A Medication Adherence Monitoring System for Pill Bottles Based on a Wearable Inertial Sensor

Chen Chen, Nasser Kehtarnavaz, *IEEE Fellow*, and Roozbeh Jafari, *IEEE Senior Member*

***Abstract*—This paper presents a medication adherence monitoring system for pill bottles based on a wearable inertial sensor. Signal templates corresponding to the two actions of twist-cap and hand-to-mouth are created using a camera-assisted training phase. The act of pill intake is then identified by performing a moving window dynamic time warping in real-time between signal templates and the signals acquired by the wearable inertial sensor. The outcomes of the experimentations carried out indicate that the developed medical adherence monitoring system identifies the act of pill intake with a high degree of accuracy.**

I. Introduction

Adherence to medication regimens continues to rank as a major clinical problem in disease management. Achieving optimal medication adherence requires patients being prescribed the right medication, filling it and taking it correctly over time. This requires appropriate prescribing, effective patient-provider communication, coordination among care-providers and active engagement and participation by patients. Poor adherence to medication regimens accounts for a substantial load on health care costs in the United States. Of all medication-related hospital admissions in the United States, 33 to 69 percent are due to poor medication adherence, costing more than $100 billion annually in increased medical costs [1].

There have been a number of efforts addressing systems or devices for medication adherence. For example, a context-aware pill bottle/stand was developed in [2], which provides audio/visual alerts for taking a medication on time. However, this system operates based on the limiting assumption that the pill is in fact consumed when a pill bottle is removed from the stand. A smart medication dispenser was proposed in [3], which dispenses a predetermined medication at a predetermined time. Again, this device does not detect whether the user is actually taking the medication. Methods based on computer vision techniques have also appeared in [4-6]. Obviously, the limitation with such vision based systems is that they require the user to take a medication within the field of view of a camera and cannot monitor the user wherever the user goes. A system consisting of several sensors (motion sensor, wearable sensor and bed sensor) was proposed in [7], which is rather complex to set up and operate.

One can easily see that the availability of a low-cost and easy-to-use device for pill intake has been lacking. In this paper, an attempt has been made to introduce such a device. More specifically, a watch-like motion sensor that can identify the act of opening a pill bottle and transporting a pill from the bottle to the mouth is developed in this paper. This system incorporates a camera-assisted training phase where the user imitates the act of taking fake placebo pills, and a signal processing algorithm is designed to train the inertial sensor for the act of pill intake. After training or during the actual operation, no camera is used and only the inertial sensor monitors the act of pill intake.

The principal novelty of our solution lies in offering a low-cost wearable device thus not creating a cost burden on users/health-care providers. This solution can be incorporated into smart watches in the near future considering that the newer generations of smart watches are beginning to incorporate motion sensors and micro-controllers capable of executing signal processing tasks. The signal processing involved in our system is based on a computationally efficient implementation of dynamic time warping (DTW) that is designed for processing time series acquired from inertial sensors. A major contribution of this work lies in its camera-assisted training of the inertial sensor providing adaptation to specific users and the way they open pill containers dispensed by pharmacies in the US. This user-specific attribute leads to a high degree of accuracy for the signal processing tasks involved.

The rest of the paper is organized as follows. Section II provides the details of our developed monitoring system followed by the results and discussion in section III. The paper is concluded in section IV.

II. Developed Monitoring System

To achieve medication adherence monitoring, the user is asked to wear an inertial sensor on the right or left wrist (depending on whether right-handed or left handed) similar to a smart watch or as part of a smart watch. Fig. 1(a) shows the placement of an inertial sensor on the wrist and the sensor world coordinates. Our monitoring approach involves detecting two actions of "twist-cap" and "hand-to-mouth" that one normally goes through when taking a pill out of pharmacy bottles used in the US. The signals generated by the inertial sensor are used to detect these two actions. It is worth emphasizing that the detection of the second action (hand-to-mouth) is activated only after the detection of the first action (twist-cap), i.e. only after the bottle cap is detected to have been opened. Note that our focus in this paper is on pill bottles, see Fig. 1(b), that are dispensed by pharmacies in the US.

The authors are with the Electrical Engineering Department, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: {Chen.Chen12; kehtar; rjafari}@utdallas.edu).

A. *Training Phase or Signal Template Setup*

Kinect is a low-cost RGB-Depth camera (see Fig. 1(c)) introduced by Microsoft for human-computer interface applications [8]. The Kinect SDK software [9] allows tracking 20 body joints as illustrated in Fig. 1(d). In order to identify the portions of relatively long duration inertial sensor signals that correspond to the two actions of twist-cap and hand-to-mouth, users are asked to go through a training phase by sitting/standing in front of a Kinect camera. The Kinect camera is then used to automatically determine the start and end of the inertial signals corresponding to the two actions by tracking the joint positions via the SDK software. This software is programmed to detect the twist-cap action by using the positions of the left and right wrists denoted by $P_{lw}(x_{lw}, y_{lw}, z_{lw})$ and $P_{rw}(x_{rw}, y_{rw}, z_{rw})$, and the hand-to-mouth action by using the positions of the right wrist (the roles are reversed for left-handed users), and the shoulder center denoted by $P_{sc}(x_{sc}, y_{sc}, z_{sc})$. More specifically, the pose detection which is a built-in function of the Kinect SDK is used to trigger the detection of the twist-cap or hand-to-mouth actions. In other words, the user is asked to start with his/her own pose Ψ before performing a twist-cap or hand-to-mouth action. The start and end of a twist-cap action is then determined sequentially by measuring the closeness between the two wrists via $|x_{lw} - x_{rw}|$. The procedure is provided as a pseudo-code in Algorithm 1. At the same time, the inertial sensor signals between the time stamps t_s and t_e are obtained to form a template of a user-specific twist-cap action. The detection of the hand-to-mouth action is achieved similarly. The start is determined by $0 < y_{hc} - y_{rw} \leq \mu$ and the end is determined by $y_{hc} - y_{rw} > \mu$, where $\mu = 15cm$ was experimentally found to work well across different users.

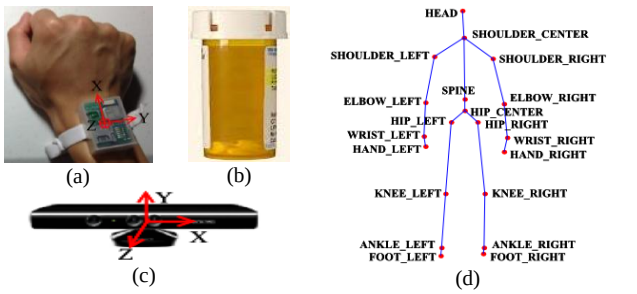


Figure 1. (a) Inertial sensor placement and the corresponding world coordinates, (b) twist-cap pill bottle, (c) Kinect camera and the corresponding world coordinates, (d) skeleton joints tracked by Kinect

This training phase allows creating inertial signal templates for the two actions of "twist-cap" and "hand-to-mouth". Basically, the Kinect camera is used during a training phase in order to obtain the inertial sensor signal segments which correspond to the two actions of interest by automatically time stamping the start and end of the actions. Fig. 2 shows an example of the segments of a sensor signal (x-axis acceleration) automatically determined by using the Kinect camera.

The training phase consists of the user taking a fake placebo pill a few times (*e.g.*, 5 times) in front of a Kinect camera the way he/she naturally does. Templates of the two actions (twist-cap and hand-to-mouth) are then generated by taking averages of the signal segments that are automatically identified by the Kinect camera. Notice that all the signal segments of an action are re-sampled to have the same normalized length before averaging. Also, it is important to note that the training is user-specific which has a major impact on the monitoring accuracy reported later.

Algorithm 1 Pseudocode for twist-cap detection

Initialization: twist_start = false, twist_end = false, $\sigma = 15cm$
if (Ψ pose detected) *// only if Ψ pose is detected, the signal processing begins*
  twist_start = true *// twist-cap detection*
  if (twist_start = true)
    if ($|x_{lw} - x_{rw}| \leq \sigma$) *// indicating two hands are together*
      record time stamp t_s for the inertial sensor signals
      twist_end = true
    end if
  end if
  if (twist_end = true)
    if ($|x_{lw} - x_{rw}| > \sigma$) *// indicating two hands are separated*
      record time stamp t_e for the inertial sensor signals
    end if
  end if
end if

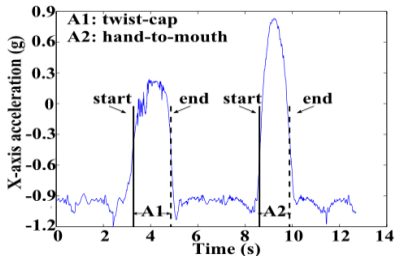


Figure 2. Segmented sensor signal using Kinect camera during training phase

B. *Operation Stage or Signal Template Matching*

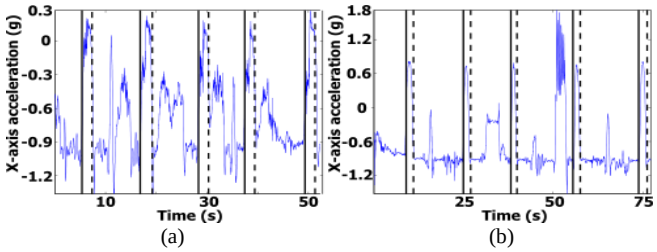
During the operation stage, the Kinect camera is removed and only the inertial sensor is used. This is because it is not practical to use the Kinect camera during actual operation as it is a stationary platform and it is not worn by the user while the inertial sensor is worn and carried by the user. The template signal gets matched to the signal from the inertial sensor in real-time. A sliding or moving window is used to do the matching with the template by utilizing the dynamic time warping (DTW) technique. DTW is known to be an effective matching algorithm for measuring similarity between two time series which may have different lengths or durations [10]. The window size is chosen to be the average length of the segmented signals during the training phase. Let w indicate the window size. For example, the sliding window can be shifted by $w/4$ with the overlap of $3w/4$ between neighboring windows.

C. *Visual Verification or Ground Truth Generation*

To evaluate the performance of our developed system, a visual verification is done by using a video camera during the operation stage to record videos of the actions. The software developed in [11] is used to align the readings from the accelerometer with the video recording as the ground truth. Examples of the segmented signals obtained by visual verification of the two actions are shown in Fig. 3. The time stamps or sample indices indicating the start and end of the two actions are used to serve as the ground truth.

D. Pill Intake Detection

Since the twist-cap and hand-to-mouth actions take place sequentially for a pill intake, the system first tries to detect the twist-cap action. If the twist-cap action is detected, the system then tries to detect the hand-to-mouth action within a user-specified time duration (30 seconds considered in our experiments). It is reasonable to expect that the hand-to-mouth action gets performed after the twist-cap action within this time duration. If the hand-to-mouth action is not detected within this time duration, the system status returns to the detection of the twist-cap action. Note that when the system is attempting to detect the hand-to-mouth action after the twist-cap action has occurred, the detection of the twist-cap action also runs at the same time. If a new twist-cap action is detected within the time duration of 30 seconds, a new 30-second time duration for hand-to-mouth detection gets initiated. This sequential detection of the two actions leads to a high detection accuracy for the act of pill take.



(a) (b)

Figure 3. Visual verification of (a) "twist-cap" and (b) "hand-to-mouth" actions using recorded video: solid and dashed vertical lines indicate the start and end of a "twist-cap"/ "hand-to-mouth" action (only the x-axis acceleration signal is displayed in the figure)

III. MONITORING RESULTS AND DISCUSSION

A. Template Selection

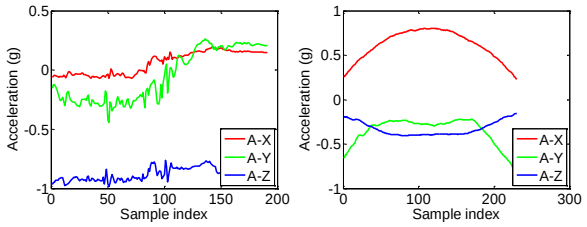
An experiment was first carried out by examining all the signals from an inertial sensor developed in the ESSP Lab including its 3-axis acceleration signals and its 3-axis angular velocity signals with the sampling rate of 200Hz [12]. The detection accuracies of the two actions are shown in Table I. In this table, TP denotes true positive, TN true negative, FP false positive, FN false negative, and 20 TP and 20 TN cases were considered for each action. The accuracy was found according to

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad . \qquad (1)$$

It was found that the acceleration signals provided more discriminatory power than the angular velocity signals for the two actions involved in pill intake. Therefore, the acceleration signals were used for the matching step. In other words, the template considered was a matrix of size $3 \times N$, where N denotes the length of the time series. Here, it is worth pointing out that it is possible to use only one signal template, *e.g.*, acceleration of the x-axis signal, for matching purposes if it is desired to achieve lower computational complexity for real-time signal processing. Note that each user may perform the two actions of pill intake differently. In other words, the template signal reflects the way a user performs pill intake or is user specific. Fig. 4 displays the template signals (3-axis accelerations) of a user for the two actions of "twist-cap" and "hand-to-mouth", respectively.

TABLE I. TWIST-CAP AND HAND-TO-MOUTH DETECTION ACCURACY USING DIFFERENT INERTIAL SIGNALS

Template	Accuracy	FP	FN
Twist-cap			
Acc-X	92.5%	3	0
Acc-Y	82.5%	5	2
Acc-Z	92.5%	3	0
Acc-(X, Y, Z)	95.0%	2	0
Gyro-X	80.0%	7	1
Gyro-Y	75.0%	8	2
Gyro-Z	82.5%	6	1
Gyro-(X, Y, Z)	80.0%	6	2
Acc + Gyro	95.0%	2	0
Hand-to-mouth			
Acc-X	97.5%	1	0
Acc-Y	77.5%	7	2
Acc-Z	90.0%	4	0
Acc-(X, Y, Z)	97.5%	1	0
Gyro-X	87.5%	5	0
Gyro-Y	72.5%	7	3
Gyro-Z	85.0%	6	0
Gyro-(X, Y, Z)	92.5%	2	1
Acc + Gyro	97.5%	1	0



(a) (b)

Figure 4. Templates of acceleration signals for (a) "twist-cap" and (b) "hand-to-mouth" actions (A-X, A-Y and A-Z denote accelerations obtained from X, Y and Z axes of the accelerometer, respectively)

B. Detection Accuracy

To examine the detection accuracy, we carried out two sets of experiments. In the first set of experiments, five subjects (3 males and 2 females) were asked to perform the twist-cap action and the hand-to-mouth action sequentially and as naturally done for 20 times over a 20-minute time duration. The pill bottle used is the one normally dispensed by pharmacies in the US with one tightness level. No subjects suffered from hand jitters or tremors. A dataset was collected by considering 20 correctly done pill intake to form a positive set. In the second set of experiments, the act of pill intake was done incorrectly on purpose. More specifically, the following two realistic scenarios were considered: (i) the subjects were asked to twist the bottle cap, then close the bottle and put it away, and (ii) the subjects were asked to twist the cap of a water bottle, then drink the water. These realistic scenarios were considered to form the negative set.

Fig. 5 shows the DTW distances obtained by matching the templates to the signals within each sliding window. Solid and dashed vertical lines indicate the start and end of a twist-cap action, respectively, which were found by visual inspection or verification of the recorded video frames. In other words, the visual inspection was used to provide the ground truth actions. Circles indicate the DTW distances for the windows. The solid horizontal line denotes a detection threshold. A DTW distance smaller than the threshold indicated a twist-cap or a hand-to-mouth action. For easier visual appearance of the detection outcome, the DTW distances smaller than the threshold were assigned a value of 1, and 0 otherwise, to

generate Fig. 6. This way, 1 indicated the action took place and 0 indicated the action did not take place. In order to increase the robustness of the detection, the majority vote over a number of consecutive 1's was considered. The results reported here correspond to the majority voting over 3 consecutive 1's. The optimal thresholds for detecting the two actions were obtained from the training data. Fig. 7 illustrates how the accuracy varied with different thresholds for the twist-cap and hand-to-mouth actions. Based on this figure, a threshold of 60 was thus chosen during testing or operation.

Each of the five subjects went through the training phase before the operation phase. Each subject carried out two aforementioned experiments. In the second set of experiments, each scenario was performed 10 times. Table II shows the detection accuracies for the pill intake (twist-cap and hand-to-mouth) corresponding to each subject. As can be seen from this table, our developed medical adherence monitoring system generated no FN and FP for the positive set, primarily due to our user-specific training phase. Our system generated low FPs for the negative set. Specifically, our system was able to reject all the negative cases associated with the first scenario (no hand-to-mouth action) since pill intake was associated with the sequential detection of the two actions. The FPs in Table II were caused by the second scenario which was drinking water from a water bottle. Since the hand during drinking stays to the mouth longer than during pill intake, the DTW was able to reject most of the negative cases associated with the second scenario.

TABLE II. DETECTION ACCURACIES OF PILL INTAKE (TWIST-CAP AND HAND-TO-MOUTH) DETECTION FOR FIVE SUBJECTS

	Positive set		Negative set	
Subject	**FP**	**FN**	**FP**	**FN**
Subject 1	0	0	0	0
Subject 2	0	0	1	0
Subject 3	0	0	2	0
Subject 4	0	0	0	0
Subject 5	0	0	1	0

IV. CONCLUSION

In this paper, a medication adherence monitoring system for pill intake from pharmacy bottles has been introduced based on a low-cost wearable inertial sensor. The Kinect depth camera was used to automatically generate templates for signal matching during a training phase. For actual operation, only the inertial sensor was used to identify the two actions of "twist-cap" and "hand-to-mouth" that are associated with the act of pill intake. The experimental results demonstrated that our proposed monitoring system can detect pill intake with high degree of accuracy. In our future work, we plan to utilize an inexpensive passive RFID tag attached to pill containers in order to provide enhanced reliability of the developed medical adherence monitoring system. Furthermore, we intend to extend the method discussed in this paper to cover other forms of medication containers including foil-wrapped pills, syrups containers and cream tubes.

REFERENCES

[1] NEHI, "Improving patient medication adherence: a $290 billion opportunity," http://www.nehi.net/bendthecurve/sup/documents/Medication_Adherence_Brief.pdf

[2] A. Agarawala, S. Greenberg, and G. Ho, "The context-aware pill bottle and medication monitor," Technical Report, Department of Computer Science, University of Calgary, Calgary, Canada, 2004.

[3] J. Pak and K. Park, "Construction of a smart medication dispenser with high degree of scalability and remote manageability," *Journal of Biomedicine and Biotechnology*, vol. 2012, 2012.

[4] H. H. Huynh, J. Meunier, J. Sequeira, and M. Daniel, "Real time detection, tracking and recognition of medication intake," *World Academy of Science, Engineering and Technology*, vol. 60, pp. 280–287, December 2009.

[5] G. Bilodeau and S. Ammouri, "Monitoring of medication intake using a camera system," *Journal of Medical Systems*, vol. 35, no. 3, pp. 377–389, June 2011.

[6] F. Hasanuzzaman, X. Yang, Y. Tian, Q. Liu, and E. Capezuti, "Monitoring activity of taking medicine by incorporating RFID and video analysis," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 2, no. 2, pp. 61–70, July 2013.

[7] J. Lundell, T. L. Hayes, S. Vurgun, U. Ozertem, J. Kimel, J. Kaye, F. Guilak, and M. Pavel, "Continuous activity monitoring and intelligent contextual prompting to improve medication adherence," in *Proceedings of 29th IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBS)*, Lyon, France, August 2007, pp. 6286–6289.

[8] C. Chen, K. Liu and N. Kehtarnavaz, "Real-time human action recognition based depth motion maps," *Journal of Real-Time Image Processing*, August 2013, doi: 10.1007/s11554-013-0370-1, print to appear in 2014.

[9] http://www.microsoft.com/en-us/kinectforwindowsdev/Start.aspx

[10] D. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," *KDD Workshop*, Seattle, WA, vol. 10, no. 16, pp. 359–370, April 1994.

[11] O. Dehzangi, Z. Zhao, M. Bidmeshki, J. Biggan, C. Ray, and R. Jafari, "The impact of vibrotactile biofeedback on the excessive walking sway and the postural control in elderly," in *Proceedings of ACM International Conference on Wireless Health*, November 2013.

[12] M. Bidmeshki and R. Jafari, "Low Power Programmable Architecture for Periodic Activity Monitoring," in *Proceedings of ACM/IEEE International Conference on Cyber-Physical Systems*, Philadelphia, PA, April 2013, pp. 81-88.

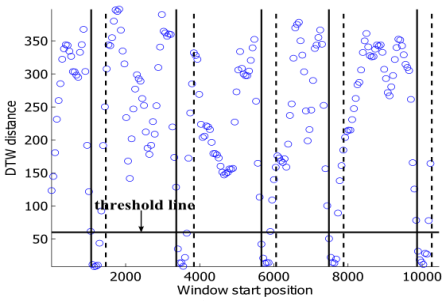


Figure 5. DTW distances for the twist-cap action: x-axis index indicates sample number of the sliding window start position

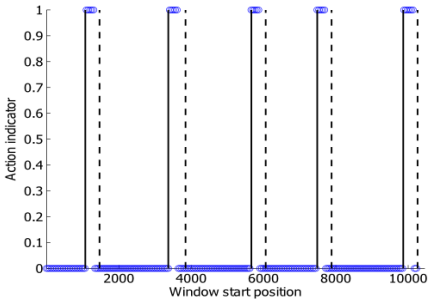


Figure 6. Twist-cap action indicator: 1 indicates a twist-cap action took place and 0 indicates no twist-cap action took place

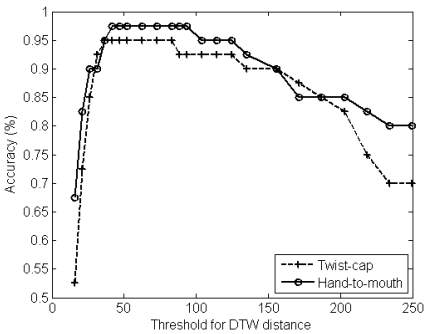


Figure 7. Detection accuracies for various thresholds