequency, duration and type of social interactions of the patients with one another and their caregivers. Interaction with others is generally considered a positive and necessary part of our daily life. Changes in interaction patterns can reflect mental and physical status of a person. Naturally, the level of social interaction of a person depends on a wide range of factors, such as his/her health condition, his/her personal preference, and aptitude for social interaction. More important, most social interactions are observable. This makes it possible for detecting them using an automatic system.

This paper explores the feasibility of building such a sensor-based analyzer to detect social interactions in a nursing home environment. Automatic detection of social interaction in a nursing home requires a set of physical and algorithmic sensors. For example, we can use an RF (Radio Frequency) sensor to track the location of each patient or a speech detector (an algorithm) from audio signals. However, the development and deployment of physical and algorithmic sensors are not trivial tasks. Furthermore, attaching physical sensors on bodies of patients is not practical. To this end, we employ a Wizard of Oz approach that enables the effectiveness study of various combinations of sensors and multiple models from a wide range of plausibly simulated sensors.

One important goal of this study is to obtain critical knowledge of detecting social interactions without physically developing and deploying ineffective or unnecessary sensors. This study also aims to find out the intrinsic structure among social interaction events and answer the following questions: how to construct necessary sensors to analyze social interactions, and how far from building these sensors we are using current technologies. Due to the fact that human beings infer interaction activities mainly from audio and visual cues, we are able to simulate potential useful sensors using the knowledge of human experts from audio and visual channels. Therefore, this study can be performed on the basis of long-term digital audio and video recording of a nursing home environment. We first evaluate the importance of each individual sensor and then employ a variety of machine learning techniques to create statistical models to identify interactions between people using simulated sensor data.

2. Related work

Social interaction consists of multiple individual human activities among multiple people. The work presented in this paper is closely related to location awareness and

human activity analysis, which have been addressed by many researchers in different areas such as multimedia processing, pervasive computing, and computer vision.

Various wearable sensors have been developed in recent years to address person tracking and activity analysis in the ubiquitous computing area. Global Position System [24], active bat location system [15], and PlusOn time modulated ultra wideband technology [34] provide location measures from meter to centimeter precision. Some wearable sensors have been applied to health monitoring [23], group interaction analysis [16], and memory augmentation [29].

Elderly individuals are usually unwilling to adapt to even tiny changes in environment, including wearable sensors in their clothes. Some non-contact sensors are considered to be more practical in our task. Power line network [4] and Ogawa's monitoring system use switches and motion detectors to track human activities indoors. The data provided by switches and motion sensors are reliable and very easy to process. However, they cannot provide detailed information. For example, a motion sensor can only tell that there is a person in the monitored area but cannot tell the exact location.

A vision-based system can non-obtrusively provide location information. Many computer vision algorithms have been developed for not only recovering 3D locations of a person, but also providing detailed appearance information of the person and his/her activities. Koile et al [21] at MIT proposed a computer vision system to monitor the indoor location of a person and his/her moving trajectory. The Living Laboratory [20] was designed by Kidd, et. al. for monitoring the actions and activities of the elderly. Aggarwal, et. al. [1] has reviewed different methods for human motion tracking and recognition. Various schemes such as single or multiple camera schemes, and 2D and 3D approaches, have been broadly discussed in this review.

A large number of algorithmic sensors have been proposed to detect activities from audio and visual signals, including gait recognition [3], hand gesture analysis [11], facial expression understanding [10], sitting, standing and walking analysis [23], and speech detection [26]. Hudson, et. al examined the feasibility of using sensors and statistical models to estimate human interruptibility in an office environment [18]. These sensors are still mostly research challenges today, but can be potentially applicable in the future. Combinations of these sensors for analyzing human behaviors have been applied in some constrained environment, such as meeting rooms [36] and sports fields [19].

Figure 1 Examples of interaction patterns in a nursing home.

3. Data collection and preprocessing

Four cameras and four audio collectors were carefully placed in two rooms and a hallway of a nursing facility. Recording was performed from 9am to 5pm for 10 days.

Overall, 320 hours were recorded at the nursing facility. Each video and its corresponding audio channels were digitalized and encoded into an MPEG-2 stream in real time and recorded onto hard disks through a PC. The video data was captured and finally recorded in 24-bit color with a resolution of 640x480 pixels at 30 frames per second. The audio data was recorded at 16-bit 44.1KHz. Figure 1 illustrates some examples of interaction patterns from the data. In this paper, only the hallway videos are manually ground-truthed and used for analysis.

Since we only focus on multi-person activities, we developed a preprocessing algorithm to segment audio/video streams into shots, and classify the shots into three classes: non-activity, individual activity, and multi-person activity using audio and video event detection techniques.

3.1 Video event detection

For the video channel, we use a background subtraction algorithm to detect frames that contain human activities. To speed up this detection process, only video from one camera in the network is used. The background of a frame is obtained by the adaptive background method [33]. We employ a threshold to extract pixels that have high differences between the current frame and its background. To remove noise, we group extracted pixels into regions and only keep those regions that contain more than 15 pixels. We consider the frame f to contain a visual interaction event $V_f=1$ if any of the following rules is satisfied; otherwise $V_f=0$:

1. There are two or more regions in the frame.
2. There is region that does not touch the bottom the frame, whose width to height ratio is more than 0.7.

We choose these thresholds to detect as many interactions as possible without inducing excess false alarms. The output of the detection is reported every second. For a second of NTSC video, we output the percentage of visual cues in its 30 frames as:

$$C_v = \frac{1}{30}\sum_{f=1}^{30} v_f$$

3.2 Audio event detection

To detect events using an audio stream, we use a very simple power-based method similar to the one proposed by Clarkson and Pentland in [6][7]. This method adaptively normalizes signal power to zero mean and unity variance using a finite-length window; segments where the normalized power exceeds some threshold are designated "events." [6] and [7] describe an ambulatory system which could be exposed to arbitrary acoustic environments; adaptive normalization allows such a system to compensate for unusually loud or quiet environments and still detect events reliably. Our task differs from that system in that we have a *stationary* system where changes in power level really do indicate events and not just changes of venue. As such, instead of adaptive normalization, we use global normalization. That is, a single mean and variance is calculated for each two-hour recording and the globally-normalized power is threshold to detect events a_f.

In this implementation, we extracted 16-bit mono audio from the audio-video stream, and used analysis windows 200*ms* in length with a 50% overlap. This window length results in a frame rate of 10 frames per second, which is more than adequate to detect events using the power-based approach. After signal power is calculated and normalized, it is passed through a simple 3-frame averaging filter for smoothing. We then apply the power threshold; any segment which exceeds the threshold is designated an event. We also stipulate a minimum event time of 1 second in order to filter out isolated auditory transients. The confidence of audio event per second is defined as:

$$C_a = \frac{1}{10}\sum_{f=1}^{10} a_f$$

3.3 Fusing video and audio event detection

Video and audio streams are synchronized and segmented into one-second non-overlapping patches. The final event detection of each patch combines the video event confidence and audio event confidence linearly:

$$C_d = \alpha C_v + (1-\alpha)C_a$$

We consider a one-second patch to contain an interaction if its confidence C_d is higher than 0.5.

To evaluate the preprocessing algorithm, we labeled 10 hours of video/audio data. Using only video detection, we extract 33.3% of the entire video as candidate interaction shots, which is listed in Table 1. In order to not miss any interactions, we only filter out the one-second-long video segments with zero confidence.

Table 1 Results of event detection from video.

	Total Event Time	Event Time as % of Total Signal
No activity	13711	38.1%
Individual	6700	18.6%
Multi-person	15589	33.3%

Table 2 Results of event detection from audio.

Threshold	Total Event Time	Event Time as % of Total Signal
1.1	6705	18.6%
1.6	5582	15.5%
2.1	4327	12.0%

Using only audio detection with varying thresholds, we obtain the results listed in Table 2. The table shows the total event time and percentage of data in the recordings using three thresholds.

Table 3 Preprocessing results based on the ground-truth.

	Recall	Precision	Process speed
Video	98%	13%	real time
Audio	71%	28%	10% real time
Multimodal	92%	21%	

By fusing the audio (threshold 1.6) and video results, we extracted total 9435 seconds from the entire 10 hours data. In this way, 85 our of 91 interactions in the ground truth are covered by the candidate shots, which obtain reasonable recall and precision

in terms of event time as listed in Table 3. The audio has a lower recall due to the presence of silent interactions such as walking assistance of a wheelchair-bound patient. The audio precision is actually higher in general than is reported here. The hallway environment is a poor representative of audio precision, as many events that are audible in the hallway are off-camera and not in the ground-truth labels; thus audio event detection generates many false alarms. Even so, our results show that by fusing audio and video results, we can achieve more than 90% recall and 20% precision. We project even better precision when we test our fused system over the entire set of the data. The multi-person activity shots are then manually labeled using events selected by a group of doctors. Our study focuses on detecting interactions in multi-person activities, since social interaction is mutual or reciprocal action that involves only two people.

4. Sensor simulation

A sensor is usually defined as a device that receives a signal or stimulus and responds to it in a distinctive manner. As we mentioned in the introduction, we consider both physical and algorithmic sensors in this study. For example, in order to investigate the temporal referencing probability of detecting an interaction, we consider "temporal interaction reference" as an algorithmic sensor, which is a detection result of another 1-second interval related to the current interval. On the other hand, to reduce number of candidate sensors, we omit some sensors that are impossible to implement from current technologies, such as speech recognition and facial expression understanding. There are some compromises between the technology capability and the medical request. We keep some sensors, for instance "hand trembling", which are very important for human experts but are questionable for real implementation. In detail, we select 21 events from the Pittsburgh Agitation Scale, which are listed in Table 4 and their occurrences in temporal neighborhoods as simulated sensors:

Table 4 Sensors defined on events and temporal neighborhood.

Approaching	Leaving		- 5s	
Standing	Hand trembling		- 4s	
Talking	Pushing a wheelchair		- 3s	
Shaking hands	Passing		- 2s	
Hand touch body slowly	Sitting		- 1s	
Hand touch body normally	Walking	×	0s	= Sensors
Hand touch the body quickly	Hand in hand		+ 1s	
Hugging	Kiss		+ 2s	
Face turning	Kick		+ 3s	
Walking (moving) together	Sitting down		+ 4s	
Temporal interaction reference			+ 5s	

We label each shot second by second. The range of the temporal neighborhood is chosen from 5 seconds ahead to 5 seconds behind the current one. Overall we obtained 230 (21×11-1) simulated sensors, including 21 events times and 11 temporal neighbors, except the "temporal interaction reference (T-reference)" in the current

interval, which is not considered as a sensor. All the sensors are labeled as binary events since there is no ambivalent in human experts' judgments during the labeling. We can see that one-second recording content may contain more than one direct or derived event detected by the simulated sensors.

To know which sensors would be most useful, we first analyze the effectiveness of individual sensors in detecting social interactions. The first measure that we use to study individual sensors is information gain [30]. Information gain indicates the potential power of each sensor in predicting an interaction. The details of how this technique works will not be covered in this paper. Table 5 lists top 28 sensors selected by information gain with respect to a correct prediction of a social interaction.

Table 5 Top 28 sensors selected by information gain technique.

1	T-reference-1	8	Walking 0	15	Talking-2	22	Approaching+1
2	T-reference+1	9	T-reference-5	16	Walking+2	23	Walk together 0
3	T-reference-2	10	T-reference+4	17	Talking-3	24	Walking+3
4	T-reference+2	11	Walking+1	18	Talking+2	25	Talking-5
5	T-reference-3	12	Walking-1	19	Approaching 0	26	Approaching-1
6	T-reference+3	13	T-reference+5	20	Walking-2	27	Talking+3
7	T-reference-4	14	Talking+1	21	Talking-4	28	Leaving 0

The table shows that the T-reference of an interaction has obvious temporal consistency. Most interactions take longer than one second, and this consistency information is so important that it occupies the top 7 ranks with respect to the information gain scores.

Besides the temporal consistency, it also shows that the sensors of walking and talking are very important cues associated with an individual person; and relative location, such as approaching, leaving, walking together, and hand gesture are important between two persons. These sensors are important even in our daily experience. However, some sensors, such as "hand normal" and "pushing", which are also obvious evidence of an interaction, have very low ranks in information gain. They are either co-occurrences with some high rank sensors or omitted by the information gain technique due to their small number of examples.

Information gain takes an empirical risk to rank the sensors, which can be biased when training samples are redundant in some interaction patterns. For example, a long sequence of standing conversation will lead to higher ranks for talking and standing than that of a short sequence. It tends to omit the sensors with small numbers of examples in the training set, even though these sensors are very powerful in predicting social interactions. To avoid this kind of bias, we also analyze the power of each sensor using a structural risk based support vector machine (SVM) method [2]. This method trains an SVM using a subset of the training set from all sensors, and then eliminates sensors with low weight in representing the decision hyper-plane. Because the decision hyper-plane is trained to maximize the margin between the closest positive support vectors and negative support vectors, repeated patterns in the training set don't affect the result. Therefore, it is robust to the training set which contains a biased number of training examples for different sensors.

Table 6 lists the top 28 sensors selected by the SVM method. These 28 sensors cover most events in our total 21 events. Only "sitting" and "passing" are not included. This selection is more reasonable since the high rank sensors, such as "walk-

Table 6 Top 28 sensors selected by SVM.

1	T-reference+1	8	Pushing+4	15	Pushing-3	22	Face turning 0
2	T-reference-1	9	Hand in hand 0	16	Walking+2	23	Walk together 0
3	Walk together 0	10	Kick 0	17	Face turning+1	24	Shaking hand+5
4	Hand normal 0	11	Hand slow 0	18	Approaching 0	25	Pushing+3
5	Talking 0	12	Hand-trem 0	19	Pushing-4	26	Hug+2
6	Pushing 0	13	T-reference-2	20	Hand normal+3	27	Standing+2
7	Talking+1	14	Leaving 0	21	Walk together+4	28	T-reference+2

together", "hand touch body normally", "talking", and "pushing", are obvious evidence of an interaction. The sensors with the top 2 ranks are still "T-reference" in the closest neighborhoods. This indicates that the 1-second interval is small and precise enough for analyzing social interactions in a nursing home environment. In comparison with the information gain results, the sensor "talking" is a common important sensor selected by both methods. The "walking" sensor is replaced by "walk together" and "pushing". They all overlap the sensor "walking", but provide more specific information. Hand related sensors are also ranked higher, which indicates that social interaction may benefit from developing better hand analysis sensors.

Temporal information is included in our simulated sensors. We evaluated the effectiveness of temporal orders by averaging the two selection results together and computing the histogram of the temporal orders. Figure 2 illustrates the effectiveness of temporal order in detecting social interactions.

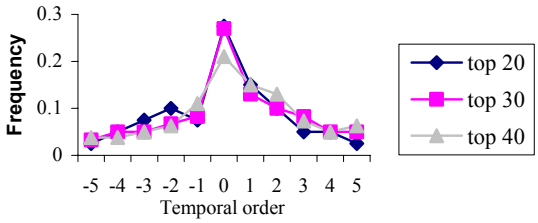

Figure 2 Effectiveness of temporal order.

The effectiveness of the temporal order drops quickly as the time span from the current time increases from zero. The effect of events more than 3 seconds away from the current one is very limited and can provide little useful information for analyzing social interactions. Sensor selection only analyzes the effectiveness of individual sensors; in the next section we will investigate the power of combinations of sensors using statistical models.

5. Detection models

It should be noted that there are some overlaps among simulated sensors, e.g., "walking together" implies "walking". The first goal of this section is to explore proper statistical models to detect social interactions. We consider that the detection of a social interaction is a binary classification problem: interaction and non-interaction. The other goal of this section is to further investigate the associations be-

tween different sensors. This will enable us to replace some impracticable sensors with combinations of sensors that can be more easily developed. Since we have considered including temporal information in the simulated sensors, the interaction detection problem can be simplified as a problem to classify the sensor outputs of each 1-second interval into two classes, indicating interaction and non-interaction respectively.

To find a proper model for classifying interactions, we evaluated various machine learning algorithms: decision tree [28], naive Bayesian [22], Bayes network [17], logistic regression [14], support vector machine [35], adaboost [25], and logitboost [12]. We will not describe details of these algorithms in this paper; interested readers can find these details in the references.

The evaluations are shown in Table 7. We use equal size training and testing data. Standard 5-fold cross-validation is performed to find optimal parameters for each model. We then perform the resulted optimal models on the testing set to report the numbers in Table 7.

Table 7 Performance of interaction detection using different models.

	With T-reference			Without T-reference		
Model	**Prec.**	**Recall**	**F-measure**	**Prec.**	**Recall**	**F-measure**
Decision tree	99.5%	99.2%	99.3%	97.1%	**96.4%**	**96.8%**
Naive Bayesian	98.4%	92.9%	95.6%	96.3%	90.1%	93.1%
Bayes network	98.4%	93.0%	95.6%	96.3%	90.4%	93.3%
Logistic reg.	99.6%	98.7%	99.2%	96.5%	94.5%	95.5%
SVM	99.5%	**99.5%**	**99.5%**	**98.0%**	95.1%	96.5
adaboost	**99.7%**	99.1%	99.4%	95.4%	93.9%	94.6%
logitboost	**99.7%**	99.1%	99.4%	96.0%	95.6%	95.8%

We can see that under the ideal conditions (all sensors output correct result without any ambiguity), all these models obtain good detection results. To our surprise, the simplest method, decision tree, employs only four kinds of sensors: "T-reference", "talking", "walking", and "leaving", but achieves very good performance. None of these sensors except "T-reference" requires complex visual and audio analysis in comparison to sensors such as "face turning" and "hand in hand". It seems possible that social interaction can be detected by just developing good "talking", "walking", and "leaving" sensors. It is true if the "T-reference" sensor can be successfully derived from these three kinds of sensors.

To remove the effect of the temporal information of the derived sensor "T-reference", we assume that the "T-reference" sensor is not available to its neighbors. We remove all "T-reference" sensor outputs from feature vectors and evaluate the above methods. The results are also listed in Table 7. After removing the "T-reference" sensor, the performance drops about 3-5%, which indicates that we can achieve around 90% accuracy in detecting current interaction with the temporal information of interaction decisions in neighborhoods. As we assume outputs of other sensors are under ideal conditions, the real accuracy of the current "T-reference" sensor output is expected to be about 90% of the average accuracy of all the other sensors' outputs. The decision tree still achieved the best performance even without the "T-reference" sensors. However, the resulting decision tree includes all kinds of sensors. The top 10 sensors are:

Rank	Sensor	Rank	Sensor
1	Talking	6	Hand in hand
2	Walk together	7	Standing
3	Walking	8	Leaving
4	Pushing	9	Approaching
5	Hand normal	10	Passing

A drawback of the decision tree is that it is sensitive to the noise in sensor outputs. In practice, outputs of sensors might be ambiguous or even incorrect. Some of the sensor outputs have to be represented by probabilities, e.g., 60% "talking" or 30% "hand in hand". The uncertainties of sensor outputs can only be determined from real data of experiments. What we can do in a simulation is to add some noise into outputs of sensors. Table 8 lists results of adding 20% noise (20% sensors have wrong outputs) into the data without "T-reference" sensors.

Table 8 Performances of interaction detection using different models with 20% noise.

Model	Prec.	Recall	F-measure
Decision tree	90.0%	90.4%	90.2%
Naive Bayesian	88.6%	75.3%	81.4%
Bayes network	88.1%	77.6%	82.5%
Logistic regression	90.1%	93.5%	91.8%
SVM	91.4%	95.3%	93.3%
adaboost	89.6%	93.8%	91.6%
logitboost	90.1%	95.6%	92.8%

The performance of the decision tree decreases from 96.8% (F-measure) to 90.2%, or loses 6.6% accuracy. At the same time, the performance of the SVM model decreases from 96.5% to 93.3%, or only loses 3.2% accuracy. Notably, the recall of the SVM only decreases 0.5% with 20% noise. The logitboost model also proved to be robust against noise; the recall remains the same after adding noise. The F-measure loses only 3% accuracy. This indicates that the SVM model is potentially more robust than the decision tree model in real applications.

It should be noted that the noise level of 20% is an empirical assumption. Real sensors will have different accuracies. According to our preliminary implementations of some real sensors, walking related sensors introduce around 15% noise on average, the "standing" sensor only has 6% noise, the "talking" sensor produces about 30% noise, and face and hand related sensors are still under development. If the noise range of face and hand related sensors are 40%-60%, 20% noise on average for all sensors is reasonable.

6. Conclusions

This paper presents a study of feasibility of sensor-based analysis of social interaction patterns in a skilled nursing facility. We have analyzed the capabilities of various individual sensors for detecting social interactions. The relative location related sen-

sors, hand related sensors, talking sensors, and temporal consistency information are ranked high priorities in the task of detecting interactions.

We have also compared various statistical models to explore overlapped spaces of multiple sensors. The experimental results have indicated that the decision tree model could achieve more than 99% accuracy with only three kinds of sensors: “talking”, “walking”, and “leaving”, plus temporal information under noise free conditions. This indicates the possibility of achieving good interaction detection performance by developing perfect “talking”, “walking”, and “leaving” sensors, instead of developing complex ones, such as face and hand gesture sensors. We also demonstrated the robustness of various models when noisy sensors are considered. The SVM model and the logitboost model have been proven to be more robust against noise than other models. Based on the promising results from this study, we will develop working systems with real sensors. We will further classify social interaction patterns and evaluate those systems and algorithms.

Reference

[1] Aggarwal, J. K., Cai, Q. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, Vol. 73, pp. 428-440, 1999.

[2] Brank, J., Grobelnik, M., Milic-Frayling, N. and Mladenic, D. Feature selection using linear support vector machines. MSR-TR-2002-63, Microsoft research 2002.

[3] Bregler, C. Learning and Recognizing Human Dynamics in Video Sequences. In CVPR, pages 568-574, 1997.

[4] Brumitt, B., Krumm, J., Meyers, B. and Shafer, S. Ubiquitous computing and the role of geometry. In Special Issue on Smart Spaces and Environments, volume 7-5, pages 41-43. IEEE Personal Communications, October 2000.

[5] Carp, F. Assessing the environment. Annul review of gerontology and geriatrics, 14, pages: 302-314, 1994.

[6] Clarkson, B. and Pentland, A. Framing Through Peripheral Perception. Proc. of ICIP, Vancouver, September 2000.

[7] Clarkson, B. and Pentland, A. Unsupervised Clustering of Ambulatory Audio and Video. Proc. of the ICASSP, Phoenix, 1998.

[8] Emler N., Gossip, reputation, and social adaptation. In R.F.Goodman and A. Ben-Ze'ev (Eds.) Good Gossip, pages.117-138. Wichita, Kansas, USA: University Press of Kansas 1994

[9] Eppig, F. J. and Poisal, J. A. Mental health of medicare beneficiaries: 1995. Health Care Financing Review, 15, pages: 207-210, 1995.

[10] Essa, I. and Pentland, A. Facial expression recognition using a dynamic model and motion energy. In Proc. 5th Intl. Conf. on Computer Vision, pages 360--367, 1995.

[11] Freeman, W. T. and Roth, M. Orientation histograms for hand gesture recognition. In International Workshop on Automatic Face and Gesture Recognition, pages 296-301, June 1995.

[12] Friedman, J., Hastie, T. and Tibshirani, R. Additive logistic regression: a statistical view of boosting. Annals of Statistics, 28:307--337, 2000.

[13] German, P.S., Rovner, B.W., Burton, L.C., Brant, L.J. and Clark, R. The role of mental morbidity in the nursing home experience. Gerontologist, 32(2): 152-158, 1992.

[14] Hastie, T. and Tibshirani, R. Nonpararmetric logistic and proportional odds regression. Applied statistics 36:260-276, 1987.

[15] Harter, A., Hopper, A., Steggles, P., Ward, A. and Webster, P. The anatomy of a context-aware application. In Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pages 59-68, Seattle, WA, August 1999.
[16] Holmquist, L., Falk, J. and Wigstrm, J. Supporting group collaboration with interpersonal awareness devices. *Personal Technologies*, 3:13–21, 1999.
[17] Hooyman, N.R. and Kiyak, H.A. Social Gerontology: A Multidisciplinary Perspective. 6th ed., Allyn and Bacon 2002.
[18] Hudson, S., Fogarty, J., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. and Yang, J. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pages 257-264 2003.
[19] Jug, M., Pers, J., Dezman, B. and Kovacic, S. Trajectory based assessment of coordinated human activity. In *ICVS 2003*, pages 534–543, 2003.
[20] Kidd, C. D., Orr, R., Abowd, G. D., Atkeson, C. G., Essa, I. A., Macintyre, B., Mynatt, E. and Starner, T. E. and Newstetter, W. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. *Proc. of CoBuild '99*, pp.191-198, 1999.
[21] Koile, K., Tollmar, K., Demirdjian, D., Shrobe, H. E., Darrell, T. Activity Zones for Context-Aware Computing. Ubicomp 2003, pp. 90-106, 2003.
[22] Kononenko I., Semi-naive bayesian classifier. In Proceedings of sixth European Working Session on Learning, pages 206-219. Springer-Verlag, 1991.
[23] Lee, S. and Mase, K. Activity and location recognition using wearable sensors. In *1st IEEE International Conference on Pervasive Computing and Communications*, pages 24–32, 2002.
[24] Lubinski, R. Dementia and communication. Philadelphia: B. C. Decker, 1991.
[25] Margineantu, D. D. and Dietterich, T. G. Pruning adaptive boosting. In 14th Int. Conf. on Machine Learning, pages 211-218. Morgan Kaufmann, 1997.
[26] Martin, A., Karrray, L. and Gilloire, A. High Order Statistics for Robust Speech/Non-Speech Detection. In Eusipco, Tampere, Finland, Sept. 2000, pp. 469--472.
[27] Nelson, J. The influence of environmental factors in incidents of disruptive behavior. Journal of Gerontological Nursing 21(5):19-24, 1995.
[28] Quinlan, J. R. C4.5: programs for machine learning. Morgan Kaufmann 1993.
[29] Rhodes, B. The wearable remembrance agent: A system for augmented memory. In *Proceedings of the 1st International Symposium on Wearable Computers*, pp: 123–128, 1997.
[30] Schraudolph, N. and Sejnowski, T. J. Unsupervised discrimination of clustered data via optimization of binary information gain. In Stephen Jose Hanson, Jack D. Cowan, and C. Lee Giles, editors, Advances in Neural Information Processing Systems, volume 5, pages 499-506. Morgan Kaufmann, San Mateo, 1993.
[31] Sloane, P. D., Mitchell, C. M., Long, K. and Lynn, M. TESS 2+ Instrument B: Unit observation checklist – physical environment: A report on the psychometric properties of individual items, and initial recommendations on scaling. University of North Carolina 1995.
[32] Steele, C., Rovner, B. W., Chase, G. A. and Folstein, M. Psychiatric symptoms and nursing home placement in Alzheimer's disease. *American Journal of Psychiatry*, 147(8): pp.1049-1051, 1990.
[33] Stauffer, C. and Grimson, W. E. L. Adaptive background mixture models for real-time tracking. Proc. of CVPR 1999.
[34] Time Domain Corporation, 7057 Old Madison Pike, Huntsville, AL 35806. PulsON Technology: Time Modulated Ultra Wideband Overview, 2001.
[35] Vapnik, V.N. The nature of statistical learning theory. Springer Verlag, New York, 1995.
[36] Zhang, D., Li, S. Z., Gatica-Perez, D. Real-Time Face Detection Using Boosting Learning in Hierarchical Feature Spaces. 17th International Conference on Pattern Recognition 2004.