

Use of Low-Resolution Infrared Pixel Array for Passive Human Motion Movement and Recognition

Yordanka Karayaneva
Coventry University
Coventry, United Kingdom
karayany@coventry.ac.uk

Samuel Baker
Coventry University
Coventry, United Kingdom
bakers29@coventry.ac.uk

Bo Tan
Coventry University
Coventry, United Kingdom
ac5162@coventry.ac.uk

Yanguo Jing
Coventry University
Coventry, United Kingdom
ac2716@coventry.ac.uk

The daily monitoring of ageing population is a current issue which can be effectively tackled by applying daily activity monitoring via smart sensing technology. The purpose of the monitoring is mostly aimed at collecting health conditional related activity awareness and emergency events detection. This is a pilot study that uses low pixel resolution infrared sensors for nonintrusive human activity detection and recognition without body attachments and taking of individual image. In this work, we design and implement a multiple IR sensors system and a serial experiment to verify the availability of applying low-resolution IR data for human activity recognition for both single and multiple target scenarios in the healthcare context. In the experimental setup, the sensor system achieves 82.44% accuracy in general and reaches 100% accuracy rate for some particular activities. The work proves that the low-resolution IR information is an effective metric for human activity monitoring in healthcare applications.

Infrared sensors. Human activity detection. Classification methods. Elderly care homes.

1. INTRODUCTION

The ageing population is a worldwide problem that is expected to become more prominent shortly. According to a recent United Nations report, the number of people aged 80 or over on the planet is projected to grow to 434 million in 2050 (United Nations, 2015). In contrast with 2015, their number was estimated to be 125 million. Ageing society is one of the 4 Grand Challenges in the UK's Industrial Strategy. The crisis in the health sector which faces lack of available staff complicates the problem further (Slawson, 2017). A cost-effective solution for monitoring resident's daily activities will be the key to approach the issue. Thus, the implementation of smart sensing technology which can monitor human daily activities will have a massive opportunity in future healthcare. Jing et al. (2017) propose a novel monitoring system which uses a huge diversity of sensors for human activity recognition in extra care homes. Thus, the information can be provided to staff and relatives for peace of mind while keeping the privacy of the residents. In the context of healthcare, researchers have tested optic cameras (Tabar, Keshavarz & Aghajan, 2006), wearable sensors (M. Hossain, Pal & S. Hossain, 2015), radio frequency sensors (Tan et al., 2015) and environmental sensors (Popescu et al., 2008) to tackle different challenges, but those solutions suffer from lack of privacy, uncomfortableness, envision or low accuracy issues. Thus, we initiated a pilot study of using the low resolution infrared pixel array for nonintrusive, privacy free and accurate human activity monitoring solution. Low pixel resolution Infrared (IR) sensors are between of PIR and high-resolution IR camera.

It is aimed at monitoring the residents in a nonintrusive manner without attachment on the body and taking sensitive individual images. IR sensors detect infrared radiation which is emitted by the human body with a temperature greater than the absolute zero. Thus, humans can be detected, but they cannot be identified because of the very low pixel resolution. The work in this paper addresses the following problems of applying the low pixel resolution IR data for human activity recognition in healthcare applications:

- 1) Prove the availability of low-resolution IR sensor to detect the human target and identify different activities.
- 2) Using classic classification approaches to test utilisation of low-resolution IR images in the real experimental environment.
- 3) The potential for detection and recognition of multiple targets.
- 4) Figure out the potential constraint factors of applying the new type of sensors in healthcare scenarios.

We design and deploy the experiments based on the Grid-EYE[®] sensors developed by Panasonic[®] to verify our hypotheses. Classical machine learning methods are applied for recognition with 82.44% accuracy rate in average for the three sensors for the three methods while for some of the gestures the recognition even reaches 100%. The rest of the paper is organised as follows: Section 2 briefly introduces the related works in this field. Sensor system design, experiment setup and dataset properties are described in detail in Section 3. We elaborate the machine learning methods and the related performance and discussion in Section

4. In Section 5, the conclusion and future works are displayed.

2. RELATED WORK

The use of low pixel resolution infrared sensors is growing due to their advantages over traditional and more widespread technologies such as cameras, radio frequency sensors and wearables. In regards to traditional cameras, the major issue with these is the invasion of privacy which needs to be avoided especially in the residential context. In addition to this, cameras are difficult to be effective during night time which further increases their inapplicability in the researched terms. Radio Frequency sensors are another type of sensors used for similar purposes as IR sensors. They have advantages including huge coverage of distance which is up to 50 m and their ability to travel through objects such as walls (Tan, Woodbridge & Chetty, 2016). In spite of the popularity of radio frequency sensors, they have some evident disadvantages. The architecture is more complex due to the high number of sensors that are needed to comprise a triangle or square/rectangle shape. Logically, the charged electricity power is very huge which on the other hand increases cost (Ajami and Carter, 2015). These sensors use radio waves that are transmitted in the entire system which raises the issue of health safety which is highly related to this research. Older adults are seen as vulnerable, and due to this, the use of such architecture is not strongly advisable (RF Wireless World, 2012). Human activity detection can also be achieved via wearable sensors. These provide simplicity regarding implementation and support, but as their name implies, the individual needs to wear it or carry it with them (Zhang and Sawchuk, 2013). Thus, this can be seen as demanding for elderly people, which makes the technology highly undesirable.

Smart sensing technology is applied in a recent study where five activities have been used including falling (Mashiyama, Hong and Ohtsuki, 2015). One Grid-EYE sensor is attached to the ceiling of two rooms with the participation of five subjects. In spite of the satisfactory results, only one classification method is implemented which is the popular Support Vector Machine (SVM) with Radial Basis Function kernel. Therefore, it is difficult to make conclusions on the suitability of other machine learning classifiers. There do not exist any reasons regarding the selection of the method and the causes of its success. Furthermore, the composed dataset can be regarded as insignificant due to its simplicity and size.

Basu and Rowe (2014) used Grid-EYE® sensor to predict room occupancy in common occasions such as meetings. Despite the dissimilarity of experiment's purpose, the nature of execution is

strongly related to this research. As in the previous project, a novel dataset is created with over 900 scenes with moving subjects in a room. SVM is used for classification, and the results are poorer as on average the accuracy is around 80%.

A far-infrared sensor array is employed by Honoso et al. (2015) for 13 human activities detection. The difference with Grid-EYE® sensor is the grid format which is four times the 8 × 8 array and becomes 16 × 16. A thermo-spatial sensitive histogram is implemented for recognition which is 70% on average for all activities and for only one the recognition is 100%.

Three classical classification methods are used in this study – Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbours (K-NN) to prove the availability of low resolution IR pixel data. SVM is a very effective method with notable results for sensing which is observed in the studies above. SVMs use a classification hyperplane which discriminates between different classes (Gunn, 1998). Stochastic Gradient Descent training can be applied to the linear SVM for improved performance (Buttou, 2012). RF represents an ensemble of decision trees where the input vector is placed in each tree. "Voting" takes place in the forest, and the tree that receives the majority of the votes defines the label (Breiman & Cutler, 2014). K-NN is known for separation of the classes based on their input data. When new input data needs to be a classifier, the highest similarity with a particular class is considered and eventually given the same label (Sutton, 2012).

3. SYSTEM ARCHITECTURE

3.1. Grid-EYE sensors set up

The sensor system is developed using Panasonic Grid-EYE® evaluation boards, which feature a thermopile-based sensor array arranged in an 8 × 8 format. The sensors are driven and controlled using LabVIEW, with each sensor set to a refresh rate of ten frames per second.

The LabVIEW® program communicates with each evaluation board using a Serial interface providing real-time control over the behaviour of the sensor as well as access to the raw temperature values measured at each pixel. These raw temperature values are processed and used to generate a low-resolution 8 × 8 image representing temperature values of the target positioned in front of the sensor. This process is repeated until a suitable catalogue of gesture readings are collected as training data.

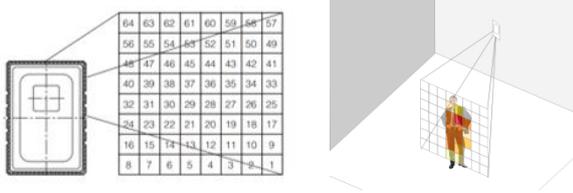


Figure 1: The structure of an array of thermopile sensors within the grid eye package

Gesture data was captured using two sensor arrangements - both are displayed in the image below:

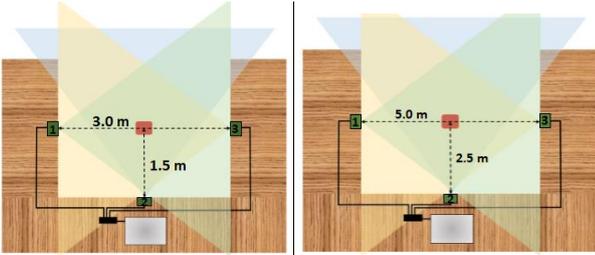


Figure 2: Small and large layout of three sensors

For the initial experiment, all three sensors were positioned upright, equally 1.5 meters away from a space assigned as the area where gestures were to be performed; all three sensors were elevated roughly 1 meter from the ground. After completing each gesture experiment, the surfaces used to support each sensor were moved further apart to an equal distance of 2.5 meters between each sensor and the space assigned for gesture capture.

At the beginning of the gesture capture experiment, the ambient temperature of the surrounding room was measured to be 18 degrees centigrade, during the time taken to perform and capture gestures the ambient temperature of the surrounding room rose to 21 degrees. Air conditioning was used to reduce the ambient temperature of the room during the experiment phase to reduce background temperature noise when capturing gesture data.

3.2. Dataset for human activity recognition

The dataset is composed of 15 gestures in total which are distinguished in two categories. The two categories are based on the number of participants, while the first contains body gestures with one subject, and the second category consists of gestures with two subjects.

Table 1: Gestures classified into two categories

	One subject		Two subjects
A1	Sit Down	B1	Sitting
A2	Stand Still	B2	Sitting; Movements
A3	Sit Down; Stand Up	B3	Sitting; Standing
A4	Stand Up	B4	Movements
A5	Move Left, Right	B5	Standing
A6	Move Forward, Backward	B6	Standing; Movements
A7	Walking Diagonally 1	B7	Walking Diagonally
A8	Walking Diagonally 2		

To prepare the raw data for recognition, certain steps need to be taken including the conversion to a suitable data format and unification of the number of the frames in the files. As the files have already been converted to the *.csv format, their number of lines need to be equalized. Each line represents a single frame with temperature values where greater values logically imply the presence of a subject. One second from the scene is equal to ten frames. There exists a huge diversity among the duration in the different files, and hence, the file with the smallest number of frames needs to be taken for a sample. Due to the fact that it contains 27 frames, the remaining set of files needs to be pruned for an equal number of frames. To do this, the minimum number of frames (f_{min}) needs to be subtracted from the number of frames in the file (f) in cases when $f \neq f_{min}$:

$$f_{remove} = f - f_{min} \quad (1)$$

The frame removal can be conducted in two manners: by random or by similar frames.

4. EVALUATION

Three classical methods for machine learning were used for classification: Random Forest, Support Vector Machine with Stochastic Gradient Descent training and k-Nearest Neighbours. Evaluation test has been performed in order to compare the accuracy for the experiments involving one subject in the big area and two subjects which were performed in the big area as well. The purpose of this experiment was to discover any differences in the performance based on the number of human subjects. Expectedly, the one subject experiment had higher accuracy than the experiment with two subjects as it can be observed on Figure 3.

As it can be observed on Figure 4, the two scenarios for one subject in the small area and one subject in the big area are compared for the three sensors. The purpose of this experiment is to observe any notable differences in the performance regarding the size of the room including the classifiers themselves. The experiment which involved one subject in the small area can be recognised with better accuracy due to the fact that the human's temperature is higher which makes the gesture on the low resolution more tangible.

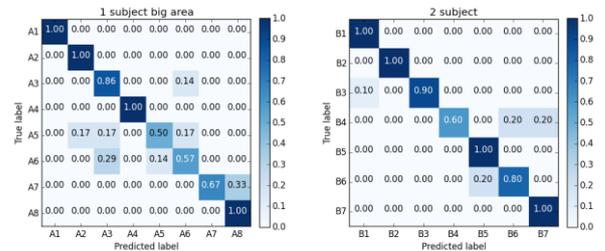


Figure 3: Comparison between 1 subject large area and two subjects performance

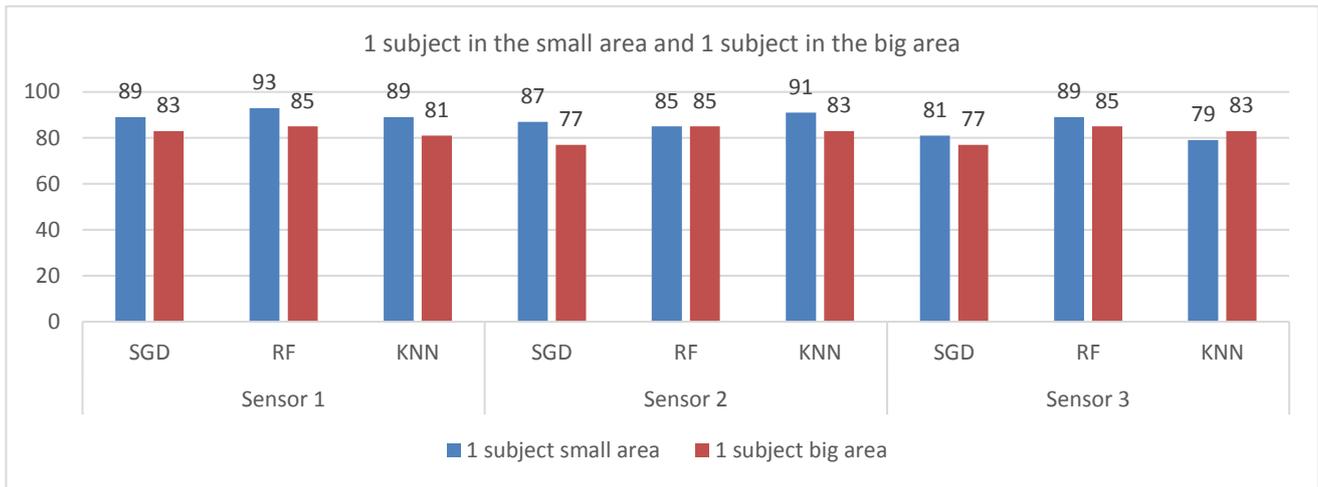


Figure 4: Comparison of the accuracy between the two scenarios for one subject

5. DISCUSSION

Table 2: 1 subject in the small layout

	SVM	RF	KNN
Sensor 1	97.92%	95.83%	85.42%
Sensor 2	89.58%	89.58%	87.50%
Sensor 3	85.42%	89.58%	81.25%

Table 3: 1 subject in the large layout

	SVM	RF	KNN
Sensor 1	77.08%	85.42%	83.33%
Sensor 2	79.17%	85.42%	93.75%
Sensor 3	72.92%	83.33%	81.25%

Table 4: 2 subjects

	SVM	RF	KNN
Sensor 1	88.10%	90.47%	80.95%
Sensor 2	92.86%	85.71%	83.33%
Sensor 3	71.43%	88.10%	76.19%

Table 2 presents the performance for one subject in the small layout while Table 3 is related to one subject in the large layout. Then, Table 4 corresponds to the performance of the scenario with two subjects.

Regarding the separate experiments themselves, the experiments for one subject conducted in the small area logically shows higher accuracy than the ones conducted in the large area. This is plausible as with distance growth, the detected heat radiation is lower which complicates the gesture recognition. In connection with one subject in the large area compared with two subjects experiments, the recognition is higher for two subjects. The accuracy of sensor 2 is tangibly higher as it is located in front of the subjects while sensor 1 and sensor 3 have the likelihood of observing the two subjects as only

one. Performances of the combination of the three sensors have been made where any two sensors are taken together and all three sensors as well. The performance is outlined in Table 5 as follows:

Table 5: Accuracy of combination of sensors

Method	S1; S2	S1; S3	S2; S3	S1; S2; S3
SVM	84.78%	81.88%	75.36%	94.20%
RF	90.57%	88.40%	87.68%	91.30%
KNN	81.88%	76.81%	84.04%	87.68%

6. CONCLUSION AND FUTURE WORK

In this pilot work, three classification methods, SVM, RF and KNN have been applied to a low-resolution IR dataset from three synchronized Grid-EYE® sensors for human target activity recognition with aims to be used in healthcare applications. According to the experiments carried out in this study, we can easily spot that the recognition performance is promising for practical usage for both small and large geometry size. Accuracy rates from the three sensors vary between 71% and 97% overall for all gesture classes. Random Forest is the most successful algorithm among the used methods showcasing 100% for eight (8) classes from the three sensors. Overall for the entire application, the accuracy has been estimated to be 82.44% which is promising in the field of low resolutions images recognition.

In regards to future work, the current research will be enhanced by collecting more data in order to have more examples to train the model. Deep learning can be applied to the project which has shown promising results in previous studies involving the use of IR images. Moreover, the project will be piloted in care home settings for evaluation.

7. REFERENCES

- United Nations. 2015. World Population Ageing 2015. Department of Economic and Social Affairs, Population Division.
- Slawson, N. 2017. UK social care sector in crisis due to staff shortages. (March 2017). Retrieved April 5, 2018, from <https://www.theguardian.com/society/2017/mar/08/uk-social-care-crisis-staff-shortages>
- Jing, Y., Eastwood, M., Tan, B., Konios, A., Hamid, A., and Collinson, M. (2017). An intelligent well-being Monitoring System for Residents in Extra Care Homes. International Conference on Internet of Things and Machine Learning (IML 2017). Liverpool John Moores University. October 17 – 18, 2017. Published by ACM. DOI:10.1145/3109761.3109769
- Maleki Tabar, Ali & Keshavarz, Arezou & Aghajan, Hamid. (2006). Smart home care network using sensor fusion and distributed vision-based reasoning. doi: 10.1145/1178782.1178804.
- M. N. A. Hossain, A. Pal and S. K. A. Hossain, "A wearable sensor based elderly home care system in a smart environment," 2015 18th International Conference on Computer and Information Technology (ICCIT), Dhaka, 2015, pp. 329-334. doi: 10.1109/ICCITechn.2015.7488091
- B. Tan et al., "Wi-Fi based passive human motion sensing for in-home healthcare applications," 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), Milan, 2015, pp. 609-614. doi: 10.1109/WF-IoT.2015.7389123
- Popescu, M.; Li, Y.; Skubic, M.; Rantz, M. An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008), Vancouver, BC, Canada, 21–24 August 2008; pp. 4628–4631.
- B. Tan, K. Woodbridge and K. Chetty, "A wireless passive radar system for real-time through-wall movement detection," in IEEE Transactions on Aerospace and Electronic Systems, vol. 52, no. 5, pp. 2596-2603, October 2016. doi: 10.1109/TAES.2016.140207
- Ajami, S. and Carter, MW. 2013. The advantages and disadvantages of Radio Frequency Identification (RFID) in Health-care Centers; approach in Emergency Room (ER). Pak J Med Sci (2013), 443-448 DOI: [http://dx.doi.org/10.12669/pjms.291\(Suppl\).3552](http://dx.doi.org/10.12669/pjms.291(Suppl).3552)
- RF Wireless World. 2012. IR vs RF | difference between IR remote and RF remote controls. (2012) Retrieved April 8, 2018, <http://www.rfwireless-world.com/Terminology/Difference-between-IR-remote-and-RF-remote.html>
- Zhang, M. and Sawchuk, AA. 2013. Human Daily Activity Recognition With Sparse Representation Using Wearable Sensors IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS (May 2013), 553-560
- Mashiyama, S., Hong, J., and Ohtsuki, T. 2015. Activity Recognition Using Low-Resolution Infrared Array Sensor. IEEE ICC 2015 SAC – Communication for E-Health (2015), 495-500
- Basu C. and Rowe, A. 2014. Tracking Motion and Proxemics using Thermal-sensor Array. Carnegie Mellon University report.
- Hosono, T., Takahashi, T., Deguchi, D., Ide, I., Murase, H., Aizawa, T., and Kawade, M. (2014). Human Tracking Using a Far-Infrared Sensor Array and a Thermo-Spatial Sensitive Histogram. 262-274. 10.1007/978-3-319-16631-5_20.
- Gunn, S. (1998). Support Vector Machines for Classification and Regression. University of Southampton, pp.5-17.
- Bottou, L. (2012). Stochastic Gradient Descent Tricks. Springer, [online] 7700, pp.430-445.
- Breiman, L. and Cutler, A. (2014). Random Forests. [online] Statistics at UC Berkeley.
- Sutton, O. (2012). Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction. pp.1-3.