Automatic Task Analysis Based on Head Movement

Robert Makepeace, *Member, IEEE* and Julien Epps, *Member, IEEE*

***Abstract*— Analysis of movement using accelerometers mounted on the torso or limbs has shown good potential for the recognition of physical activities. However many contemporary lifestyle tasks are sedentary, and less is known about how these can be automatically characterized using movement signals. This paper proposes possibly the first system that employs head movement for recognizing different levels of mental activity and for discriminating between various kinds of sedentary and non-sedentary tasks. Results from analysis of a 20-participant database show that head movement is surprisingly indicative of cognitive load and discriminative between different task types, as well as exhibiting some sensitivity to the instant of task change. Given the possibility for wearing hats or glasses with embedded inertial measurement units, this suggests a range of interesting future applications, including monitoring of sedentary daily activities, and developing rough estimates of mental exertion.**

I. Introduction

Since its introduction to biomedical engineering, accelerometry has proven invaluable to a range of applications, from movement classification [1] to energy expenditure [2] and walking speed estimation [3], and this has been spurred on further by the advent of consumer wearable computing. In the vast majority of previous investigations, accelerometers have been mounted on the person's hip, torso or in some cases their limbs [4], and the type of analysis mainly concerns gross motor movements. Recent years have also seen great growth in studies focusing on automatically detecting particular types of human activity or tasks. Key examples relevant to this paper include detecting typical daily tasks [4], classification of workshop-style tasks [5], discrimination between 'light' and 'heavy' assembly activities [6] based on accelerometer signals, and detection of transitions between physical tasks [7].

Head and neck movement has previously been a subject of interest in biomedical injuries, for example Giansanti [8] applied Inertial Measurement Unit (IMU) sensors to model head movements. However, in their work, movement reconstruction from their proposed models proved difficult due to sensor inaccuracies. Ideally, head-mounted sensors should be placed as close as possible to the centre of gravity to capture the true movement. Biomedical literature suggests that the centre of mass for the head/neck system is at eye level towards the back of the skull [9]. There has also been some preliminary work on task analysis using head movement, although this was done using multiple cameras, which limits the mobility of the system [10].

Robert Makepeace is with the School of Electrical Engineering and Telecommunications, UNSW Australia (e-mail: robmakepeace@hotmail.com).

Julien Epps is with the School of Electrical Engineering and Telecommunications, UNSW Australia, and the ATP Laboratory, National ICT Australia (e-mail: j.epps@unsw.edu.au).

In the area of mental task analysis, considerable previous research has been devoted to the classification of pre-defined heterogeneous mental tasks using electroencephalography (EEG), e.g. [11], and the classification of cognitive load level using behavioral signals, e.g. [12, 13], which provides guidance for the study of mental tasks in the current context.

The investigation of automatic task analysis using head movement reported herein is motivated by the convergence of a number of key considerations. Firstly, although there has been significant investigation of movement related to physical activity, there has been very little related to mental activity. Secondly, many tasks of interest, and particularly mental tasks, are sedentary, involving mainly fairly limited movements of the head and upper torso. This partly motivates the third consideration, which is that systems based on head movement are relatively less studied to date. Finally, there is a key practical motivation in the growing ubiquity and acceptance of wearable devices and most recently in particular head-worn devices. Here we investigate (i) how accurately a range of different task types can be classified using head movement, (ii) whether head movement can estimate the level of cognitive load, and (iii) how well the instant of task transition can be detected from head movement.

II. Database

A. Task Design

In the absence of any existing database for task analysis using head movement signals, an ethics-approved experiment was constructed to elicit and record normal behavioral responses to a range of different task types. Twenty volunteer university students participated after giving their formal consent, and completed the following tasks wearing the recording device attached to a hat (as seen in Figure 1):

(a) a familiarization task designed to distract participants from the fact their head movement was being monitored; (b) following a practice iteration, in a random order participants (i) watched a video, (ii) read a document, (iii) wrote on a sheet of paper, (iv) attempted to solve a Rubik's cube, (v) walked around the room and (vi) held a discussion with the experimenter. Each task lasted two minutes, with a buzzer requiring participants to switch to the next task; (c) arithmetic tasks of low (e.g. 14+3), medium (e.g. 56+25) and high (e.g. 528+693) difficulty were presented visually in a random order, and after each task participants rated the perceived level of difficulty on a 10-point scale as a secondary means of verifying the cognitive load level ground truth, as per [12, 13, 14, 15, 16]. Calculation accuracy and response time (time to solution) were also recorded; (d) an open-ended recording of the participants, to attempt to simulate their typical real world behavior. Participants were recorded for 10 minutes and their activities annotated, and were instructed only that where possible they should include some tasks from within set (b) and some from outside (eg phone use, internet browsing, homework, etc). Task (d) was annotated by the experimenter to provide the five-second periods during which task transition occurred and the ground truth for the task type.

B. Accelerometer Configuration and Recordings

A recording device was designed and constructed to capture head movement using an Inertial Movement Unit (IMU), an Arduino microcontroller, Bluetooth communications and a battery (see Figure 1). The device recorded linear acceleration, angular velocity and magnetic field strength at a rate of 20Hz, i.e. nine separate signal values together with a time stamp for each sample. A similar device was also attached to the hip of the participants, to provide a reference for comparison with previous studies. The database consisted of time stamped and synchronized raw data for the head and hip devices with annotated classes of activity, task transition times and cognitive load level (arithmetic only). In total the database comprises 120s of data per activity per person for the tasks in (b) above, around the same for the arithmetic tasks ((c) above – note that the task duration was varied based on the time taken to compete the arithmetic questions), and 10 minutes per person of open-ended recording of the participants to attempt to simulate their typical real world behavior ((d) above).

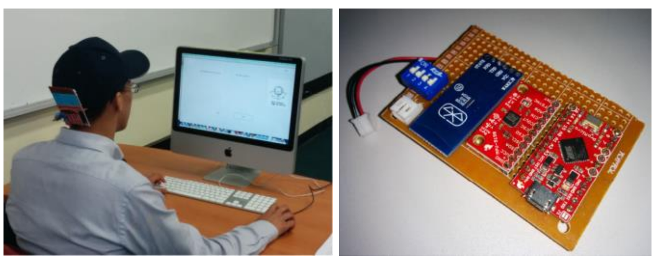

Figure 1. Experimental setup for an example desk-based task, with hat-mounted recording equipment (left), and close-up of the Bluetooth/IMU/microcontroller hardware (70×40×10mm) (right)

III. Methods

A. Pre-processing and Feature Extraction

Following signal acquisition, the raw signals were median filtered (length 3 samples) and low pass filtered (cut-off frequency 0.33 Hz) to remove artifacts. Sensor fusion was applied, to compute the relative orientation of the device with respect to the earth. The proposed system used gyroscopic rotation compensated by gravity and magnetic directions due to drift errors, and the Mahony algorithm was chosen for the purpose; this uses a feedback loop with integral and proportional error terms. From the orientation, the actual velocities and positions can be computed.

Initial experiments suggested that the energy of head movements is mostly in the frequency range less than 1 Hz, so the analysis window size was set to 1s, with 50% overlap (confirmed during preliminary experiments). A wide range of features were considered, based on the principle of characterizing the statistical properties of each of the accelerometer, gyroscope and magnetometer axis signals (i.e. in the *x*, *y* and *z* directions) and their first-order differences across one frame. These first-order differences are referred to as ‘jerk’:

$$J_x[n] = x[n] - x[n-1], \tag{1}$$

with similar definitions for the y and z directions, for either acceleration or rotation. Some features also had their long term mean removed, to provide greater sensitivity to local changes.

The statistical features were the mean, standard deviation, skewness, kurtosis, percentiles, interquartile range and correlations between pairs of signals. In general, mean and percentile (10%/90%) statistical features were most often found to be discriminative for the problems investigated.

During extensive preliminary experiments on five other participants across five task types, it was found that the magnetometer features were noisy and unhelpful for classification, and they were excluded from subsequent feature sets, however magnetometer data was retained for sensor fusion. Features extracted over longer periods of time were also considered, however these provided no benefit over features extracted from short-term frames.

B. Classification

Since many different classification approaches have been previously employed for movement and task classification systems, several different classifiers were compared, using the Weka toolkit. These included Naïve Bayes, OneR, Decision Tree (J48), Support Vector Machine (SVM) and *k*-Nearest Neighbors (IBK). Due to the small size of the database, it was expected that classification methods with a low parameter count would generalize better to unseen evaluation data.

During preliminary experiments, majority voting among the post-classification decisions across five frames was trialed, which provided improvements in classification accuracy except during task transitions. Various forms of classifier fusion were also tested, and showed promise, but were not pursued due to concerns of over-fitting.

All results shown in Section IV are based on a 90%/10% split (i.e. 18 participants/2 participants), in a 10-fold cross-validation configuration.

C. Task Recognition

Similarly to [4], we investigated the problem of task classification accuracy, but here based on head movement. Seven task types were used: tasks (b)(i)-(vi) from Section IIA, together with an arithmetic task. This experiment used two minutes of recorded data for each activity for each participant.

D. Cognitive Load Estimation

Similarly to experiments on eye activity [13] and EEG [16], we investigated a three-class cognitive load classification problem based on the low, medium and high arithmetic tasks described in Section IIA.

E. Task Transition Detection

Similarly to experiments on task transition [14], we investigated the two-class problem of determining whether a particular interval of time (5 seconds in this case) contained a transition between two different tasks. This is a detection problem, and requires the detection of transitions between *any* pair of different tasks from among (b)(i)-(vi) from Section IIA and an arithmetic task.

RESULTS AND DISCUSSION

A. *Task Recognition*

For this task, the most discriminative features were acceleration and gyroscope jerk mean, and mean-removed head acceleration. Posture information was generally found to be important for discriminating the head orientation, e.g. towards the table or computer screen (see Figures 2 and 3 for example). Features specific to particular axes seem important for discriminating the type of activities that are somewhat similar, e.g. reading, writing and Rubik's cube solution. The hip acceleration, when included, as well as head acceleration, was very helpful for detecting walking activity, as expected from previous studies.

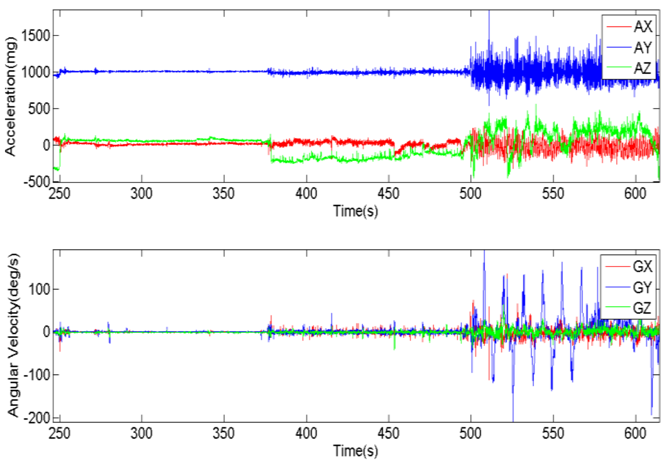

Figure 2. Example accelerometer (top) and gyroscope (bottom) raw signals for three activities: left – watching, middle – Rubik's cube and right – walking. Note the distinct differences in the amount of movement (amplitude), head position (offset) and patterns in movement over time.

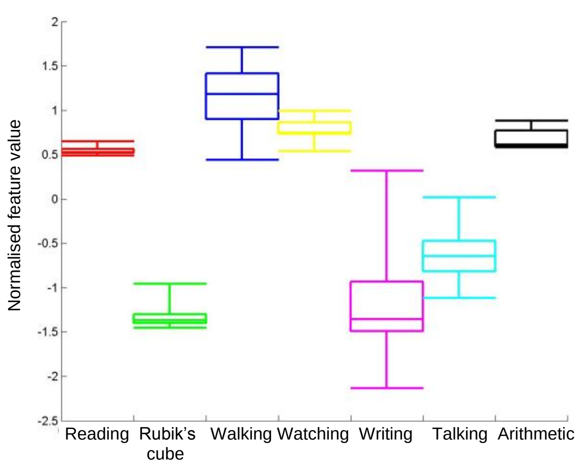

Figure 3. Example feature distributions for different task types, for participant 2, based on the *z*-axis mean acceleration feature. Horizontal lines show (from bottom) minimum, 25% quantile, median, 75% quantile and maximum. Distributions like this from individual participants often have narrow distributions, as the head stays in one area during the task

Classification accuracies for the 7-class problem were generally above 60% for the classifiers tested (see Fig. 4), however the example-based *k*-Nearest Neighbor approach (IBK) seemed to perform better than others, yielding a participant-independent accuracy of 82%, and a participant-dependent accuracy of 90%.

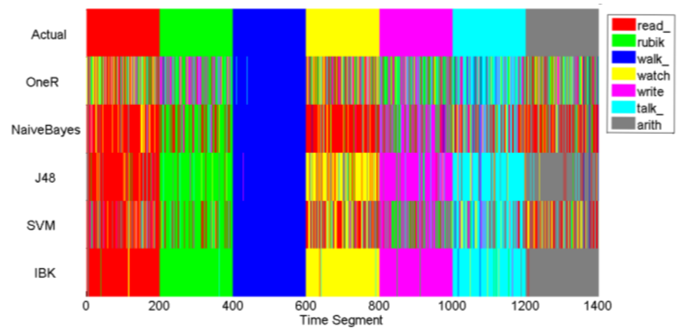

Figure 4. Visual representation of classification accuracy: The top row represents the ground truth across the seven different tasks, while the lower rows represent the corresponding classifier decisions

Associated with these results is an f-measure of 0.84, which is comparable with f-measures from the Opportunity Challenge (recognition of daily life activities comprising five physical locomotion modes), in the range 0.83 – 0.9 [17].

B. *Cognitive Load Estimation*

As seen in Table 1, the experimental paradigm was successful in consistently inducing three substantially different levels of load, as indicated by the self-ratings, calculation accuracy (%) and average time taken to complete the arithmetic solution.

TABLE I. AVERAGES AND STANDARD DEVIATIONS OF INDUCED COGNITIVE LOADS, OVER ALL PARTICIPANTS

Level	**Self-rating (1-10)**	**Accuracy %**	**Response time**
Low	1.42 ± 0.60	0.99 ± 0.02	3.46 ± 0.50
Medium	3.86 ± 1.41	0.97 ± 0.04	8.34 ± 2.24
High	6.62 ± 1.23	0.85 ± 0.17	18.94 ± 6.04

Among features discriminative of cognitive load, the dominance of the mean-removed signals reflects the importance of scale of movements in their relative contexts for individuals. The lower loads produced more active participants, who responded quickly to the new questions and entered the responses. Higher loads tended to reflect more unconscious movements related to thinking about the problem.

The 3-class cognitive load classification accuracy was found to be 67% (participant-independent) and 82% (participant-dependent), using a Nearest-Neighbor classifier (IBK). The significantly higher participant-dependent accuracy may reflect differences in the perceived difficulty of a particular level of arithmetic task or different strategies adopted between different participants, and it may be reasonable to expect that cognitive load would be calibrated to the individual in any clinically deployed system.

The cognitive load classification accuracies are reasonably high considering that cognitive activity typically does not produce very much movement (although the mouse movement during question answering may possibly be a factor here), and compare favorably with other behavioral signals like eye activity (50% for similar 3-class task) [13], and speech (53% for different kind of 3-class task without speaker-specific feature warping [15]). Significantly higher accuracies (98%) have been found for EEG-based estimation of load level across 7 classes of arithmetic task [16], but EEG is more invasive and not always suitable for mobile use.

C. *Task Transition Detection*

The transition features mainly attempt to capture the sudden shifts in motion during the task changes that were induced during data collection. These events are characterized by posture shifts and head movements reflected in the features selected. Distinguishing these movements is difficult as the non-transition state may also contain a wide range of activities. As for cognitive load classification, the mean-removed features dominated, reflecting an increase in relative movement as a person disengages from an activity and moves to start the next activity.

Accuracies for task transition detection (i.e. 2-class classification) were quite high, producing precision results of 93.7% and 97.3% and recall results of 62.3% and 99.7% for non-transition and transition classes respectively, using a Nearest-Neighbor classifier.

These accuracies are significantly higher than previous work on task transition detection using eye activity, which found precision and recall results around 61% and 77% respectively [14]. On one level, this is not surprising, as accelerometry has proven very effective at distinguishing different physical activities [7]. However considering that this experiment includes transitions between physically similar tasks such as reading, writing, watching, arithmetic and Rubik's cube solution, based only on *head movement*, detection accuracies above 90% are highly encouraging.

IV. CONCLUSION

This paper has investigated the use of a head-mounted inertial measurement unit for characterizing the activities that comprise not only physical activities typical of many studies found in the biomedical engineering literature, but sedentary mental activities, which often comprise a significant portion of a patient's day. This study has found that features based on accelerometer and gyroscope jerk are quite sensitive to the type of task performed, and features reflective of posture and unconscious movement (during mental activity) are quite sensitive to changes in mental load, at least for an arithmetic task. Encouragingly, some trends observed seemed to generalize across many participants. The reasonably high classification accuracies for classifying different types of tasks, many of which involve relatively small-scale movements, and the possibility for assessing the level of mental load and transition between tasks, seems to be a call for further research.

Together with expanding the scope of the study to consider other activity types and a larger number and more diverse range of participants, it would be interesting to perform long-term studies of mental activity, to see how well these results generalize. Future work will also concentrate on developing a real time prototype system, possibly combined with other sensors. Inertial measurement units are non-invasive, can be processed in real time using contemporary processors, and have already been shown to be feasible for embedding in a range of wearable consumer devices.

ACKNOWLEDGMENT

This work was partly funded by the US Army International Technology Center (Pacific). NICTA is funded by the Australian Government as represented by the Department of Broadband, Communication and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

REFERENCES

[1] Karantonis, D. M., *et al*. "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring.", *IEEE Trans. Inf. Tech. in Biom.*, 10.1 (2006): 156-167.

[2] Ward, D. S., *et al*. "Accelerometer use in physical activity: best practices and research recommendations." *Medicine & Science in Sports & Exercise* 37 (2005): S582-8.

[3] Le Masurier, G. C., and Tudor-Locke, C.. "Comparison of pedometer and accelerometer accuracy under controlled conditions." *Medicine and Science in Sports and Exercise* 35.5 (2003): 867-871.

[4] Bao, L., and Intille, S. "Activity recognition from user-annotated acceleration data.", *IEEE Pervasive Computing* (2004): 1-17.

[5] Ward, J., et al. "Activity recognition of assembly tasks using body-worn microphones and accelerometers." *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28.10 (2006): 1553-1567.

[6] Hansson, G-Å., et al. "Precision of measurements of physical workload during standardised manual handling. Part II: Inclinometry of head, upper back, neck and upper arms." *Journal of Electromyography and Kinesiology* 16.2 (2006): 125-136.

[7] Ali, R., *et al*. "Detection and Analysis of Transitional Activity in Manifold Space." *IEEE Trans. Inf. Tech. Biom*., 16.1 (2012): 119-128.

[8] Giansanti, D., *et al*. "Is it Feasible to Reconstruct Body Segment 3-D Position and Orientation using Accelerometric Data?", *IEEE Trans. Biomedical Engineering*, 50(4) (2003):476-483.

[9] National Aeronautics and Space Administration, Man-Systems Integration Standards. Number v.3 in NASA-STD, 1995.

[10] Madabhushi, A. and Aggarwal, J. K. "Using Head Movement to Recognize Activity", in *Int. Conf. on Pattern Rec.*, 4 (2000):698-701.

[11] Chen, D., and Vertegaal, R. "Using mental load for managing interruptions in physiologically attentive user interfaces", in *Proc. CHI*, 24.29 (2004).

[12] Chen, F., *et al*. ″Multimodal Behaviour and Interaction as Indicators of Cognitive Load″, *ACM Trans. Intell. Inter. Sys.*, 2.4 (2012) 22.

[13] Chen, S., and Epps, J., ″Automatic Classification of Eye Activity for Cognitive Load Measurement with Emotion Interference″, *Computer Methods and Programs in Biomedicine*, 110.2 (2013): 111-124.

[14] Chen, S., Epps, J., and Chen, F. ″Automatic and Continuous User Task Analysis using Eye Activity″, in *Proc. Int. Conf. on Intelligent User Interfaces* (2013): 57-66.

[15] Yap, T., *et al*. ″Formant Frequencies Under Cognitive Load: Effects and Classification″, *EURASIP J. Advances in Signal Proc.*, 2011.

[16] Zarjam, P., *et al*. ″Estimating cognitive workload using wavelet entropy-based features during an arithmetic task″, *Computers in Biology and Medicine*, 43.12 (2013): 2186-2195.

[17] Chavarriaga, R., *et al*. "The opportunity challenge: A benchmark database for on-body sensor-based activity recognition", *Pattern Recognition Letters*, 34.15 (2013): 2033-2042.