Video Article

Evaluation of a Smartphone-based Human Activity Recognition System in a Daily Living Environment

Edward D. Lemaire[1], Marco D. Tundo[2], Natalie Baddour[2]

[1]The Ottawa Hospital Research Institute, University of Ottawa

[2]Department of Mechanical Engineering, University of Ottawa

Correspondence to: Edward D. Lemaire at elemaire@toh.on.ca

Abstract

An evaluation method that includes continuous activities in a daily-living environment was developed for Wearable Mobility Monitoring Systems (WMMS) that attempt to recognize user activities. Participants performed a pre-determined set of daily living actions within a continuous test circuit that included mobility activities (walking, standing, sitting, lying, ascending/descending stairs), daily living tasks (combing hair, brushing teeth, preparing food, eating, washing dishes), and subtle environment changes (opening doors, using an elevator, walking on inclines, traversing staircase landings, walking outdoors).

To evaluate WMMS performance on this circuit, fifteen able-bodied participants completed the tasks while wearing a smartphone at their right front pelvis. The WMMS application used smartphone accelerometer and gyroscope signals to classify activity states. A gold standard comparison data set was created by video-recording each trial and manually logging activity onset times. Gold standard and WMMS data were analyzed offline. Three classification sets were calculated for each circuit: (i) mobility or immobility, ii) sit, stand, lie, or walking, and (iii) sit, stand, lie, walking, climbing stairs, or small standing movement. Sensitivities, specificities, and F-Scores for activity categorization and changes-of-state were calculated.

The mobile versus immobile classification set had a sensitivity of 86.30% ± 7.2% and specificity of 98.96% ± 0.6%, while the second prediction set had a sensitivity of 88.35% ± 7.80% and specificity of 98.51% ± 0.62%. For the third classification set, sensitivity was 84.92% ± 6.38% and specificity was 98.17 ± 0.62. F1 scores for the first, second and third classification sets were 86.17 ± 6.3, 80.19 ± 6.36, and 78.42 ± 5.96, respectively. This demonstrates that WMMS performance depends on the evaluation protocol in addition to the algorithms. The demonstrated protocol can be used and tailored for evaluating human activity recognition systems in rehabilitation medicine where mobility monitoring may be beneficial in clinical decision-making.

Video Link

The video component of this article can be found at http://www.jove.com/video/53004/

Introduction

Ubiquitous sensing has become an engaging research area due to increasingly powerful, small, low cost computing and sensing equipment [1]. Mobility monitoring using wearable sensors has generated a great deal of interest since consumer-level microelectronics are capable of detecting motion characteristics with high accuracy [1]. Human activity recognition (HAR) using wearable sensors is a recent area of research, with preliminary studies performed in the 1980s and 1990s [2-4].

Modern smartphones contain the necessary sensors and real-time computation capability for mobility activity recognition. Real-time analysis on the device permits activity classification and data upload without user or investigator intervention. A smartphone with mobility analysis software could provide fitness tracking, health monitoring, fall detection, home or work automation, and self-managing exercise programs [5]. Smartphones can be considered inertial measurement platforms for detecting mobile activities and mobile patterns in humans, using generated mathematical signal features calculated with onboard sensor outputs [6]. Common feature generation methods include heuristic, time-domain, frequency-domain, and wavelet analysis-based approaches [7].

Modern smartphone HAR systems have shown high prediction accuracies when detecting specified activities [1,5,6,7]. These studies vary in evaluation methodology as well as accuracy since most studies have their own training set, environmental setup, and data collection protocol. Sensitivity, specificity, accuracy, recall, precision, and F-Score are commonly used to describe prediction quality. However, little to no information is available on methods for "concurrent activity" recognition and evaluation of the ability to detect activity changes in real-time [1], for HAR systems that attempt to categorize several activities. Assessment methods for HAR system accuracy vary substantially between studies. Regardless of the classification algorithm or applied features, descriptions of gold standard evaluation methods are vague for most HAR research.

Activity recognition in a daily living environment has not been extensively researched. Most smartphone-based activity recognition systems are evaluated in a controlled manner, leading to an evaluation protocol that may be advantageous to the algorithm rather than realistic to a real-world environment. Within their evaluation scheme, participants often perform only the actions intended for prediction, rather than applying a large range of realistic activities for the participant to perform consecutively, mimicking real-life events.

Some smartphone HAR studies [8,9] group similar activities together, such as stairs and walking, but exclude other activities from the data set. Prediction accuracy is then determined by how well the algorithm identified the target activities. Dernbach *et al.* [9] had participants write the activity they were about to execute before moving, interrupting continuous change-of-state transitions. HAR system evaluations should assess the algorithm while the participant performs natural actions in a daily living setting. This would permit a real-life evaluation that replicates daily use of the application. A realistic circuit includes many changes-of-state as well as a mix of actions not predicable by the system. An investigator can then assess the algorithm's response to these additional movements, thus evaluating the algorithm's robustness to anomalous movements.

This paper presents a Wearable Mobility Monitoring System (WMMS) evaluation protocol that uses a controlled course that reflects real-life daily living environments. WMMS evaluation can then be made under controlled but realistic conditions. In this protocol, we use a third-generation WMMS that was developed at the University of Ottawa and Ottawa Hospital Research Institute [11-15]. The WMMS was designed for smartphones with a tri-axis accelerometer and gyroscope. The mobility algorithm accounts for user variability, provides a reduction in the number of false positives for changes-of-state identification, and increases sensitivity in activity categorization. Minimizing false positives is important since the WMMS triggers short video clip recording when activity changes of state are detected, for context-sensitive activity evaluation that further improves WMMS classification. Unnecessary video recording creates inefficiencies in storage and battery use. The WMMS algorithm is structured as a low-computational learning model and evaluated using different prediction levels, where an increase in prediction level signifies an increase in the amount of recognizable actions.

Protocol

This protocol was approved by the Ottawa Health Science Network Research Ethics Board.

1. Preparation

1. Provide participants with an outline of the research, answer any questions, and obtain informed consent. Record participant characteristics (*e.g.*, age, gender, height, weight, waist girth, leg height from the anterior superior iliac spine to the medial malleolus), identification code, and date on a data sheet. Ensure that the second smartphone that is used to capture video is set to at least a 30 frames per second capture rate.
2. Securely attach a phone holster to the participant's front right belt or pant waist. Start the smartphone application that will be used to collect the sensor data (*i.e.*, data logging or WMMS application) on the mobility measurement smartphone and ensure that the application is running appropriately. Place the smartphone in the holster, with back of the device (rear camera) facing outward.
3. Start digital video recording on a second smartphone. For anonymity, record the comparison video without showing the person's face, but ensure to record all activity transitions. The phone can be handheld.

2. Activity Circuit

1. Follow the participant and video their actions, on the second smartphone, while they perform the following actions, spoken by the investigator:
    1. From a standing position, shake the smartphone to indicate the start of the trial.
    2. Continue standing for at least 10 sec. This standing phase can be used for phone orientation calibration [14].
    3. Walk to a nearby chair and sit down.
    4. Stand up and walk 60 meters to an elevator.
    5. Stand and wait for the elevator and then walk into the elevator.
    6. Take the elevator to the second floor.
    7. Turn and walk into the home environment.
    8. Walk into the bathroom and simulate brushing teeth.
    9. Simulate combing hair.
    10. Simulate washing hands.
    11. Dry hands using a towel.
    12. Walk to the kitchen.
    13. Take dishes from a rack and place them on the counter.
    14. Fill a kettle with water from the kitchen sink.
    15. Place the kettle on the stove element.
    16. Place bread in a toaster.
    17. Walk to the dining room.
    18. Sit at a dining room table.
    19. Simulate eating a meal at the table.
    20. Stand and walk back to the kitchen sink.
    21. Rinse off the dishes and place them in a rack.
    22. Walk from the kitchen back to the elevator.
    23. Stand and wait for the elevator and then walk into the elevator.
    24. Take the elevator to the first floor.
    25. Walk 50 meters to a stairwell.
    26. Open the door and enter the stairwell.
    27. Walk up stairs (13 steps, around landing, 13 steps).
    28. Open the stairwell door into the hallway.

29. Turn right and walk down the hall for 15 meters.
30. Turn around and walk 15 meters back to the stairwell.
31. Open the door and enter the stairwell.
32. Walk down stairs (13 steps, around landing, 13 steps).
33. Exit the stairwell and walk into a room.
34. Lie on a bed.
35. Get up and walk 10 meters to a ramp.
36. Walk up the ramp, turn around, then down the ramp (20 meters).
37. Continue walking into the hall and open the door to outside.
38. Walk 100 meters on the paved pathway.
39. Turn around and walk back to the room.
40. Walk into the room and stand at the starting point.
41. Continue standing, and then shake the smartphone to indicate the end of trial.

3. Trial Completion

1. Stop the video recording smartphone and ask the participant to remove and return the smartphone and holster. Stop the data logging or WMMS application on the smartphone. Copy the acquired motion data files and the video file from both phones to a computer for post-processing.

4. Post-processing

1. Synchronize timing between the video and the raw sensor data by determining the time when the shake action started. This shaking movement corresponds to a distinct accelerometer signal and video frame. Check for synchronization error by subtracting the end shake time from the start shake time, for sensor and video data sources. Time differences should be similar between the two data sets.
2. Determine actual change-of-state times from the gold-standard video by recording the time difference from the start shake time to the video frame at the transition between activities. Use video editing software to obtain timing to within 0.033 sec (*i.e.*, 30 frames per second video rate). Use WMMS software to generate comparable changes-of-state from the sensor data.
3. Generate two data sets, one with true activities and the second with predicted activities, by labeling the activity for each video frame (based on the change of state timing) and then calculating the predicted activity at each video frame time from the WMMS output. For WMMS performance evaluation, calculate true positives, false negatives, true negatives, false positives between the gold-standard activity and WMMS predicted activity. Use these parameters to calculate sensitivity, specificity, and F-score outcomes measures.
   Note: A tolerance setting of 3 data windows on either side of the window being analyzed can be used for determining change-of-state outcomes, and 2 data windows for classification outcomes. For example, since 1 second data windows were used for the WMMS in this study, 3 sec before and after the current window were examined so that consecutive changes within this tolerance are ignored. The consideration was that changes of state that happen in less than 3 sec can be ignored for gross human movement analysis since these states would be considered transitory.

Representative Results

The study protocol was conducted with a convenience sample of fifteen able-bodied participants whose average weight was 68.9 (± 11.1) kg, height was 173.9 (± 11.4) cm, and age was 26 (± 9) years, recruited from The Ottawa Hospital and University of Ottawa staff and students. A smartphone captured sensor data at a variable 40-50 Hz rate. Sample rate variations are typical for smartphone sensor sampling. A second smartphone was used to record digital video at 1280x720 (720p) resolution.

The holster was fastened to the participant's right-front belt or pant without further standardization of the location. This demonstrated a natural method for placing the device in the hostler on the hip. With the device placed in the holster and the data logger application running, each person traversed the circuit once, at a self-selected pace. The circuit was not described in advance to the participant and proceeding activities were spoken by the investigator sequentially during the trial.

The WMMS consisted of a decision-tree with upper and lower boundary conditions, similar to work by Wu, *et al.*[13]. The revised classifier used a 1 sec window size and features from the linear acceleration signal (sum of range, simple moving average, sum of standard deviation) and gravity signal (difference to Y, variance sum average difference)[15]. Three classification sets were calculated for evaluation: (i) mobility or immobility, (ii) sit, stand, lie, or walking, and (iii) sit, stand, lie, walking, climbing stairs, or small standing movement. Activities of daily living were labeled as small movements. Representative results are shown in **Table 1**.

Classification	TP	FN	TN	FP	Sensitivity (%)	Specificity (%)	F1-Score (%)
Classification Set 1	350	55	8701	91	86.30 ± 7.2	98.96 ± 0.6	86.17 ± 6.3
Classification Set 2	359	47	8660	131	88.35 ± 7.80	98.51 ± 0.62	80.19 ± 6.36
Classification Set 3	423	75	8540	159	84.92 ± 6.38	98.17 ± 0.62	78.42 ± 5.96
Classification Set 1							
Immobile to Mobile	177	19					
Mobile to Immobile	171	36					

During Mobile			3990	73			
During Immobile			4711	18			
Classification Set 2							
Stand to Walk	134	17					
Walk to Stand	137	26					
Walk to Sit	29	0					
Sit to Walk	30	0					
Walk to Lie	11	4					
Lie to Walk	15	0					
During Stand			2872	73			
During Sit			644	9			
During Lie			447	9			
During Walk			4697	40			
Classification Set 3							
Stand to Walk	70	7					
Walk to Stand	74	14					
Walk to Sit	29	0					
Sit to Walk	30	0					
Walk to Lie	15	0					
Lie to Walk	15	0					
Walk to Small Move	68	7					
Small Move to Walk	61	13					
Walk to Stairs	13	2					
Stairs to Walk	13	2					
Small Move to Small Move	35	30					
During Stand			1584	25			
During Sit			643	10			
During Lie			447	15			
During Walk			4398	56			
During Stairs			246	0			
During Brush Teeth			190	12			
During Comb Hair			158	2			
During Wash Hands			152	6			
During Dry Hands			119	4			
During Move Dishes			93	5			
During Fill Kettle			190	5			
During Toast Bread			70	1			
During Wash Dishes			250	18			

Table 1. Results for change-of-state determination; including, true positives (TP), false negatives (FN), true negatives (TN) false positives (FP), total changes-of-state, sensitivity, specificity, and F1-Score. During refers to TN and FP for changes-of-state during the specified action.

From **Table 1**, the mobile versus immobile classification set had a sensitivity of 86.30% ± 7.2% and specificity of 98.96% ± 0.6%, while the second prediction set had a sensitivity of 88.35% ± 7.80% and specificity of 98.51% ± 0.62%. For the third classification set, sensitivity was 84.92% ± 6.38% and specificity was 98.17 ± 0.62. F1 scores for the first, second and third classification sets were 86.17 ± 6.3, 80.19 ± 6.36, and 78.42 ± 5.96, respectively.
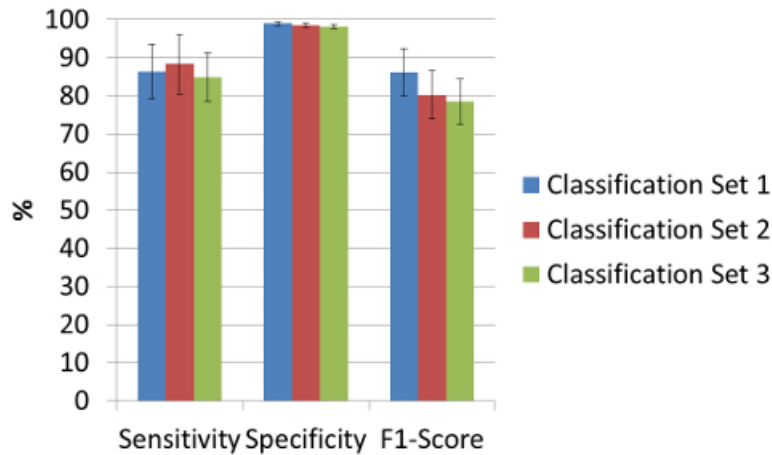
Figure 1. Changes-of-state sensitivity, specificity, and F1-Score for three classification sets.

Discussion

Human activity recognition with a wearable mobility monitoring system has received more attention in recent years due to the technical advances in wearable computing and smartphones and systematic needs for quantitative outcome measures that help with clinical decision-making and health intervention evaluation. The methodology described in this paper was effective for evaluating WMMS development since activity classification errors were found that would not have been present if a broad range of activities of daily living and walking scenarios had not been included in the evaluation.

The WMMS evaluation protocol consists of two main parts: data acquisition under realistic but controlled conditions with an accompanying gold standard data set and data post-processing. Digital video was a viable solution for providing gold standard data when testing WMMS algorithm predictions across the protocol activities. Critical steps in the protocol are (i) to ensure that the gold standard video captures the smartphone shake since this permits synchronization of the gold standard video with data acquired from the participant-worn phone and (ii) to ensure that the gold-standard video records all the transitions performed by the trial participant (*i.e.*, the person recording the gold-standard video must be in the correct position when following the trial participant).

The evaluation protocol incorporates walking activities, a daily living environment, and various terrains and transitions. All actions are done consecutively while a participant-worn smartphone continuously records data from accelerometer, gyroscope, magnetometer, and GPS sensors, and a second smartphone is used to video all activities performed by the trial participant. The protocol may be modified by adapting the order of activities based on the test location, as long as a range of continuous-performed activities of daily living are incorporated. Ten to fifteen minutes was required to complete the circuit, depending on the participant. During pilot tests, some participants with disabilities could only complete one cycle, therefore single trial testing should be considered with some populations to ensure a complete data set.

Limitations of the proposed WMMS evaluation method are that timing resolution is limited to the video frame rate of the camera used to record the gold-standard comparator video and difficulty identifying distinct change-of-state timing from video for activities of daily living. Variation by several frames when identifying a change-of-state leads to differences between the gold-standard and WMMS results that could be due to interpretation of activity start rather than WMMS error. A tolerance at each change-of-state, where no comparisons are made, can be implemented to help account for these discrepancies.

Generally, increasing the number of activities being classified and the categorization difficulty (*i.e.*, stairs, small movements) reduced the average sensitivity, specificity, and F1 score. This may be anticipated since increasing the number of activities increases the chance for false positives and false negatives. Evaluation protocols that only use activities that are advantageous to the algorithm will produce results that are misleading and are unlikely to produce similar results when evaluated under real-world conditions. Hence, the significance with respect to existing methods is that the protocol will result in more conservative results for WMMS systems than previous reports in the literature. However, the results will better reflect outcomes in practice. The proposed method of WMMS evaluations can be used to assess a range of wearable technologies that measure or assist human movement.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

References

1. Lara, O.D., Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys Tutorials.* **15** (3), 1192 - 1209 (2013).
2. Foerster, F., Smeja, M., Fahrenberg, J. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior.* **15** (5), 571 - 583 (1999).
3. Elsmore, T.F., Naitoh, P. Monitoring Activity With a Wrist-Worn Actigraph: Effects of Amplifier Passband and threshold Variations. *DTIC Online.* (1993).
4. Kripke, D.F., Webster, J.B., Mullaney, D.J., Messin, S., Mason, W. Measuring sleep by wrist actigraph. *DTIC Online.* (1981).
5. Lockhart, J.W., Pulickal, T.,Weiss, G.M. Applications of mobile activity recognition. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing.* 1054 - 1058 (2012).
6. Incel, O.D., Kose, M., Ersoy, C. A Review and Taxonomy of Activity Recognition on Mobile Phones. *BioNanoScience.* **3** (2), 145 - 171 (2013).
7. Yang, C.-C., Hsu, Y.-L. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors.* **10** (8), 7772 - 7788 (2010).
8. He, Y., Li, Y. Physical Activity Recognition Utilizing the Built-In Kinematic Sensors of a Smartphone. *International Journal of Distributed Sensor Networks.* Article ID 481580, 10 pages (2013).
9. Vo, Q.V., Hoang, M.T., Choi, D. Personalization in Mobile Activity Recognition System using K-medoids Clustering Algorithm. *International Journal of Distributed Sensor Networks.* Article ID 315841, 12 pages (2013).
10. Dernbach, S., Das, B., Krishnan, N.C., Thomas, B.L., Cook, D.J. Simple and Complex Activity Recognition through Smart Phones. *2012 8th International Conference on Intelligent Environments (IE).* 214 - 221 (2012).
11. Hache, G., Lemaire, E.D., Baddour, N. Mobility change-of-state detection using a smartphone-based approach. *2010 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA).* 43 - 46 (2010).
12. Wu, H.H., Lemaire, E.D., & Baddour, N. Change-of-state determination to recognize mobility activities using a BlackBerry smartphone. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, *EMBC.* 5252 -5255 (2011).
13. Wu, H., Lemaire, E.D., & Baddour, N. Activity Change-of-state Identification Using a Blackberry Smartphone. *Journal of Medical and Biological Engineering.* **32** (4), 265-271 (2012).
14. Tundo, M.D., Lemaire, E., Baddour, N. Correcting Smartphone orientation for accelerometer-based analysis. *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA).* 58 - 62 (2013).
15. Tundo, M. D. Development of a human activity recognition system using inertial measurement unit sensors on a smartphone, at <http://www.ruor.uottawa.ca/handle/10393/30963>. (2014).