Monitoring Household Activities and User Location with a Cheap, Unobtrusive Thermal Sensor Array

Peter Hevesi [1], **Sebastian Wille** [2], **Gerald Pirkl** [1], **Norbert Wehn** [2], and **Paul Lukowicz** [1]

[1]German Research Center for Artificial Intelligence, Email: {firstname.lastname}@dfki.de

[2]University of Kaiserslautern, Germany, Email: {lastname}@eit.uni-kl.de

ABSTRACT

We demonstrate that a cheap (30USD) small, low power 8x8 thermal sensor array can by itself provide a broad range of information relevant for human activity monitoring in home and office environments. In particular the sensor can track people with an accuracy in the range of 1m (which is sufficient to recognize activity relevant regions), detect the operation mode of various appliances such as toaster, water cooker or egg cooker and actions such as opening a refrigerator, the oven or taking a shower. While there are sensing modalities for each of the above types of information (e.g. current sensors for appliances) the fact that they can all be detected by such a simple sensor is highly relevant for practical activity recognition systems. Compared to vision (or thermal imaging systems) the system has the advantage is being less privacy invasive allowing it for example to monitor bathroom activities (as shown in one of our evaluation scenarios). The paper describes the sensor, the methods used for activity detection and the evaluation.

Author Keywords

activity recognition; thermal sensor; household monitoring

ACM Classification Keywords

I.5.4 Pattern recognition: Applications

INTRODUCTION

Activity recognition in household or office environments is an important research area of ubiquitous computing with many highly relevant applications such as health, elderly care or energy conservation. To date significant advances were made towards reliable recognition of various activities in highly instrumented, "smart home" environments [1]. Typically such environments include elaborate indoor location systems, the augmentation of various objects and appliances and often a range of on body sensors (motion sensors, RFID reader etc.). By contrast, activity recognition in environments where only limited augmentation is possible remains an open research issue. At the heart of this issue is the search for sensor systems that are on one hand cheap, unobtrusive and easy to install, while at the other hand providing reliable information about as many activities as possible. In this paper we demonstrate that medium size arrays of infrared (temperature) receivers are a promising candidate.

Single "scalar" infrared sensors have long been used as motion detectors. There also exists a lot of research on thermal imaging cameras e.g. to locate people. Recently devices situated between the two have started becoming available: cheap arrays of infrared sensors with good sensitivity and a resolution in the range of 100 pixels (e.g. 8x8). We propose to mount such an array for example on the ceiling (figure 2) to to monitor heat sources present in the room. The basic idea is, that in a typical indoor environment, localized, significant changes in temperature distribution are unlikely to occur accidentally (although exceptions such as e.g. the sun heating up a surface may happen). Instead they correspond to events and activities such as the presence of a person, switching on an appliance, opening the fridge door or a window etc. (see figure 1). We demonstrate that such events have a characteristic signature in terms of equilibrium temperature, rise time/slope and duration (see figure 1). Within a typical household or office environment where the number of potential heat sources is limited such a signature in general allows to reliably identify the corresponding event. Thus the vision is that a single device mounted ceiling can provide information about:

1. The presence and location (to within about 1m) of people.
2. Operation of various appliances such as toaster, water boiler, or a stove.
3. Events such as opening the refrigerator, carrying around a hot drink or the difference between cooking a soup (limited by 100 °C temperature of water) or frying something in a pan (given by the higher temperature of oil).

While other established sensors exist that can provide information about many above events, we propose a single, cheap (in our case 30 USD), small (slightly larger than a 1 Euro coin including microcontroller, optics and RF transmitter, see figure 2 left), low power ($< 5mA$ for the sensor, $< 30mA$ for the entire node including RF part) sensor node that provides all this information. We argue that this is a significant step in the direction of activity recognition with minimal augmentation of the environment and the ability to cost effectively "upgrade" existing homes with at least some degree of activity recognition.

Related Work

Various sensor setup have been proposed and implemented in smart homes for activity recognition and providing even a

UbiComp'14, September 13-17, 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-2968-2/14/09...$15.00.
http://dx.doi.org/10.1145/2632048.2636084

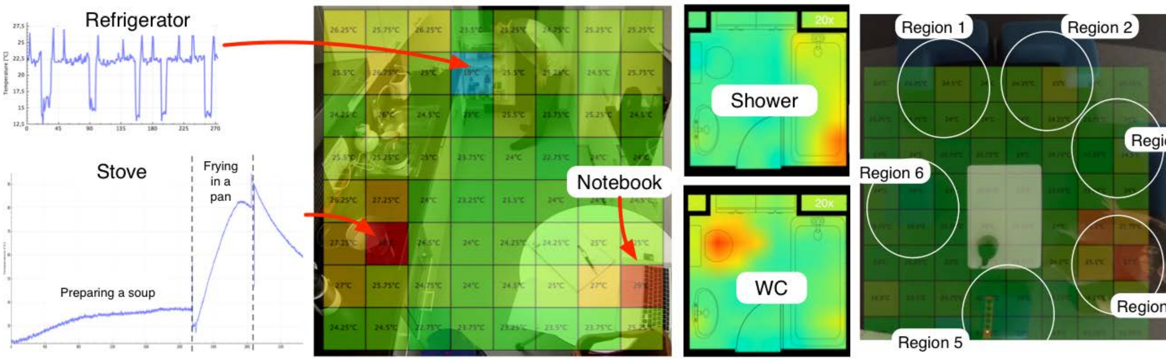

Figure 1. The figure shows from right to left: (1) top view of the living room scenario overlaid with values of current sensor array and the defined regions (seats), (2) overlay of the temperature sensor over the bathroom layout used for testing the sensor in privacy sensitive environment, (3) a snapshot of the sensor array overlaid over an image of the kitchen with different events taking place and a person present, (4) Left side shows the temperature curve of the stove and the fridge related pixels (see text for explanation).

rough overview goes beyond the scope of this paper. A well known example is [7], an overview can be found in [1]. In all cases the combinations of events that we propose to recognize require combinations of several, often complex sensors to be detected. Obviously a camera can in principle provide all the information from a single source and there has been a lot of work on video based activity recognition in household environments (e.g. [6]). While it has its advantages it also has a number of issues ranging from privacy through large amount of training and issues with light conditions, background cluttering etc. As an alternative infrared cameras have also been used [2] which reduces some of the problems but still relies on actually recognizing human shape and motion. Infrared sensing has been used for indoor location in two ways. In [8], infrared beacons carried by the user are used. In [4] the IR emissions of a human body are triangulated. With respect to recognizing the usage of electrical appliance current monitoring is an established technique [5]. However it recognizes only the appliances, would not work for example for a gas stove and things like opening the oven or distinguishing between water pot and a frying pan.

The closest to our work is [3] that uses a thermal imaging camera to "disaggregate total power usage from a master power meter into individual appliance power usage.". On a methodological level the main difference is that [3] builds on image segmentation techniques and the correlation between thermal change and master power meter output to extract information from a rich source (320x240 video stream) whereas we leverage heat emission models to infer events from a very sparse information source (8x8 array of IR sensors covering an entire room). On application level this leads to a system that (1) is very cost effective (30USD for a sensor as opposed to up to 1000 USD for a thermal camera), (2) is suitable for privacy sensitive environments where a full fledge camera would not be acceptable (as we demonstrate by including bathroom activities in our evaluation) and (3) is not restricted to events that produce a distinct signal in a master power meter (e.g. difference between boiling water and heating a pan with oil on a stove, or detection of the shower).

HARDWARE

The sensor node we are using is shown in figure 2 left. The actual sensor is a AMG8831 "Grid Eye" from Panasonic with 8x8 pixels, $0.25\,^{\circ}\mathrm{C}$ sensitivity, 10Hz sampling rate and integrated optics with a 60 degree field of view. We have combined this sensor with an Atmel ATmega324P microcontroller and a RF12B with 868MHz transmitter (up to 115.2 kbps). A matching receiver can be connected to a PC to collect the data. The platform is 26.5x44x6mm and consumes between < 15mW when only acquiring data from the infrared array and around 100mW when transmitting data. An on board motion sensor can be used to trigger the system from sleep mode further reducing the power consumption and allowing it to potentially operate over weeks on a single battery. Since the idea is to just stick the node to the ceiling the sensor could also be powered from the lamp outlet.

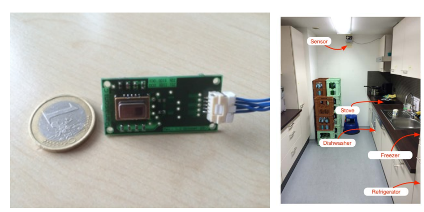

Figure 2. Left: the sensor used for the experiments. Right: the kitchen environment. The sensor is mounted with a mechanism that allows to adjust the angle in the edge where the ceiling meets the back wall.

ANALYSIS AND RECOGNITION METHODS

Core Ideas

The signals produced by the sensor can be seen in figure 1. Each of the 64 pixels corresponds to an area within the room. The assignment of pixels to areas depends on the position of the sensor and can be easily computed as an optical projection. The sensor assigns this area a temperature value, which however is not the true temperature of any physical object in that area. Instead it is the average temperature within the area modified by the distance to the sensor with the distance essentially introducing a constant scaling factor.

When nothing happens in a room the temperature distribution will be largely homogeneous and will vary over time only slowly (as the room as a whole heats up or cools down). Any localized, quick (over seconds or minutes) change corresponds to an event such as switching on an appliance, opening the fridge or a person entering. Such events can be described by a set of parameters like (1) the minimum and (2) maximum temperature, (3) the slope of the temperature rise, (4) the slope of the temperature return to ambient value, and (5) the time scale of the event. In figure 1 bottom left this is illustrated on the example of the stove. We can see that cooking a pot of water (e.g. for a soup) has a slow slope (large mass) and a low saturation temperature (given by the boiling point of water). Frying something in a pan has a much steeper slope and higher saturation value (given by the thermal properties of oil). Events such as taking off the pot (revealing the even hotter stove plate) and putting on a cold pan have an immediate effect with their own characteristic temperature.

Analysis Approach

We differentiate between (1) events that are determined by a simple temperature threshold (person tracking and bathroom activities closely related to tracking, or temporally well located events such as opening the fridge) and (2) events where the temporal dynamics of the temperature (length slope etc.) is needed for the recognition (most appliances). For the former the recognition works on the basis of individual frames. For the later we work on segments that encompass all the relevant phases of the event.

To remove sensor noise a low pass filtering of the temperature curves is necessary for both approaches, however the filter grade (how strong the signal is filtered) can be different for other appliances. E.g. for long time events like cooking a soup it is better to use a filter with a low cut off frequency which filters the values where the sight was covered. Meanwhile events like opening the refrigerator could get lost because they are short-time temperature changes, thus they must not be filtered with the same filter settings.

Time Independent (Frame) Recognition

For every time step and every element of the sensor array the absolute temperature, temperature change and temperature difference with respect to the average room temperature is computed from the filtered signal. The resulting feature vector is then used as input to a standard statistical classifier trained using appropriate ground truth. For person location we essentially perform an independent binary classification (person present, no person present) on groups of pixels corresponding to each location region. For the bathroom activities, which involve certain temperature distributions over groups of pixels, a classification is performed on the entire vector.

Time Dependent (Segment) Recognition

For appliances that are characterized by a temperature curve rather then an single value (or spatial value distribution) classification in single frames is not possible. Instead we must look at a sequence of values that certain pixels or pixel groups take over time. In doing so we leverage the fact that an appliance event in general consists of a startup phase (when the appliance has been switched on and the appliance temperature is rising), an operation phase when its temperature remains over a certain threshold, and a cool down phase after it has been switched off and the temperature sinks towards the ambient value. Thus, to identify time periods where an event could possibly be taking place we first look for distinct peaks in the signal of each single sensor and identify the startup, the operation and the cool down phases. Mathematically it's equivalent with finding the zero crossings of the first derivative where the second derivative is positive. An appropriate filtering of the signal is needed for that. During the analysis a Butterworth low pass filter of second grade is used.

In the next step for every found segment a feature vector is generated which contains the following parameters: (1) minimum, (2) mean and (3) maximum of the temperature change in the segment (differential quotient), (4) minimum, (5) mean and (6) maximum of the temperature in the segment, (7) sum of the temperature values in the segment (energy), (8) position of the maximum temperature in the segment, (9) position of the smallest temperature difference (largest fall), (10) position of the largest temperature difference (slope), (11) length of the segment, (12) pixel index of the current temperature curve (information about the location). A standard statistical classifier is then applied to the features to determine what event (appliance) it corresponds to. In a final steps parallel events found in neighbouring sensors are fused to account for the fact that appliance may at times be positioned in such a way that they generate a signal in more then one sensor.

EVALUATION

Experimental Setup

The evaluation has taken place in four environments. First there is a small, long but narrow kitchenette in our lab where the sensor is mounted on the wall looking sidewards (since the ceiling is too low to cover the entire room) as shown in figure 2 right. In this room there are built-in appliances like refrigerator, freezer, stove and dishwasher. Movable appliances such as toaster, water boiler, egg cooker are also used there. The whole scenario is longer than 8 hours but it was recorded in parts with lengths of between 20 and 60 minutes to reduce risk of data loss. In a post processing step the small parts are merged to one big scenario for further analysis. Overall around 300 events corresponding to different appliances (see in figure 5 for a breakdown) took place at random times. Times between appliance usage events are labelled as zero class segments. The kitchen was normally used during the recordings, people walked in and out and operated the appliances. The notable events are written down. After the recording the event labels are matched to the corresponding pixels and time slots manually.

The second is a bathroom with a toilette, basin, a cupboard and a bath with a shower. This layout can be seen in Figure 1 in the third column. The sensor was mounted on the ceiling in the center of the room. This recording is 50 minutes long and contains events like sitting on the WC, standing at the basin (washing hand, teeth etc.), standing at the cupboard, taking a shower and hair drying. The events in the scenario are completely natural. The labels are recorded simultaneously by a second person standing out of the visible area. The evaluation (and labelling) take place on the basis of 500 msec long frames of which there is a total of 6144 in the data set.

The third environment is our social area with a couch table where we focus on people tracking (for 30min). The sensor was mounted on the ceiling in a height of about 2,5m. A second kitchen scenario (as our fourth location) is also recorded for people tracking where the sensor is on the ceiling in a height of 3 m. For people location we have some 2600 labelled frames in which both the living space and the kitchen were divided into 6 areas $0.5m^2$ or less in size.

Evaluation

In each environment we first extract from the recorded data 10 % of each event class as training set and use them to train the recognition systems as described above. We use standard WEKA PART decision tree (for the appliances) and KNN (for all other events) classifiers.

Classified as —>	a	b	c	d	e	f	Nr. of frames	Preci sion	Recall
a: zero class	1564	2	6	2	0	0	1574	0,994	0,994
b: wc	1	792	3	0	0	0	796	0,994	0,995
c: basin	4	2	636	3	2	2	649	0,985	0,980
d: cupboard	4	1	1	587	1	1	595	0,988	0,987
e: hair drying	1	0	0	1	677	0	679	0,996	0,997
f: taking shower	0	0	0	1	0	1850	1851	0,998	0,999
							average:	0,992	0,992

Figure 3. Bath activity recognition results

Location	Number of frames	Precision	Recall
Kitchen	1972	0,986	0,979
Living room	2735	0,929	0,98

Figure 4. Person tracking evaluation results

Classified as —>	a	b	c	d	e	f	f1	f2	f3	g	h	i	Nr. of events	Preci sion	Recall
a: zero class	341	9	2	4	7	0	0	0	0	5	7	8	383	0,85	0,89
b: toaster	11	46	0	0	0	0	0	0	0	0	0	0	57	0,821	0,807
c: egg cooker	4	0	17	2	0	0	0	0	0	0	0	0	23	0,850	0,739
d: water boiler	6	1	0	14	0	0	0	0	0	0	0	0	21	0,700	0,667
e: stove	8	0	0	0	22	0	0	0	0	0	0	0	30	0,759	0,733
f: stove (location assumed as known)	0	0	0	0	0	30	0	0	0	0	0	0	30	1,000	1,000
f1: stove - cooking water in a pot (location assumed as known)	0	0	0	0	0	0	12	0	0	0	0	0	12	1,000	1,000
f2: stove - frying in a pan (location assumed as known)	0	0	0	0	0	0	0	8	0	0	0	0	8	1,000	1,000
f3: stove - cooling (location assumed as known)	0	0	0	0	0	0	0	0	10	0	0	0	10	1,000	1,000
g: dishwasher	5	0	1	0	0	0	0	0	0	10	0	0	16	0,667	0,625
h: refrigerator	12	0	0	0	0	0	0	0	0	0	57	1	70	0,877	0,814
i: freezer	12	0	0	0	0	0	0	0	0	0	1	57	70	0,864	0,814
													average:	0,799	0,751

Figure 5. Appliance and event recognition results in the kitchen. Note that for the class f and its subclasses (stove activities) we assumed that the stove position on the sensor array is already known, and we want only recognised the current activity like (1) cooking water in a pot, (2) frying in a pan, (3) let the stove cool down.

For testing the recorded data is feed into our system as a continuous stream. For the events that are recognized on a frame basis (user location, bathroom events) our system outputs a classification in each time step (number of the area in which the user was located or the type of bathroom event). If no event is spotted (no bathroom event or no user) a zero class label is returned. This classification is compared with the ground truth to create the confusion matrix (for the bathroom events, see figure 3) and the precision recall values (for the person location see figure 4). For the events that are spotted on segment basis the system returns the start and end of frame number of a segment as well as the estimated class. Periods during which no segment belonging to one of the relevant classes is spotted are labelled as zero class segments. To create the confusion matrix in figure 5 we compare the returned boundaries with the ground truth and consider the system output to be correct if the class is the same and the event timing has at least 50% overlap.

Results

The result are summarized in tables 3, 4 and 5. We can see that the precision recall rates are mostly over 90% for the frame based events and between 70% and 80% for the appliances. In particular the people tracking results are in the high nineties. We believe that most of the wrong classifications that we observe are due to the "thermal pollution" of the environment caused by switching many devices in a short time period. Increasing the number of samples used for the training had only minimal effect.

Discussion

The results confirm the hypothesis that simple classifier algorithms trained on a small number of samples is sufficient to recognize a variety of events and locate people. The main limitation of the system that we have run across during the experiments is the interaction between close by devices. If two devices are within one pixel they can obviously not be separated. The same is true for devices that generate significant heat and are allocated to neighbouring pixels. In an extreme case a frying pan cooking for the long tie on a small kitchen can temporarily obscure any other events.

CONCLUSION

The core conclusion of our work is that arrays of infrared detectors are a simple and cheap yet rich source of information for household activity recognition. Note that the aim of this short paper was not to develop a complete activity recognition system and test it under real life conditions. Instead it was to empirically demonstrate the potential and limitations of a sensor modality that has so far not been used in this form in activity recognition allowing other researchers to build on our result.

Our own next steps are to combine the sensor with sound analysis to perform long term complex activity recognition in real world environments. We will also expand the experiments to other locations (different kitchen, bathroom) including more events (e.g. tracking a hot cup, detecting vacuum cleaning, etc. for which we have already acquired initial promising signals).

ACKNOWLEDGMENT

This work was supported by "Landesschwerpunkt" AmSys and Asandoo GmbH (asandoo.com). Funding of this joint research project (CoCoRec) by the German Federal Ministry of Education and Research of Germany is gratefully acknowledged.

REFERENCES

1. Cook, D. J., and Das, S. K. How smart are our environments? an updated look at the state of the art. *Pervasive and mobile computing 3*, 2 (2007), 53–73.

2. Han, J., and Bhanu, B. Human activity recognition in thermal infrared imagery. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, IEEE (2005), 17–17.

3. Ho, B.-J., Kao, H.-L. C., Chen, N.-C., You, C.-W., Chu, H.-H., and Chen, M.-S. Heatprobe: a thermal-based power meter for accounting disaggregated electricity usage. In *Proceedings of the 13th international conference on Ubiquitous computing*, ACM (2011), 55–64.

4. Kemper, J., and Hauschildt, D. Passive infrared localization with a probability hypothesis density filter. In *Positioning Navigation and Communication (WPNC), 2010 7th Workshop on*, IEEE (2010), 68–76.

5. Patel, S. N., Robertson, T., Kientz, J. A., Reynolds, M. S., and Abowd, G. D. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *UbiComp 2007*. Springer, 2007, 271–288.

6. Rohrbach, M., Amin, S., Andriluka, M., and Schiele, B. A database for fine grained activity detection of cooking activities. In *CVPR, 2012 IEEE*, IEEE (2012), 1194–1201.

7. Tapia, E. M., Intille, S. S., and Larson, K. *Activity recognition in the home using simple and ubiquitous sensors*. Springer, 2004.

8. Want, R., Hopper, A., Falcao, V., and Gibbons, J. The active badge location system. *ACM Transactions on Information Systems (TOIS) 10*, 1 (1992), 91–102.