IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 12, 2017, accepted February 28, 2017, date of publication March 29, 2017, date of current version May 17, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2689067

Using a Smartwatch to Detect Stereotyped Movements in Children With Developmental Disabilities

YEONGJU LEE AND MINSEOK SONG, (Member, IEEE)
Department of Computer Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Minseok Song (mssong@inha.ac.kr)

This work was supported by the Korea Health Technology R&D Project through the Korea Health Industry Development Institute, Ministry of Health and Welfare, South Korea, under Grant HI14C0765.

ABSTRACT It is important to determine when and why stereotyped movements indicative of developmental disabilities occur in order to provide timely medical treatment. However, these behaviors are unpredictable, which renders their automatic detection very useful. In this paper, we propose a machine learning system that runs on a smartwatch and a smartphone to recognize stereotyped movements in children with developmental disabilities. We train a classifier by tagging data from an accelerometer and a gyroscope in a smartwatch to one of six stereotyped movements made by children and recognized by special educational needs teachers. This classifier can then recognize when a child wearing a smartwatch is making one of the stereotyped movements. These schemes were implemented as a suite of apps used by parents and caregivers. In tests on children and young people with developmental disabilities, the system achieved an average recognition accuracy of 91% when individual training data was used.

INDEX TERMS Activity recognition, assisted living, machine learning.

I. INTRODUCTION

Developmental disabilities affect a diverse group of people suffering from chronic conditions caused by mental or physical impairments. They typically experience difficulties with a number of issues, including language, mobility, learning, self-help, and independent living. In the United States, it has been estimated that approximately 15% of children aged 3 to 17 have one or more developmental disabilities, and that this percentage is increasing over time [1]. For example, the last 12 years have witnessed an increase of 17.1% [1].

Developmental disabilities can be detected early in life, and usually persist indefinitely [2]. One of the major issues for children with such disabilities is the care required from parents, especially because these children have trouble with self-advocacy and defending their own rights [3]. A recent survey suggested that, on average, parents of children with developmental disabilities spend more than 12 hours each weekday, and over 36 hours across each weekend, taking care of their children [3]. Thus, relieving so significant a big burden on these parents is important.

Children with developmental disabilities typically exhibit stereotyped movements: they may perform the same actions over and over again, utter the same phrase repeatedly, and constantly insist on the same routine [4]. Severe stereotyped movements are likely to be interpreted as aggressive, and can result in injury to the child or others. Less severe movements are still likely to be considered socially inappropriate, which complicates the child's social integration. Stereotyped movements also interfere with the acquisition of new skills by disturbing the child's concentration. Unfortunately, it is often unwise to prevent or interrupt these movements, as the child can then become anxious, agitated, or aggressive [5].

Stereotyped movements can be treated in various ways, but all of them require the collection of accurate data to determine when and under what circumstances the movements occur [6]. Therapists have typically measured this behavior using methods such as rating scales while observing the child directly or on video, but this approach is often inaccurate and certainly time consuming. Rating scales are subjective, and fail to capture differences in the form, amount, and duration of behaviors across children. The problems can be addressed by automated detection of stereotyped movements.

Many attempts have been made in the literature to recognize postures assumed by the human body, including in

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

activities such as ambulation and exercise. The most widespread approach involves using machine learning algorithms to convert sensory data into higher-level activity information [7], [8]. Due to the widespread uptake of smartwatches, many people already wear accelerometers and other sensors on their persons, offering a greater number of opportunities for activity detection. This is particularly relevant to children with developmental disabilities, who tend to release unfamiliar things from their bodies but may tolerate watch-shaped things, and many already wear identity bracelets [9].

Smartwatches also provide facilities absent from other wearable sensors, such as multi-modal user interfaces and Bluetooth connectivity. Smartwatches potentially offer the facility to detect stereotyped movements that are otherwise unobtrusive. However, to the best of our knowledge, no study has been reported on the use of commercially available smartwatches for monitoring children with developmental disabilities.

In this paper, we investigate the suitability of smartwatches for detecting stereotyped movements in children with disabilities. We propose an activity recognition scheme divided into training and recognition phases: in the training phase, a classifier is constructed to establish the relationship between sensor data and activities, and in the recognition phase, this classifier decides the activity that is taking place. We go on to present the results from a study of seven children with developmental disabilities, where we look for six well-known stereotyped behaviors. Our scheme was implemented as a suite of apps designed for use by parents and caregivers.

The rest of this paper is organized as follows: We review related work in Section II, propose activity recognition schemes in Section III, and provide implementation details in Section IV. We assess the effectiveness of these schemes by means of a case study in Section V, and draw conclusions in Section VI.

II. RELATED WORK

There are two well-known approaches to the recognition of human physical activity, involving external and wearable sensors [10]. External sensors need to be placed appropriately to record the poses or motions being recorded, and no data can be obtained if the subject moves out of the prescribed range. Wearable sensors have long been used for motion capture in automation, and have more quotidian applications [10].

Bao and Intille [11] placed bi-axial accelerometers at five points on a subject's body to recognize 20 common physical activities in daily life. They concluded that the thigh and wrist were the best locations for sensors for this application. Yang [12] developed an activity recognition system that uses a single three-axis accelerometer in a cellphone to identify when the subject was sitting, standing, walking, running, driving, and cycling. Tapia [13] proposed a system for recognizing physical activities based on multiple body-worn sensors and a decision-tree classifier where feature vectors are chosen based on information gain to recognize simple physical activities.

Lukowicz *et al.* [6] used body-worn microphones and accelerometers, placed at different locations on a subject's body, to recognize workshop activities. They combined data from accelerometers with those concerning the intensity of signals from microphones to associate environmental sounds with the subject's activity. Chung *et al.* [14] proposed a system that uses features obtained from a chest-mounted three-axis accelerometer to classify running and walking activities. Naranjo-Hernández *et al.* [15] employed a three-axis accelerometer located on the back of the subject to monitor human activities with the aim of promoting a healthy lifestyle for elderly people. Curone *et al.* [16] developed a classifier that detects activities irrespective of the orientation of sensors. Ravi *et al.* [17] compared a number of classification algorithms using various feature vectors used for activity recognition. However, all these schemes were mainly developed to recognize ordinary human activities, because of which their feasibility for the recognition of stereotyped movements is questionable.

Accurate recognition of stereotyped movements in subjects with developmental disabilities is difficult because these movements follow unpredictable patterns that may not be repeated. Albinali *et al.* [18] recognized stereotyped motor movements using a wireless three-axis accelerometer and a gyroscope on a patient's chest and wrist. They observed and recorded stereotyped hand flapping and body rocking in both laboratory and classroom environments. Gonalves *et al.* [19] compared the effectiveness of the Microsoft Kinect sensor for gesture recognition algorithms with a system with a three-axis accelerometer attached to the subject's right arm for the detection of stereotyped movements recognized using statistical methods. Plötz *et al.* [20] used a data logger connected wirelessly to sensors on the subject's limbs to detect clinically significant behaviors. However, none of these methods makes use of smartwatches for activity recognition.

Several schemes have been developed for activity recognition using smartwatches. Xu *et al.* [21] proposed a system to recognize finger, hand, and arm gestures using a bespoke smartwatch. They also showed that their smartwatch could recognize characters written using the wearer's index finger. Shen *et al.* [7] traced arm motion using a system based on inertial sensors and a model of the anatomy of the arm to construct a hidden Markov model.

Smartwatches can be used to detect risky situations and dangerous physical activities with the aim of preventing accidents. Lee *et al.* [22] developed a system for detecting driver drowsiness based on a smartwatch. It correlates levels of drowsiness with driving behavior obtained from sensors in the smartwatch. Lee *et al.* [23] (not the same group) developed a location tracking system that correlated domestic activities with the room occupied by the subjects and movements characterized by sensor data from a smartwatch. Lee and Song [24] tested the accuracy of machine learning algorithms for normal persons without developmental disabilities, but conducted no detailed analysis including precision, recall, and Cohen's kappa. To the best of our knowledge, no study has handled the detection of stereotyped movements using a case study on children with developmental disabilities through smartwatches.

III. ACTIVITY RECOGNITION

A. BASIC IDEA

Most smartwatches are now equipped with sensors, of which the most common are gyroscopes and accelerometers, which provide data we use to recognize a variety of activities. Our method of activity recognition involves a training phase followed by a recognition phase. In the training phase, sensor data is collected while stereotyped movements are acted out by teachers specializing with children with developmental disabilities. We consider six stereotyped movements: clenching a fist, waving a hand, swinging an arm, raising an arm, lowering an arm, and throwing [2]. Other movements are classified into a garbage set.

Feature values are obtained from the sensor input data to reduce the effect of irrelevant and redundant data, which increases learning accuracy. To find the best features to use, we determined the information gain, which represents the level of certainty in a particular prediction, for each of 41 feature vector calculations. Table 1 summarizes the feature vectors used in our study [13]. Based on this empirical analysis, we selected 1) the number of sensors, 2) the number of features, and 3) the feature vector types to use.

TABLE 1. Feature vectors [13].

Sensor	Feature Vector	Meaning
Three-axis accelerometer	acl-DCTotalMean	Computed over the summation of all signals over all axes
	acl-DCPostureDist	The differences between the mean values of the X-Y, X-Z, and Y-Z axes per sensor
	acl-DCMean	Mean or average over the signal window
	acl-DCArea	The area under the signal computed by summing samples contained in a given window
	acl-ACAbsArea	The area under the absolute value of the signal computed by simply summing the samples inside a given window
Gyroscope	gy-ACTotalAbsArea	Computed over the summation of all the signals over all axes
	gy-ACAbsArea	The area under the absolute value of the signal computed by simply summing the samples inside a given window
	gy-ACEnergy	Computed from FFT coefficients
	gy-ACEntropy	Measures the degree of "disorder" in the band-pass filtered signals
	gy-ACSkew	A measure of the "peakedness" of signals over a given window

A classifier was then constructed for activity recognition. We chose a decision tree algorithm [25] because it shows a good tradeoff between accuracy and computation time [11], [12], [18].

B. TRAINING PHASE

Training consists of three phases:

1) Data collection: Signals from the three-axis accelerometer with a sensing range of +/-2g and a gyroscope corresponding to each activity were collected at 20 Hz. A moving average technique was used to reduce noise; a 10-point moving average filter was used to match the number of processed data items in real-time.
2) Feature extraction: The processed data was split into time windows with a 50% overlap [18], and the value of each feature vector was calculated from every 10 data points. Thus, four feature vectors were calculated for each input second. The feature vectors used are shown in Table 1.
3) Classifier construction: A classifier was built from the feature vector values and the corresponding activities.

We performed preliminary experiments on three persons to determine the sensor types, feature vectors, and machine learning algorithm to use.

1) CHOICE OF FEATURE VECTORS

Feature vectors have been widely used in activity recognition [12], [13], [21]. We derive an information gain value for each feature vector to determine how well the given feature characterizes an activity [26].

Let N^f be the number of feature vectors, let N_i^I be the number of instances of feature vector i, and let $N_i^{\oplus}$ be the number of instances when stereotyped behavior is deemed to have occurred. If $N_i^{\ominus} = N_i^I - N_i^{\oplus}$, the entropy of all feature vectors E can be calculated as follows:

$$E = -\frac{\sum_i^{N^f} N_i^{\oplus}}{\sum_i^{N^f} N_i^I} \log_2 \frac{\sum_i^{N^f} N_i^{\oplus}}{\sum_i^{N^f} N_i^I} - \frac{\sum_i^{N^f} (N_i^{\ominus})}{\sum_i^{N^f} N_i^I} \log_2 \frac{\sum_i^{N^f} (N_i^{\ominus})}{\sum_i^{N^f} N_i^I}. \quad (1)$$

We can then derive the information gain G_i for feature vector i as follows:

$$G_i = E - \left(-\frac{N_i^{\oplus}}{N_i^I} \log_2 \frac{N_i^{\oplus}}{N_i^I} - \frac{N_i^{\ominus}}{N_i^I} \log_2 \frac{N_i^{\ominus}}{N_i^I}\right). \quad (2)$$

We considered 41 feature vectors [13], based on both the three-axis accelerometer and gyroscope data, for six activities, and calculated the average information gain for each activity. Table 2 shows the six best feature vectors for each sensor type and their information gain values.

TABLE 2. Top 10 feature vectors.

Sensor	Feature Vector	Information Gain	Rank in Total
Three-axis Accelerometer	acl-DCTotalMean	3.14	1
	acl-DCPostureDist	3.02	2
	acl-DCMean	2.85	4
	acl-DCArea	2.76	5
	acl-ACAbsArea	2.01	10
Gyroscope	gy-ACTotalAbsArea	2.97	3
	gy-ACAbsArea	2.43	6
	gy-ACEnergy	2.28	7
	gy-ACEntropy	2.11	8
	gy-ACSkew	2.04	9

The accuracy of recognition is likely to increase if more feature vectors are used, but determining these vectors requires computation. Thus, there is a tradeoff between accuracy and time. We performed experiments to determine the effect of the number of feature vectors computed for each sensor on recognition accuracy. We collected training data from three subjects for each activity, and then calculated the average recognition accuracy using the three-axis accelerometer and the gyroscope.

TABLE 3. Recognition accuracy depending on the number of feature vectors(%).

Activity	Number of feature vectors for each sensor				
	1	2	3	4	5
Clenching	83.3	88.6	91.4	92.8	94.0
Swinging Arm	85.6	91.2	93.2	93.9	94.3
Throwing	88.5	90.4	92.1	93	94.1
Waving Hand	83.9	88.1	93.3	93.9	94.7
Raising Arm	87.3	91.8	93.8	95.1	95.5
Lowering Arm	86.1	88	90.3	91.6	91.9

TABLE 4. Average computation time depending on number of feature vectors(s).

	Number of feature vectors for each sensor				
	1	2	3	4	5
Computation time	0.0418	0.0785	0.1264	0.1639	0.2187

Table 3 shows the effect of the number of feature vectors used on recognition accuracy, and Table 4 shows the average feature vector computation time, on a Samsung Galaxy S7 smartphone. We found that the time difference between activity occurrence and its detection may have an undesirable effect on activity notification when activity determination time exceeds 0.15 seconds. We therefore used three feature vectors for each sensor type.

2) DETERMINATION OF SENSOR TYPE

We compared the accuracy of recognition of stereotyped movements obtained using the accelerometer and the gyroscope on a smartwatch, as well as a combination of both sensors. The results are shown in Table 5. Using both sensors yielded the best performance, which was to be expected because the three-axis accelerometer and the gyroscope measure different characteristics of motion. We used both sensors in the experiments reported in the remainder of this paper.

TABLE 5. Recognition accuracy against sensor types(%).

	Three-axis Accelerometer	Gyroscope	Both
Clenching	84.3	82.7	91.4
Swinging Arm	82.1	81.3	93.2
Throwing	83.7	85.6	92.1
Waving Hand	85.6	87.1	93.3
Raising Arm	88.9	84.3	93.8
Lowering Arm	79.2	78.5	90.3

3) DETERMINATION OF CLASSIFIER

We tried four different classifiers from the WEKA Toolkit [27]: the naive Bayesian (NB) classifier, the *k*-nearest neighbor (*k*NN) classifier, the support vector machine (SVM), and the C4.5 decision tree (DT). For the SVM technique, a radial basis function (RBF) kernel was used, and feature vectors were normalized to between 0 and 1. The accuracy of each algorithm for each activity is shown in Table 6. On the basis of these results, we chose to use the C4.5 decision tree classifier [26].

TABLE 6. Recognition accuracy for each classifier (%).

Classifier	NB	kNN	SVM	DT
Clenching	90.5	86.4	90.1	91.4
Swinging arm	89.4	88.1	92.2	93.2
Throwing	90.4	91.7	93.3	92.1
Waving hand	88.8	90.1	88.0	93.3
Raising arm	90.7	89.5	91.5	93.8
Lowering arm	87.8	85.5	92.9	90.3
Average	89.6	88.6	91.3	92.4

C. RECOGNITION PHASE

As shown in Table 4, the recognition required a good deal of computation. It was therefore performed on a smartphone connected to a smartwatch, as shown in Fig. 1. Data from the smartwatch sensors was sent to the smartphone. This data was averaged, the feature vectors were calculated, and the classifier determined the child's activity.

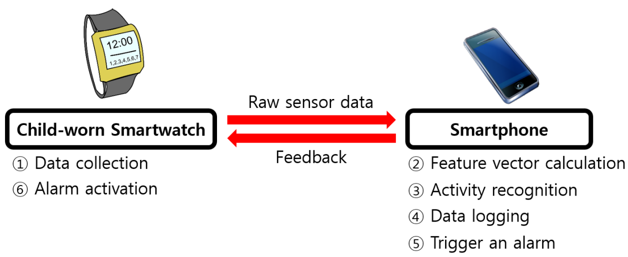


FIGURE 1. Process of the recognition phase.

IV. IMPLEMENTATION

We implemented our scheme on the Android operating system running on a SONY SWR50 smartwatch and a Samsung Galaxy S7. As shown in Fig. 1, our implementation can be divided into five functionalities: data collection, feature vector calculation, activity recognition, data logging, and alarming. To gather sensor data, a smartwatch needs to support the data collection functionality, but most functionalities are built on a smartphone to support the high computation required for activity recognition. The smartwatch was connected to a smartphone via Wi-Fi, a cellular network, and Bluetooth, so that gathered data could be delivered to the caregiver's smartphone. The children could be alarmed by way of sound or vibration when specific activities were detected by the smartphone.

Our implementation can be summarized as follows:

1) Data collection: It involved collecting raw sensor data from sensors in a smartwatch at a rate of 20 Hz and delivering them to the smartphone.

2) Feature vector calculation: By processing raw sensor data, it calculated feature vectors.
3) Activity recognition: A classifier recognized activities that matched feature vectors.
4) Data logging: Tuples of (time, activity) were recorded when activities were recorded.
5) Alarming: When specific activities were detected, they sent an alarm message to the smartwatch, which allowed users to be notified through a vibration.

Fig. 2 shows the interfaces for training and recognition. The screen shown in Fig. 2 (a) allows a user to select one of nine activities for training. The screen shown in Fig. 2 (b) appears, and the user can start, pause, or finish training as shown in Fig. 2 (b). After training, the feature vectors and their labels recorded on the smartphone are output to a machine learning tool [27] to produce the classifier.

FIGURE 2. Examples of user interfaces of our implementation. (a) Training phase: selection of activity. (b) Training phase: collection of clenching training dataset. (c) Recognition phase: start of recognition. (d) Recognition phase: clenching activity is detected.

At the start of the recognition phase, the screen shown in Fig. 2 (c) is displayed on the smartphone. Recognition starts when the button is clicked. When a stereotyped movement occurs, a graphic of the movement is displayed on the screen. Fig. 2 (d) shows the image for clenching. This can be used to alert caregivers. An interface for preventing the frequency with which different stereotyped movements were recognized is also provided.

V. EXPERIMENTAL RESULTS

A. EXPERIMENTAL ENVIRONMENT

We tested our system at a summer camp for young people diagnosed with developmental disabilities, as shown in Fig. 3. Our experiments were designed to evaluate the: (1) accuracy of activity recognition, (2) factors affecting accuracy, such as number of feature vectors and sensors, and (3) the effect of different ways of constructing the classifier.

FIGURE 3. A picture of the case study.

Seven young people (three females and four males) between the ages of 13 and 24 (17.3 on average, with a standard deviation of 3.9) participated in the experiment for two days. The participants were guided by specialized teachers to perform six stereotyped movements (clenching a fist, swinging arms, throwing things, waving hands, raising arms, and lowering arms) in arbitrary order. In this way, on the first day of the experiment, 37,028 items of training data were collected, and 17,436 feature vectors were computed to train the classifier. We compared the results from each classifier by assessing the same movement by two special needs teachers. These teachers were asked to resolve any difference in their assessments, but in fact differed only 1% of the time.

We use the words positive and negative to refer to the classifier's predictions, and true and false to refer to the decision made by the special needs teacher. Thus, each test had four possible outcomes—(true, positive), (true, negative), (false, positive), and (false, negative)—and tp, tn, fp, and fn were, respectively, the number of tests for each outcome.

Precision and recall [28] were then calculated as follows:

$$\text{Precision} = \frac{tp}{tp + fp} \tag{3}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{4}$$

Cohen's coefficient κ was used to measure the probability of agreement between teachers and the classifier. A value of κ above 0.8 is known to indicate very good agreement [29]. Let p_o be the relative agreement between observations, and p_e be the hypothetical probability of chance agreement; then, κ can be expressed as $\kappa = (p_o - p_e)/(1 - p_e)$.

We compared the accuracy of results obtained by three classifiers: one trained on data from the same subject, a second trained on all data, and a third trained on all data except those for the given subject. Along with having two teachers assess each participant's movements as they were made, the movements were recorded on video for subsequent assessment.

B. EXPERIMENT 1: RESULTS FROM CLASSIFIERS TRAINED WITH THE SUBJECT'S DATA

Tables 7 and 8 are, respectively, the confusion matrices [30], [31] for the immediate and video analyses for a classifier trained with each subject's own training data. The overall recognition accuracy, obtained by dividing tp by the number of tests, was 86.31% derived for immediate analysis, and 91.13% for video analysis.

TABLE 7. Confusion matrix for overall recognition result by immediate analysis in Experiment 1(C: Clenching, S: Swinging Arm, T: Throwing, W: Waving Hands, R: Raising Arm, and L: Lowering Arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	456	35	15	5	6	4
	S	0	393	11	49	12	0
	T	5	9	254	7	22	0
	W	21	29	7	551	15	0
	R	0	19	33	7	431	2
	L	1	12	10	7	0	77

TABLE 8. Confusion matrix for overall recognition result by video analysis in Experiment 1(C: clenching, S: swinging arm, T: throwing, W: waving hands, R: raising arm, and L: lowering arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	474	28	6	1	2	1
	S	0	418	0	33	6	0
	T	1	3	278	5	16	1
	W	8	22	1	578	6	0
	R	0	15	37	3	455	1
	L	0	11	8	6	1	80

We also calculated the values of precision, recall, and κ. Table 9 shows the results of immediate analysis and Table 10 of the video analysis. The video analysis gave better results for all activities, with an average precision of 0.89 and a recall of 0.88.

The accuracy, precision, and recall values for each subject are shown in Table 11. The average accuracy was 86.69%.

TABLE 9. Precision, recall, and Cohen's kappa coefficient values of immediate analysis in Experiment 1.

	Precision	Recall	κ
Clenching	0.94	0.88	0.79
Swinging Arm	0.79	0.85	0.87
Throwing	0.77	0.86	0.85
Waving Hands	0.88	0.88	0.78
Raising Arm	0.87	0.88	0.92
Lowering Arm	0.93	0.72	0.78
Mean	0.86	0.85	0.83

TABLE 10. Precision, recall, and Cohen's kappa coefficient values of Video Analysis in Experiment 1.

	Precision	Recall	κ
Clenching	0.98	0.93	0.82
Swinging Arm	0.84	0.91	0.91
Throwing	0.84	0.91	0.86
Waving Hands	0.92	0.93	0.88
Raising Arm	0.94	0.89	0.93
Lowering Arm	0.96	0.75	0.85
Mean	0.91	0.89	0.88

TABLE 11. Accuracy, precision, recall, and κ values per participant in Experiment 1(video analysis).

ID	Accuracy (%)	Precision	Recall	κ
1	89.7	0.85	0.95	0.79
2	84.5	0.84	0.89	0.89
3	75.1	0.8	0.83	0.79
4	87.1	0.86	0.91	0.94
5	86.3	0.89	0.88	0.91
6	92.1	0.96	0.93	0.94
7	92.0	0.93	0.95	0.94
Mean	86.2	0.87	0.9	0.88

Participant 3 showed worse performance than the others. This participant tended not to follow the teacher's instructions, and her movements were ambiguous. By excluding Participant 3, the average accuracy, precision, and recall were improved to 88.62%, 0.89, and 0.91, respectively.

We think that video analysis yielded better results because it is possible to replay a questionable movement as many times as necessary with this medium. Note that the recognition of the "swinging arm" and "throwing" movements was relatively poor; these were whole-arm movements that are difficult to detect using wrist-worn sensors.

C. EXPERIMENT 2: EXPLOITING OTHER TRAINING DATA

Sometimes, it is difficult to acquire training data from a subject, which makes it necessary to use a classifier constructed from other training data. Therefore, we conducted one-left-out experiments, where the movements of each subject were assessed by a classifier trained on data from the other six subjects.

Tables 12 and 13 show confusion matrices for this experiment, and Tables 14 and 15 show the precision, recall, and κ. The average accuracy of these classifiers was 73.67% for immediate analysis and 76.83% for video analysis, lower than the average accuracy of classifiers trained on the subjects' own data, as we might expect.

TABLE 12. Confusion matrix for one-left-out result by immediate analysis (C: clenching, S: swinging arm, T: throwing, W: waving hands, R: raising arm, and L: lowering arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	429	56	14	31	10	12
	S	15	338	32	67	16	13
	T	9	14	214	20	78	18
	W	52	76	9	446	16	3
	R	1	12	52	2	455	1
	L	0	35	43	2	0	102

TABLE 13. Confusion matrix for one-left-out result by video analysis (C: clenching, S: swinging arm, T: throwing, W: waving hands, R: raising arm, and L: lowering arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	441	52	12	25	9	13
	S	10	352	27	64	15	13
	T	8	14	229	17	69	16
	W	48	71	8	457	14	4
	R	1	3	39	2	478	0
	L	0	26	41	2	1	112

TABLE 14. Precision, recall, and Cohen's kappa coefficient value for each activity in immediate analysis in Experiment 2.

	Precision	Recall	κ
Clenching	0.85	0.78	0.71
Swinging Arm	0.64	0.71	0.66
Throwing	0.59	0.61	0.69
Waving Hands	0.79	0.74	0.72
Raising Arm	0.79	0.87	0.67
Lowering Arm	0.68	0.56	0.64
Mean	0.72	0.71	0.68

TABLE 15. Precision, recall, and Cohen's kappa coefficient value for each activity in video analysis in Experiment 2.

	Precision	Recall	κ
Clenching	0.87	0.8	0.72
Swinging Arm	0.68	0.73	0.73
Throwing	0.64	0.65	0.74
Waving Hands	0.81	0.76	0.76
Raising Arm	0.82	0.91	0.71
Lowering Arm	0.71	0.62	0.68
Mean	0.76	0.75	0.72

Precision and recall for the "lowering arm" measurement were significantly less than those in Experiment 1. This was because the manner of this movement was particularly variable among subjects. Compared with Experiment 1, κ fell to between 0.64 and 0.76.

D. EXPERIMENT 3: USING THE SAME CLASSIFIER FOR EACH PARTICIPANT

In this experiment we constructed a classifier based on all the participants' training data, and then used the classifier to recognize the movements of all participants. The results, given in Tables 16, 17, 18, and 19, showed a small improvement over those of Experiment 2, probably because each subject's

TABLE 16. Confusion matrix for Experiment 3 by immediate analysis (C: clenching, S: swinging arm, T: throwing, W: waving hands, R: raising arm, and L: lowering arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	533	52	17	19	9	8
	S	7	444	23	67	16	6
	T	8	13	284	14	54	9
	W	40	58	9	608	18	1
	R	0	19	49	5	529	1
	L	0	25	28	5	0	129

TABLE 17. Confusion matrix for Experiment 3 by video analysis (C: clenching, S: swinging arm, T: throwing, W: waving hands, R: raising arm, and L: lowering arm).

		Predicted					
		C	S	T	W	R	L
Actual	C	552	45	10	13	5	7
	S	5	468	13	55	11	6
	T	4	9	309	12	45	8
	W	29	50	4	633	11	2
	R	0	12	45	3	557	0
	L	0	20	26	5	1	112

TABLE 18. Precision, recall, and Cohen's kappa coefficient value for each activity in immediate analysis in Experiment 3.

	Precision	Recall	κ
Clenching	0.91	0.84	0.81
Swinging Arm	0.73	0.79	0.75
Throwing	0.69	0.74	0.74
Waving Hands	0.85	0.83	0.76
Raising Arm	0.85	0.88	0.93
Lowering Arm	0.81	0.64	0.75
Mean	0.81	0.79	0.79

TABLE 19. Precision, recall, and Cohen's kappa coefficient value for each activity in video analysis in Experiment 3.

	Precision	Recall	κ
Clenching	0.94	0.87	0.83
Swinging Arm	0.77	0.84	0.73
Throwing	0.76	0.8	0.77
Waving Hands	0.88	0.87	0.81
Raising Arm	0.88	0.9	0.94
Lowering Arm	0.83	0.68	0.79
Mean	0.84	0.83	0.81

training data were used. The accuracy did not significantly improve the "clenching" and "raising arms" movements, which could be because they were made in a similar manner by all subjects.

VI. CONCLUSIONS

In this study, we showed that a commercially available smartwatch in communication with a smartphone can be used to detect and classify stereotyped movements in children and young people with developmental disabilities by employing a movement recognition scheme with training and recognition phases. In the training phase, a classifier is constructed to encode the relationship between a particular stereotyped movement and data from a gyroscope and a three-axis accelerometer in the smartwatch. In the recognition phase, this classifier determines when stereotyped movements occur.

We conducted a case study on seven young people with developmental disabilities to look for six well-known stereotyped movements: clenching a fist, waving hands, raising, lowering, or swinging of arms, and throwing things. When the classifier was trained using movements identified from videos (rather than on the spot, which was less consistent), the classifications yielded an accuracy between 77% and 91%, a precision between 0.76 and 0.91, a recall between 0.75 and 0.89, and values of Cohen's kappa between 0.72 and 0.88.

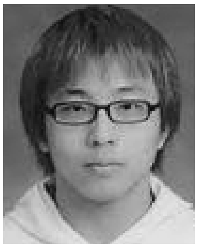
Using training data for users separately improves the accuracy of recognition, probably because the form of stereotyped movements is significantly different from child to child; however, we found that more restricted movements, such as clenching a fist, could be detected effectively with a classifier trained on data obtained from many subjects.

At present, we are extending the proposed method to a context-aware video surveillance system that records video clips when stereotyped movements occur. For future work, we plan to sort out stereotyped movements that correspond to the parts of usual activity to determine risky situations. The use of mobile devices in this application means that power consumption needs to be considered; thus, we also plan to make our scheme adaptable to changing energy budgets.

REFERENCES

[1] C. A. Boyle *et al.*, "Trends in the prevalence of developmental disabilities in US children, 1997–2008," *Pediatrics*, vol. 127, no. 6, pp. 1034–1042, 2011.
[2] S. Shin *et al.*, "Development of treatment casebook and cyber parent education program about each problem behavior," Seoul Nat. Hospital, Seoul, Korea, Med. Rep. 8, Dec. 2014.
[3] B. Choi, Y. Yoo, S. Lim, and Y. Cho, "Survey of children with disabilities and their familes," Korea Disabled People's Develop. Inst., Res. Rep. 13–21, Nov. 2013.
[4] H. S. Singer, "Motor stereotypies," *Seminars Pediatric Neurol.*, vol. 16, no. 2, pp. 77–81, 2009.
[5] V. A. Holm, "Developmental disabilities: Delivery of medical care for children and adults," *J. Amer. Med. Assoc.*, vol. 262, no. 20, pp. 2935–2936, 1989.
[6] P. Lukowicz *et al.*, "Recognizing workshop activity using body worn microphones and accelerometers," in *Proc. Int. Conf. Pervasive Comput.*, 2004, pp. 18–32.
[7] S. Shen, H. Wang, and R. R. Choudhury, "I am a smartwatch and I can track my user's arm," in *Proc. ACM Int. Conf. Mobile Syst., Appl., Services*, 2016, pp. 85–96.
[8] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Proc. Int. Conf. Pervasive Comput.*, 2004, pp. 158–175.
[9] D. Forte-Pathroff, "Child loss prevention system and method of use," U.S. Patent 5 423 574, Jun. 13, 1995.
[10] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 1–33, 2014.
[11] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. Int. Conf. Pervasive Comput.*, 2004, pp. 1–17.
[12] J. Yang, "Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones," in *Proc. Int. Workshop Interact. Multimedia Consum. Electron.*, 2009, pp. 1–10.
[13] E. M. Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure," Ph.D. dissertation, Dept. Archit. Planning, Program Media Arts Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.
[14] W.-Y. Chung, A. Purwar, and A. Sharma, "Frequency domain approach for activity classification using accelerometer," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1120–1123.
[15] D. Naranjo-Hernández, L. M. Roa, J. Reina-Tosina, and M. Á. Estudillo-Valderrama, "SoM: A smart sensor for human activity monitoring and assisted healthy ageing," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 11, pp. 3177–3184, Nov. 2012.
[16] D. Curone, G. M. Bertolotti, A. Cristiani, E. L. Secco, and G. Magenes, "A real-time and self-calibrating algorithm based on triaxial accelerometer signals for the detection of human posture and activity," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 1098–1105, Jul. 2010.
[17] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. Conf. Innov. Appl. Artif. Intell.*, 2005, pp. 1541–1546.
[18] F. Albinali, M. S. Goodwin, and S. S. Intille, "Recognizing stereotypical motor movements in the laboratory and classroom," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2009, pp. 71–80.
[19] N. Gonçalves, J. L. Rodrigues, S. Costa, and F. Soares, "Automatic detection of stereotyped hand flapping movements: Two different approaches," in *Proc. IEEE Int. Symp. Robot Human Interact. Commun.*, Sep. 2012, pp. 392–397.
[20] T. Plötz, N. Y. Hammerla, A. Rozga, A. Reavis, N. Call, and G. D. Abowd, "Automatic assessment of problem behavior in individuals with developmental disabilities," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2012, pp. 391–400.
[21] C. Xu, P. H. Pathak, and P. Mohapatra, "Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch," in *Proc. Int. Workshop Mobile Comput. Syst. Appl.*, 2015, pp. 9–14.
[22] B. L. Lee, B. G. Lee, and W. Y. Chung, "Standalone wearable driver drowsiness detection system in a smartwatch," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5444–5451, Jul. 2016.
[23] S. Lee, Y. Kim, D. Ahn, R. Ha, K. Lee, and H. Cha, "Non-obstructive room-level locating system in home environments using activity fingerprints from smartwatch," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2015, pp. 939–950.
[24] Y. Lee and M. Song, "Recognizing problem behaviors of children with developmental disabilities using smartwatch," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr. 2016, pp. 1001–1002.
[25] C. Ferri-Ramírez, P. Flach, and J. Hernández-Orallo, "Multi-dimensional ROC analysis with decision trees," Dept. Comput. Sci., Univ. Bristol, Bristol, U.K., Tech. Rep. 17-Jan-2002.
[26] T. M. Mitchell, *Machine Learning*. New York, NY, USA, McGraw-Hill, 1997.
[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[28] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
[29] D. G. Altman, *Practical Statistics for Medical Research*. London, U.K.: Chapman & Hall, 1995.
[30] F. Kohavi and F. Provost, "Glossary of terms," *J. Mach. Learn.*, vol. 30, nos. 2–3, pp. 271–274, 1998.

[31] D. Minnen, T. Westeyn, T. Starner, J. A. Ward, and P. Lukowicz, "Performance metrics and evaluation issues for continuous activity recognition," in *Proc. Perform. Metrics Intell. Syst. Workshop*, 2006, pp. 4–11.



Yeongju Lee received the B.S. and M.S. degrees in computer engineering from Inha University, South Korea, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree from the Department of Computer Science and Information Engineering, Inha University. His current research interests include embedded software and ubiquitous computing.

Minseok Song (M'07) received the B.S., M.S., and Ph.D. degrees in computer engineering from Seoul National University, South Korea, in 1996, 1998, and 2004, respectively. Since 2005, he has been with the Department of Computer Science and Information Engineering, Inha University, Incheon, South Korea, where he is currently a Professor. His research interests include embedded systems and multimedia systems.