Sparse Feature Extraction for Activity Detection Using Low-Resolution IR Streams

Sparse Feature Extraction for Activity Detection Using Low-Resolution IR Streams

Yordanka Karayaneva*, Sara Sharifzadeh*, Yanguo Jing* Kevin Chetty† and Bo Tan‡
* *Coventry University, UK* † *University College London, UK* ‡ *Tampere University, Finland*

***Abstract*—This paper is about activity detection for a new type of low resolution image data. Such ultra-low-resolution infrared (IR) images are suitable for activity detection and monitoring in private areas, e.g. elderly car and modern healthcare applications. The focus is on analysis of sequences of frames including single subjects doing daily activities such as walking, body swing, sitting, or standing. Due to lack of spatial dependencies of the pixels in ultra-low resolution images, the pixels are considered as independent variables. Therefore, the analysis is based on temporal variation of the pixels in vectorized sequences of several frames. This results in a high dimensional feature space and an (N¡¡P) problem. Therefore, two different sparse analysis strategies are employed and compared; sparse discriminant analysis (SDA) and sparse principal component analysis (SPCA). Four different classification strategies namely, Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbours (KNN) and Logistic Regression (LR) are used to evaluate the extracted sparse features. The best classification accuracy is 92% using the sparse features. Furthermore, comparison of the classification results in noisy condition using the two sparse feature extraction techniques and their corresponding nan-sparse techniques, namely, Fishers canonical variables known as reduced rank LDA (RRLDA) and PCA, shows the superiority of sparse techniques in terms of accuracy.**

I. Introduction

The activity detection is the analysis of human movement behaviour and their situation in the scene. This is achieved using sensors data (usually images) in smart pervasive environments. One important application of activity detection is

According to an article published by the Work Foundation in 2016 [1], 15.4 million people in England suffer from one or more chronic medical conditions including dementia, stroke, cardiovascular or musculoskeletal disease. The current treatment costs is 70% of the UK National Health Service (NHS) budget [2] and will increase with the growing aging population. The cost of continuous management and medical treatment for long term health conditions in clinical and residential environments has become an increasing economic pressure. To address this situation, novel solutions for out-patient long-term monitoring in residential settings are required for health monitoring and early diagnosis. This will lead to a more efficient national healthcare model.

In this context, the Internet of Things (IoT) is considered as a core technology to maintain pervasive connectivity of monitoring devices or sensors. It also supports transmission of recorded human activity or residential environmental data to local data centre or cloud. There have been researches and practical projects bridging the IoT and healthcare, like SPHERE [3] for resident house monitoring and exemplary sensor system in [4], [5] for elderly care house activity monitoring. Based on the experiences gained from these research or practices, we notice that the environmental data such as smart meter, water reading, temperature, humidity, door sensors, are suitable for the Narrow-Band (NB) and low data rate nature of the IoT networks. However, the environmental data are just indirect indexes of human activity, which conduct low activity estimation accuracy. There are other types of sensors like camera, Radio Frequency (RF) sensors like Radar [6], [7] and wearable accelerators that acquire rich data, accurately captured from human activities. However, the short battery life, poor comfort and low compliance of wearables, especially amongst the elderly, are limiting factors in the applications of wearable sensors. On the other hand, contactless sensors like cameras and radar are facing ***i).*** high data rate and ***ii).*** computational load problems in IoT networks.

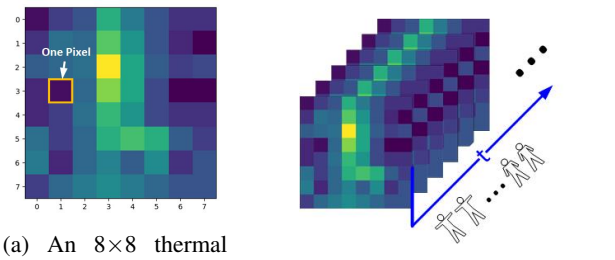
Video streaming is a challenging task for even wide-band networks, not to mention the IoT networks which often rely on NB and short duration IEEE 802.15.1 (Bluetooth) or multiple hop 802.15.4 (Zigbee) protocols. Similar to the video data, transmission of the raw RF I/Q samples which are usually in the range of mega to giga samples per second via IoT networks, is almost mission impossible. An alternative strategy is to push the processing into the the general processor (for videos) or dedicated designed circuit (for radar) on local nodes (edge) and only forward the post-processing information to data centre or cloud. This strategy reduces the IoT networks load, while increases the complexity and cost of the sensor node. Furthermore, the images of the camera often generate privacy issues when deployed in residential environment. From above statement, we can see that there is still large space to seek IoT friendly solution for daily activity monitoring in the healthcare context.

Therefore, we propose an ultra-low resolution infrared sensor-based human body gesture capturing system and the corresponding sparse feature analysis methods to prove the feasibility of a high performance and IoT-

friendly activity monitoring solution. The ultra-low resolution image comes from an infrared sensor array which provides 64 thermal pixels (can be equivalent to degree of gray) in 8×8 two dimension layout (an example image can be seen in Figure 1). Compared to the video stream with RGB information, the processing time and transmission cost of 64 pixels can be ignored. Compared to the RF sensing, which relies on complex algorithms for target detection and movement estimation [8] using the raw I/Q samples, the low-resolution IR images are easier to use; the target and its movement are visualised directly in the temporal sequence of high temperature pixels. The low-resolution IR image provides two obvious features ***i)*** low-rate but rich information data; ***ii)*** simple data for light weight on-device artificial intelligent (AI) that promote the activity monitoring in healthcare IoT. There are some pilot studies [9], [10], [11] on low resolution images for activity monitoring in healthcare projects however, these works have not dived into the detailed features of this new type of data. Also, literature [12], [13] showed the low resolution IR image possible for fall detection and occupation recognition, but not yet extend to generic daily activities. This paper remarks the following three aspects to prove the feasibility of the new sensor:

- **Generic Daily Activities:** This work targets on recognition performance of generic daily activities by using low resolution IR image stream, which are more challenging than the occupation recognition or the "Y/N" fall detection, because of the multinomial classification problem and potential weak data variance between activities.
- **Sparse Feature:** The sparse feature extraction techniques, SDA and SPCA methods are applied to limit the variance of the classification models in high dimensional feature space and also transform the features into a low dimensional more interpretable feature space. The sparse features of this new type of sensor data are verified effectively using four widely used classifiers like support-vector machines (SVM), random forest (RF), k-nearest neighbors (KNN) and logistic regression (LR).
- **System and Real Data:** We design and build a single-sensors proof-of-concept (PoC) system for real time human activity monitoring. Real experimental data including daily activities of single targets are collected and processed with the proposed method to prove the robustness of the recognition results.

The rest of the paper is organized by following: Section II focuses on the system and experiment design; The sparse feature analysis, model selection and classification results are discussed in Section III; We conclude the work in Section IV.



(a) An 8×8 thermal pixel matrix (one frame)

(b) A stream of frames, corresponding an activity

Fig. 1: Example of one snapshot of IR image (a) and low rate IR stream

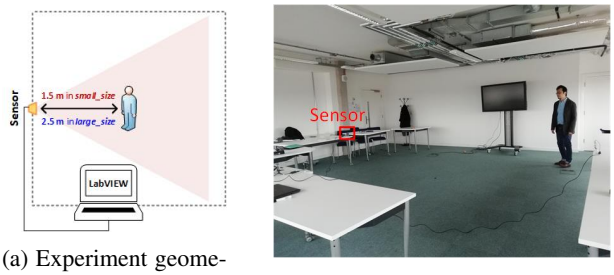
II. Monitoring System Design and Experiment

A. Low Resolution Infrared Sensor and System Design

The low resolution IR array used in this paper is Panasonic Grid-EYE (AMG8832) [14] which comprises 64 IR sensors with 8×8 two dimension square layout. Each IR sensor in Grid-EYE can sense temperature from 0∼80°C with 0.25°C resolution and ±2.5°C deviation (this means maximum $\frac{80}{0.25} = 320$ discrete value readings). The claimed maximum detection distance is 5 meters. Each detected thermal information is presented as a pixel in the 8×8 matrix, and each matrix can be considered as a frame as shown in Figure 1a. For the experimental system, we used NI LabVIEW® which is hosted in a laptop PC to coordinate three Grid-EYE sensors AMG8832EK (evaluation board) via UART ports which is embedded in the ATSAMD21 micro-controller. The AMG8832EK can output 10 frames per second (fps) as a stream shown in Figure1b. 10 bits will be enough to demonstrate one out of 320 discrete temperature values. Then, the data rate of one single Grid-EYE array, will be only $(10\,bits \times 64 \times 10\,fps = 6400)$ bits per second (bit/s). This is suitable for onboard PAN1740 802.15.1 module to transmit, if the sensor is to be used in NB IoT network. Please note that the Grid-EYE is not the only choice, there are also 4×4 IR array [15] and Omron 16×16 array [16] can be used according situations. The premier single sensor data acquiring code is an open source LabVIEW® code [17] from Panasonic®. Our designed LabVIEW software also provides real time data log function and interface for future online data processing. In Section IV, the IR image stream is used for sparse feature analysis and recognition performance evaluation as a proof of the concept.

B. Experiment Description

1) ***Geometry Description****:* For designing the experiments related to activity detection, one of the early required settings is the spatial location of the sensor. Based on series of initial test results, a ***large view_size single-sensor*** test scenario was considered. For this aim,



(a) Experiment geometry

(b) The real experiment venue

Fig. 2: The geometry size and real experiment venue

one sensor was positioned upright 2.5 meters away from the observation area assigned for activities. Moreover, the sensor was elevated roughly 1 meter from the ground. The geometry layout and real experiment venue can be seen in Figure 2.

*2) **Environment Description**:* As the IR sensor is thermal sensitive, the premier factor we considered during the experiment is temperature. It is obvious that lower ambient temperature will lead to a better human body profile imaging, which is often around 36°C. But it has to be compliant with the normal room temperature which is in the range of 20∼24 °C to make the experiment realistic. In our case, the ambient temperature of the surrounding room was measured to be 18°C at the beginning, and rose to 21°C during the time taken to perform the experiment. Clothing is another uncertainty that impacts the result. In this experiment, our participants were dressed with normal household clothes like cotton shirt, knits and jackets as shown in Figure 2 to make the data more realistic.

C. Data Set

The data set is composed of 8 single-subject activities from three participants. Details are shown in Table I. Each activity is performed 10 times by each participant. The number of files captured using the sensor is 240 in total. Each recorded frames timing is relatively within the range of 3∼20 seconds. In order to equalise the number of frames for all activities and subjects, 4 seconds or 40 frames are considered as a base; if the data is more than 4 seconds, extrapolation is used for equalization while in cases where the data is less, interpolation is applied. Then, the final data set is a cube of size (240×40×64), where 240 is the number of observations and 40x64 shows the number of total recorded frames by the number of pixels of the vectorized IR image (8x8). For the analysis, further concatenation of all frames pixels is performed yielding a 2D matrix data of size 240x2560.

III. Data Analysis for Activity Recognition

The choice of methodologies for the classification problem is influenced by the size of the acquired data set

TABLE I: Investigated Activities

Activities		Activities	
A1	Sit Down	A5	Move Left, Right
A2	Stand Still	A6	Move Forward, Backward
A3	Sit Down; Stand Up	A7	Walking Diagonally 1
A4	Stand Up	A8	Walking Diagonally 2

in this paper. Having only $n = 240$ samples compared to $p = 2560$ variables makes an ill-posed problem that might lead to over fitting issue. Therefore, sparse feature extraction strategies are considered to reduce the variance of the model and transform the features into a lower dimensional feature space. Two sparse feature transformation techniques are utilised: the unsupervised SPCA and the supervised SDA feature extraction.

A. Sparse Feature Transformation Techniques

*1) **Sparse Principal Components Analysis (SPCA)**:* PCA is one of the most commonly used data-processing and dimensionality reduction technique. It reduces the dimensionality of data by aiming to retain most of the variance. It transforms interrelated variables to uncorrelated principal components so that, the first component retains the highest variance of the original variables [18]. PCA can be computed via the Singular-Value Decomposition (SVD) of a data matrix. Let the size of X be $n \times p$, where p and n are the numbers of observations and variables, respectively. Then, the SVD of X will be $X = UDV^T$, where U and V are the row and column matrices and D is a diagonal matrix so that, D_{ii}^2/n describes the sample variance of the i^{th} PC. Then, $Z = UD$ is the matrix of principal components (PCs) and the columns of V represent the corresponding loadings of the PCs and are known as eigen vectors.

The main drawback of PCA is that it represents linear combinations of the p variables and the loadings are usually nonzero. SPCA takes advantage of the fact that PCA can be written as a regression-type optimization problem, with a quadratic penalty (see equation (1)). Thus, by applying both L1 (Manhattan distance) and L2 (Euclidean distance) elastic-net constraints, sparse loadings can be achieved which improves the interpretation of the PCs. In other words, the remaining non-zero coefficients of the Eigen vectors correspond to the important variables, having more contribution in data variation. Adding both L1 and L2 constraints removes the lasso (based on L1) limitation about the maximum number of selected variables n. If we have a model with p variables and n samples, the lasso can cancel out maximum n number of variables from the model which is unsatisfactory. For example, in this work, having only 240 samples and 2560 variables, cancelling only the effect of 240 variables in the loadings, remains still $2560 - 240 = 2320$ non-zero variables in each loading that keeps the model variance high and might cause

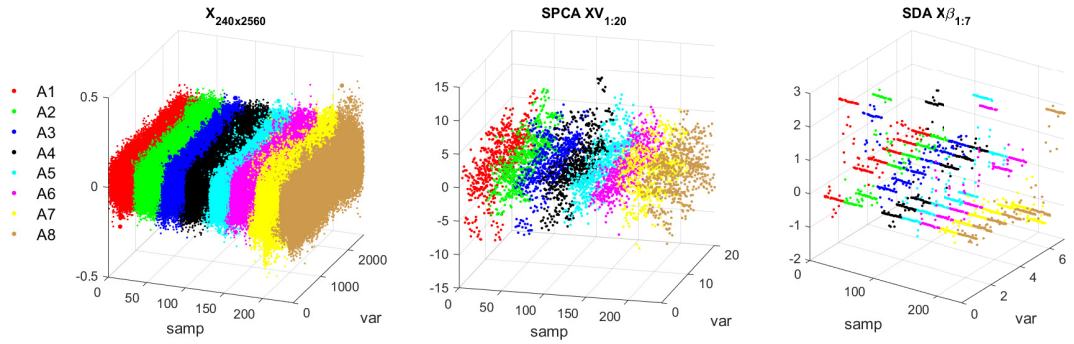
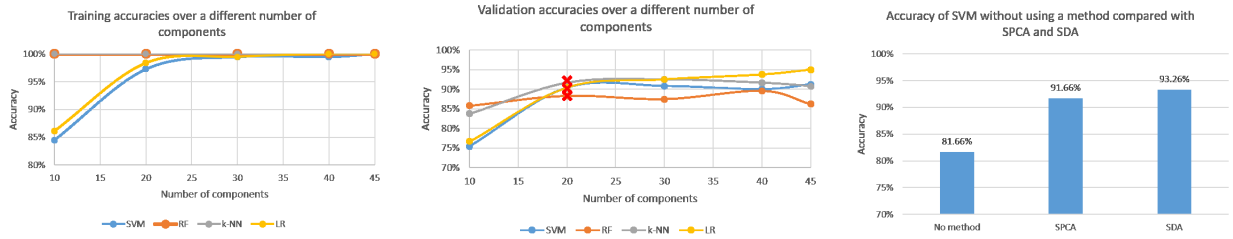


Fig. 3: Visualization of the original data matrix and the transformed samples using SPCA and SDA



(a) Average of training accuracies over 10 folds of CV loop, used for SPCA component selection

(b) Average of validation accuracies over 10 folds of CV loop, used for SPCA component selection

(c) Comparison of SVM accuracy using original features, transformed features based on SPCA and SDA

Fig. 4: Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA

overfitting problem.

Regarding the mathematical description of SPCA, given Σ as the covariance matrix of X, $A = [\alpha_1, \alpha_2, ..., \alpha_p]$, $Y^{**} = \Sigma^{\frac{1}{2}}\alpha_j$ and $X^{**} = \Sigma^{\frac{1}{2}}$, the SPCA objective function can be written as:

$$\hat{V_j} = \arg\min_{V} \|Y^{**} - X^{**}V\|^2 + \lambda \|V\|^2 + \lambda_1 \|V\|_1 \quad (1)$$

where λ and λ_1 are the L1 and L2 regularization parameters and V_j is the j^{th} estimated sparse Eigen vector, $j = 1, ..., p$. This is a bi-convex objective function on A and V that can be solved iteratively. For more information about the method, refer to [19].

Model Selection: Once the sparse Eigen vectors are computed, an appropriate number of them needs to be chosen to achieve the best accuracy. One of the techniques in this case is based on K-fold CV which is a model selection strategy [22]. It is a widely used technique that splits training data into K almost equal size folds and uses a subset of them ($K-1$ folds) for training models. The single remaining fold is considered as validation set and used to evaluate the accuracy (or error) of the model on candidate set(s) of parameters. Then, for all possible K different number of validation sets, this will be repeated, yielding a multidimensional accuracy (error) map. By averaging this map over the K iterations results, the optimal model can be identified; that corresponds to the maximum average performance or minimum error. The dimensionality of the average map depends on the number of candidate parameters. The desired parameter in this work, is the number of SPCA PCs. Therefor, a one dimensional average validation map is formed. The average training accuracy map is usually compared to the validation map to ensure no over-fitting problem occurs resulting in a significant difference between the training and validation performance (error). The mathematical description for the average estimate of CV performance is as follows:

$$CV(\hat{f}, c) = \frac{1}{N}\sum_{i=1}^{N} Prf(y_i \hat{f}^{-k(i)}(x_i, c)) \quad (2)$$

where, $k : \{1, ..., N\} \longmapsto \{1, ..., K\}$ is the indexing function that indicates the fold to which observation i belongs to and, $\hat{f}^{-k}(x, c)$ is the fitted function, when the k^{th} fold is removed and the c^{th} candidate parameter is used. The $CV(\hat{f}, c)$ function gives an estimate of the performance over the range of candidate parameters and helps to find the optimum parameter $\hat{c}$ for the model.

*2) **Sparse Discriminant Analysis (SDA)**:* Linear Discriminant Analysis (LDA) is one of the important supervised linear classification and dimensionality reduction techniques. Considering a classification problem with Q classes, LDA assumes a Gaussian distribution for the samples of each class to find $Q-1$ discriminative linear hyperplanes. Using the Reduced Rank LDA framework, it is also possible to transform data into a

lower dimensional feature space of size $Q-1$. This can be achieved based on the Fisher's discriminant analysis using Rayleigh quotient criteria as shown in equation 3. The Fisher's *canonical variables* or *discriminant coordinates* are used to transform the feature into a lower dimensional feature space [20]. Fisher's problem maximizes the Rayleigh quotient:

$$\max_{a} \frac{a^T B a}{a^T W a} \tag{3}$$

where W is the within-class covariance matrix and B is the covariance matrix of the class centroid matrix M, also referred as Between-class covariance matrix. a is the canonical variable. The aim of this method is to increase the distances between the class means as much as possible in order to improve the separability.

However, in $n \ll p$ condition, LDA cannot be applied directly because the within-class covariance matrix W of the features is singular. The SDA algorithm uses an L1 norm constraint similar to lasso objective function to compute sparse transformation factors β_q, $q = 1, 2, ..., Q-1$ to transform the original data X to a new low dimensional space $X\beta_{1:m}, m \leq Q-1$. The notation $1:m$ shows selection of the first m vectors. Each β_q vector length is 2560. The objective function is based on optimal scoring as shown below:

$$\min_{\beta_q} \left\{ \frac{1}{n} \left\| Y\theta_q - X\beta_q \right\|^2 + \gamma \beta_q^T \Omega \beta_q + \lambda \left\| \beta_q \right\|_1 \right\} \tag{4}$$

where, Y denotes an n×Q matrix of dummy variables for the Q classes; Y_{iq} is an indicator variable showing whether the i^{th} observation belongs to the q^{th} class. θ_q is a Q-vector of scores. This sparse matrix is used for transformation of the data matrix into the lower dimensional space $Z = X\beta_{1:m}$, $m \leq Q-1$. Ω is a positive definite matrix added to compensate for singularity of the with-in class covariance W. λ and γ are non-negative tuning parameters. More information about the algorithm can be found in [19].

B. Classification

To test the quality of SPCA and SDA features of low resolution IR image stream, four commonly used classifiers are applied to conduct the recognition task; ***i)*** SVM trained using Stochastic Gradient Descent (SGD); ***ii)*** LR; ***iii)*** KNN based on 4 neighbours; and iv) 80 36-depth trees RF.

C. Numeric Results

The data set was divided randomly into training (75%) and test (25%) for experiments; $X_{tr\_180\times2560}$ and $X_{ts\_60\times2560}$.

*1) **SPCA**:* Using the model selection strategy explained before, the optimum number of PCs $\hat{c}$ is found based on 10 fold CV. The candidate sets of number of components is $[10, 20, 30, 40, 45]$. One-dimensional training and validation accuracy maps are computed. They are shown in Figures (4a) and (4b) for the four classification methods. Then, based on the validation results, 20 PCs are selected for transforming data, since the first prominent peak corresponds to 20 components in validation map. Then, the transformation to the reduced space is performed based on $Z_{tr} = X_{tr}V_{1:\hat{c}}$, $\hat{c} = 20$. The $1:\hat{c}$ notation shows the first 20 Eigen vectors. This is a significant reduction in the dimensionality of data. The test data is also transformed similarly.

*2) **SDA**:* Using SDA a sparse $\beta_{2560\times7}$ is obtained and used for transforming data $Z_{tr} = X_{tr}\beta$.

*3) **Classification Analysis**:* To explore the effectiveness of the feature transformation strategies, first, the original data of the 8 different classes and their corresponding transformed samples using SPCA and SDA are visualised in Figure 3. The general variation range of the samples as well as the variations between different classes has increased by transformation (please note the z axis range). This helps to improve the classification accuracy. The comparison of the classification accuracies shown in Figure 4c confirms this. As shown, using the sparse feature extraction methods, the accuracy improves considerably.

The four classification techniques have been applied on the training and test sets. In regards to training set, the model's accuracy is 100% for all methods, and then it is compared to the results on the unseen data. The accuracy on both training and testing sets can be observed in figure 5.

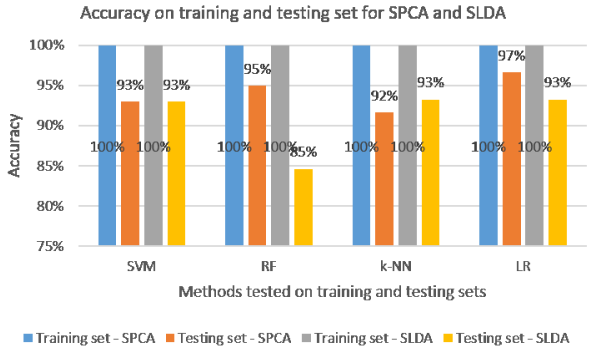


Fig. 5: Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA

D. Discussion

Figure 6 shows the absolute values of the first SDA sparse component β_1 and first SPCA Eigen vector V_1. While the level of sparsity is not the same for the two vectors, the achieved training and test accuracies

shown in figure 5 are mostly in the same range for the two techniques. Furthermore, the largest non-zero coefficients over the 40 frames or 4 seconds of activity, are mainly distributed in the central frames, showing in most activities, middle frames contribute in classification more significantly compared to the two extremes. This matches to the fact that in most early and last frames, the subjects are less active.

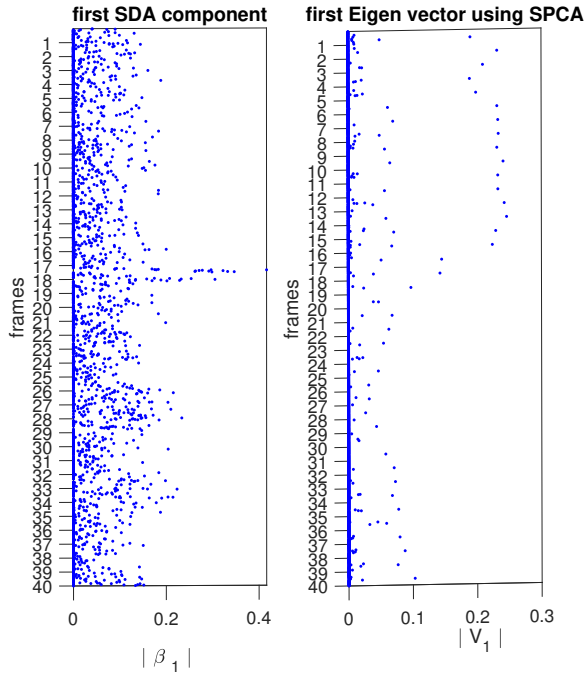


Fig. 6: Visualization of the first sparse component of SDA β_1 and first sparse Eigen vector of SPCA V_1

IV. Conclusion

In this paper, we proposed a new type of sensor and sparse feature selection and transformation techniques for human activity monitoring in the healthcare IoT context. Using the sparse feature transformation techniques have proved high classification accuracy, as seen in the Section III. The sparse directions found by SDA and SPCA can help to find essential frames for activity recognition. Experimental results demonstrate that by applying sparse feature transformation techniques on ultra-low resolution IR images time-series, robust recognition models can be developed. Such models are suitable for narrow band transmission and potential light-weight on-board machine learning. Based on the current work, we still need to examine the following tasks in the future work: the strategy for the system to benefit from multiple sensors, feature extraction method on distributed sensor nodes and machine learning strategy for distributed sensors.

References

[1] Z. Bajorek, A. Hind, and S. Bevan. "The impact of long term conditions on employment and the wider UK economy". Work Foundation, 2016.

[2] G. Iacobucci. Nhs in 2017: Keeping pace with soci- ety. BMJ: British Medical Journal (Online), 356, 2017.

[3] N. Zhu, T. Diethe, M. Camplani, L. Tao, A. Burrows, N. Twomey, D. Kaleshi, M. Mirmehdi, P. Flach, and I. Craddock. Bridging e-health and the internet of things: The sphere project. IEEE Intelligent Systems, 30(4):39–46, 2015.

[4] Y. Jing, M. Eastwood, B. Tan, A. Konios, A. Hamid, and M. Collinson. "An intelligent well-being monitoring system for residents in extra care homes". In Proceedings of the 1st International Conference on Internet of Things and Machine Learning. ACM, New York, NY, USA, 2017.

[5] M. Eastwood, A. Konios, B. Tan, Y. Jing, A. Hamid, "Conditional Random Field Feature Generation of Smart Home Sensor Data using Random Forests", IEEE The International Microwave Biomedical Conference, Nanjing, China, 2019

[6] 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), pp. 609-614, Milan, 2015.

[7] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li and R. Piechocki, "Exploiting WiFi Channel State Information for Residential Healthcare Informatics," in IEEE Communications Magazine, vol. 56, no. 5, pp. 130-137, May 2018.

[8] B. Tan, K. Woodbridge and K. Chetty, "A real-time high resolution passive WiFi Doppler-radar and its applications," 2014 International Radar Conference, Lille, pp. 1-6, 2014.

[9] Y. Karayaneva, S. Baker, B. Tan, and Y. Jing. "Use of low-resolution infrared pixel array for passive human motion movement and recognition". In Proceedings of the 32nd International BCS Human Computer Interaction Conference. UK, Article 143, 5 pages, 2018

[10] L. Tao, T. Volonakis, B. Tan, Y. Jing, K. Chetty, M. Smith. "Home Activity Monitoring using Low Resolution Infrared Sensor", arXiv preprint arXiv:1811.05416, Nov 13, 2018

[11] L. Tao, T. Volonakis, B. Tan; Z. Zhang; Y. Jing; M. Smith; "3D Convolutional Neural network for Home Monitoring using Low Resolution Thermal-sensor Array", IET TechAAL, 2019

[12] S. Mashiyama, J. Hong and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication, pp. 2109-2113, Washington, DC, 2014.

[13] S. Savazzi, V. Rampa, S. Kianoush, A. Minora and L. Costa, "Occupancy Pattern Recognition with Infrared Array Sensors: A Bayesian Approach to Multi-body Tracking, IEEE International Conference on Acoustics, Speech and Signal Processing, United Kingdom, pp. 4479-4483, 2019.

[14] Panasonic. Infrared Array Sensor Grid-EYE [Online]. Available: http://pewa.panasonic.com/assets/pcsd/catalog/grid-eye-catalog.pdf

[15] P. Wojtczuk, A. Armitage, T. Binnie, and T. Chamberlain. Pir sensor array for hand motion recognition. In Proc. 2nd Int. Conf. on Sensor Device Technologies and Applications, pages 99–102, 2011.

[16] Omron. OMRON Develops the World's First 16x16 Element MEMS Non-Contact Thermal Sensor for Use in Human Presence Sensors Utilizing Wafer-Level Vacuum Packaging Technology [Online]. Available: https://www.omron.com/media/press/2013/05/e0529.html

[17] Panasonic. Labview Software Source Code [Online]. Available: https://eu.industrial.panasonic.com/sites/default/pidseu/files/labview_0.zip

[18] L.I. Smith, "A tutorial on principal components analysis." Cornell University. 51, 52. USA 2002

[19] H. Zou, T. Hastie, R. Tibshirani, "Sparse Principal Component Analysis," Journal of Computational and Graphical Statistics, 15:2, 265-286, 2006

[20] T. Hastie, R. Tibshirani, J. H. Friedman, "The elements of statistical learning: data mining, inference, and prediction". 2nd ED, New York: Springer. 2009

[21] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse discriminant analysis," Technometrics, 53(4): 406-413, 2011
[22] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction," 2nd ED. Springer, pp.241-244., 2001