Tensor Factorisation and Transfer Learning for Sleep Pose Detection

Sara Mahvash Mohammadi[a], Samaneh Kouchaki[b], Saeid Sanei[c], Derk-Jan Dijk[d], Adrian Hilton[a] and Kevin Wells[a]

[a]Faculty of Engineering and Physical Science, University of Surrey, Guildford, UK
[b]Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[c]School of Science & Technology, Nottingham Trent University, Nottingham, UK
[d]Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

***Abstract*—In this study, a novel hybrid tensor factorisation and deep learning approach has been proposed and implemented for sleep pose identification and classification of twelve different sleep postures. We have applied tensor factorisation to infrared (IR) images of 10 subjects to extract group-level data patterns, undertake dimensionality reduction and reduce occlusion for IR images. Pre-trained VGG-19 neural network has been used to predict the sleep poses under the blanket. Finally, we compared our results with those without the factorisation stage and with CNN network. Our new pose detection method outperformed the methods solely based on VGG-19 and 4-layer CNN network. The average accuracy for 10 volunteers increased from 78.1% and 75.4% to 86.0%.**

***Index Terms*–** Tucker decomposition, Transfer learning, Sleep, Pose identification

I. INTRODUCTION

Sleep posture has a considerable impact on sleep quality which in turn affects health as well as quality of life. Several studies have revealed a close relationship between body position and sleep features and disorders. As an example, a connection between body posture and sleep apnea frequency has been seen by [1]. Sleeping on the side results in fewer apneas and postural adjustment is deemed to work as a treatment in some patients. Sleep positions may also have connections with dream characteristics and sleep quality. As [2] suggests, subjects sleeping on their left sides may experience more nightmares than when sleeping on their right sides. Another study in [3] shows how posture detection and adjustment reduces the risk of pressure ulcers development.

Polysomnography (PSG) is the gold standard to evaluate the sleep quality. PSG uses a combination of measurements including EEG, electrooculography (EOG), electromyography (EMG), in addition to other measurements such as respiration and heart rate. These physiological variables have been employed in different studies for characterising sleep such as sleep stage detection [4], [5]. Although PSG can provide a detailed study of sleep, it is expensive and intrusive. Patients are required to sleep in a dedicated sleep facility and therefore, PSG is not suitable for in-home monitoring. Such extensive multi-sensor data are analysed manually which is cumbersome and time consuming. Actigraphy is often used as an alternative approach to investigate sleep; it uses non-invasive wearable sensors (e.g. watch) to monitor the activity and sleep/wake cycles using accelerometer. However, these products still demand for a device to be attached to the subject's wrist and does not allow for a comprehensive characterisation of movements and body posture.

Non-contact sleep monitoring solutions are an alternative to PSG and actigraphy. These methods have the ability to overcome the aforementioned obstacles and this can improve sleep quality and patient experience. Non-contact methodologies can be summarised into three main categories: instrumented mattresses, video-based monitoring, and sensor fusion based approaches. Within the first category, Hoque et al. [6] suggested using WISP accelerometers tagged to the bed to monitor four different sleep postures. Others [7] suggested that supine, prone, and lateral positions can be recognised with the aid of pillow sensors and bed sheets that contain sensor arrays. Alternatively, video-based methods [8], [9], [10], [11], [12], assume that a single video depth camera can detect different sleep positions. One approach [11] relies on Kinect v2 skeleton tracking for posture recognition. However, this method requires that patients'bodies should not be covered by sheets, etc. Sensor fusion approaches by means of which different sleep postures can automatically be analysed through use of a depth camera combined with instrumented pressure mattress, have also been proposed [13], [14], [15], [12]. This latter category is considered complex and high cost which limits its use. On the other hand, our approach exploits remotely monitoring sleep patterns using a single infrared (IR) camera, providing an easy, objective, cost-effective and ubiquitous approach for sleep pose detection.

This study has two main aims. Firstly, the application of tensor factorisation for enabling detection of specific group level data signatures that exist in the data and reduce occlusion effect. Moreover, the aim of such techniques is to factorise the data into lower dimensional factors so that a compact basis is formed. If this is achieved properly, the original data can be demonstrated in a concise manner that is also more noise immune and better generalised. One such dimensionality reduction method, tensor factorisation has been used recently in various fields including feature extraction, computer vision, and classification [16][17][18]. A tensor is known as a multi-way data shown through a multidimensional

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

array. Each tensor direction is described as a mode or way. Tensor factorisation protects the underlying structure of data [19]. Multi-way factorisation techniques are more favourable than matrix factorisation methods due to their uniqueness of the optimal solution. Moreover, they consider the multi-way structure of the data (here images) while the data structure could not be kept due to the need for data flattening in two-way techniques. Tensor factorisation provides us with a pattern extraction that well generalises across common modes. Different multi-way factorisation models exist including Tucker decomposition [20][21] which is a generalisation of singular value decomposition (SVD) to higher order arrays. Through a Tucker model, a tensor is decomposed into different factors and a core tensor of lower dimensions than the original data. In this way, feature extraction of the compressed data that was originally high dimensional is feasible thanks to Tucker decomposition. It also has popular applications in classification in order to determine patterns in group-level data and to reduce the dimensions.

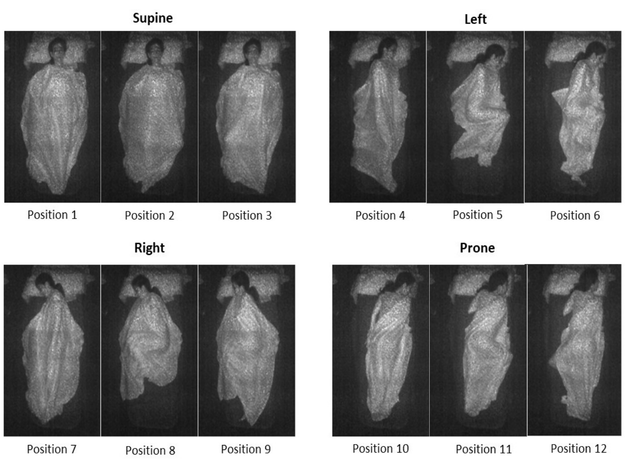

Fig. 1. IR images of twelve sleep postures illustrated by one of the volunteers.

The second goal in this paper is using deep learning (DL) for classification of the different sleep poses. In association with the ongoing advancement in the area of DL, convolutional neural networks (CNNs), in particular, greatly assist in image classification and processing. Training a CNN relies on large datasets compared to tasks in specialised fields which are often based on very small datasets. This, however, causes overfitting. To address this, pre-trained networks are in use, whereby a trained network from another task for which enough data is available exists. In principle, the weights of a previously trained network are repurposed for the new tasks. In this paper, a non-invasive camera-based method is suggested to accurately distinguish between different body postures occluded by a blanket. The IR images are pre-processed and then imported to a pre-trained network to classify twelve different sleep postures.

In section2, the aim is to discuss the dataset used and the pre-processing procedure the data has gone through to make way for tensor decomposition and classification. Section 3 explains the theory behind tensor analysis and factorisation accompanied by a description for CNN network structure. Section 5 illustrates the results and section 6 draws the conclusion.

II. DATASET

Data Collection: We used Microsoft Kinect depth camera in order to capture the IR images with a resolution of 640×480 pixels at 30 frames per second. For this work, depth data was ignored, and we tried to see if simple 2D IR images can be analysed. The dataset includes 10 healthy volunteers consisting of three female and seven males took part in our experiment[1]. Their Body Mass Index (BMI) ranged between 21.3 kg/m^2 and 27.8 kg/m^2. Volunteers were trained by our team to mimic twelve various sleep positions, each of which was maintained for 10 seconds (Fig. 1).

Pre-processing: The Kinect IR speckle pattern was filtered out to cancel obfuscating effects. To this end, a 2D median filter was applied to the raw IR images before applying the model. Median filters of different orders have been tested and the most optimised result was obtained using order 6. In addition, since the pre-trained network utilised RGB images, all grey scale images were converted into 8-bit RGB. This enabled us to directly insert them into the network. To avoid border effects, the images were all cropped in a way that a surrounding box was formed around the body shape and the mattress.

III. METHODOLOGY

Tensor Factorisation: For each subject in this work, the available frames are transferred into a 3-way tensor. Each tensor for individual subject demonstrates width, height, and segment information as depicted in Fig. 2. To extract the common information amongst K subjects, the tensors are linked together along the third dimension (Fig. 3). This results in a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{600\times800\times I_K}$ where I_K is the overall number of segments of the chosen subjects.

Tucker mode was employed to factorise the tensor into two modes (width, height) and a core tensor. Each mode was adjusted to further to include 100 components. As a result the decomposition can be represented as:

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{U}_w \times_2 \mathbf{U}_h \tag{1}$$

where $\underline{\mathbf{G}} \in \mathbb{R}^{100\times100\times I_K}$ reveals the core tensor, $\mathbf{U}_w \in \mathbb{R}^{600\times100}$ and $\mathbf{U}_h \in \mathbb{R}^{800\times100}$ are the width and height modes respectively. These modes allow us to extract the important features for a subject i by taking out its core tensor:

$$\underline{\mathbf{G}}^{(i)} = \underline{\mathbf{X}}^{(i)} \times_1 \mathbf{U}_w^T \times_2 \mathbf{U}_h^T \tag{2}$$

where $\underline{\mathbf{G}}^{(i)}$ represents the core tensor achieved by projecting the data $\underline{\mathbf{X}}^{(i)}$ of subject i onto the space spread by $\mathbf{U}_w$ and

[1]All participants were given an information sheet and gave written consent for their images and data being used for research work and subsequent publications, consistent with our faculty policy at the time.

$\mathbf{U}_h$ [19]. After obtaining the factors and the core tensor for each subject, the factors that are helpful for classification can be kept and the reconstructed images can be used for classification. Our analysis showed that by considering the factor dimensions within [2:50] could reduce the blanket effect and improve the performance.

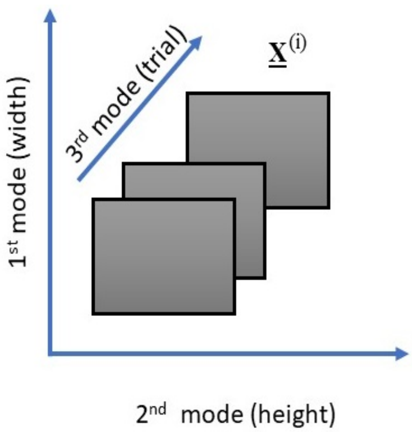

Fig. 2. Representation of 3-way tensor built for each subject.

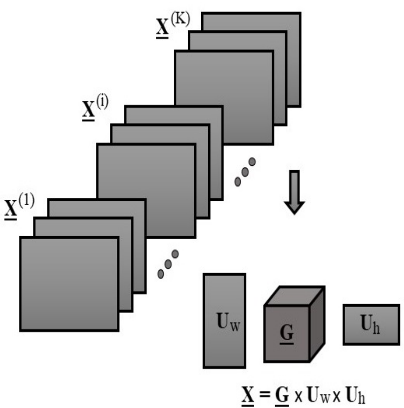

Fig. 3. Concatenating the data along the third dimension and factorising the resulting tensor through the Tucker decomposition.

Transfer Learning: The network we have used is based on pretrained deep neural network VGG-19 which can be publicly accessed. This convolutional network is pre-trained on the ImageNet dataset (with more than 1.2 million RGB images) of natural objects which has extensive use in different classification approaches. Layer weights achieved from the object classification tasks were were employed to place twelve sleep poses into the appropriate classes. The size of last fully connected (FC) layer of this network is 1000. In order to classify our twelve sleep poses, all the layers except the last one are frozen. In order for this network to be compatible with our dataset which has fewer classes, the last layer of the VGG-19 network was removed. A new FC layer was added to the network followed by softmax layers, but with fewer parameters, since we have less classes to classify for a given current task. Finally, all the layers of VGG-19 network with the exception of last layer are reused and only the last layer is updated (see Fig. 4). The learning rate was also empirically set and optimised to 0.0001.

VGG-19: The VGG-19 model has 19 layers, formed by 16 convolutional layers (3×3 filters) with rectifier (ReLU), five 2×2 max pooling (stride = 2), followed by three FC layers and dropout between them. The first two FC layers have 4096 features while the last one only 1000. The last FC layer is followed by a softmax layer.

Train & Test: A test dataset was chosen from the initial dataset by splitting the data before training, so that no training image symbol was utilised during the test. Cross validation (CV) known as an approach for model selection that functions by dividing the dataset into a number of complementary subsets known as folds. In a k-fold CV the data are partitioned into k subsets. Next, each subset is employed once for model validation. However, the rest k-1 subsets are utilised for training the classifier. In this work, 10-fold cross validation with a ratio of 90%:10% for training and testing respectively was used to separate the data. The DL algorithm then classified each single image into one of twelve possible sleep pose classes.

IV. RESULT AND DISCUSSION

After the factorisation stage and inputting images into the network, they were downsampled to the same dimensions as the images used for the pretrained network: $224 \times 224 \times 3$. Considering a total number of 10 volunteers, a 10-fold cross validation technique ('leave-one-subject-out' (L-O-S-O)) was used to validate the classifier. The mean detection accuracy for each pose over the test data with 4-layer CNN, without tensor decomposition, and by including tensor decommission can be seen in Table 1. We have also demonstrated the accuracy of the classifier for each volunteer over twelve poses for the proposed method, for the retrained VGG-19 network, and for our previous work in [22] during which a conventionally trained single CNN network with four layers was used. We have achieved the average accuracy of 0.86, 0.78, and 0.74 over 10 volunteers for the tensor factorisation combined with VGG-19 method, the method solely based on VGG-19, and CNN network respectively (see Fig. 5).

V. CONCLUSIONS

In this work we incorporated the pre-trained VGG-19 network as a classifier together with an effective method for refining the data. The refining procedure is based on a known method, tensor factorisation. Tucker decomposition model was employed to extract meaningful features from our IR data. The factorisation not only provides all the necessary features of the data for the classification of sleep poses, but also reduces the noise to improve the classification performance. Due to the small dataset, transfer learning was used to classify different sleep poses occluded by blanket. We compared the accuracy

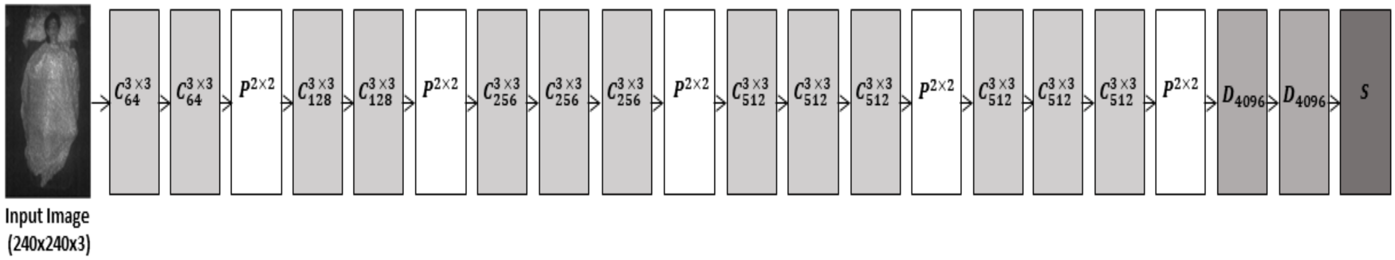

Fig. 4. Architecture of the CNN for pose detection. $C_k^{h\times w}$ illustrates a convolutional layer where k is activation maps with a dimension of $h \times w$. P shows a pooling layer. D_n represents FC layer with n kernels and S shows a softmax layer.

TABLE I

THE AVERAGED DETECTED ACCURACY (%) FOR EACH TWELVE POSES OVER 10 VOLUNTEERS WHEN CNN NETWORK, VGG-19 NETWORK, AND WHEN TENSOR FACTORISATION + VGG-19 ARE APPLIED.

Sleep Poses	**CNN**	**VGG-19**	**VGG-19 & Tensor**
Pos1	79	79	89
Pos2	89	92	96
Pos3	80	84	91
Pos4	80	81	81
Pos5	70	79	85
Pos6	65	73	81
Pos7	80	80	86
Pos8	60	70	78
Pos9	56	60	73
Pos10	72	72	88
Pos11	80	84	94
Pos12	84	84	90
Mean Accuracy	74	78	86

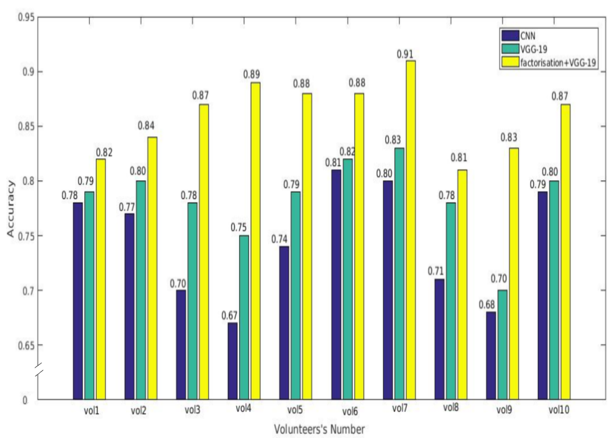

Fig. 5. Comparing the accuracy for classification of 12 poses over 10 volunteers amonge the proposed tensor factorisation combined with VGG-19 network, retrained VGG-19, and 4-layer CNN network.

of our method with the one without factorisation step and with the single CNN network and observed notable improvement in terms of accuracy when using the tensor factorisation. Investigation of body poses during natural sleep and comparison to actigraphy watch and PSG derived measures of sleep quality can be considered as future work.

REFERENCES

[1] C. F. George, T. Millar, and M. Kryger, "Sleep apnea and body position during sleep," *Sleep*, vol. 11, no. 1, pp. 90–99, 1988.

[2] M. Y. Agargun, M. Boysan, and L. Hanoglu, "Sleeping position, dream emotions, and subjective sleep quality," *Sleep and Hypnosis*, vol. 6, pp. 8–13, 2004.

[3] S. Ostadabbas, R. Yousefi, M. Nourani, M. Faezipour, L. Tamil, and M. Q. Pompeo, "A resource-efficient planning for pressure ulcer prevention," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1265–1273, 2012.

[4] S. M. Mohammadi, S. Kouchaki, M. Ghavami, and S. Sanei, "Improving time–frequency domain sleep EEG classification via singular spectrum analysis," *Journal of neuroscience methods*, vol. 273, pp. 96–106, 2016.

[5] S. M. Mohammadi, S. Enshaeifar, M. Ghavami, and S. Sanei, "Classification of awake, rem, and nrem from eeg via singular spectrum analysis," in *37th IEEE Annual International Conference of Engineering in Medicine and Biology*, 2015, pp. 4769–4772.

[6] E. Hoque, R. F. Dickerson, and J. A. Stankovic, "Monitoring body positions and movements during sleep using wisps," in *Wireless Health*. ACM, 2010, pp. 44–53.

[7] T. Harada, A. Sakata, T. Mori, and T. Sato, "Sensor pillow system: monitoring respiration and body movement in sleep," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000, pp. 351–356.

[8] M. C. Yu, H. Wu, J. L. Liou, M. S. Lee, and Y. P. Hung, "Multiparameter sleep monitoring using a depth camera," in *International Joint Conference on Biomedical Engineering Systems and Technologies*. Springer, 2012, pp. 311–325.

[9] X. Liu and S. Payandeh, "Toward study of features associated with natural sleep posture using a depth sensor," in *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016, pp. 1–6.

[10] T. Grimm, M. Martinez, A. Benz, and R. Stiefelhagen, "Sleep position classification from a depth camera using bed aligned maps," in *IEEE International Conference on Pattern Recognition (ICPR)*, 2016, pp. 319–324.

[11] J. Lee, M. Hong, and S. Ryu, "Sleep monitoring system using kinect sensor," *International Journal of Distributed Sensor Networks*, 2015.

[12] Y. Y. Li, Y. J. Lei, L. C. L. Chen, and Y. P. Hung, "Sleep posture classification with multi-stream cnn using vertical distance map," in *International Workshop on Advanced Image Technology*. IEEE, 2018, pp. 1–4.

[13] W. Huang, A. A. P. Wai, S. F. Foo, J. Biswas, C.-C. Hsia, and K. Liou, "Multimodal sleeping posture classification," in *IEEE International Conference on Pattern Recognition (ICPR)*, 2010, pp. 4336–4339.

[14] V. Metsis, D. Kosmopoulos, V. Athitsos, and F. Makedon, "Non-invasive analysis of sleep patterns via multimodal sensor input," *Personal and ubiquitous computing*, vol. 18, no. 1, pp. 19–26, 2014.

[15] M. Papakostas, J. Staud, F. Makedon, and V. Metsis, "Monitoring breathing activity and sleep patterns using multimodal non-invasive technologies," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2015, p. 78.

[16] P. Comon, "Tensor decompositions, state of the art and applications. arxiv preprint," *arXiv preprint arXiv:0905.0454*, vol. 48, 2009.

[17] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[18] A. Shashua and T. Hazan, "Non-negative tensor factorization with applications to statistics and computer vision," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 792–799.

[19] L. Spyrou, S. Kouchaki, and S. Sanei, "Multiview classification of brain data through tensor factorisation," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015, pp. 1–6.

[20] R. A. Harshman, M. E. Lundy *et al.*, "Parafac: Parallel factor analysis," *Computational Statistics and Data Analysis*, vol. 18, no. 1, pp. 39–72, 1994.

[21] P. M. Kroonenberg and J. De Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.

[22] S. M. Mohammadi, M. Alnowami, S. Khan, D.-J. Dijk, A. Hilton, and K. Wells, "Sleep posture classification using a convolutional neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1–4.