This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30044624

Reproduced with the kind permissions of the copyright owner.

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Copyright : 2005, IEEE
Abstract

In this paper we explore an emotive, multi-modal smart house. The smart house is an instance of a monitoring application, inspired by the need to provide semi-autonomous assisted living for elderly and infirm people. A particular aspect of smart environments relevant to the care of the elderly is the detection of potential hazards. A hazardous situation represents an abnormal activity or event. Consequently, to detect abnormality we model normality, that is, the normal activities associated with a user’s interaction with the environment. We use the concept of anxiety as a measure of normality modelled with a probabilistic approach. The anxiety is associated with a hazardous device using a fusion of multi-modal data. The data is gathered from simple sensors, and from information derived from the audio domain indicating the presence of an activity within the environment. We present the results for the anxiety for a number of activity sequences, both normal and abnormal. The pervasive nature of the audio data enabled the detection of activity when interactions between a user and device didn’t occur, successfully preventing false hazardous situations from being detected.

1. INTRODUCTION

This paper is concerned with the development of smart environments for the assisted living of elderly people. A particular aspect of smart environments relevant to the care of the elderly is the determination of hazards within the environment in real time. Typical hazards are appliances being left unattended, such as unattended electrical devices, or abnormal environmental states such as a front door being left open and unattended. It is important to note that estimation of a hazard must be made in the context of the user’s normal behaviour, and it is this requirement that makes the problem difficult. Potential hazards can be determined in many ways across a broad spectrum of modalities. At one end of the spectrum, we can have a fully monitored home with multiple cameras augmented by other sensors in the environment. At the other end of the spectrum, the home can be augmented with simple sensors such as pressure mats, simple microphones (for sound level sensing) and reed switches. All these sensors have modalities that have advantages and disadvantages. For example pressure pads give simple discrete on/off events usually with 100% reliability. Video processing is highly flexible and consequently can be used to detect many different aspects of activity. However the current state of the art in video processing is not reliable enough and the processing required is complex. Although microphones can be regarded as simple sensors when used as a sound level device e.g. for detecting loud noises, it can be used to detect more complex audio events in various ways. They can also be used to localise sound events [2] but this requires complex processing. Ultimately, the challenge is the integration of most if not all of these different devices seamlessly into a framework for activity and hazard monitoring. When using audio and video sensors, one must be aware of the privacy issues raised and hence it is important to use these devices to only detect events and not record the video and sound for later playback. In this paper we explore the integration of sound, and simple sensors such as pressure pads, reed switches and X10 devices into a multi-modal framework for hazard detection in assisted living.

Most systems for activity monitoring are interested in event detection after the fact, for example every couple of hours [7]. Although this is valid in many cases, real time response in needed for many hazards at home e.g. leaving the bath to run over. Other approaches to such problems is to use temporal models of activity recognition with hidden Markov models [3] and build representations for normality from which abnormality can be inferred. However, the variability in behaviour patterns both within and across individuals makes it difficult to detect consistent patterns. Additionally the computational complexity makes it difficult to use in real time, particularly when multiple as well as multi-modal sensors are used as state space grows exponentially with the number of devices and states. An alternative novel approach was recently proposed [14] where each device, with the potential to be hazardous, is represented as an agent. A measure of anxiety was associated with each agent, representing the potential hazard represented by the device. The proposed agent based approach is device centric and the state space is automatically factored, thus making the approach scalable and applicable in small as well as large pervasive environments. The previous work on the agent based approach has concentrated on using simple sensors such as pressure pads and reed switches.

This paper incorporates a very important modality, namely sound and in particular the detection of what are defined to be sound events indicative of normal activities by the occupant. Our approach is to model background noise, from which...
foreground sounds, i.e. sound events indicating activity, can be detected. Background noise is defined as consisting of typical regularly occurring sounds such as the fridge being on and traffic. Events based on sound are different to those generated by pressure pads, reed switches and accelerometers, is that there is no spatial information available. In fact sound is picked up wherever the activity is occurring such as away from other sensors and devices and hence can detect activity when people are not interacting with devices but still moving around. The characteristics of sound implies that it provides a pervasive and contextual source of data from which we get many foreground events, unlike other discrete devices, and this paper explores the integration of these foreground sound events into the anxiety framework.

The significance of this paper is the use of audio as a pervasive method for examining the activities of an user over a wide area. This is in contrast to simple sensors whose use is limited by practical issues such as their placement.

The novelty is threefold: first the use of robust detection of foreground audio events for activity detection in assisted living; second, the integration of event detection by audio into the multi-modal anxiety framework seamlessly demonstrating the extended functionality of the model; and third, we extend the probabilistic framework of the anxiety model used for multi-modal sensor information in a smart home context.

The layout of this paper is as follows. Section 2 discusses related work in the areas of audio surveillance and smart home environments, and outlines the relevant background information for the “anxious home” and the audio background modelling. Sections 3 details audio activity detection and our multi-modal approach to the determination of anxiety. Section 4 describes the experimental process and results of the implementation of the multi-modal anxiety approach.

2. BACKGROUND

A. The Anxious Home

The previous approach to hazard detection [14] within the smart house environment modelled the hazards within the environment by learning the patterns of interaction between a user and numerous simple sensors. This method can cope with many activities but relies on an occupant moving around and regularly interacting with devices monitored by sensors during normal activities. This was extended to include simple loud sound detection as well as wearable devices such as accelerometers and Personal Digital Assistants (PDAs) [1]. Loud noises were taken to indicate potential emergencies but no attempt was made to identify the noise. The accelerometers were used on the body to detect the occupant’s stance and whether they had fallen. The PDA was used to obtain a response from the user for one of these events and an emergency invoked if no response was received from the occupant.

The approach follows an agent based methodology. Each device with the potential to be a hazard is represented as an agent, with many agents representing the various potentially hazardous devices within the environment. Each agent continuously computes a function that represents how “anxious” it is when it is active (switched on) and hence has the potential to get into a hazardous state. This anxiety measure essentially represents how worried the device is given that it is being ignored by the occupant. The more the occupant interacts with the device or is found to be nearby, the less anxious the device is. An agent based approach allows us to treat each device independently, as well as allowing us to estimate the overall state of the house or any room from the anxieties of each of the devices. In other words, the more devices that are active, each with its own anxiety value, the more anxious the house or room would become. An extreme case would be cooking the dinner and running the bath at the same time. Anxiety would be kept low if the occupant kept moving between the stove and the bath to check the progress of the cooking and bath filling, but become high if the occupant ignored one or both devices.

Anxiety is an emotion in the human sense. Recently, much work has been carried out into emotional computing [11], mainly to enable computers to communicate with humans. It is argued that decision making by humans requires emotions as well as rational thought for fast decision making. Hence modelling human emotions in computer decision making may improve the performance of computers for such activities.

Devices are considered as either hazardous or passive. A hazardous device refers to devices that have to be attended to while they are in a hazardous state. Examples of hazardous devices include the stove, fridge, and taps. The second class, passive devices, consist of devices for which there is no hazardous state, such as pressure pads, and reed switches on cupboards. For example, no hazardous state is associated with a reed switch on a cupboard as no hazard is introduced by leaving the cupboard open. A measure of anxiety for each hazardous device is formulated using the patterns of activity associated with the device, with the grouping of the anxiety measures for hazardous devices forming the anxious home. Anxiety is formulated with statistical models consisting of a model representing typical interactions of a user with the device, and a model representing interaction with other devices, both hazardous and passive, while the device is active. The anxiety for the hazardous device is then determined using a probabilistic framework generated using the expected time periods between an interaction with the hazardous device, and the other devices within the environment. As this approach does not model sequences directly, complex activity patterns along with variations in activity sequences are accounted for. The anxiety is then used as an indicator to determine hazardous situations and provide feedback to a user.

B. Smart House Activity Monitoring

There has been considerable research using simple sensor data to monitor activity. “Stove Guard” has current and motion sensors that can turn off the stove after a certain time if it is on (current flowing), and if it is unattended (no motion detected). Combinations of simple sensors have been used

1 www.absoluteautomation.com/stoveguard — accessed Feb’05.
for recognising activity in houses [12], [4], [9]. Glascock and Kutzik [7] use a small number of infra-red sensors for coarse activity monitoring that is mainly suited for making sure someone has taken medication, eaten etc. which only requires events to be reported at two hour intervals. Recently, a system [8] has been proposed that uses a form of anxiety that rises if there is little activity in the house, determined by the use of a number of standard sensors, such as window and door sensors. If the lack of activity is unusual, based on learned data, an alarm can be raised. The work presented in this paper differs from the above in that we are interested in the interactions of various devices enabling richer semantics to be inferred and monitored in real time enabling prompt responses to abnormal behaviour.

C. Audio Surveillance and Monitoring

Audio analysis methods for surveillance and monitoring have predominantly centred on the detection of specific audio cues, or sound events. Such methods include a tele-monitoring system for the detection of sound events such as cries for help [13], and the detection of alarm sounds [6]. Cowling [5] proposed a method to develop a taxonomy for the classification of environmental sounds for the purpose of audio surveillance. While these methods focus on the detection of specific sound events, our approach extends audio surveillance by deriving contextual information from the analysis of the audio signal.

D. Audio Background Modelling

In contrast with [1], the foreground sound events determined in this paper are determined by a difference in the characteristics of the audio, which is a more robust approach in comparison with sound level sensing. An online, adaptive Guassian Mixture Model (GMM) is used to model background audio, as detailed in [10]. This method was augmented by combining fragmented background models using entropy calculated between the GMM distributions, resulting in a more robust determination of the background.

3. Multi-Modal Anxiety Determination

In experimentation, the anxiety is determined for each hazardous device independently. In the context of modelling the interactions associated with the hazardous device for which the anxiety is being determined, the remaining hazardous devices are considered to be passive.

To incorporate audio into the statistical model for anxiety we treat audio activity as a passive device, indicating a non-hazardous interaction between the user and environment. Therefore the presence of audio activity, i.e. the presence of a segment of foreground audio, is processed in a similar manner to a sensor. The beginning of a foreground segment of audio indicates the activation, or on state, of the audio activity, which then reverts to the off state on completion of the activity, the transition from foreground audio to background. In determining the anxiety for the hazardous device \(d^i\), we use a number of statistical models. For the device \(d^i\) we define the Self Interaction Duration model (SID), \(P_{SID}^d(\tau_1)\), which denotes the probability density distribution of the time intervals between interactions with \(d^i\), where \(\tau_1\) denotes the time between interactions. The corresponding cumulative distribution is represented by \(P_{SID}^d(t - t_o)\), where \(t_o\) is the time \(d^i\) was last interacted with, and \(t\) the current time. The closer this probability gets to one, the more anxious device \(d^i\) becomes.

We also define the Interaction Event model (IE), \(P_{IE}^{d,i}\), which denotes the probability of interaction between the user with passive device \(d^j\) while device \(d^i\) is in a hazardous state. For example, \(P_{IE}^{d,i} = 0.9\) means that 90% of the time device \(i\) is in a hazardous state, the user interacts with device \(j\).

For each passive device two statistical models are defined, the Inter Interaction Duration model (IID), and the Inter Activity Duration model (IAD). The description of the time intervals between interacting with the passive device \(d^j\) and then the hazardous device \(d^i\) given that \(d^i\) is in a hazardous state. The corresponding cumulative distribution is represented by \(P_{IID}^{d,i}(t - t_o)\), where \(t_o\) is the time \(d^i\) was last interacted with, and \(t\) is the time \(d^i\) was last interacted with.

The anxiety associated with a device \(d^i\) is attenuated if a user interacts with devices in the environment associated with device \(d^i\). Consequently, we modify the probability associated with the hazardous device \(P_{SID}^d\), to reflect these interactions. The scaling factor associated with each device in the environment is defined as

\[
S^{d,i,j}(t_o, t - t_e) = 1.0 - P_{IAD}^d(t - t_o) \times (1.0 - P_{IID}^d(t - t_e))
\]

The value \(1.0 - P_{IAD}^d(t)\) represents the probability that a user will interact with device \(d^i\) after the current time \(t\), given that device \(d^j\) was interacted with at time \(t_e\). The value \(1.0 - P_{IID}^d(t)\) represents the probability of interacting with device \(d^j\), at a time \(t_e\), given an interaction with device \(d^i\), at a time \(t_o\).

The probabilities are then incorporated to determine an overall anxiety associated with device \(i\):

\[
\text{Anxiety}_{overall}^d(t) = P_{SID}^d(t - t_o) \times \prod_{e_j} S^{d,i,j}(t_o, t - t_e)
\]

where \(e_j\) is an event for device \(d^j\) and assuming that \(e_j\) is independent of each other.

As the anxiety is modelled for each hazardous device independently, the unification of the anxieties of need to be considered. Currently the device with the highest anxiety is used to represent the overall anxiety associated with the house. A value for the anxiety of 1.0 indicates that something that has never been seen before has occurred. In keeping with the pessimistic nature of the anxiety, a lower threshold is set, at which point the user is asked if everything is okay (normal).
4. Experimentation

A. Experimental Environment

To explore these and other ideas, we developed a Smart House laboratory environment. The laboratory is populated with a number of devices to simulate those that would be found in a typical house. The house has several rooms: a kitchen, lounge and bedroom (figure 1 shows two of the rooms). The kitchen includes a small electric stove, microwave oven, fridge, dishwasher, cupboards, a kitchen table and chair. Each device is augmented with sensors to detect interaction by the occupant. Reed switches detect the opening and closing of doors (e.g. the fridge, dishwasher, microwave, cupboards), while pressure mats detect the proximity of the occupant to the doorways. For hazardous devices, pressure mats are positioned on the floor in front of each device to detect proximity and hence potential interaction by the occupant.

An omnidirectional microphone was attached to the roof at the centre of the room by the Room 1: overhead camera to capture the audio associated with the activities in the smart house. The audio activity was determined by synchronising the start and end time stamps for foreground sections of the audio signal with the logs obtained from the sensor data.

B. Activity Data

Sensor and audio data were collected for a number of test sequences, consisting of a normal scenario in conjunction with a number of abnormal scenarios. To speed up data capture, the scenarios were acted at greater than normal speed.

The normal scenario consisted of 35 sequences depicting activities associated with making breakfast. We focus on the anxiety with respect to the stove. The anxiety increases over time if no interaction with the stove is determined, and reduces to zero upon interaction. Variations were present in both the sequence and duration of events, and the presence of certain events within sequences. The interactions with monitored objects included the fridge, stove, microwave, dishwasher, toaster, and a cabinet. Audio activities present that were not associated with a monitored object were predominantly associated with eating breakfast (e.g. cutlery). Further audio activities occurred due to interaction with the devices monitored by sensors, e.g. doors slamming. Table 1 displays a typical example of a normal scenario. Seven sequences were generated depicting abnormal scenarios to test the determination of the anxiety for events not seen within the training sequences.

C. Audio Processing

The normal scenario data was captured in two sessions. Each session consisted of a number of contiguous sequences within a single audio signal. The contiguous capture of sequences was necessary to enable the adaption of the background audio within the environment. The first session consisted of 1.67 hours of audio comprising 21 sequences. The second session contained 14 sequences in 1.33 hours of audio. The hazardous sequences were captured within a single audio of 1.73 hours in duration. The audio signal was captured at 44.1kHz, 16bit, mono, wave format. The background was modelled at a clip size of 1s. Approximately 98.4% of the background audio was modelled correctly. This was determined using the number of foreground events detected by the algorithm over a combined total of 25.4 minutes of inactivity recorded at the beginning and end of each data capture sequence. For the purposes of determining the background accuracy, all audio within the periods of inactivity was considered to be background, including spurious noises occurring outside the smart-house environment.

D. Results

To determine the effect of incorporating audio into the calculation of anxiety, 32 normal sequences were used to model the anxiety, and the remaining 3 sequences were used for testing the attenuation of anxiety due to the presence of audio activity.

To examine the anxiety for the seven hazard scenarios, the 35 normal sequences were used to generate statistical models. The subsequent models were then used to determine the anxiety for the hazardous scenarios.

The anxiety for all cases was determined using equation 2, i.e. the same formulation was used for both testing and training. The statistical models were generated from all interactions with all devices and consequently all interactions within the test sequences were used to determine anxiety.

1) Attenuating Anxiety Due to Audio: Three normal sequences were used to determine the behaviour of the anxiety for the stove with and without audio to explore how audio enhances the anxiety measure. Figures 2 and 3 show two

<table>
<thead>
<tr>
<th>Table 1: Example of a Normal Breakfast Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
</tr>
<tr>
<td>1. Get ingredients from the fridge</td>
</tr>
<tr>
<td>2. Turn on stove and start cooking</td>
</tr>
<tr>
<td>3. Make coffee</td>
</tr>
<tr>
<td>4. Get cereal and check on stove</td>
</tr>
<tr>
<td>5. Eat cereal</td>
</tr>
<tr>
<td>6. Put cereal back and check on stove</td>
</tr>
<tr>
<td>7. Put ingredient back in the fridge</td>
</tr>
<tr>
<td>8. Turn off stove</td>
</tr>
<tr>
<td>9. Eat cooked breakfast</td>
</tr>
</tbody>
</table>

Fig. 1: One room of the Smart House laboratory environment.
normal test sequences, calculating the anxiety both with and without the audio activity data. The lower graph for each figure displays the timeline for the sequence of events. The upper graph displays the anxiety associated with the stove calculated according to $P_{SID}$ (solid line), and the anxiety attenuated due to interactions with other devices within the environment (dashed line).

In Figure 2(a), the anxiety of the stove when no other device interactions are considered ($P_{SID}$) and the anxiety when device interactions are taken into account (dashed line) are shown. The user switches the stove on at 36s, and anxiety starts to increase until 95s, at which time the user interacts with the stove, reducing the anxiety to 0. The user keeps interacting with the stove until 101s, as a result of which the anxiety stays at 0. The anxiety then rises from 101–126s when the user moves away to do other tasks before returning to the stove (126–131s), again reducing the anxiety to 0. The anxiety rises again until 136s, at which point the user turns the stove off. The dashed line displays the attenuated anxiety. At 45s and 92s the anxiety is attenuated due to interactions with the cabinet. Figure 2(b), shows the same sequence but this time the audio is included. Whilst the unattenuated anxiety (solid line) behaves as described before, the attenuated anxiety (dashed line) shows that the anxiety has been greatly attenuated as compared to figure 2(a), because of the contextual audio.

Figures 3(a) and (b) show the results for a differing sequence. In Fig 3(a), between 59 – 92s, the lack of interaction with monitored devices causes the anxiety to rise unattenuated. Figure 3(b) displays a greater level of attenuation due to the presence of audio activity within this section of the sequence.

Both figures display variations of the breakfast scenario, evidenced by the differing interactions with the stove. Despite the differences in the sequence of the activities, the anxiety still produces meaningful results.

From both figures it is evident that the audio activity results in a higher attenuation of the anxiety in comparison with just using the sensor data, as would be expected as the person is in the room. The degree of audio activity present is indicated by the timeline (lower graph). For both figures, the audio occurs more frequently in comparison with the remaining monitored devices.

2) Abnormality: Presence and Absence of Activity: Figure 4 shows the anxiety, determined using the combined audio and sensor data, for two examples of abnormal scenarios. The anxiety was determined for four abnormal scenarios characterised by a lack of activity while the stove was in a hazardous state. Figure 4(a) depicts a normal breakfast scenario from 0 – 225s with the last interaction with the stove at 216s. The user subsequently leaves the room without turning off the stove. Note the absence of audio as the user left the room. The $P_{SID}$ of the stove rises to 1.0 at 299s, with the attenuated anxiety reaching 1.0 at 311s due to the user interacting with the environment after the last interaction with the stove, from 216 – 225s.

The anxiety was then determined for three abnormal scenarios characterised by the presence of activity within the room, and a lack of interaction with the stove while it was in a hazardous state. Figure 4(b) depicts an abnormal sequence in
the presence of activity. The normal breakfast scenario occurs from 0 \(-\) 168s, with 168s being the time of last interaction with the stove. The user subsequently forgets to turn off the stove after cooking, but remains active within the room, both interacting with sensors and producing audio activity. The \(P_{SID}\) reaches 1.0 at 249s, and the attenuated anxiety reaches 1.0 at 302s. In this case, it is the absence of interaction with the stove that results in an increase in the activity due to the use of the \(P_{IAD}\) distribution. The presence of activity within the environment results in an increased time for the anxiety to reach a value of 1.0.

In both cases the lack of audio activity meant the anxiety rose to 1.0, meaning an alarm would be raised.

5. Conclusion

In this paper we have proposed a method for hazard detection in smart environments using a fusion of multi-modal data within an emotive computing framework. Previous approaches to determining abnormal activity have centred around activity recognition fusing data collected from various sensors. We approach the problem from a different perspective. Rather than using activity recognition, we determine normality with respect to the patterns of interaction associated with each hazardous device. We use a probabilistic approach that enables the modelling of complex interaction without being reliant on the sequence of interactions.

The significance of this work lies in two key areas. Firstly, the fusion of multi-modal data for hazard determination in smart house environments. Secondly, this work represents an extension of audio analysis in the field of surveillance and monitoring. Audio is a powerful cue that can be mapped to higher level semantic analysis. The advantage of audio analysis in comparison with simple sensors lies in the contextual and pervasive nature of the audio data. Audio also offers an advantage over video analysis due to the lower processing overheads. Audio background modelling classifies background audio according to the dominant characteristics of the audio over a period of time. Foreground audio classification is therefore determined by a difference in the audio signal from the background. We argue that the novel sounds, i.e., the foreground, are sounds associated with an activity. This links the audio with a higher level semantic meaning.

The anxiety was determined for a number of normal and hazardous sequences, producing meaningful results in both cases. The inclusion of the audio activity resulted in a more meaningful attenuation of the anxiety as sounds are made by people even though they are not interacting with a monitored device.

Future work involves attenuating the effect of known sound sources within the environment, such as TV.

REFERENCES