Patient's Body Motion Study using Multimodal RGBDT Videos

Haque, Mohammad Ahsanul; Kjeldsen, Simon S.; Arguissain, Federico G.; Brunner, Iris; Nasrollahi, Kamal; Andersen, Ole Kæseler; Nielsen, Jørgen F. ; Moeslund, Thomas B.; Jørgensen, Anders

Published in:
Advances in Visual Computing

DOI (link to publication from Publisher):
10.1007/978-3-030-03801-4_48

Publication date:
2018

Document Version
Early version, also known as pre-print

Patient's Body Motion Study using Multimodal RGBDT Videos

Mohammad A. Haque[1,4], Simon S. Kjeldsen[2], Federico G. Arguissain[3], Iris Brunner[2], Kamal Nasrollahi[1], Ole K. Andersen[3], Jørgen F. Nielsen[2], Thomas B. Moeslund[1], and Anders Jørgensen[1]

[1] Visual Analysis of People Lab, Aalborg University, Denmark
[2] Regionshospitalet Hammel Neurocenter, Denmark
[3] SMI, Dept. of Health Science and Technology, Aalborg University, Denmark
[4] Corresponding author's email: `mah@create.aau.dk`

Abstract. Automatic analysis of body movement to identify physical activity of patients who are at bed rest is crucial for treatment or rehabilitation purposes. Existing methods of physical activity analysis mostly focused on the detection of primitive motion/non-motion states in unimodal video data captured by either RGB or depth or thermal sensor. In this paper, we propose a multimodal vision-based approach to classify body motion of a person lying on a bed. We mimicked a real scenario of 'patient on bed' by recording multimodal video data from healthy volunteers in a hospital room in a neurorehabilitation center. We first defined a taxonomy of possible physical activities based on observations of patients with acquired brain injuries. We then investigated different motion analysis and machine learning approaches to classify physical activities automatically. A multimodal database including RGB, depth and thermal videos was collected and annotated with eight predefined physical activities. Experimental results show that we can achieve moderately high accuracy (77.68%) to classify physical activities by tracking the body motion using an optical flow-based approach. To the best of our knowledge this is the first multimodal RGBDT video analysis for such application.

Keywords: Physical activity · Multimodal · RGBDT · Video · Rest activity · Patient on bed.

1 Introduction

Tracking of physical activity (PA) or body motion has received great interest and use in the health care sector, mainly because of the importance of assessing the amount of PA that is performed by healthy individuals and by patients with different conditions [1–3]. The information provided by PA tracking is regularly used to develop recommendations and guidelines for promoting a healthy lifestyle, and to improve the outcome of rehabilitation therapies. Particularly, tracking PA while subjects are lying on the bed has been widely used in sleep analysis and also in many other health care applications like breathing, epilepsy, vital signs and activity monitoring [4, 5]. Technological modalities such as actigraphy and global positioning systems are emerging tools in the health care sector that provide objective measurements of PA [6]. Particularly, actigraphy is an accelerometer-based method which is widely used to track PA [7, 8]. Actigraphy has also been used to quantify rest activity-cycles and to monitor sleep in recumbent persons, in patients with severe traumatic brain injury and in intensive care patients [9]. These studies show that actigraphy is a useful tool in these situations. However, it is very likely that wakefulness is underestimated in these bedridden, inactive awake patients possibly due to the low specificity of actigraphy in detecting wakefulness and the poor agreement between actigraphy devices placed in different positions of the body [10]. Besides the use of actigraphy, video-based analysis of PA or body motion has drawn notable interest over the last three decades due to inherent unobtrusiveness, low cost and ubiquity [11, 12].

When a subject is lying on a bed, tracking PA or body motion has been mostly focused on detecting the presence of motion and non-motion states in temporal dimension [13–15]. However, certain health care applications would benefit from detecting more than just motion and non-motion states. In fact, human body motion is a broad concept that includes as many different movements as the human body can exhibit [12]. While some of these motions mostly represent the primitive configuration change of the body parts over time (e.g. movement of fingers or head), some motions can be interpreted as different perceptual classes like pose, gesture and expression.This concept suggests that meaningful body motion information could be extracted by analyzing perceptual physical activity classes in situations where patients are lying on hospital beds. For example, neurorehabilitation institutions that monitor patients with Acquired Brain Injury (ABI) would take advantage of an automatic detection and identification of spontaneous motor activity in an otherwise paralytic extremity. This information could be crucial for the prognosis of these patients [16, 17]. In such a scenario, different PAs associated with motor activity need to be classified in order to document patient's activity level and variation of activities. However, there are two categories of challenges associated with this kind of application: a) methodological challenges and b) resource-related challenges.

Methodological challenges are associated to the procedure of tracking body motion or PA in videos. Three popular methods have been previously used in the literature: frame differencing, optical flow, and block matching [14]. Frame differencing is used to identify areas with motion, and optical flow and block matching are used to estimate local displacement. While considering video data, the differences in choices between different video sensors such as RGB, thermal and depth were also observed [4]. Resource-related challenges are mostly associated with lack of availability of databases to evaluate the performance of a methodological proposal for tracking PA [5].

In this paper, we envision to develop a system that will facilitate patient monitoring in a "patient in bed" scenario of a neurorehabilitation or care-giving institution. Thus, as a preliminary step, we propose here a multimodal vision-based approach to detect and classify body motion in an experimental setting with healthy volunteers that mimicked a real scenario of "patient on bed". Unlike the traditional methods that mostly focused on detecting primitive motion/non-motion states in unimodal video data captured by either of RGB or depth or thermal sensors, our contributions are:

– Recognize the scenario of "patient on bed" to categorize relevant PAs.

- Assess different motion analysis and machine learning approaches to detect and classify PAs.
- Collect a multimodal RGBDT video database with healthy volunteers and exploit its multi-modality to achieve better classification accuracy.

To the best of our knowledge this is the first multimodal RGBDT video analysis for such applications.

The rest of the paper is organized as follows. Section 2 explains the relevant scenario and categories of PAs considered in this study. Section 3 describes the database along with acquisition and processing challenges. Section 4 illustrates the proposed methodology of detecting and classifying PAs and Section 5 shows the performance of the system on the collected database. Finally, Section 6 concludes the paper.

2 Scenario and Categorization of the PAs

The main motivation for this study is to develop a tool to monitor patients with ABI who are admitted to Hammel Neurorehabilitation and Research Center (HNRC), Denmark. This is a specialized neurorehabilitation hospital, which is responsible for rehabilitation of patients with ABI from Western Denmark. The developed tool could potentially be used to help HNRC health workers in neurorehabilitation to automatically identify spontaneous motor activity during hours without direct monitoring. It could also be used to document and monitor the overall 24-hour motor activity of bedridden patients with disorders of consciousness. This information would be not only very useful in the prognosis of these patients [17], but also with regard to providing the right intervention at the right time.

To initiate the process of developing a video-based, body motion analysis tool that could be implemented in a hospital setting, we first performed the present experimental study on healthy subjects. We first categorized the possible body movement events, that are typically observed by the clinical staff in patients lying in a hospital bed. We then divided them into two categories, based on the presence or absence of a blanket that covered the subject's lower body (Fig. 1). In each category we put five different movements based on the following body segments: foot, wrist, head, leg and arm. Two movements of the whole body were also considered: body turning and moving from lying down in the bed to sitting on the edge of the bed. These movements cover different possible small, moderate and large body movements performed by patients in a real scenario (Fig. 1). Employing healthy volunteers and a predefined sequence of movements allowed us to evaluate the performance of the computer vision-based data processing methodology in a controlled-fashion.

3 The Database

The database was collected from healthy volunteers including both men and women. All volunteers gave written informed consent before participating. To mimic a genuine hospital setting, a patient room at HNRC was outfitted with a hospital bed and relevant monitoring equipment. Volunteers were asked to lie in the bed in a supine position. The

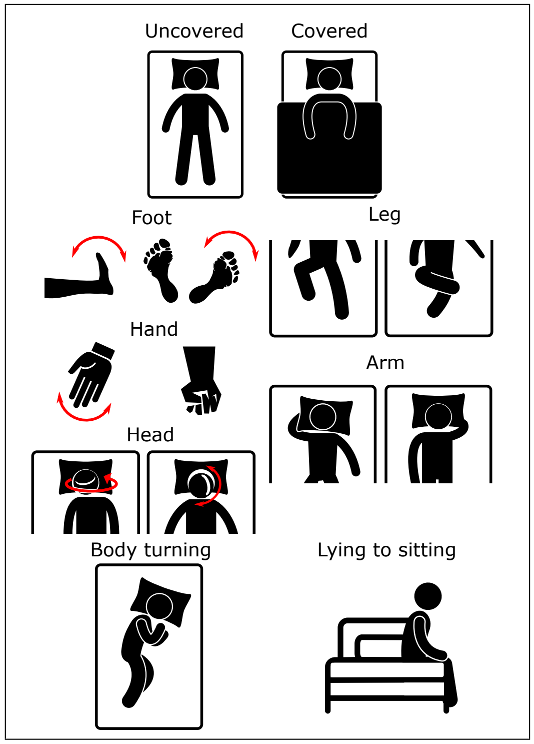

Fig. 1. Different PA performed by healthy volunteers in a hospital setting while mimicking the ”patient on bed” scenario

recordings were performed under regular lighting conditions, and the room temperature was that of the hospital. In order to replicate the hospital setting in the best possible way, we took into account some of the common challenges that are observed at the HNRC ward in terms of standard video monitoring of the rooms. First, it was noticed that cameras are generally located in the ceiling or up in the walls, so they do not interfere with the work of the staff members. Other challenging aspects that have to be taken into account are the positioning of the body relative to the field of view of the camera, the presence of monitoring and life-support equipment, the furniture and the lighting conditions of the room. Other practical challenges are the regular visits of staff members to the room to tend the patients needs. Nevertheless, the presence of a second person in the room was not taken into account in the present analysis. The overall scenario of a hospital room with data collection setup is shown in Figure 2.

Volunteers performed a sequence of relevant predefined movements described in the previous section. To mimic the presence of a possible blanket covering the patients, half of the sequences were performed with a blanket covering the lower body and the trunk. The order of the two conditions (covered, uncovered) was alternated and randomized between subjects. After receiving a full explanation, the subjects were instructed

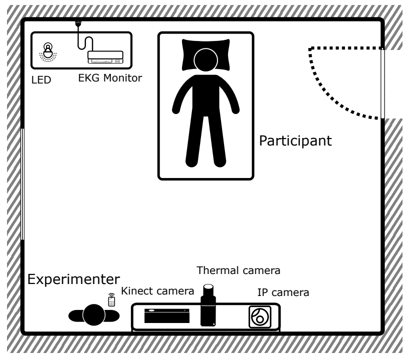

Fig. 2. Data collection setup in a patient room at HNRC

to lie in the bed still and wait for the instructions of the experimenter. Each movement of the sequence was pre-announced verbally by the experimenter, who gave the verbal command 'go' to cue the subject to perform the movement. It was emphasized that the subjects could choose to move relatively free, only constraining their movement to the body segment that was announced before. Subjects were also asked to make movements that resembled those that they would do while resting in bed e.g. scratching hair, crossing legs, putting arm behind their head, etc. Participants performed 70 movements in total (2 conditions, each repeated in 5 sequences of 7 movements). As baseline measurements (non-motion data), we used the frames from the pause between two consecutive movements.

Video cameras were mounted in a wall shelf, in front of the patient and facing the foot-board of the bed. Three video cameras were set to record simultaneously: a Microsoft Kinect V2 for depth, an axis 214 PTZ RGB camera [18], and an axis Q1922 thermal camera. The final database contained the videos (one video contains one event of PA) annotated with eight PA categories (including a non-motion category). RGB and thermal video resolution was 640x480 with fps=30, and depth video resolution was 512x424 with fps<30. The database contained 1127 annotated videos (duration 1-7 seconds) from 9 subjects. Each video contained one motion event.

4 The Methodology of Body Motion Analysis

In this section, we describe the step-by-step procedure to identify perceptual PA from the body motions in a "patient in bed" scenario.

4.1 Data Preprocessing

The multimodal video data was captured by different cameras with different video frame rates. In order to synchronize the video data across modalities, we followed the time-stamps of the depth frames and kept only the corresponding frames from the other modalities by following [19]. Before going to motion tracking we needed to reduce the search space in the video frames. This was possible since the patient was lying on the bed while performing PA and the remaining area of the room was not our concern while tracking. We assumed that the bed area in the hospital room is fixed in a specific room setting. Therefore, we cropped the whole bed area as the Region of Interest (ROI). Fig. 3 shows the cropped bed region in each modality while a subject is lying on the bed.

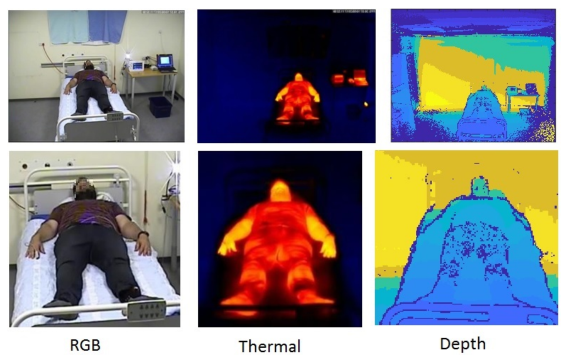

Fig. 3. Original video frames (top row) and the cropped bed area as the ROI (bottom row) in different modalities

4.2 Body Motion Tracking

While detecting motion/non-motion states, the body motion exhibited globally in the ROI could be used as total surface motion. Any presence of surface motion after noise elimination could be considered as a trigger to the presence of body motion [14]. However, as we were considering the classification of 8 perceptual PA categories, global surface motion was not sufficient. Instead, we needed to find the local motions within the ROI and the intensity of motion as well. This was needed because in some situations the movements of different body segments might fall in the same region. For example, a movement of the wrist may occur in a similar region as a movement of the arm. However, the intensities of movement are not same, as the range of movement for the wrist is significantly lower than that of the whole arm. Thus, in order to track the movement and its intensity, we employed a point-based motion tracking approach which uses a method called Good Features to Track (GFT) to detect appropriate pixels in the ROI and a method called Lucas-Kanade-Tomasi (LKT) feature tracker to track those points [20, 21]. The process is described below.

The pattern of pixel intensities changes due to body movement in consecutive video frames. This intensity change can be expressed by an affine motion model in order to form a tracking algorithm. Let I and J be two consecutive video frames. The two quantities $I(\mathbf{x}) = I(x, y)$ and $J(\mathbf{x}) = J(x, y)$ present the intensity values of the two images at the coordinate $\mathbf{x} = [x, y]^T$. If $\mathbf{p} = [p_x, p_y]^T$ is an image point on the first frame I, tracking this point in the next frame J is a feature tracking problem. The goal of feature tracking is to find the location $\mathbf{q} = \mathbf{p} + \delta = [p_x + {}_x, p_y + {}_y]^T$ on the second frame J, such as $I(\mathbf{p})$ and $J(\mathbf{q})$ are similar. The vector $\delta = [\delta_x, \delta_y]^T$ is called optical flow or image velocity at the point $\mathbf{x} = [x, y]^T$. In a practical scenario the notion of similarity between two points in two frames are defined by using the so-called neighborhood sense or window of pixels. Thus, tracking a window of size $w_x \times w_y$ in the frame I to the frame J can be defined on the point velocity parameter δ by minimizing a residual function f_{GFT} as follows:

$$f_{GFT}(\delta) = \sum_{x=p_x}^{p_x+w_x} \sum_{y=p_y}^{p_y+w_y} \left(I(\mathbf{x}) - J(\mathbf{x}+\delta)\right)^2 \tag{1}$$

where $(I(\mathbf{x}) - J(\mathbf{x}+\delta))$ stands for $(I(x,y) - J(x+\delta_x, y+\delta_y))$. The velocity parameter δ is a function of the image position $\mathbf{x}$, and variation in δ are often noticeable even within the small window $w_x \times w_y$ used for tracking. Thus, there can be different displacements within the same window. In order to address this matter an affine motion field was proposed in [21] as follows:

$$\delta = \vartheta \mathbf{x} + \alpha \tag{2}$$

where $\vartheta = \left[\begin{smallmatrix} \vartheta_{xx} & \vartheta_{xy} \\ \vartheta_{yx} & \vartheta_{yy} \end{smallmatrix}\right]$ is a deformation matrix and $\alpha = [\alpha_x, \alpha_y]^T$ is the translation of the feature windows center in video frames. Thus, tracking a point (or window) from image I to image J means determining the 6 parameters of ϑ and α. Bouguet [20] proposed a minimization scheme of f_{GFT} by using a LKT feature tracker [22]. According to the observation in [23, 24], the quality of estimate by this tracker depends on three factors: the size of the window, the texturedness of the image frame, and the amount of motion between frames.

When the ROI is detected after preprocessing, we first detected the appropriate feature points (pixels) in the ROI including body region by GFT. However, tracking only the best few points selected globally over the whole ROI by GFT exhibits the problem of having no points to track on some of the body-parts. This happens due to low-texturedness in some portion of the body in the video frames, which the GFT automatically discard as bad feature points in comparison to high-textured areas. In order to solve this problem, we divided the whole ROI into a grid of certain size (20×20 in the experiment) and then employed the GFT to each of the grid cells separately. Thus, we got certain number of feature points to track in each of the grid cells representing certain body-parts. As a measure of tracking the body motion, we then tracked those pixels in the subsequent video frames by LKT tracker. The tracks provided a motion map in the ROI where the motion of a specific part of the body was reflected by higher intensity level. The motion map has three dimensions: first axis contains the video frame index, second axis contains the tracked feature points' index, and the third axis show the intensity of motion of the feature points over consecutive video frames. However, instead of putting the difference of the location of feature points to represent intensity of body motion, we simply kept the locations of feature points over time in the original map. Thus, among these three dimensions, first axis implied temporal dimension over video, second axis implied spatial dimensions over feature points in the frames, and the third axis implied intensity.

4.3 Feature Extraction for Activity Type Classification

The motion maps for the PA events were different in size, because of the varying length of the individual videos representing individual PA events. Furthermore, the numbers of feature points tracked in the videos by GFT varied also from video to video. Thus, if we consider a two dimensional grid of feature points vs video frames with the intensity of movement in the third dimension, the resulting motion map is not only varying in the intensities in the third dimension, but also in other two dimensions. Keeping this in mind, we extracted the following three categories of features from the motion map to discriminate between the PA categories.

First Difference of the Motion Map (FDiff) This feature is extracted by calculating the first difference of feature map over third dimensions. As the third dimension kept the location of the feature points over time, the first difference gave us the true intensity of the body motion. We employed Euclidean distance metric to calculate the difference between locations. However, this intensity map had a varying size in the first two dimensions as discussed before. So, we employed a grid of 50×50 along with bicubic interpolation in the third dimension to augment the map and obtain our first feature $FDiff$.

Principal Components (PCA) Principal component analysis is widely used to reduce the size of data (or feature) by employing an orthogonal transformation [5]. While reducing the feature size one can effectively select the most discriminative features based on the energy represented by each feature. We started from the $FDiff$ features to extract PCA features from the motion map. We then employed the transformation on the spatial dimension and obtained the features that kept 90% of the discriminative energy.

Primitive Radon Features (PRadon) Radon transform is an integral transform that computes projections of an image matrix along specified directions and that is widely used to reconstruct images from medical computed tomography scan [25]. As our motion map was similar to a two-dimensional image and our focus is to discriminate in the motion map by the location and intensity of PA events, we employed Radon transform on the first difference of the feature map as we did for $FDiff$. The resulting Radon image had 180 line-directions in the spatial dimension, while keeping the original temporal dimension of the motion map. We then resized the temporal dimension with a grid of 180×50 along with bicubic interpolation in the third dimension and obtained the $PRadon$ features.

Radon Distance Features (DiffRadon) we have previously showed that the point-to-point distance of a Radon image also has an effective discriminating property [25]. Thus, we employed a pairwise-distance method between every possible pairs of pixels

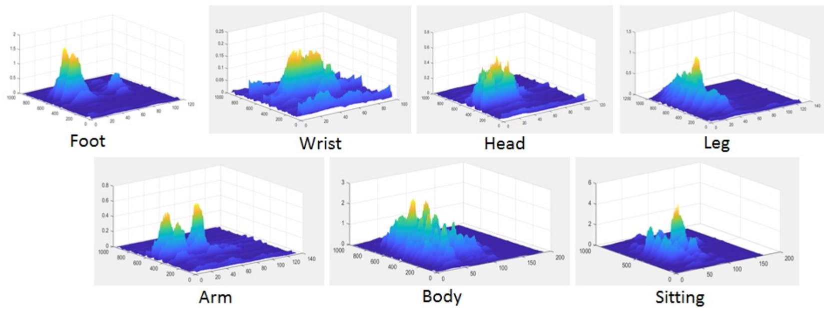

Fig. 4. Example motion intensity map for different PA from a subject's RGB videos

in the PRadon feature vector. The resulting matrix produced a $DiffRadon$ feature vector.

4.4 Classification of Activities

The obtained feature vectors were then fed into a classification framework to identify the PA categories from which each specific video event belonged to. In this paper, we examined 6 well-known shallow learning classifiers to evaluate the discrimination performance: K-Nearest Neighbor with $k = 3$ (KNN), Linear Discriminant Analysis (LDA), Support Vector Machines with linear kernel (SVM), Decision Tree (DT), Naive Bayes (NB), and Generalized Linear Model (GLM) [5].

5 Experimental Environment and Results

The proposed methodology was implemented in MATLAB 2017. We extracted all four categories of features and classifiers described before to find the best performing scenario. The experiment was conducted on individual video modalities first, and then on a feature level concatenation for early fusion of video modalities. We used 80/20 percent train/test ratio of the data.

From the track of the feature points over video, we obtained a motion map, which kept the location of feature points in the consecutive video frames. When we took the first difference of this motion map, we got the intensity of movement in the motion map. Fig. 4 shows the example motion maps for 7-categories of PAs of a subject. It can be observed that the location and intensity of motion varies in different PA occurrences. In fact, this constitutes the rationale of generating such motion maps to find discriminatory features for different PAs.

Table 1 shows the performance of the different extracted features with different classifiers for individual video modalities. Furthermore, Table 2 shows the results when different modalities are fused at the feature level. It can be observed that Radon transform-based features ($PRadon$ and $DiffRadon$) showed better discriminating performance than $FDiff$ and PCA both in individual and fusioned cases. On the other hand, both

Table 1. The performance of different features and classifiers for video modalities (row-best is highlighted by $\mathbf{bold}$).

Features	KNN	DT	NB	SVM	LDA	GLM
RGB (accuracy in %)						
FDiff	35.82	42.47	**49.68**	49.29	30.14	45.96
PCA	55.51	43.94	59.51	**60.00**	56.82	57.47
PRadon	68.25	61.45	12.50	**72.66**	66.74	71.57
DiffRadon	65.76	63.72	61.90	58.63	63.09	**68.02**
Thermal (accuracy in %)						
FDiff	36.21	32.65	**47.49**	44.83	33.38	45.56
PCA	**56.57**	33.29	12.50	46.49	46.40	48.73
PRadon	58.52	63.34	12.50	63.18	41.37	**64.10**
DiffRadon	63.11	58.13	49.43	32.72	59.05	**64.95**
Depth (accuracy in %)						
FDiff	18.22	24.51	**29.19**	28.54	21.68	26.70
PCA	28.26	22.97	32.20	**37.83**	32.21	34.31
PRadon	38.38	42.32	12.50	46.19	**50.81**	46.67
DiffRadon	41.37	37.65	33.67	**47.12**	44.26	45.24

$FDiff$ and PCA showed relatively better performance than the other features when considering the NB and SVM classifiers performance. Notably, a moderately high classification performance was achieved by the GLM classifier using the PRadon features (77.68%). When comparing the different video modalities, we observed that depth features showed lower discriminating ability than the other two modalities(1). Furthermore, RGB and thermal features not only showed better performance individually than the other modalities, but they also provided the best performance when combined (77.68%). Finally, the combination of depth with RGB and thermal features did not increase the performance, which may due to the lack of texturedness in the depth frames.

6 Conclusions

Body motion analysis of a subject lying on a bed may have important applications for diagnostic and rehabilitation purposes. In this paper, we investigated that notion by first defining the taxonomy of some physical activities commonly exhibited by patients while lying on a bed. We collected a multimodal RGBDT video database of those activities from healthy subjects by mimicking a hospital room scenario in a neurorehabilitation center. Relevant feature extraction and machine learning methods were also investigated. The presented work is the first step towards the development of a multimodal video-based tool to assess patient movements in a patient on bed scenario. From the obtained results, new questions for the future work arise. E.g.: can we improve the accuracy of the PA classification by, e.g., employing a human body shape model? Will the system work in real-time, and/or is it possible to make an automated PA log over time by video sensing? The present work is just the beginning of the quest to find the solutions to these challenges.

Table 2. The performance of different features and classifiers for fusioned video modalities (row-best is highlighted by **bold**).

Features	KNN	DT	NB	SVM	LDA	GLM
RGBD (accuracy in %)						
FDiff	30.73	**53.13**	49.30	50.02	34.28	47.72
PCA	52.33	48.28	**63.01**	51.83	48.99	53.77
PRadon	59.20	59.34	12.50	**72.29**	70.30	70.26
DiffRadon	60.40	60.90	61.39	55.84	65.89	**70.20**
RGBT (accuracy in %)						
FDiff	39.70	44.22	**54.97**	49.29	40.32	50.10
PCA	61.19	45.48	12.50	60.82	57.48	**62.62**
PRadon	69.38	60.39	12.50	74.93	71.24	**77.68**
DiffRadon	61.81	67.89	66.04	55.74	68.72	**72.41**
DT (accuracy in %)						
FDiff	27.81	37.33	42.81	**48.08**	31.17	42.62
PCA	42.16	35.19	12.50	46.17	40.61	**47.03**
PRadon	50.50	54.17	12.50	59.40	51.95	**61.12**
DiffRadon	53.91	52.91	49.18	40.98	59.25	**63.96**
RGBDT (accuracy in %)						
FDiff	32.62	46.68	**52.56**	48.24	38.43	47.49
PCA	53.57	41.67	12.50	57.29	52.03	**57.75**
PRadon	61.31	66.60	12.50	**75.28**	72.62	74.58
DiffRadon	69.53	66.25	67.18	51.40	**73.10**	71.46

References

1. Hamer, M., Kivimaki, M., and Steptoe, A., "Longitudinal patterns in physical activity and sedentary behaviour from mid-life to early old age: a substudy of the Whitehall II cohort," *J Epidemiol Community Health* **66**, 1110–1115 (Dec 2012).
2. Telama, R., "Tracking of Physical Activity from Childhood to Adulthood: A Review," *Obes Facts* **2**, 187–195 (2009).
3. Talkowski, J. B., Lenze, E. J., Munin, M. C., Harrison, C., and Brach, J. S., "Patient participation and physical activity during rehabilitation and future functional outcomes in patients after hip fracture," *Arch Phys Med Rehabil* **90**, 618–622 (Apr 2009).
4. Sathyanarayana, S., Satzoda, R. K., Sathyanarayana, S., and Thambipillai, S., "Vision-based patient monitoring: a comprehensive review of algorithms and technologies," *Journal of Ambient Intelligence and Humanized Computing* (Nov 2015).
5. Klonovs, J., Haque, M. A., Krueger, V., Nasrollahi, K., Andersen-Ranberg, K., Moeslund, T. B., and Spaich, E. G., [*Distributed Computing and Monitoring Technologies for Older Patients*], SpringerBriefs in Computer Science, Springer International Publishing, Cham (2016).
6. Gluck, S., Chapple, L. S., Chapman, M. J., Iwashyna, T. J., and Deane, A. M., "A scoping review of use of wearable devices to evaluate outcomes in survivors of critical illness," *Crit Care Resusc* **19**, 197–204 (Sep 2017).
7. Martin, J. L. and Hakim, A. D., "Wrist actigraphy," *Chest* **139**, 1514–1527 (Jun 2011).
8. Sadeh, A., "The role and validity of actigraphy in sleep medicine: an update," *Sleep Med Rev* **15**, 259–267 (Aug 2011).

9. Beecroft, J. M., Ward, M., Younes, M., Crombach, S., Smith, O., and Hanly, P. J., "Sleep monitoring in the intensive care unit: comparison of nurse assessment, actigraphy and polysomnography," *Intensive Care Med* **34**, 2076–2083 (Nov 2008).
10. Kamdar, B. B., Kadden, D. J., Vangala, S., Elashoff, D. A., Ong, M. K., Martin, J. L., and Needham, D. M., "Feasibility of Continuous Actigraphy in Patients in a Medical Intensive Care Unit," *Am. J. Crit. Care* **26**, 329–335 (Jul 2017).
11. Roebuck, A., Monasterio, V., Gederi, E., Osipov, M., Behar, J., Malhotra, A., Penzel, T., and Clifford, G. D., "A review of signals used in sleep analysis," *Physiol Meas* **35**, 1–57 (Jan 2014).
12. Poppe, R., "Vision-based human motion analysis: An overview," *Computer Vision and Image Understanding* **108**(1), 4 – 18 (2007). Special Issue on Vision for Human-Computer Interaction.
13. Heinrich, A., Geng, D., Znamenskiy, D., Vink, J. P., and de Haan, G., "Robust and sensitive video motion detection for sleep analysis," *IEEE Journal of Biomedical and Health Informatics* **18**, 790–798 (May 2014).
14. Cabon, S., Pore, F., Simon, A., Ugolin, M., Rosec, O., Carrault, G., and Pladys, P., "Motion estimation and characterization in premature newborns using long duration video recordings," *IRBM* **38**(4), 207 – 213 (2017).
15. Heinrich, A., van Heesch, F., Puvvula, B., and Rocque, M., "Video based actigraphy and breathing monitoring from the bedside table of shared beds," *Journal of Ambient Intelligence and Humanized Computing* **6**, 107–120 (Feb 2015).
16. Ilyas, C., Nasrollahi, K., Moeslund, T., Rehm, M., and Haque, M., [*Facial Expression Recognition for Traumatic Brain Injured Patients*], vol. 4, 1, SCITEPRESS Digital Library (2018).
17. Duclos, C., Dumont, M., Blais, H., Paquet, J., Laflamme, E., de Beaumont, L., Wiseman-Hakes, C., Menon, D. K., Bernard, F., and Gosselin, N., "Rest-Activity Cycle Disturbances in the Acute Phase of Moderate to Severe Traumatic Brain Injury," *Neurorehabil Neural Repair* **28**, 472–482 (Jun 2014).
18. Haque, M. A., Nasrollahi, K., and Moeslund, T. B., "Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera," in [*2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*], 443–448 (Aug 2013).
19. Haque, M. A., Bautista, R. B., Noroozi, F., Kulkarni, K., Laursen, C. B., Irani, R., Bellantonio, M., Escalera, S., Anbarjafari, G., Nasrollahi, K., Andersen, O. K., Spaich, E. G., and Moeslund, T. B., "Deep multimodal pain recognition: A database and comparison of spatio-temporal visual modalities," in [*2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*], 250–257 (May 2018).
20. Bouguet, J., "Pyramidal implementation of the lucas kanade feature tracker," *Intel Corporation, Microprocessor Research Labs* (2000).
21. Shi, J. and Tomasi, C., "Good features to track," in [*1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*], 593–600 (Jun 1994).
22. Baker, S. and Matthews, I., "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision* **56**, 221–255 (Feb 2004).
23. Haque, M. A., Irani, R., Nasrollahi, K., and Moeslund, T. B., "Heartbeat rate measurement from facial video," *IEEE Intelligent Systems* **31**, 40–48 (May 2016).
24. Haque, M. A., Irani, R., Nasrollahi, K., and Moeslund, T. B., "Facial video-based detection of physical fatigue for maximal muscle activity," *IET Computer Vision* **10**(4), 323–329 (2016).
25. Haque, M. A., Nasrollahi, K., and Moeslund, T. B., [*Heartbeat Signal from Facial Video for Biometric Recognition*], 165–174, Springer International Publishing, Cham (2015).