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—In this paper, we propose an ultra-low-resolution infrared (IR) images based activity recognition method which is suitable for monitoring in elderly care-house and modern smart home. The focus is on the analysis of sequences of IR frames, including single subject doing daily activities. The pixels are considered as independent variables because of the lacking of spatial dependencies between pixels in the ultra-low resolution image. Therefore, our analysis is based on the temporal variation of the pixels in vectorised sequences of several IR frames, which results in a high dimensional feature space and an "n≪p" problem. Two different sparse analysis strategies are used and compared: Sparse Discriminant Analysis (SDA) and Sparse Principal Component Analysis (SPCA). The extracted sparse features are tested with four widely used classifiers: Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbours (KNN) and Logistic Regression (LR). To prove the availability of the sparse features, we also compare the classification results of the noisy data based sparse features and non-sparse based features respectively. The comparison shows the superiority of sparse methods in terms of noise tolerance and accuracy.

I. INTRODUCTION

The impact of activity detection in health sector such as elderly care homes is quite high. For example, 15.4 million people in England suffer from one or more chronic medical conditions including dementia, stroke, cardiovascular or musculoskeletal disease [1]. The current treatment cost is 70% of the UK National Health Service (NHS) budget [2]. The cost of continuous management and medical treatment for long term health conditions in clinical and residential environments has become an increasing economic pressure. Therefore, novel solutions for out-patient long-term monitoring in residential settings are urged [3].

Current popular activity monitoring solutions include camera, wearable sensor, radio frequency (RF) sensors [4]. The camera-based activity recognition systems allow in-depth analysis and modeling of the activity due to the high resolution images. However, the highly varied types of activity in real world is a challenge for generalisation of the predictive models. Furthermore, the used of camera-based systems in home environments can be intrusive due to the privacy concerns. The wearable sensors offer low cost activity monitoring solutions while, accompanying with low battery life and unwillingness-to-wear problem [5]. Also, wearable sensors have diffi-

culty to differentiate between activities which are similar. RF based sensors shows the advantage on non-intrusive, but suffer from the interference, complicated signal processing and geometry dependence. Thus, we introduce ultra-low resolution infrared images which provide non-intrusive and privacy preserving activity recognition solution. For this aim, a single-sensors proof-of-concept (PoC) system for real time human activity recognition is designed and built. The ultra-low resolution images are formed based on an infrared sensor array structure, consisting of only 64 thermal pixels in 8×8 two dimension layout (an example image can be seen in Figure (2). This low-rate rich information data is specially suitable for light weight on-device artificial intelligent (AI) with great impact of activity monitoring in artificial intelligent (AI) aided healthcare.

Compared to the RGB camera images, their spatial resolution is much lower. This influences the analysis strategies for activity detection. From a methodological perspective, activity recognition techniques can be categorised into two main groups [6]; the manual (hand-crafted) feature extraction followed by classification and the Neural Networks (NN)-based techniques. A comprehensive review of these techniques are presented in II. In these works, the analysis is based on the image arrays and the spatial proximity of the neighbouring pixels are considered [5], [7], [8]. Due to the characteristic of low spatial dependency of the pixels in ultra-low-resolution IR images, they are considered as independent variables in this paper. This is similar to [9]. Therefore, the image arrays are vectorised and the image vectors of video streams for each activity are concatenated. This results in high dimensional variable space. So that, compared to the number of samples n, the number of variables p are quite high "n≪p". Another characteristic of the low resolution IR sensor is the thermal working range, that is in the range 20∼24°C and the rise in temperature adds noise to the pixel values. These two characteristics make sparse discrimination strategies ideal for activity recognition.

In this paper, two sparse feature extraction techniques, Sparse Discriminant Analysis (SDA) and Sparse Principal Component Analysis (SPCA) are employed to limit the variance of the classification models in high dimensional feature space in high temperatures, 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: support-vector machines (SVM), random forest (RF), k-nearest neighbors (KNN) and logistic regression (LR). In addition, the sparse features are analysed with noised data. A comparison of the recognition accuracy is performed between each of the SDA and SPCA techniques and their corresponding non-sparse feature extraction methods, Reduced Rank LDA (RRLDA) [10] and Principal Component Analysis (PCA). Reviewing the existing pilot studies [9], [11], [12] on low resolution images for activity monitoring in healthcare projects however, no detailed analysis of the features of this new type of data based on sparse techniques and their tolerance to noise was found.

The rest of the paper is organized by following: Section II reviews the previous studies on activity recognition using camera images and low resolution IR sensors. Section III explains the materials and methods which includes a brief explanation of data acquisition, the sparse feature extraction, model selection and classification techniques as well as noise analysis are described. The experimental results are presented in section IV and a conclusion is provided in Section V.

II. RELATED WORK

A. Handcrafted Features

In this section the analysis strategies for vision-based activity detection are reviewed. In the case of camera images, one group of early methods for action recognition was based on manual (handcrafted) action representations [6]. Many of these types of methods were developed before the emergence of deep learning.

1) Statistical Features: The common approach was extraction of local statistics from both spatial saliences and action motions and then combining them over sequence of frames (video) and finally applying discriminative classifiers such as SVM [13]. For example, the local image features or interest points over temporal sequences of video frames were used to detect local structures in space-time [14]. The Histogram of Oriented Gradient (HOG) was used in [15], as appearance descriptors for human detection in static images together with motion features extracted by oriented histograms of various kinds of local differences or differentials of optical flow. A mixture of Gaussians together with Kalman filters were used in [16], [17].

2) Use of Depth Information: The emergence and popularization of depth cameras allowed the use of depth images and added the pose information to the analysis. Examples of such techniques are the use of single depth images for recognition of human body pose [18] and the use of RGBD images that provides both color and depth information together with deep neural networks for human motion recognition [19].

3) Trajectory Based Features: Trajectory based action representation also proved great success using manual shallow representations. The trajectories were obtained by tracking densely sampled points using optical flow and features such as HOG, Histogram of Optical Flow (HOF) and motion boundary histogram (MBH) [20] or multi-layer stacked Fisher vector (FV). [21] helped improve the human action recognition. In contrast to the above methods, global approaches consider the whole action video as the framework input and perform pooling or filtering on the video pixel level [22], [23].

The main limitation of the effective manual strategies is the high computational complexity at both training and test step due to the use of local features requiring densely sample strategy which is an issue for real-time applications [6].

B. Learning-Based Features

Another category of action recognition approach is the Learning-based action recognition. In contrast to the manual action features that normally operate at video pixel levels and low-level statistics in spatial body shapes or temporal motions, these advanced approaches work on top of manual action features or establish end-to-end action recognition frameworks from the pixel-level to action categories.

1) Classic Learning Approaches: Some learning-based approaches utilise advance modelling strategies such as Genetic Programming (GP). For example in [24], a population of primitive 3D operators, such as 3D-Gabor filter and wavelet coefficients were utilised to evolve the motion features from both colors and optical flow fields. Then, evolutions are performed over this population using either crossover or mutation strategies through reproduction with single or pair parents. TextonBoost is another strong object detection algorithm. In this technique, joint models for texture, layout, and context are used in a conditional random field framework so that a strong classifier is formed using a group of weak classifiers [25].

In Dictionary learning-based methods sparse representations of the input data is learned in the form of a linear combination of basis dictionary atoms. The bag-of-words (BoW) model which is a particular type of dictionary learning uses a combination of basis dictionary atoms instead so that, each input is assigned to a single basis dictionary atom. In [26], using BoW with sparse representation improved the discrimination power.

2) Neural Networks Approaches: A major group of activity recognition strategies are based on Neural Networks (NN). Many of the early NN-based techniques were formed at discrimination and modeling step and the features were learned from manually extracted features [27], [28]. On the other hand, Deep NN (DNN) strategies perform convolution at spatio-temporal dimensions at pixel level. Then unlike manual feature extraction,

DNN does intellectual leaning and contains hierarchical feature extraction layers that contain much more trainable parameters than manual architectures. However, they require more training data. Some DNN based strategies such as [29] were based on an unsupervised approach for learning spatio-temporal features using a Gate Restricted Boltzmann Machine (GRBM)-based convolutional architecture. Other supervised strategies, employed 3D Convolutional Neural Networks (CNN) along spatio-temporal dimensions [30] or capture the features' temporal evolution over time by adapting CNN to sequential data and form Recurrent NN (RNN) [31], [32] with one hidden layers of Long Short-Term Memory (LSTM) [33].

The above explained activity detection methods are based on camera systems and streams of 2D or 3D images in higher resolution than IR images used in this paper. The ultra low-resolution IR images were used in some works. Researchers have applied SVM, RF and KNN on low-resolution raw pixels data and achieved 84.2% overall accuracy [12]. In [34], a Y/N fall detection strategy was developed based on a KNN framework using spatio-temporal pixel level statistics. A Bayesian framework was developed for multi-body tracking in [35]. In another recent contribution, DNN was applied on raw vectorised pixels [9].

III. MATERIALS AND METHODS

The methodology process as seen in figure 1 starts with data acquisition from the IR sensor. Then, data preparation is performed in regards to frames equalisation. The next step is to select the number of SPCA and SDA components based on a 10-fold cross-validation. Finally, sparse feature extraction and classification with four methods are performed.
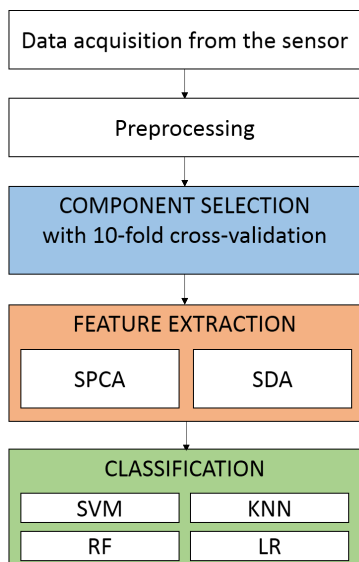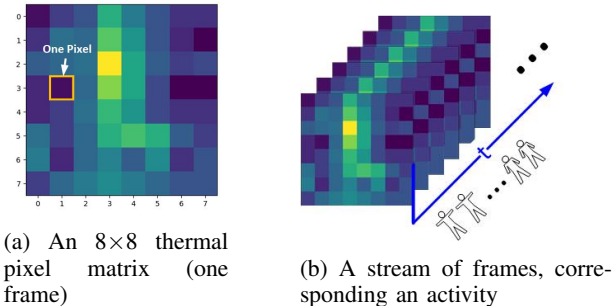

Fig. 1: Methodology flowchart

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

(b) A stream of frames, corresponding an activity

Fig. 2: Example of one snapshot of IR image (a) and low rate IR stream

A. Data Acquisition

The low-resolution IR array used in this paper was acquired from Panasonic Grid-EYE(AMG8832) [36] 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. 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 2a. The system can output 10 frames per second (fps) as a stream shown in Figure2b.

Regarding the spatial location of the sensor field of view, a large view setting was considered so that, 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 3.

The IR sensor is thermal sensitive. The background temperature is generally less than human body. In image profiles, the higher pixel values correspond to higher temperature. Then, a lower environmental temperature will lead to a better human body profile imaging, which is often around 36°C. 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 3 to make the data more realistic.

B. Data Preparation

The data set is composed of 8 single-subject activities from three participants. Details are shown in Table I. The segmentation of individual activities frames from total streams can be performed in different ways such as background subtraction. However, in this paper the focus is on recognition of individual activities rather than

(a) Experiment geometry
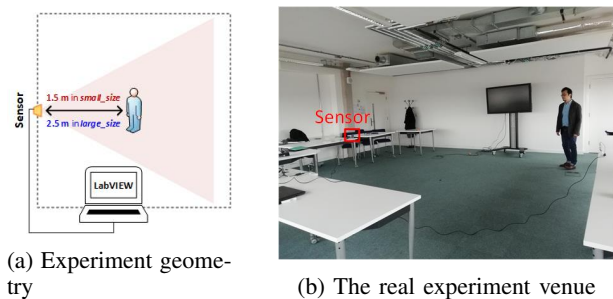

(b) The real experiment venue

Fig. 3: The geometry size and real experiment venue

analysis of continuous video streams. Therefore, each activity is recorded separately. It 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 ($240 \times 40 \times 64$), where 240 is the number of observations and 40x64 shows the number of total recorded frames by the number of pixels of the vectorised IR image (8×8). For the analysis, further concatenation of all frames pixels is performed yielding a 2D matrix data of size 240×2560. Then, having only

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

$n = 240$ samples compared to $p = 2560$ variables makes an ill-posed problem that might lead to overfitting 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.

C. Sparse Feature Transformation Techniques

1) Sparse Principal Components Analysis (SPCA): PCA is one of the most commonly used data-processing and dimensionality reduction techniques. 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 [37].

The main drawback of PCA is that it represents linear combinations of the p variables and the loading 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 loading can be achieved. This improves the interpretation of the PCs and solves the overfitting issues. In other words, the remaining non-zero coefficients of the Eigen vectors correspond to the important variables, having more contribution in data variation.

Let the size of X be $n \times p$, where p and n are the numbers of observations and variables, respectively. 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 jth 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 [38]. Once the sparse Eigen vectors are computed, an appropriate number of them needs to be chosen to achieve the best accuracy. K-fold CV is used as a model selection [39] strategy to find the optimum number of SPCA PCs in this paper.

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 RRLDA 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 2. The Fisher's *canonical variables* or *discriminant coordinates* are used to transform the feature into a lower dimensional feature space [10]. Fisher's problem maximizes the Rayleigh quotient:

$$\max_a \frac{a^T B a}{a^T W a} \quad (2)$$

where W is the within-class co-variance matrix and B is the co-variance matrix of the class centroid matrix M, also referred as Between-class co-variance 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 co-variance 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.
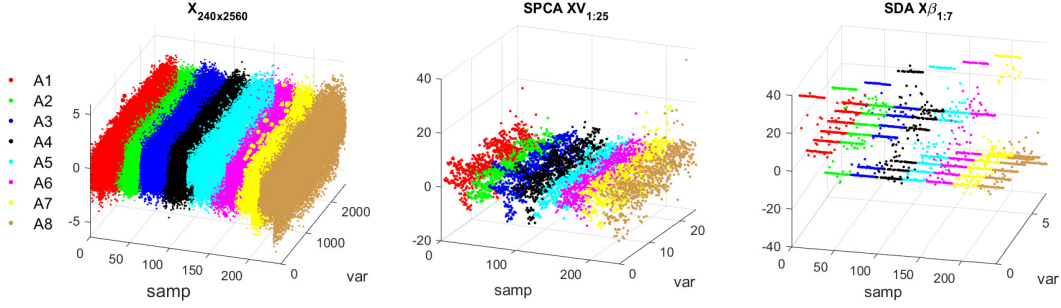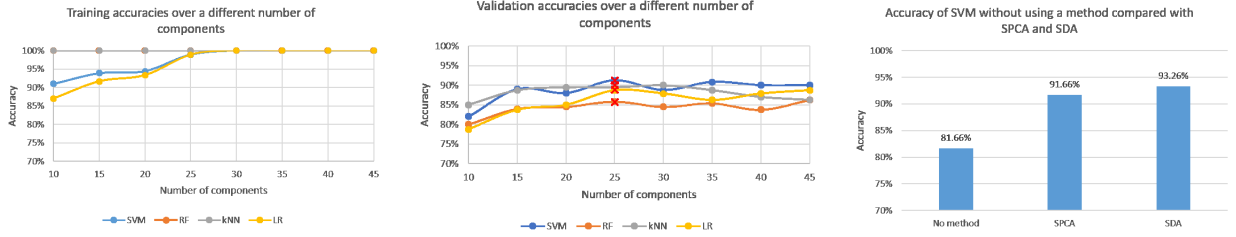
Fig. 4: Visualization of the original data matrix and the transformed samples using SPCA and SDA

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

(b) Average of validation accuracy 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. 5: Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA

Each β_q vector length is 2560. The objective function is based on optimal scoring as shown below:

$$\min_{\beta_q} \left\{ \frac{1}{n} \left\| Y\theta_q - X\beta_q \right\|^2 + \gamma \beta_q^T \Omega \beta_q + \lambda \left\| \beta_q \right\|_1 \right\} \quad (3)$$

where, Y denotes an n×Q matrix of dummy variables for the Q classes; Y_{iq} is an indicator variable showing whether the ith observation belongs to the qth 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 [40].

D. Classification

In order 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.

IV. EXPERIMENTAL RESULTS

The data was randomly divided into training (75%) and testing (25%) for experiments; $X_{tr\_180\times2560}$ and $X_{ts\_60\times2560}$. That ensures that the population samples are split fairly in terms of their variation between training

and testing sets. The number of components selected for SPCA based on K-fold CV is 25. In regards to SDA, the maximum number of components $K - 1$ where K is the number of classes, is retained. The four machine learning methods SVM, RF, k-NN and LR are used on the transformed matrices extracted from the two feature extraction methods. Next, they are tested again on the data with added noise in order to evaluate the noise tolerance of sparse techniques.

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 $[10, 15, 20, 25, 30, 35, 40, 45]$. One-dimensional training and validation accuracy maps are computed. They are shown in Figures (5a) and (5b) for the four classification methods. Then, based on the validation results, 25 PCs are selected for transforming data, since the first prominent peak corresponds to 25 components in the validation map. Then, the transformation to the reduced space is performed based on $Z_{tr} = X_{tr}V_{1:\hat{c}}$, $\hat{c} = 25$. The $1 : \hat{c}$ notation shows the first 25 Eigen vectors. This is a significant reduction in the dimensionality of data. The testing 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 cor-

responding transformed samples using SPCA and SDA are visualised in Figure 4. 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 improve the classification accuracy. The comparison of the classification accuracy shown in Figure 5c 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 testing sets. In regards to the 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 6. As SDA is a supervised feature extraction method, it shows better accuracy than the unsupervised SPCA with the exception of RF classifier only.
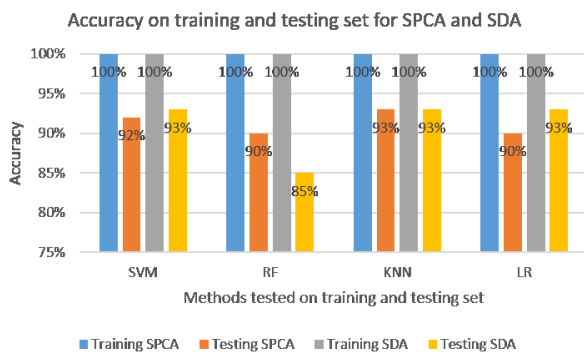
Fig. 6: Accuracy of SVM, RF, KNN and LR on both training and testing sets for SPCA and SDA

4) Noise Analysis: In order to evaluate the tolerance of the trained models to noise, randomly distributed positive noise with $\mu = 0$ and $\sigma = 0.1, 0.5, 1$ is added to the testing data and the classification performance for the testing data is computed. The performance is compared for the sparse feature extraction methods and their corresponding non-sparse techniques. So that, testing data classification using RRLDA and SDA as well as PCA and SPCA are compared. As shown in figure 7, the RRLDA performance degrades more considerably compared to SDA for higher values of noise. That is also the case when comparing PCA and SPCA. It also shows that unsupervised PCA strategy is more tolerant to noise compared to the supervised strategy based on RRLDA.

A. Discussion

Figure 8 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 testing accuracies shown in figure 6 are mostly in the same range for the two techniques. Furthermore, the largest non-zero
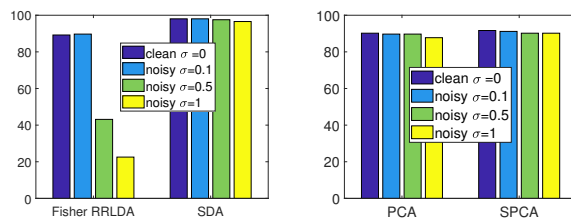
Fig. 7: Comparison of the clean and noisy test data classification performances for RRLDA and SDA as well as PCA and SPCA

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.

In regards to the limitations of the work, a more larger set of subjects can be further included in the dataset. This will ensure that the model can generalise between a diverse population of people.
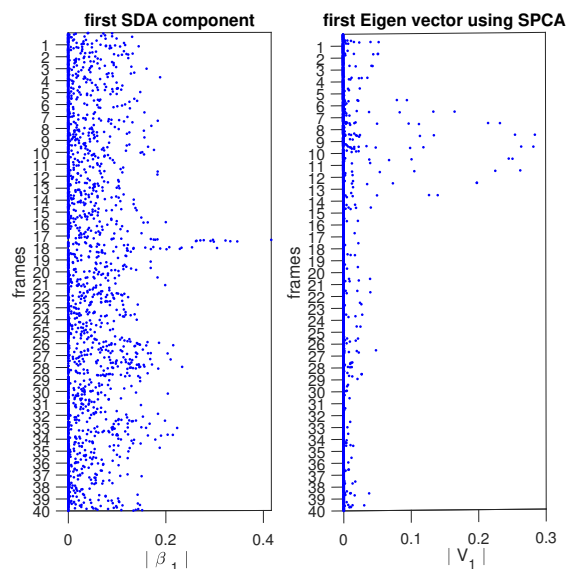
Fig. 8: Visualization of the first sparse component of SDA β_1 and first sparse Eigen vector of SPCA V_1

V. CONCLUSION

In this paper, activity detection for ultra-low-resolution IR images of 8×8 is performed using sparse feature extraction techniques, SDA and SPCA. Such low-resolution images are suitable for human activity monitoring when the privacy-preserving and the non-intrusive features are concerned. The results demonstrated that using the sparse feature transformation techniques achieved higher classification accuracy and better tolerance to noise compared to non-sparse strategies. To further prove

the applicability of the low-resolution IR image in healthcare applications, the factors like multiple targets, multiple perceiving angles will be considered with larger subjects size in our future work.

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 society". BMJ: British Medical Journal (Online), 356, 2017.

[3] N. Zhu, T. Diethe, M. Camplani, L. Tao, A. Burrows, N. Twomey, D. Kaleshi, P. Mirmehdi, P. Flach, and I. Craddock, "Bridging e-health and the internet of things: The sphere project". IEEE Intelligent Systems, vol. 30, no. 4, pp. 39-46, 2015.

[4] 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.

[5] L. Chen, C. Nugent, "Ontology-based activity recognition in intelligent pervasive environments" Int. J. Web Inf. Syst. pp. 410-430, 2009.

[6] F. Zhu, L. Shao, J. Xie, and Y. Fang, "From handcrafted to learned representations for human action recognition: a survey" Image and Vision Computing, 2016.

[7] S. Ranasinghe, F. Al Machot and C. Mayr, H., "A review on applications of activity recognition systems with regard to performance and evaluation." International Journal of Distributed Sensor Networks, 2016.

[8] P. Wang, W. Li, Ph. Ogunbona, J. Wan, S. Escalera, "RGB-D-based Human Motion Recognition with Deep Learning: A Survey", Computer Vision and Image Understanding, pp. 118-139, 2018.

[9] 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.

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

[11] 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, November 2018.

[12] 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.

[13] C.-C. Chang, C.-J. Lin, "LIBSVM: a library for support vector machines", ACM Trans. Intell. Syst. Technol., vol. 2, Article No. 27, 2011.

[14] I. Laptev, "On space-time interest points", International Journal of Computer Vision, vol. 64, no. 2-3, pp. 107-123, 2005.

[15] N. Dalal, B. Triggs, C. Schmid, "Human Detection Using Oriented Histograms of Flow and Appearance." European Conference on Computer Vision (ECCV '06), Graz, Austria, May 2006.

[16] K. Wongpatikaseree, M. Ikeda, M. Buranarach, T. Supnithi, A. O. Lim and Y. Tan, "Activity recognition using context-aware infrastructure ontology in smart home domain," 2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems, Melbourne, VIC, 2012.

[17] L. Fiore, D. Fehr, R. Bodor, A. Drenner, G. Somasundaram, and N. Papanikolopoulos. Multi-camera human activity monitoring, Journal of Intelligent and Robotic Systems, vol.52, no.1, 2008.

[18] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore, "Real-time human pose recognition in parts from single depth images", Communications ACM, vol. 56, no.1, pp. 116-124, 2013.

[19] P. Wang, W. Li, P. Ogunbona, "Rgb-d based human motion recognition with deep learning: A survey", Computer Vision and Image Understanding (CVIU), 2018.

[20] H. Wang, C. Schmid, "Action recognition with improved trajectories", IEEE International Conference on Computer Vision, 2013.

[21] X. Peng, C. Zou, Y. Qiao, Q. Peng, "Action recognition with stacked Fisher vectors", European Conference on Computer Vision, Springer, pp. 581-595, 2014.

[22] X. Zhen, L. Shao, X. Li, "Action recognition by spatio-temporal oriented energies", Information sciences, vol. 281, pp. 295-309, 2014.

[23] L. Shao, X. Zhen, D. Tao, X. Li, "Spatio-temporal Laplacian pyramid coding for action recognition", IEEE Transactions on Cybernetics, vol. 44, no. 6, pp. 817-827, 2014.

[24] L. Shao, L. Liu, X. Li, "Feature learning for image classification via multi objective genetic programming", IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 7, pp. 1359-1371, 2014.

[25] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," IJCV, vol. 81, no. 1, 2009.

[26] T. Guha, R.K. Ward, "Learning sparse representations for human action recognition", Pattern Analysis and Machine Intelligence, Transaction on IEEE, vol. 34, no. 8, pp. 1576-1588, 2012.

[27] H.-J. Kim, J.S. Lee, H.-S. Yang, "Human action recognition using a modified convolutional neural network", Advances in Neural Information Processing Systems, Springer, 2007.

[28] H. Jhuang, T. Serre, L. Wolf, T. Poggio, "A biologically inspired system for action recognition", IEEE International Conference on Computer Vision, 2007.

[29] G.W. Taylor, R. Fergus, Y. LeCun, C. Bregler, "Convolutional learning of spatio-temporal features", European Conference on Computer Vision, Springer 2010.

[30] S. Ji, W. Xu, M. Yang, K. Yu, "3D convolutional neural networks for human action recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, 2013.

[31] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition", IEEE Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 855-868, 2009.

[32] Y. Du, W. Wang, L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[33] F.A. Gers, N.N. Schraudolph, J. Schmidhuber, "Learning precise timing with LSTM recurrent networks", Journal of Machine Learning Research, vol. 3, no.1, pp. 115-143, 2003.

[34] 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.

[35] 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.

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

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

[38] H. Zou, T. Hastie, R. Tibshirani, "Sparse Principal Component Analysis," Journal of Computational and Graphical Statistics, vol. 15, no. 2, pp. 265-286, 2006.

[39] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction," 2nd ED. Springer, pp. 241-244., 2001.

[40] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse discriminant analysis," Technometrics, vol. 53, no. 4, pp. 406-413, 2011.