Hybrid Model–Based Motion Recognition for Smartphone Users

Beomju Shin, Chulki Kim, Jae Hun Kim, Seok Lee, Changdon Kee, and Taikjin Lee

This paper presents a hybrid model solution for user motion recognition. The use of a single classifier in motion recognition models does not guarantee a high recognition rate. To enhance the motion recognition rate, a hybrid model consisting of decision trees and artificial neural networks is proposed. We define six user motions commonly performed in an indoor environment. To demonstrate the performance of the proposed model, we conduct a real field test with ten subjects (five males and five females). Experimental results show that the proposed model provides a more accurate recognition rate compared to that of other single classifiers.

Keywords: Hybrid model, motion recognition, decision tree, artificial neural network, smartphone.

Manuscript received Nov. 18, 2013; revised Mar. 28, 2014; accepted June 16, 2014.

This work was supported by the KIST Institutional Program (Project No. 2E24812) and also supported by Institute of Advanced Aerospace Technology at Seoul National University.

Beomju Shin (bjshin1984@snu.ac.kr) was with the Sensor System Research Center, Korea Institute of Science and Technology, Seoul, Rep. of Korea.

Chulki Kim (chulki.kim@kist.re.kr), Jae Hun Kim (jaekim@kist.re.kr), Seok Lee (slee@kist.re.kr), and Taikjin Lee (corresponding author, taikjin@kist.re.kr) are with the Sensor System Research Center, Korea Institute of Science and Technology, Seoul, Rep. of Korea.

Changdon Kee (kee@snu.ac.kr) is with the Department of Mechanical and Aerospace Engineering, Seoul National University, Rep. of Korea.

I. Introduction

A user's activity or status is essential information for pervasive computing that interacts with users and collects information for improving the quality of human life. The smartphone is a device that can be used to recognize a user's motion by sensing and collecting data from various kinds of embedded sensors [1]. Recognized motion can be used to improve human-centered services, intelligent buildings, location-based services (LBS), and user-context awareness [2]. In particular, recognized motion is useful to enhance the positioning accuracy of LBS in indoor environments where Global Positioning System signals are blocked. In [3], a Wi-Fi-based positioning system that combines physical motion recognition is presented. Pedestrian dead reckoning combined with motion recognition is presented in [4]–[5]. However, in the case of false motion recognition, positioning performance results are less accurate; thus, an advanced classifier that could give a more accurate recognition rate is required.

There have been several researches regarding motion recognition through the use of smartphones [3], [6]. In [6], a decision tree (DT) using simple accelerometer features was implemented. In [3], on the other hand, a support vector machine (SVM)–based classifier for smartphone users was created. Use of the DT in [6] gives a higher performance compared with use of the SVM in [3]. However, in [3], the performance of the SVM-based classifier is better than that of the DT. Usually, each of the algorithms (DT and SVM) is dependent upon defined motion states, selected features, or the position and orientation of the smartphone. Hence, the performance of each algorithm cannot be guaranteed when such factors are likely to vary.

To produce favorable results, an advanced classification

 © 2014 ETRI

technique is required, such as that used in an intelligent hybrid classifier. In this work, we present a hybrid model classifier consisting of DTs and an artificial neural network (ANN) ensemble. We define six user motions commonly performed in indoor environments. In addition, we define two motion groups — motions are assigned according to whether a proximity sensor is enacted by a motion. Next the ANN ensemble chooses a performed motion in a selected group. The experimental results show that the proposed classifier gives higher recognition rates than the single classifiers in [3], [6], [7], and [8].

II. System Description

Many researchers understand the limitations of single classifiers and how their performance is dependent on various experimental conditions. Also, in many applications, the process data to be analyzed can be too large for a single classifier to handle [9]. To overcome such limitations, we propose a hybrid model classifier. Here, the basic concept is derived from the notion that a combined opinion is more reliable than an individual one [10]. A diagram of the proposed hybrid model classifier is presented in Fig. 1. The smartphone gives sensor data to the hybrid model classifier at a sampling rate of 50 Hz. We utilize the accelerometer and proximity sensor for the motion recognition. The input of the hybrid model classifier is the combination of the three-axis output from the accelerometer and the output from the proximity sensor. In Fig. 1, the format of the input data is presented. The output of the hybrid model classifier is the estimated motion of the smartphone user. The first DT, using a variance value of the accelerometer output, determines whether the user is moving. If the first DT determines that the user is moving, then the second DT will determine the motion group to which the performed motion belongs by using an output of the proximity sensor. Each ANN ensemble finally recognizes the user's motion.

We define six motions that are commonly performed by a user carrying a smartphone as they walk in an indoor environment. The defined motions, depicted in Fig. 2, are explained in Table 1.

Table 1. Defined user motions.

Motion ID	Motion explanation	Motion group
M_1	Standing	-
M_2	Walking looking at the device	MG_1
M_3	Walking talking on the device	MG_2
M_4	Walking swinging hands	MG_1
M_5	Running	MG_1
M_6	Walking with the device in pocket	MG_2

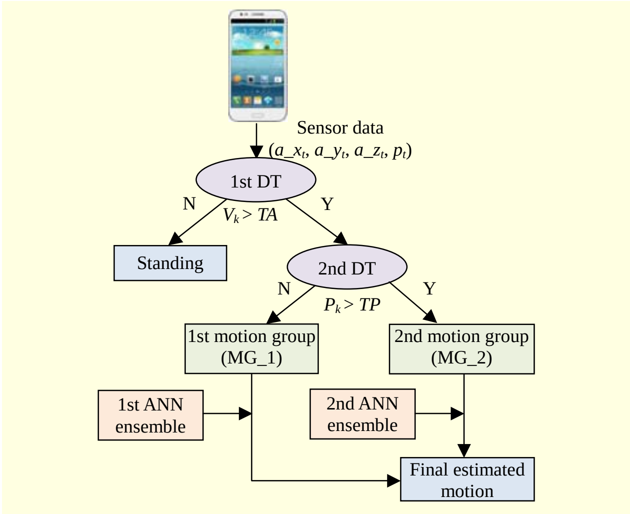

Fig. 1. Hybrid model classifier.

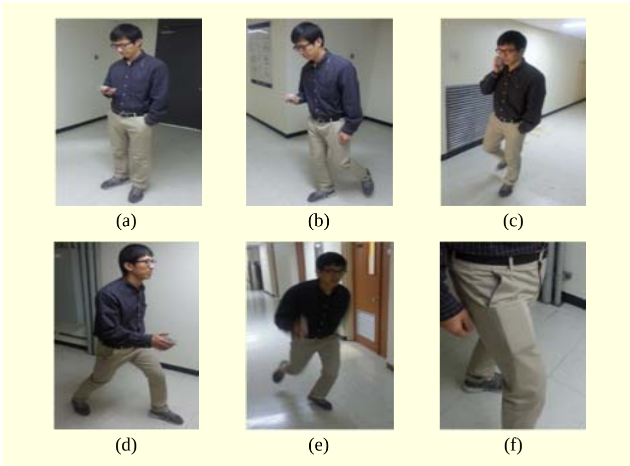

Fig. 2. Defined user motions. (a) M_1, (b) M_2, (c) M_3, (d) M_4, (e) M_5, and (f) M_6.

III. Algorithm

1. DT

A DT is a classifier that helps to represent decisions that are made in accordance with a particular feature's qualities; hence, it is also known as a qualitative classifier. For motion recognition, for which many features are required, a single DT classifier is not sufficient. On the other hand, if features can be clearly divided into certain motion groups, then a single DT in combination with other classifiers could be useful for motion recognition.

As shown in Fig. 1, we construct the first and second DTs before operating the ANN ensemble. The first DT, by

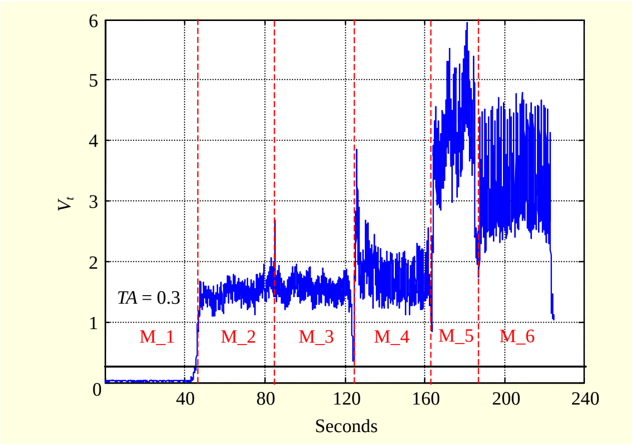

Fig. 3. Variance of accelerometer output for the six motions.

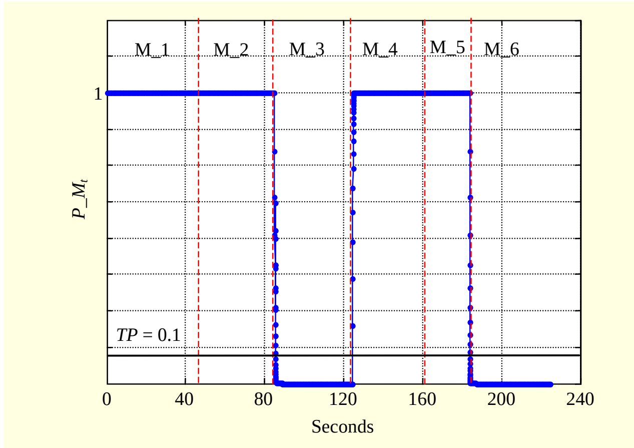

Fig. 4. Mean value of proximity sensor output for the six motions.

considering the accelerometer output, is used to determine whether the user is static or moving. Figure 3 presents the variance value (V_t) of the accelerometer output during the experiment. As expected, if the user is static, then the value of V_t is very low. On the other hand, the value of V_t is quite high when the user performs defined motions other than M_1. The first DT determines that the user is static for values of V_t that are lower than the accelerometer variance threshold (that is, TA). In our algorithm, TA is set to 0.3. The value of V_t is calculated from the following:

$$A\_N_t = \sqrt{(a\_x_t)^2 + (a\_y_t)^2 + (a\_z_t)^2}, \quad (1)$$

$$A\_M_t = \left(\sum_{k=t-n}^{t} \sqrt{(A\_N_k)^2} \right) / n, \quad (2)$$

$$V_t = \sum_{k=t-m}^{t} \sqrt{(A\_M_t - A\_N_k)^2} \;/\; m. \quad (3)$$

In (1), a_x_t, a_y_t, and a_z_t denote the accelerometer output of their respective axis at time t. The norm value of the accelerometer at time t is denoted by A_N_t. In (2), A_M_t indicates the mean value of A_N during the 0.02 × n second at time t. For example, if n = 50, then A_M_t is the mean value of A_N during the particular second at time t. In (3), V_t denotes the variance of the accelerometer during the 0.02 × m second at time t. In our experiments, n and m are set to 10 and 50, respectively.

If the first DT determines that the user is moving, then the second DT selects one of the two predefined motion groups. As shown in Table 1, each motion is assigned to either motion group one (MG_1) or motion group two (MG_2). The motions M_2, M_4, and M_5 belong to MG_1, and the remaining motions, except M_1, belong to MG_2. To assign each motion, we utilize the proximity sensor because it can detect the presence of nearby objects. For the motions depicted in Fig. 2, the proximity sensor would be enacted in the cases of M_3 and M_6. On the other hand, the proximity sensor would not respond in the cases of M_2, M_4, and M_5. The proximity sensor provides a binary near or far measurement. If the proximity sensor detects a nearby object, then it outputs the value 0. Otherwise, it outputs the value 1. To remove the possibility of false detection, we use the average value of the proximity sensor, which is calculated as follows:

$$P\_M_t = \Big(\sum_{k=t-j}^{t} P_k \Big) / j. \quad (4)$$

In (4), P_k denotes the output of the proximity sensor at time k. The mean value of P during 0.02 × j second at time t is denoted by P_M_t. In our experiments, j is set to 5. Thus, P_M_t would be the mean value of the proximity sensor during the first 0.1 seconds. If P_M is lower than the proximity sensor threshold (TP), then the second DT would conclude that the performed motion belongs to MG_2. In our algorithm, TP is set to 0.1. Figure 4 depicts the values of P_M for the various aforementioned motions. As expected, only in the cases of M_3 and M_6 is P_M lower than TP. After a certain motion group is selected, the ANN ensemble estimates the motion of user.

2. ANN Ensemble

The ANN is a classification model inspired by natural neurons. The ANN is comprised of an input layer, a hidden layer, an output layer, and weights connecting each of the nodes. The output of the ANN is dependent upon the aforementioned weights. We extract certain features from the accelerometer and use them to create an input vector. The names and definitions of these features can be found in Table 2. We utilize a backpropagation algorithm [11] for the learning of the ANN classifier. In the learning of the ANN classifier

ETRI Journal, Volume 36, Number 6, December 2014
http://dx.doi.org/10.4218/etrij.14.0113.1159

Table 2. Definition of ANN features.

Feature name	Feature definition
VarAcc	Variance of the acceleration
MeanAccX	Mean value of the acceleration *X*-axis
MeanAccY	Mean value of the acceleration *Y*-axis
MeanAccZ	Mean value of the acceleration *Z*-axis

process, final weights are decided and an output is obtained. After the completion of the ANN process, an output vector is calculated. The output vector is expressed as $\{o\_e_1, o\_ e_2, \dots, o\_e_u\}$, where u denotes the number of candidate motions. Each output element has a value between –1 and 1, and the maximum value among the output elements becomes the recognized motion.

In practice, a user's sensor data, obtained from the smartphone, varies in accordance with their physical condition and motions. This is a critical factor that can affect the performance of a single ANN classifier. To overcome this issue, we implement an ensemble of ANN classifiers. The ANN ensemble concept is derived from the widely accepted notion that "two heads are better than one." Each ANN classifier in the ANN ensemble has its own weight since an initial random weight for each was set prior to its learning. We combine the output vectors of each ANN classifier (see Fig. 5). To combine each of the ANN classifier output vectors, we transform the output vectors into probability vectors, as follows:

$$
\begin{aligned}
p_{1,v} &= \frac{e^{o\_e_{1,v}}}{e^{o\_e_{1,v}} + e^{o\_e_{2,v}} + \dots + e^{o\_e_{u,v}}}, \\
p_{2,v} &= \frac{e^{o\_e_{2,v}}}{e^{o\_e_{1,v}} + e^{o\_e_{2,v}} + \dots + e^{o\_e_{u,v}}}, \\
p_{u,v} &= \frac{e^{o\_e_{u,v}}}{e^{o\_e_{1,v}} + e^{o\_e_{2,v}} + \dots + e^{o\_e_{u,v}}},
\end{aligned}
\tag{5}
$$

where $p_{u,v}$ denotes the uth element of the probability vector of the vth ANN classifier. In other words, the output vector $\{o\_e_{1,v}, o\_ e_{2,v}, \dots, o\_e_{u,v}\}$ is transformed to the probability vector $\{p_{1,v}, p_{2,v}, \dots, p_{u,v}\}$. Then we obtain the combined probability vector as follows:

$$B = \{s_1, s_2, \dots, s_u\}, \tag{6}$$

$$s_u = \sum_{k=1}^{v} p_{u,k}, \tag{7}$$

where B denotes the combined probability of the ANN ensemble and each s is an element of the combined probability. Each element of the combined probability vector has its own labeled motion. The recognized motion is determined by the element having maximum value in the combined probability vector.

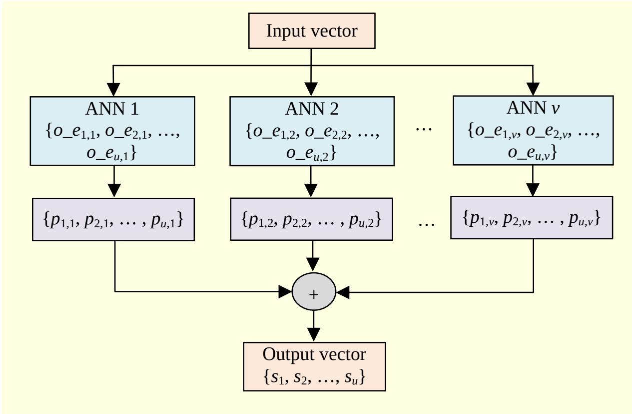

Fig. 5. ANN ensemble model.

In our proposed system, before the ANN ensemble estimates the motion, the second DT selects a motion group. Thus, the ANN ensemble's task is made easier since there are now fewer candidate motions to choose from. Nevertheless, the reason for the ANN ensemble is that the input vector of each ANN classifier is highly dependent on the characteristics of the user. In our algorithm, v is set to three throughout the experiments. However, it can be altered according to the experimental environment.

IV. Experimental Assessment

1. Data Collection and Test Setup

To verify the proposed system, real field tests were conducted in the L1 building of the Korea Institute of Science and Technology. Ten subjects (five males and five females) of different ages participated in the test. The subjects' physical characteristics and ages are presented in Table 3. For the test scenario, each subject had to walk along a corridor performing the motions M_2 to M_6 sequentially as they walked. The subject performed each motion for a distance of 50 meters, walking 250 meters in total. An application logging the sensor data was installed on a Samsung Galaxy Note (Android OS). When the subject walked, the sensor data, obtained from the accelerometer and proximity sensor, were saved in the SD card at a sampling rate of 50 Hz. Upon collecting the data and training the classifiers, we compared the performance of the hybrid model classifier with single classifiers, such as DT, ANN, and SVM. Note that the sensor data of subject M1 is only utilized for the training of all classifiers.

2. Experimental Results

In this subsection, the performance of the hybrid model classifier is analyzed on ten subjects. Table 4 presents the experimental results of the hybrid model classifier. For each of the subjects, the respective total recognition rates corresponding to each of the six predefined motions is given. The proposed classifier performs well regardless of the subject. However, in the case of M_4, the recognition rate is lower compared to that for other motions. In the proposed classifier, the M_4 is (walking swinging hands) intermittently misestimated for the M_5 (running). When the subject performs the swing motion, the patterns of the extracted feature are sometimes similar to those associated with the running motion. To solve this problem, a third DT distinguishing between M_4 and M_5 would be required.

The ANN ensemble, consisting of several individual ANN classifiers, is utilized in the proposed model. Figure 6 shows the performance of the hybrid model classifier according to the number of ANN classifiers used. The recognition rate in Fig. 6 is the mean recognition rate for all ten subjects. This mean recognition rate converges to about 98% when the number of ANN classifiers used exceeds two. Thus, we take three ANN classifiers for the ANN ensemble in the hybrid model classifier.

To evaluate the performance of the hybrid model classifier, we compare its recognition rate with those of other classifiers, such as DT [6], ANN [11], and SVM [12], with the same data sets. Table 5 presents the recognition rates of all classifiers. The proposed model provides the highest recognition rate among the classifiers. Figure 7 shows the recognition results of subject F7 for all classifiers. Blue squares and red points denote true label and estimated label, respectively. We can clearly see that false recognitions occurring in the single classifiers are

Table 3. Subjects' characteristics.

Subject	Age	Height (cm)	Weight (kg)
M1	30	182	78
M2	17	173	60
M3	28	178	83
M4	28	169	64
M5	62	168	68
F6	15	159	56
F7	29	157	48
F8	37	160	42
F9	49	160	58
F10	57	158	62

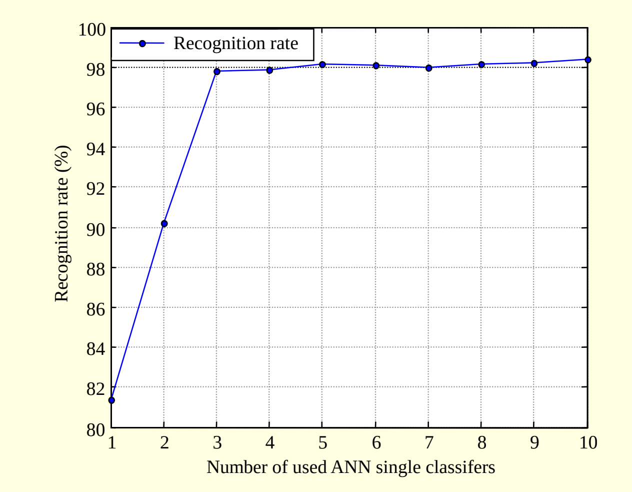

Fig. 6. Mean recognition rate of hybrid model classifier according to the number of ANN single classifiers used.

Table 4. Experimental results of hybrid model classifier.

Subject	M_1	M_2	M_3	M_4	M_5	M_6	Total
M1	100	100	100	94.7	100	100	99.1
M2	100	100	100	86.4	91.6	100	96.3
M3	100	91.1	100	91.4	100	100	97.9
M4	100	100	100	97.5	100	100	99.6
M5	100	100	95.1	70.2	100	100	94.2
F6	100	100	77.1	94.4	100	100	95.2
F7	100	100	100	92.7	100	100	98.8
F8	100	100	100	94.7	100	100	99.1
F9	100	100	100	95.4	100	100	99.2
F10	100	100	97.2	92.5	100	100	98.3
Total	100	99.1	97.0	91.0	99.2	100	97.8

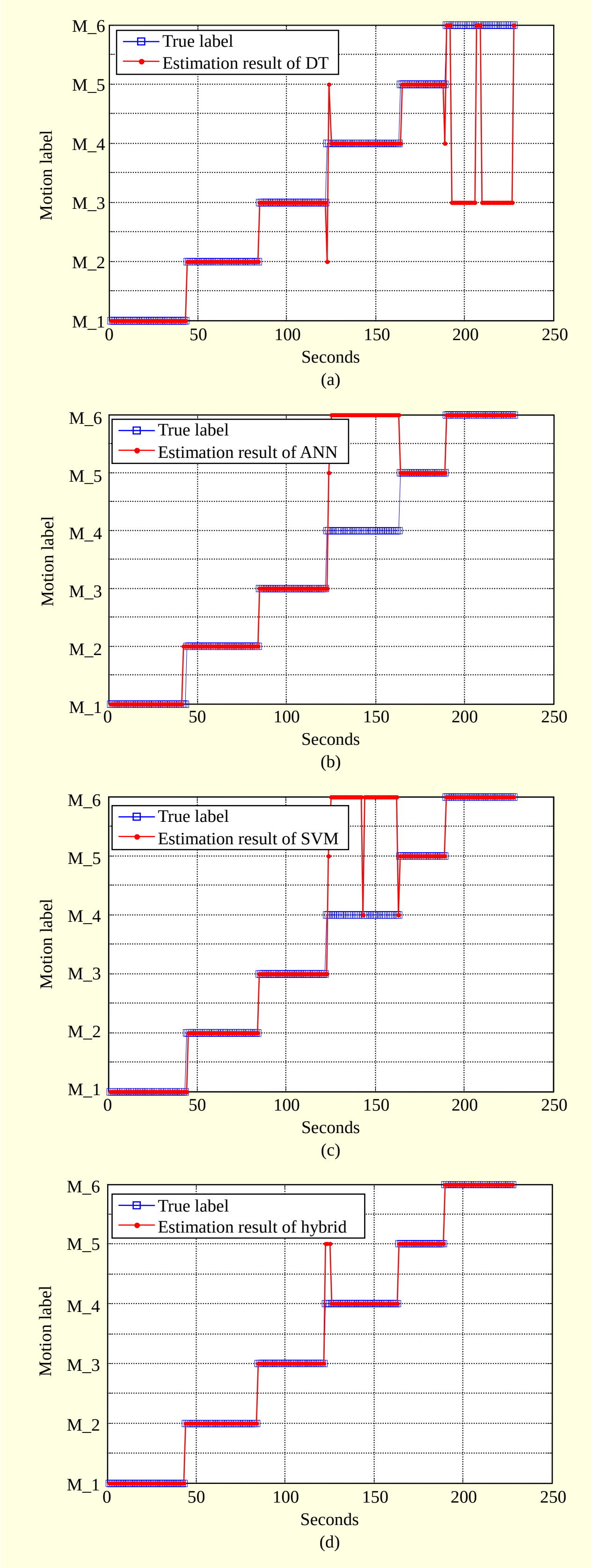

Fig. 7. Recognition results of subject F7 for all classifiers: (a) DT, (b) ANN, (c) SVM, and (d) hybrid model classifier.

Table 5. Recognition rates of all classifiers.

Subject	DT	ANN	SVM	Hybrid model classifier
M1	99.1	98.6	98.6	99.1
M2	96.2	96.2	96.2	96.3
M3	92.3	93.9	98.5	97.9
M4	97.9	98.7	98.7	99.6
M5	83.9	71.4	96.9	94.2
F6	99.5	86.8	97.2	95.2
F7	84.6	82.9	83.8	98.8
F8	84.3	78.2	98.6	99.1
F9	82.6	81.7	97.1	99.2
F10	98.6	81.5	97.2	98.3
Total	91.9	87.0	96.3	97.8

Table 6. Computational times of all classifiers.

Classifier	DT	ANN	SVM	Hybrid model classifier
Computational time (seconds)	0.7332	0.8424	2.0904	1.2012

overcome in the proposed model.

Table 6 presents the computational times of all classifiers. All algorithms were performed on a personal computer with Intel Core i7-2600 CPU. The computational time of the SVM is the longest among the classifiers. A comparison of the computational times indicates that the proposed model has more computational efficiency than that of the SVM.

V. Conclusion

In this paper, we presented a hybrid model classifier for the motion recognition of users. Single classifiers cannot guarantee high recognition rates for users of varying ages and body characteristics. A hybrid model classifier, comprising of two DTs and an ANN ensemble, was proposed to enhance such a recognition rate. We defined six motions commonly performed in indoor environments. To verify the performance of the proposed classifier, we conducted real field tests with ten subjects. We compared its performance with those of other classifiers, such as a DT, a single ANN, and a single SVM. The experimental results showed that the hybrid model classifier provided the highest results among the classifiers.

References

[1] L. Bedogni, M.D. Felice, and L. Bonori, "By Train or by Car?

Detecting the User's Motion Type through Smartphone Sensors Data," *IEEE/IFIP Int. Conf. Wireless Days*, Dublin, Ireland, Nov. 21–23, 2012, pp. 1–6.

[2] S. Reddy et al., "Using Mobile Phones to Determine Transportation Modes," *ACM Trans. Sensor Netw.*, vol. 6, no. 2, Feb. 2010, pp. 1–27.

[3] L. Pei et al., "Motion Recognition Assisted Indoor Wireless Navigation on a Mobile Phone," *Proc. ION GNSS*, Portland, OR, USA, Sept. 21–24, 2010, pp. 3366–3375.

[4] B. Shin et al., "Motion-Awareness 3D PDR System in GPS-Denied Environment Using Smartphone," *Proc. ION GNSS*, Nashville, TN, USA, Sept. 17–21, 2012, pp. 3163–3168.

[5] Y. Chon and H. Cha, "LifeMap: A Smartphone-Based Context Provider for Location-Based Services," *IEEE J. Pervasive Comput.*, vol. 10, no. 2, Apr.–June 2011, pp. 58–67.

[6] J. Yang, "Toward Physical Activity Diary: Motion Recognition Using Simple Acceleration Features with Mobile Phones," *Int. Workshop Interactive Multimedia Consum. Electron.*, Beijing, China, Oct. 2009, pp. 1–10.

[7] A. Khan et al., "Human Activity Recognition via an Accelerometer-Enabled-Smartphone Using Kernel Discriminant Analysis," *Proc. Int. Conf. Future Inf. Technol.*, Busan, Rep. of Korea, May 21–23, 2010, pp. 1–6.

[8] A. Bujari et al., "Movement Pattern Recognition through Smartphone's Accelerometer," *IEEE Consum. Commun. Netw.*, Las Vegas, NV, USA, Jan. 14–17, 2012, pp. 502–506.

[9] Y. Ding, X. Song, and Y. Zen, "Forecasting Financial Condition of Chinese Listed Companies Based on Support Vector Machine," *Elsevier J. Expert Syst. Appl.*, vol. 34, no. 4, May 2008, pp. 3081–3089.

[10] R. Polikar, "Ensemble Based Systems in Decision Making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, Sept. 2006, pp. 21–45.

[11] C. Gershenson, "Artificial Neural Networks for Beginners," submitted 2003. http://arxiv.org/ftp/cs/papers/0308/0308031.pdf

[12] L. Pei et al., "Using LS-SVM Based Motion Recognition for Smartphone Indoor Wireless Positioning," *J. Sensors*, vol. 12, no. 5, May 2012, pp. 6155–6175.

Beomju Shin received his BS and MS degrees in information and communication engineering from Sejong University, Seoul, Rep. of Korea, in 2010 and 2012, respectively. From 2012 to 2014, he worked for the Sensor System Research Center at the Korea Institute of Science and Technology, Seoul, Rep. of Korea. He is currently working toward his PhD at the School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, Rep. of Korea. His current interests include pattern recognition, machine learning, and indoor navigation systems.

Chulki Kim received his BS degree in physics from the Korea Advanced Institute of Science and Technology, Daejeon, Rep. of Korea, in 2002 and his PhD degree in physics from the University of Wisconsin-Madison, WI, USA, in 2011. He was a research staff member at the Samsung Advanced Institute of Technology, Yongin, Rep. of Korea in 2012. Since 2012, he became a researcher at the Korea Institute of Science and Technology, Seoul, Rep. of Korea. His research interests include nanomechanical single-electron transistors and sensor applications of NEMS.

Jae Hun Kim received his BS and MS degrees in electrical and computer engineering from Purdue University, NI, USA, in 1997 and 1999, respectively and his PhD in electrical engineering from Pennsylvania State University, PA, USA, in 2008. His research interests are focused on the development of sensor networks.

Seok Lee received his BS, MS, and PhD degrees in physics from Yonsei University, Seoul, Rep. of Korea in 1985, 1987, and 1994, respectively. Since 1996, he has been with the Sensor System Research Center at the Korea Institute of Science and Technology, Seoul, Rep. of Korea, where he is now a senior researcher. His research interests include integration of bio-sensors and a sensor platform with sensor networks.

Changdon Kee received his BS and MS degrees in aeronautics engineering from Seoul National University, Seoul, Rep. of Korea, in 1984 and 1986, respectively. He received his PhD degree in aeronautics and astronautics from Stanford University, Standford, CA, USA, in 1994. Since 1996 he has been with the Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul, Rep. of Korea, where he is a professor. He has more than 20 years' experience of GNSS and flight-control research.

Taikjin Lee received his BS and PhD degrees in mechanical and aerospace engineering from Seoul National University, Seoul, Rep. of Korea, in 2001 and 2008, respectively. In 2008, he was with the School of Mechanical and Aerospace Engineering, Seoul National University, Rep. of Korea, where he was a postdoctoral fellow. Since 2010, he has been with the Korea Institute of Science and Technology, Seoul, Rep. of Korea, as a senior researcher. His areas of interest are indoor navigation systems, pattern recognition, and sensor networks.