Received May 1, 2019, accepted May 23, 2019, date of publication May 30, 2019, date of current version June 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920025

Automated Non-Contact Detection of Head and Body Positions During Sleep

SINA AKBARIAN[1,2], GHAZALEH DELFI[1,2], KAIYIN ZHU[1], AZADEH YADOLLAHI[1,2], (Member, IEEE), AND BABAK TAATI[1,2,3,4], (Member, IEEE)

[1]Toronto Rehabilitation Institute, University Health Network, Toronto, ON M5G 2A2, Canada
[2]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada
[3]Department of Computer Science, University of Toronto, Toronto, ON M5T 3A1, Canada
[4]Vector Institute for Artificial Intelligence, Toronto, ON M5G 1M1, Canada

Corresponding author: Babak Taati (babak.taati@uhn.ca)

This work was supported in part by FedDev Ontario, in part by BresoTec Inc., in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant under Grant RGPIN 435653, and in part by the Toronto Rehabilitation Institute, University Health Network.

ABSTRACT Obstructive sleep apnea (OSA) is a respiratory disorder characterized by interruption to breathing during sleep. Usually, the OSA is more severe in the supine sleeping position. Recent studies also demonstrated that the head position may play an important role in the pathophysiology of the OSA. Therefore, monitoring the sleeping body and the head position has high clinical importance to optimize the treatment of the OSA. In this paper, three machine learning approaches were used to detect the head position during sleep in infrared images. In the first two methods, supervised classifiers were trained to estimate the head position based on different feature sets extracted from infrared images. In the third method, three different convolutional neural network (CNN) structures (ResNet50, MobileNet, and Darknet19) were trained to detect the head position during sleep. To detect the body position, the same CNN architectures were trained on infrared images. Overnight sleeping data (sleep duration = 5 ± 1 h) from 50 participants (age: 53 ± 15 years, BMI: 29 ± 6 kg/m2, and 30 men/20 women) with various levels of OSA severity as measured by the apnea–hypopnea index (AHI = 25 ± 29 events/h and OSA severity: 12 normal, 13 mild, 11 moderate, and 14 severe) were collected for this paper. The models were trained on the data collected in one laboratory room from half of the participants and tested on the data from the other half collected in a different laboratory room. The best performing model (Darknet19) correctly estimated the lateral versus supine head position with 92% accuracy and 94% F1-Score and the lateral versus supine body position with 87% accuracy and 87% F1-Score.

INDEX TERMS Computer vision, machine learning, position detection, sleep apnea, non-contact monitoring.

I. INTRODUCTION

Sleep apnea is a respiratory disorder characterized by interruption to breathing during sleep. Obstructive sleep apnea (OSA) is the most common type of sleep apnea which is caused by the total (apnea) or partial (hypopnea) collapse of the pharyngeal airway during sleep, blocking the flow of air to the lungs [1], affecting an estimated 10% of the adult population [2]. The severity of OSA is indicated by the Apnea-Hypopnea Index (AHI), the number of apneas and hypopneas per hour of sleep. Sleep apnea severity is categorized as: normal (AHI<5), mild (5≤AHI<15), moderate (15≤AHI<30), and severe (AHI≥30) [3]. OSA increases the risk of cardiovascular diseases, stroke, and abnormal glucose metabolism [4]. More importantly, up to 60% of patients with OSA are not compliant to the current treatment options [5]. A recent study found that the overall healthcare costs of patients with untreated sleep apnea were nearly 25% higher than those patients undergoing treatment [6]. The severity of OSA is often associated with sleeping in the supine position, meaning sleeping while facing upwards [7]. It was shown that positional therapy to change the posture from supine to lateral can reduce the severity of OSA [8]. In addition, other studies have shown that changing the head

The associate editor coordinating the review of this manuscript and approving it for publication was Thomas Penzel.

position to lateral also plays an important role in decreasing the AHI [9], [10].

Positional therapy involves wearing an item, such as a shirt with a ball sewn in the back to encourage sleeping in the lateral position. Although positional therapy is simple, its limitation is that it is difficult to know for certain whether the patient remained in the desired position throughout the night. This motivates the need for a sleeping position monitor that could provide feedback to patients and physicians. In addition, analyzing changes in sleep position could have applications in assessing sleep quality and irregular sleeping patterns [11].

One approach to monitor the body and/or head position during sleep is to attach a sensor to the body/head of the patient. There are several examples of such contact methods, e.g. using accelerometer/gyroscope [12]. These contact methods are sometimes inconvenient during sleep and could potentially alter sleep conditions due to the addition of a physical device.

The alternative is to use methods and sensors which are not in direct contact with the person to monitor their sleeping position. In one proposed solution, Liu et al. used pressure sensors under a bedsheet to produce high-resolution pressure maps, which were then analyzed to extract a set of geometrical features for sleep posture classification [13]. However, pressure sensors are expensive (>$10K), preventing this approach from acquiring large-scale popularity. To reduce the cost, Hsia et al. replaced pressure sensors with a Force Sensing Resistor (FSR) [14]. They used Bayesian classification to estimate body posture. However, reducing the number of sensors lowered the accuracy to 78%. To improve performance, Huang et al. extracted feature from FSR sensor and video camera and used support vector machines (SVM) to detect body position [15]. With recent improvements in computer vision techniques, Liu et al. were able to only use a camera to capture images/videos of individuals during sleep in order to predict sleep position [16]. They then used histograms of oriented gradients (HOG) to extract features from the images and the sleeping position was classified with SVM. However, regular cameras cannot be applied in low light sleep environment. To address this issue, a more recent solution by Liu et al. used an infrared camera to capture the videos [17]. A pre-trained convolutional neural network (CNN) – namely the convolutional pose machine (CPM) [18] – was then repurposed for sleeping position detection via fine-tuning. Although CPM obtained good performance for human pose tracking, the performance of the developed method under various sleeping conditions, such as when the body is covered or partially covered by a blanket, was not investigated.

To address these challenges, we have implemented a CNN-based image analysis algorithm to detect both body posture and head position during sleep. Also, we have validated the performance of the algorithm when the body was covered with blankets.

II. DATA COLLECTION

The data were collected in two separate rooms at the sleep laboratory of the Toronto Rehabilitation Institute. An Infrared camera (a Point Grey Firefly MV, 0.3 MP, FMVU-03MTM) was positioned about 1.4 m above the bed, which recorded infrared videos of 50 participants sleeping overnight (Age: 53±15 years, BMI: 29±6 kg/m^2, 30 men and 20 women, AHI = 25±29 events/hour, sleep duration = 5±1 hours, OSA severity: 12 normal, 13 mild, 11 moderate, and 14 severe). Videos were recorded at a resolution of 640 × 480 and 480 × 640 (depending on the camera orientation landscape/portrait during recording) at 30 frames per second. The room was illuminated by an infrared light (Raytec RM25-F-50). Simultaneously, full overnight polysomnography was recorded for clinical diagnosis of sleep apnea and AHI. The University Health Network Research Ethics Board and the University of Toronto Research Ethics Board approved this study. Participants submitted informed written consent before taking part in the study. Figure 1 illustrates (anonymized) sample image frames from this dataset.

FIGURE 1. Sample supine (left) and lateral (right) frames.

Since consecutive frames of the videos are not independent, many of them would not contain any additional information for the purposes of training a machine learning model. Including them will only increase the computational cost and might lead to over-fitting. Large movements (e.g. shifting from supine to lateral, arm or leg movement) were automatically detected by thresholding the total displacement of tracked featured points [19] over one second. For each participant, one image frame was selected at the beginning of the night and also after each large movement, resulting in a total of 4,113 image frames.

In gold standard polysomnography, sleep technicians manually annotate the body positions only, by visually inspecting the video that was recorded as part of polysomnography test.

Accordingly, a technician manually annotated the head and body positions (lateral/supine) in all captured video frames. The criteria for annotating body posture was based on the position of the shoulder. The body position was labeled as lateral if one shoulder was lifted off the bed and any part of the back (torso posterior) was visible to the left or right of the shoulder. For head annotations, the position was considered lateral if one ear could be seen and only one eye was visible. In partial turns when both eyes were visible, image was labeled as lateral if tip of nose was seen closer to outer corner of the eye compared to the inner corner. If the body or head was covered, technician made judgments based on comparing visible parts with previous or subsequent images where body/head was visible.

The recorded data of 25 participants (all in one room) were used in the training and validation sets, which includes 2,113 frames (body labels: 45% lateral and 55% supine, head labels: 71% lateral and 29% supine) and the remaining 25 (all recorded in another room) comprised the test set, including 1,998 frames (body labels: 48% lateral and 52% supine, Head labels: 64% lateral and 36% supine). The training and test data sets were chosen to be from two different rooms to make sure lighting and camera position did not impose any kind of bias on our learning. This ensures generalization, i.e. reported results already incorporate when images are captured in a different setup with a different background and camera placement.

Our dataset includes many images where sleeping position are between supine and lateral, and the face is partially or completely covered, which posed a challenge for training CNNs. To help the network learn face features from clearly distinct facial and body orientation, the mugshot dataset [20] – that does not have any occlusion, poor lighting, severe out of plane rotation, and in between supine and lateral images – was used to augment the data for training CNN models. The dataset consists of 3,248 images of subjects facing left, right and front, which were used as additional lateral (left and right) and supine (front) training images. The sample of mugshot images are shown in Figure 2.

FIGURE 2. Sample of mugshot images.

III. METHODOLOGY

We only seek to predict supine and lateral positions, since it is not clinically useful to distinguish between the left and right lateral positions. We also do not include prone position in our labels since it is rare, as reported in the literature [21]. Our dataset followed the same trend and contained only 13 prone positions out of 4,126 video frames.

A. HEAD POSITION DETECTION

We compare two machine learning approaches for detecting the head position (distinguishing supine from lateral): engineered features vs. learned features. Two different sets of engineered features were explored. In the first method (presence features), Haar feature-based cascade classifiers [22] were used to detect nine binary features: the presence/absence of left eye, right eye, both eyes, left ear, right ear, nose, mouth, frontal face, and side-view face. For instance, if the (pre-trained) Haar-cascade classifier detects a ''left eye'' in the image, the corresponding binary features would be 1 (or 0 otherwise). Each of the Presence Features is a weak indicator for the head position; e.g. an ear is more likely to be detected in the lateral position, while both eyes are more likely to be detected in the supine position. The combination of these individually weak indicators could results in a strong classifier. The overall steps of this model are shown in Figure 3. The second method (landmark features) uses the 3D coordinates of 68 facial landmarks detected via the hourglass deep convolutional network [23]. The hourglass network was fed the bounding box of the head, which was automatically detected using the Dlib library [24]. Landmark points were translated to place the nose tip at the origin and were normalized to the nose length by dividing by the nose length. Nose length was calculated as the distance between top of the nose and nose tip. The overall steps of this model are shown in Figure 3. Extracted features from each method were used to train five different binary classifiers to detect supine versus lateral head positions. Classifiers used were linear and radial basis function (RBC) support vector machine (SVM), logistic regression, multilayer perceptron, and random forest.

For logistic regression, SVMs, and random forest classifiers, Scikit-learn [25] was used and hyper parameters were tuned by coarse to fine grid search. For the multilayer perceptron classifier, TensorFlow [26] was used and hyper parameters were tuned by 40 random searches. Three-fold cross validation was used to tune hyper-parameters of each classifier. Data was divided into training and test sets based on the rooms in which videos were recorded. That is, all videos collected in one room were assigned to the training set and all videos collected in the other room were assigned to the test set. The training set was further split into training and validation partitions based on the participants. That is, all images from each participant were placed in one of the validation or train sets.

FIGURE 3. Presence Features (middle column) and Landmark Features (right column) are used to detect the head position (supine vs. lateral).

In addition to exploring engineered features for head position detection, we also trained a CNN for this task. As before, the mugshot dataset was used to augment the training set.

Artificially rotating (or translating, or sheering, or scaling) the images by a small amount is a common method employed in training deep neural networks as it improves robustness with respect to small changes in test data [27]–[29]. Here, all training images were rotated from $-30°$ to $30°$ in step of $5°$ to simulate a range of possible body orientation during sleep. Three different CNN architectures were used: 1) Darknet19 [30], 2) MobileNet [31], and 3) ResNet50 [32]. The details of the best performing architecture (Darket19, See Section IV) are shown in Table 1. Leaky rectified linear units (leaky ReLu) were used as the activation function in all of the convolutional layers in all three architectures. A linear activation function was used for all the fully connected layers. Stochastic gradient descent with a momentum of 0.9 and polynomial decay with power of four were used to optimized the cross entropy cost function during training. An initial value of 0.01 for the learning rate, 0.0001 for the decay rate, and 128 for the batch size with subdivision of 2 were used. Figure 4 illustrates the overall process.

To improve accuracy, e.g. when whole body and face are covered by a blanket, when the detection confidence score is lower than a threshold (60% in the experiments), the algorithm checks the next and previous video frames and chooses the prediction label with the highest confidence.

TABLE 1. Darknet19 architecture.

Type	Filters	Size/Stride
Convolutional	32	3×3
MaxPooling		2×2/2
Convolutional	64	3×3
MaxPooling		2×2/2
Convolutional	128	3×3
Convolutional	64	1×1
Convolutional	128	3×3
MaxPooling		2×2/2
Convolutional	256	3×3
Convolutional	128	1×1
Convolutional	256	3×3
MaxPooling		2×2/2
Convolutional	512	3×3
Convolutional	256	1×1
Convolutional	512	3×3
Convolutional	256	1×1
Convolutional	1024	3×3
MaxPooling		2×2/2
Convolutional	512	3×3
Convolutional	256	1×1
Convolutional	128	3×3
Convolutional	64	1×1
Convolutional	16	3×3
Fully connectedl	2	1×1
AvgPool		Global
Sigmoid		

B. BODY POSITION DETECTION

To estimate the body position (distinguishing supine from lateral), the recorded images were not cropped and the entire

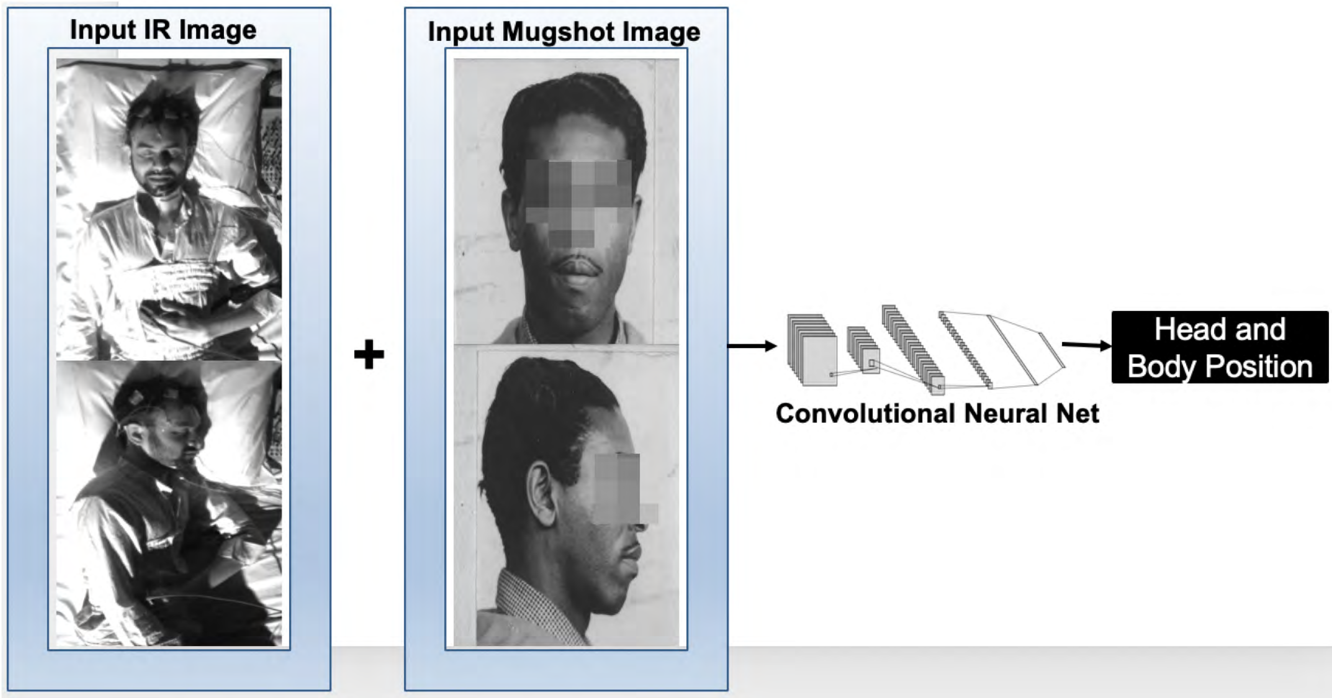

FIGURE 4. Convolutional neural network method to detect head and body position. The training set contains image data from 25 participants amounting to a total of 2,113 image frames and the mugshot dataset which includes 3,248 images. The test set consists of 1,998 image frames from the remaining 25 participants.

image was used instead. Our dataset comprises three different image sizes including 256×256 for Mugshot and 640×480 and 480×640 for our infrared images. However, training the CNN requires constant image size. All images were therefore padded (with a white background) to the same size of 640×640. The same three CNN architectures that were used for head position detection were also used for body position detection. The overall steps of this model are shown in Figure 4.

For comparison, the HOG-based method [16], developed by Liu et al., was also evaluated on our test dataset as a baseline. We evaluated both the pre-trained model (as trained by the authors) and also a re-trained model using our training data.

IV. RESULTS

In line with previously published reports [7], [9], [10], in our data of 50 participants, the difference between the participants' AHI in supine vs. lateral sleep positions was statistically significant (Wilcoxon signed-rank: W statistic = 112, $p < 10^{-5}$). As shown in the Table 2, although participants spent more time in lateral position, the supine AHI is higher than lateral AHI in all sleep apnea severity groups.

On average, participants changed their head position 15 ± 10 times and body position 9 ± 6 times during an average of 5 ± 1 hours of sleep.

For detecting the head position, results of the best classifier (random forest) trained on the engineered feature sets and result of the three CNN networks are shown in Table 3.

TABLE 2. Sleep duration and AHI in different sleep position by different severity groups.

Sleep Anea Severity	Duration in Supine (%)	Duration in Lateral (%)	Supine AHI	Lateral AHI
Normal	33.8	66.2	15.5	1.8
Mild	37.0	63.0	33.5	4.7
Moderate	55.9	44.1	34.0	10.4
Severe	38.1	61.9	58.6	47.4

TABLE 3. Head position detection results (%).

Method	Accuracy	Precision	Recall	F1-score
Presence Features	59.9	58.8	94.6	72.5
Landmark Features	66.3	71.0	85.7	77.7
MobileNet	78.3	96.8	68.5	80.3
ResNet50	80.5	96.9	72.1	82.6
Darknet19	88.3	87.3	95.7	91.3

The best performing model (Darknet19) achieved 88% accuracy, 87% precision, and 96% recall.Body position detection results are shown in Table 4. The best performing model (Darknet19) achieves similar recall (95%), but a lower precision (79%). Confusion matrices of the best models (Darknet19) are shown in Figure 5 and Figure 6 for head and body position respectively.

The best performing algorithm (Darknet19) could process the whole night (5 ± 1 hours) sleeping data of each patient in 98 ± 40 seconds on GPU(GeForce GTX 1080 Ti).

TABLE 4. Body position detection results (%).

Method	Accuracy	Precision	Recall	F1-score
HOG	50.9	48.0	88.7	62.3
ResNet50	69.3	78.8	48.2	59.8
MobileNet	74.8	78.8	64.0	70.7
Darknet19	84.9	77.8	95.4	85.7

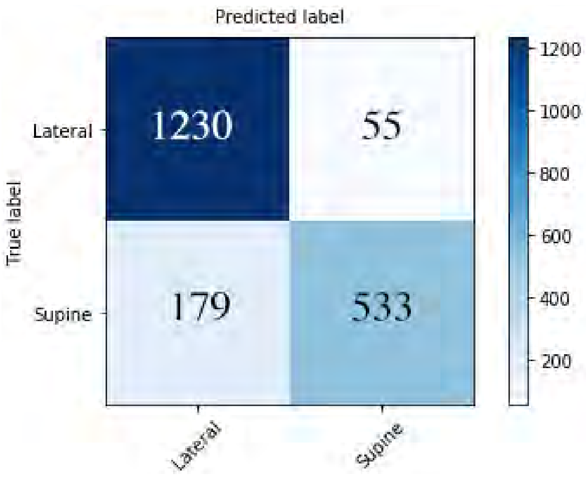

FIGURE 5. Head position labeling confusion matrix using Darknet19.

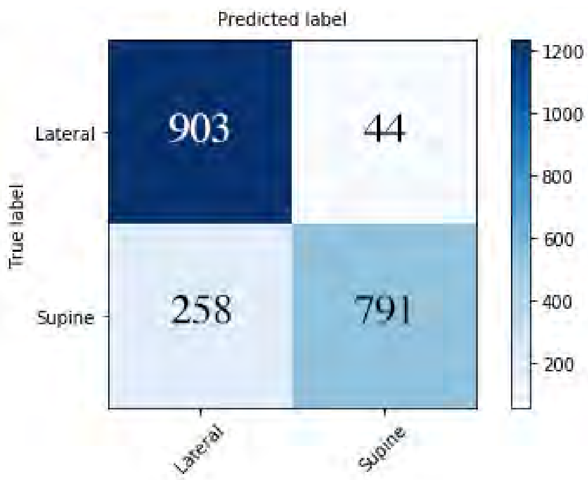

FIGURE 6. Body position labeling confusion matrix using Darknet19.

V. DISCUSSIONS

Both engineered feature sets that were explored had poor performance in detecting the head position, as compared to the CNN models. The landmark features performed better than the presence features, but they rely on an accurate detection of the face. Further analysis of the data revealed that Dlib face detector failed in 65% of the images in our dataset, of which 34% were in the supine position and the remaining 66% were in the lateral position. The algorithm returned a constant number for the cases where it failed to detect a face. The classifiers then learned to label these points as lateral based on the distribution of the data. Therefore, the reported 66% accuracy of a random forest classifier using

TABLE 5. Head position detection with and without a blanket (%).

Blanket Use	Accuracy	Precision	Recall	F1-score
No Blanket (16% of test set)	89.1	87.8	96.4	91.9
Body covered (81% of test set)	86.5	86.3	94.0	90.0
Body & head covered (3% of test set)	73.8	75.1	88.5	81.2
All	88.3	87.3	95.7	91.3

TABLE 6. Body position detection with and without a blanket (%).

Blanket Use	Accuracy	Precision	Recall	F1-score
No Blanket (16% of test set)	85.9	78.4	96.6	86.5
Body covered (81% of test set)	85.5	79.3	94.6	86.3
Body & head covered (3% of test set)	79.4	73.5	86.7	79.7
All	84.9	77.8	95.4	85.7

the landmark features over-represents the true performance of the model. By contrast, the best performing CNN model (Darknet19) achieved promising results, even in cases when the person was partially covered by a blanket. To further investigate this, the images in the test set were divided into three groups based on whether the person was covered by a blanket or not. Tables 5 and 6 show the performance of our trained Darknet19 model on these subsets. As expected, performance drops when both the head and the body are covered by a blanket, e.g., head position detection accuracy is 89% when no blanket is used, but 74% when both head and body are covered.

While the CNN models trained to directly estimate a supine vs. lateral model performed well, the CPM model repurposed to detect full body position during sleep [17] did not perform well in our data. This was primarily due to the fact that in the majority of the cases (84%) the person was partially covered by a blanket and the method was unable to recover full body position. To illustrate this issue, we show sample output of the algorithm in two sample frames (one supine and one lateral body position) in which a blanket is used (Figure 7). While the CPM model tries to detect 14 landmark points on the body (e.g. the wrists, the elbows, the shoulders, etc.), the repurposed model is only able to find the head position correctly. Across all the test images, the repurposed CPM model was able to correctly locate the head in 37% of the image frames when the person was in a lateral body position, and in 66% of the image frames when the person was in a supine body position.

To evaluate the influence of the mugshot data on our performance, we examined training the CNN models without adding the mugshot dataset. This led to increased training loss and the model did not converge. Adding the Mugshot dataset increased the number of easy examples (i.e. no occlusions, and no in-between cases) and, as shown in Section IV, resulted in a model that successfully distinguished supine

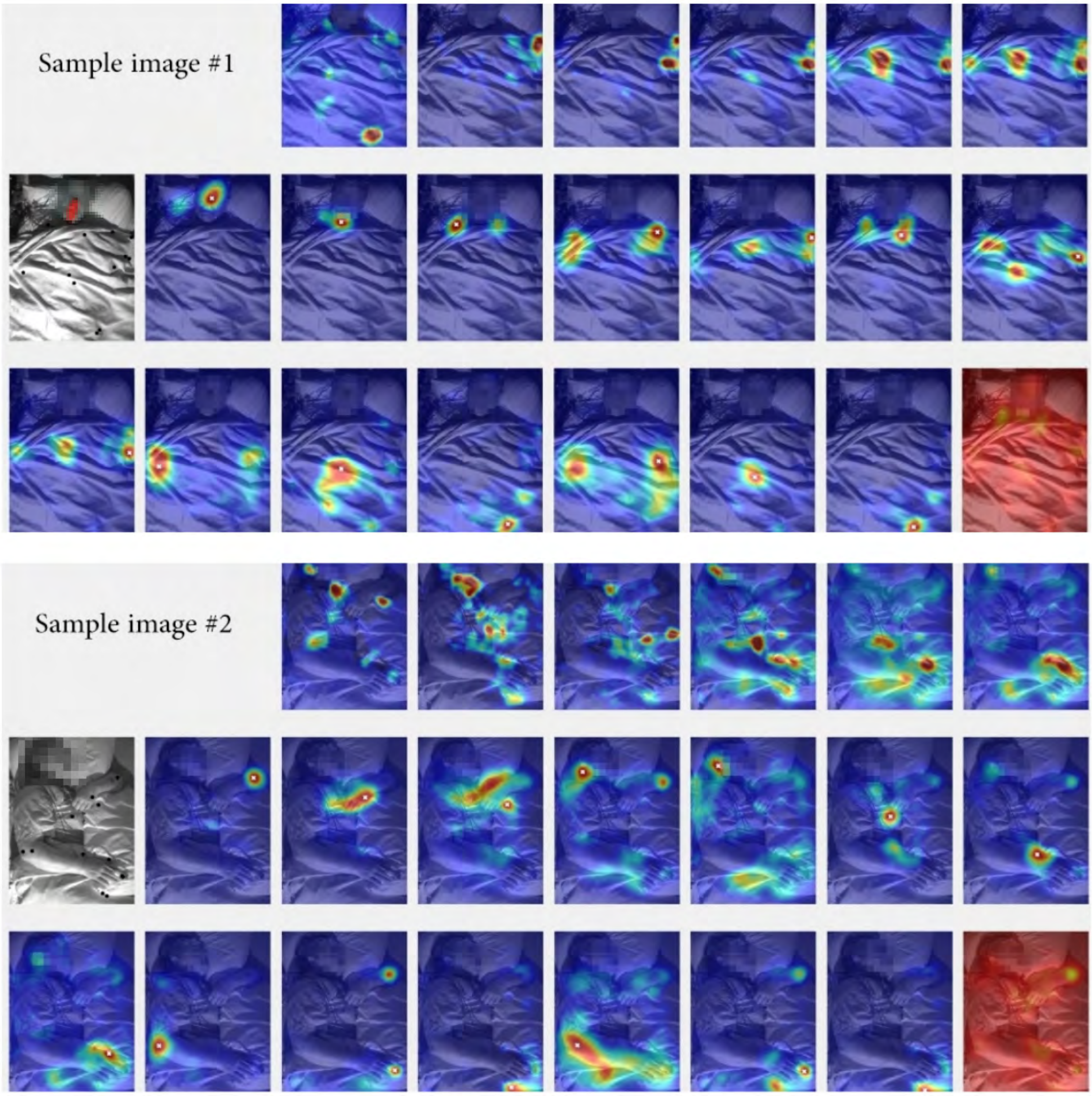

FIGURE 7. Repurposed CPM [17] performance on real data. The heat maps show detected body parts for each sample image is as follows. First row: different stage of joint detection; Second row: pose visualization, head, neck, right shoulder, right elbow, right wrist, left shoulder, left elbow; Third row: left wrist, right hip, right knee, right ankle, left hip, left knee, left ankle, background belief map.

versus lateral positions with high accuracy. We also explored whether the mugshot dataset alone was sufficient in training a model to accurately classify the sleeping position.

We trained the best performing model (Darknet19) with only the mugshot dataset (i.e. without our infrared images collected during sleep). This model achieved an accuracy of 64% for head position detection and 57% for body position detection (vs. 92% and 87% accuracy when both the mugshot dataset and our sleep dataset were used).

Our study has some limitations. One limitation is a small sample size (N = 50). Another limitation is the subjective definition of supine and lateral when the position of the head and body are in-between the two positions. In addition, determining the true head and body position was difficult in the video frames in which the person's body was entirely covered by a blanket sheet. Ignoring these extremely challenging images frames in the test set results in a large gain in accuracy from 85% to 87% for body position and from 88% to 92% for head position. The F1-Score is improved from 86% to 87% for body position and from 91% to 94% for head position.

VI. CONCLUSIONS AND FUTURE WORK

In this work, we developed an algorithm that could automatically and accurately detect both body posture and head position during sleep. The algorithm was validated vs. the clinical gold standard of manual annotation using overnight sleeping data of 50 individuals with various levels of sleep apnea severity. The algorithm performed well in detecting the head position and the body even when the body was covered by a blanket. The developed method could be used to monitor sleep positions overnight in order to provide feedback for positional therapy and to assess sleep quality and irregular sleeping patterns.

Future work involves collecting more data to further improve accuracy and also implementing the resulting algorithm in the form of a mobile application.

REFERENCES

[1] B. Chaska, J. P. Kiley, R. P. Millman, B. A. Phillips, S. D. Rogus, K. P. Strohl, P. J. Strollo, P. M. Suratt, J. K. Walsh, J. W. Weiss, D. White, B. Zepf, J. P. Kiley, S. D. Rogus, and S. T. Shero, ''Sleep apnea: Is your patient at risk,'' *Amer. Family Physician*, vol. 53, no. 1, pp. 247–253, 1996.

[2] P. E. Peppard, T. Young, J. H. Barnet, M. E. W. Palta Hagen, and K. M. Hla, ''Increased prevalence of sleep-disordered breathing in adults,'' *Amer. J. Epidemiology*, vol. 177, no. 9, pp. 1006–1014, May 2013.

[3] American Academy of Sleep Medicine Task Force and others, ''Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research,'' *Sleep*, vol. 22, no. 4, pp. 667–689, 1999.

[4] N. H. Kim, ''Obstructive sleep apnea and abnormal glucose metabolism,'' *Diabetes Metabolism J.*, vol. 36, no. 4, pp. 268–272, Aug. 2012.

[5] T. E. Weaver and A. M. Sawyer, ''Adherence to continuous positive airway pressure treatment for obstructive sleep apnea: Implications for future interventions,'' *Indian J. Med. Res.*, vol. 131, pp. 245–258, Feb. 2010.

[6] K. J. Potts, D. T. Butterfield, P. Sims, M. Henderson, and C. B. Shames, ''Cost savings associated with an education campaign on the diagnosis and management of sleep-disordered breathing: A retrospective, claims-based US study,'' *Population health Manage.*, vol. 16, no. 1, pp. 7–13, Jan. 2013.

[7] R. D. Cartwright, F. Diaz, and S. Lloyd, ''The effects of sleep posture and sleep stage on apnea frequency,'' *Sleep*, vol. 14, no. 4, pp. 351–353, Jul. 1991.

[8] O. Omobomi and S. F. Quan, ''Positional therapy in the management of positional obstructive sleep apnea—A review of the current literature,'' *Sleep Breathing*, vol. 22, no. 2, pp. 297–304, May 2018.

[9] K. Zhu, T. D. Bradley, M. Patel, and H. Alshaer, ''Influence of head position on obstructive sleep apnea severity,'' *Sleep Breathing*, vol. 21, no. 4, pp. 821–828, Dec. 2017. doi: 10.1007/s11325-017-1525-2.

[10] E. R. van Kesteren, J. P. van Maanen, A. A. J. Hilgevoord, D. M. Laman, and N. de Vries, ''Quantitative effects of trunk and head position on the apnea hypopnea index in obstructive sleep apnea,'' *Sleep*, vol. 34, no. 8, pp. 1075–1081, Aug. 2011.

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

[12] E. Hoque, R. F. Dickerson, and J. A. Stankovic, ''Monitoring body positions and movements during sleep using WISPs,'' presented at the Wireless Health, San Diego, California, Oct. 2010, pp. 44–53.

[13] J. J. Liu, Q. Wu, M.-C. Huang, N. Alshurafa, M. Sarrafzadeh, N. Raut, and B. Yadegar, ''Sleep posture analysis using a dense pressure sensitive bedsheet,'' *Pervas. Mobile Comput.*, vol. 10, pp. 34–50, Feb. 2014.

[14] C.-C. Hsia, Y.-W. Hung, Y.-H. Chiu, and C.-H. Kang, ''Bayesian classification for bed posture detection based on kurtosis and skewness estimation,'' in *Proc. 10th Int. Conf. E-Health Netw., Appl. Services*, Singapore, Jul. 2008, pp. 165–168.

[15] W. Huang, A. A. P. Wai, S. F. Foo, J. Biswas, C. Hsia, and K. Liou, ''Multimodal sleeping posture classification,'' in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 4336–4339.

[16] S. Liu and S. Ostadabbas, ''A vision-based system for in-bed posture tracking,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 1373–1382.

[17] S. Liu, Y. Yin, and S. Ostadabbas, ''In-bed pose estimation: Deep learning with shallow dataset,'' 2017, *arXiv:1711.01005*. [Online]. Available: https://arxiv.org/abs/1711.01005

[18] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, ''Convolutional pose machines,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4724–4732.

[19] M. H. Li, A. Yadollahi, and B. Taati, ''Noncontact vision-based cardiopulmonary monitoring in different sleeping positions,'' *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1367–1375, Sep. 2017.

[20] C. Watson and P. Flanagan, *NIST Special Database 18 Mugshot Identification Database*. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2016.

[21] S. J. Gordon, K. A. Grimmer, and P. Trott, ''Self-reported versus recorded sleep position: An observational study,'' *Internet J. Allied Health Sci. Pract.*, vol. 2, no. 1, Jan. 2004, Art. no. 7.

[22] P. Viola and M. Jones, ''Rapid object detection using a boosted cascade of simple features,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2001, p. 1.

[23] A. Bulat and G. Tzimiropoulos, ''How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks),'' *Int. Conf. Comput. Vis.*, vol. 1, no. 6, p. 8, Mar. 2017.

[24] D. E. King, ''Dlib-ml: A machine learning toolkit,'' *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jul. 2009.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[26] M. Abadi *et al.*, ''Tensorflow: A system for large-scale machine learning,'' in *Proc. OSDI*, vol. 16, 2016, pp. 265–283.

[27] S. Hauberg, O. Freifeld, A. B. L. Larsen, J. W. Fisher, III, and L. K. Hansen, ''Dreaming more data: Class-dependent distributions over diffeomorphisms for learned data augmentation,'' *CoRR*, vol. abs/1510.02795, Oct. 2015.

[28] M. Paulin, J. Revaud, Z. Harchaoui, F. Perronnin, and C. Schmid, ''Transformation pursuit for image classification,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3646–3653.

[29] L. Perez and J. Wang, ''The effectiveness of data augmentation in image classification using deep learning,'' *CoRR*, vol. abs/1712.04621, Dec. 2017.

[30] J. Redmon. (2016). *Darknet: Open Source Neural Networks in C*. [Online]. Available: https://pjreddie.com/Darknet/

[31] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, ''Mobilenets: Efficient convolutional neural networks for mobile vision applications,'' 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[32] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

SINA AKBARIAN received the B.S. degree in biomedical (bioelectrical) engineering from the Amirkabir University of Technology, Tehran, Iran, in 2016. He is currently pursuing the M.A.Sc. degree in biomedical engineering with the University of Toronto, Toronto, Canada. He is also a Graduate Student Researcher with the Toronto Rehabilitation Institute, University Health Network. His research interest includes the development of automated technologies for health monitoring.

GHAZALEH DELFI received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2017. She is currently pursuing the M.S. degree in biomedical engineering with the University of Toronto, Toronto, Canada. She has been a part of the Digital Signal Processing Laboratory, Sharif University of Technology, from 2014 to 2017. She is also a member of the Home and Community Team, Toronto Rehabilitation Institute, University Health Network. Her research interests include signal and image processing and computer vision.

KAIYIN ZHU received the B.A.Sc. degree in engineering science, majoring in biomedical engineering, and the M.Eng. degree in industrial engineering from the University of Toronto, Toronto, ON, Canada, in 2014 and 2016, respectively. In 2016, she joined the Toronto Rehabilitation Institute, University Health Network, as a Research Analyst. Her research interests include vision-based monitoring of sleep with the ultimate goal of developing an inexpensive, non-contact, and unobtrusive tool for the detection of sleep apnea.

AZADEH YADOLLAHI (M'02) received the B.S. degree in electrical engineering and the M.Sc. degree in biomedical engineering from the Sharif University of Technology, in 2002 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of Manitoba, Canada. From 2011 to 2014, she was a Postdoctoral Fellow with the University of Toronto, Canada, and the Toronto Rehabilitation Institute, University Health Network. She has been a Scientist with the Toronto Rehabilitation Institute, since 2014. She is also an Assistant Professor with the Institute of Biomaterials and Biomedical Engineering, University of Toronto. Her research interests include the application of signal processing in health monitoring, respiratory disorders during sleep, and medical devices.

BABAK TAATI (M'96) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 1997, the M.A.Sc. degree in engineering science from Simon Fraser University, Burnaby, BC, Canada, in 2002, and the Ph.D. degree in electrical engineering from Queen's University, Kingston, ON, Canada. From 2008 to 2009, he was the Lead Computer Vision Scientist with Feeling Software, Montreal, Canada. From 2009 to 2012, he was a Postdoctoral Fellow with the University of Toronto, Toronto, ON, Canada, and also with the Toronto Rehabilitation Institute. He has been a Scientist with the Toronto Rehabilitation Institute, University Health Network, Toronto, since 2012. He is also an Assistant Professor with the Department of Computer Science and the Institute of Biomaterials and Biomedical Engineering, University of Toronto. His research interests include the application of computer vision systems in health monitoring, rehabilitation, and diagnostic technologies.