Automated Virtual Observation Therapy

Yin-Leng Theng  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
tyiltheng@ntu.edu.sg

Lynette Ying Qin Goh  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
lyqgoh@ntu.edu.sg

Owen Noel Newton Fernando  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
fernando.onn@gmail.com

Jason Wen Yau Lee  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
jenonlee@ntu.edu.sg

Chamika Deshan  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
sooriya@ntu.edu.sg

Foo Shou Boon, Schubert  
COSMIC Centre  
Nanyang Technological University  
14 Nanyang Drive, 638332  
Singapore  
sfoo@ntu.edu.sg

ACM 978-1-4503-2474-8/14/04. Abstract  
In this paper, we present an enhanced virtual observation therapy (VOT) system to address challenges in medication adherence and the current observed therapy approach. The original approach requires both patient and healthcare worker to be physically collocated, which has several technical and practical challenges. Therefore, we developed a system that tracks the natural actions of the patient when they take medication, which provides a new experience for them. The system automates the process of recording a video of the patient’s medication taking and uploads the video log to the hospital management system. In addition, the system provides instructions to the patient from start to end of the medication taking process.

Author Keywords  
Ubiquitous Computing; Virtual observation therapy, Computer vision

ACM Classification Keywords  
H.5.2 User Interfaces: Interaction styles, Input devices and strategies; I.4.7 Feature Measurement

General Terms  
Design; Human Factors

Introduction  
Lack of medication adherence is the cause of unnecessary disease progression, disease complications, reduced functional abilities, a lower quality of life, and even death. Interventions to
support patients with chronic diseases who need to be tightly monitored include direct observation. An approach called Directly-Observed Therapy (DOT) has been introduced by the World Health Organisation, especially among patients with serious diseases like tuberculosis, to help ensure that patients adhere to their long-term and harsh treatments to prevent drug resistance and for successful treatment outcomes. Patients on DOT have to take their medication under close supervision, often being physically present, in front of medical professionals [1]. Here the distance, cost and time taken to travel for DOT are largely why some patients default from DOT treatment. Some medical providers have implemented a virtual observation therapy (VOT) alternative [2]. VOT delivers DOT via a virtual setting like Skype or other video conferencing software. Patients are required to have a computer with webcam and Internet capabilities, and on an agreed time with the medical professional, with both sides turning on their video conferencing software and the medical professional observes the patient via the computer taking the medication.

In this paper, we present an enhanced VOT system for interactive medication observation support that would address the many challenges that exist in current systems and practices. The current system requires the patient to connect with the hospital through a video link implemented in a virtual pillbox system [3]. The virtual pillbox is an instrumental pillbox system addressing needs in patient care, allowing monitoring of medication adherence on a continuous basis. Here, it delivers asynchronous medication observation mechanism where patient and healthcare worker are not required to be physically collocated at the same time. Instead, the patient has to record a video when they take the medicine and thereafter upload it to the hospital system. Healthcare workers at the hospital can then review the patient’s medication taking using the saved video logs. The functionality is implemented in the virtual pillbox system as depicted in the Figure 1. We evaluated this system in a focus group discussion with experts in chronic illnesses, recovered patients, and caregivers centered on our technological interventions. Here, we identified that patients often face a problem of the lack of user friendliness in the tracking process, for example, the system is difficult to use.

Hence, patients who are not technologically-savvy face problems in recording and uploading their videos. The process usually requires several steps:

1. Turn on the video recording window.
2. Press the start button when they are ready.
3. Take the medication.
4. Stop recording and upload.

Even patients who are technologically well-trained may not record the video when they take their medication, which leads to bad record keeping that is essential for medicolegal purposes [4].

Figure 1: Manual video recording system
Hence, we have developed a system to avoid the above mentioned problems, which sequentially tracks the natural actions of the patient when they take the medication. Our system creates a new experience of medication monitoring to the patient where the system talks to the patient and gives instructions from the start to end of the process.

**Automated Video Observation System**

![Diagram of Automated Video Observation System]

In this system, we have reduced the complexity of the video recording through an automated sequential tracking mechanism. First, facial tracking is used to identify the patient when they present themselves in front of the mobile device (camera). The system then automatically starts the video recording of the medication taking process. For the medication tracking, we conducted an extensive literature survey and discussions with patients/caregivers and healthcare providers to identify possibilities to track the medication (Figure 2). The priority is not to verify the successfulness of the medication taking, instead it is to reduce the complexity of the video recording in the current system. The challenge is to find the moment where the patient completes their current medication taking without any input from the patient. The video recording can be stopped here and then the recorded video is then uploaded to the hospital system for observation. At this point, the patient’s behavior of drinking water is identified after taking the medicine and serves as a point for tracking medication, which becomes the trigger to stop recording the video.

In Figure 2, there are a few mechanisms that allow us to track the action of water drinking. The first is a hand gesture tracking and second is the cup tracking. Due to the complexity and user unfriendliness, hand gesture tracking was taken out from the consideration. For cup tracking, we identify the possibility of using sensor based tracking and computer vision based tracking.

After weighing the pros and cons of each solution, we have decided to use vision-based cup tracking in our solution. This particular method has been selected due to three reasons, which are: (i) flexibility and adaptability to mobile devices; (ii) drinking water after taking medicine is the most common form of action; (iii) simplicity and cost-effectiveness which does not require sensors to detect movements.

There are two options in vision-based tracking which is to track the cup as shape or track the cup as a feature using the features on the cup. There are several advantages and disadvantages of both methods (see Table 1)

![Figure 3: Cup holder]
<table>
<thead>
<tr>
<th></th>
<th>Shape-Based Tracker</th>
<th>Feature-Based Tracker</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>Simple Requirements – cup</td>
<td>Tracking works with partial visibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tracking works for greater range of pose (Tracking works with different orientations also)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Does not depend on the shape of the cup</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flexibility of changing the cup by changing the tracking features</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Tracking fails with partial visibility</td>
<td>The cup should have adequate visual features (see Figure 3)</td>
</tr>
<tr>
<td></td>
<td>Tracking only works for limited pose (different cup orientation cannot track)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Only the registered shape can track</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transparent objects won’t track (glass)</td>
<td></td>
</tr>
</tbody>
</table>

By weighing the advantages and disadvantages of the 2 options, we have decided to implement our solution using feature-based tracking. We introduce a low-cost feature rich cup holder (Figure 3) where our algorithm tracks the holder that wraps around a cup. This offers the flexibility of changing to any cup. Our current algorithm works with both cylindrical and conical shaped cups. We have designed our current cup holder to have a bottom surface that facilitates the tracking from any orientation where in certain conditions only the bottom surface is visible to the camera. Even if the cup is partially visible (e.g., when patients hold the cup), our system is able to continuously track the motion as long as enough feature points remain visible. This is an advantage of the natural feature tracking technology as explained in Neumann et al [5]. Qualcomm Vuforia [6] API, which use natural feature tracking technology, is used here to track the features embroidered on the cup holder.

**System Implementation**

![Figure 4: Architecture of the automated video observation system](image)
The system architecture is depicted in Figure 4. The system is implemented in a mobile platform and integrated to the existing virtual pillbox system. Medication tracking algorithm is implemented sequentially in ten steps:

1. Start Video Capturing
2. Face Detection (Figure 5)
3. Start video recording (Figure 5)
4. Cup side surface detection (Figure 6)
5. Cup relative angle detection (Figure 6)
6. Calculate Euclidean distance between cup and face (Figure 7)
7. Cup bottom surface detection (Figure 8)
8. Cup and face intersection detection (Figure 9)
9. Detect missing of intersection (Figure 9)
10. Save and Upload video to the hospital system (Figure 10)

As we have used computer vision technology, the only input to the system is the video feed from the front camera. This enables the system to perform on most of the mobile devices in the mobile device market. 2D Cartesian coordinate system and 3D coordinate systems are used in the implementation of tracking and visualization.

**Medication tracking flow**

In order to improve the passive video logging in the conventional virtual observation systems, we have implemented a guided video observation mechanism in our system. As we sequentially track each step of the medication process, we are able to provide an interactive way of guided instructions. Here, at each tracking step, we provide instructions to patients using visual and auditory messages that can be customized to the needs of the patients.

The face detection is used for the initialization of the video recording (no face recognition is required as they will be manually viewed by a healthcare worker). Once the face is detected, the system starts video recording (Figure 5). Thereafter, video recording and face tracking continuously run until the cup is detected. We used the lightweight Haar feature-based cascade classifiers [7] for face detection.

The drinking process is tracked in four steps. First, it detects the side surface of the cup (Figure 6). Thereafter, the message is given to the user requesting drinking water. Then the system tracks the actions of user bringing the cup to mouth and water drinking in the order of following steps:

i. Calculate the absolute pitch value of the cup (Figure 6).
ii. Calculate the Euclidean distance [8] between the cup and the face (Figure 7).
iii. Detect bottom surface of the cup (Figure 8).
iv. Detect the cup and face intersection (Figure 9).

The pitch value is obtained by calculating the angle between up vectors of the camera and the cup. The angle is augmented in 3D on the video as normal vector foam [9], as shown in Figure 6. However, in this calculation, it always returns the angle relative to the mobile device. If the mobile device screen is not aligned with the vertical plane, the angle has to be corrected by adjusting the device angle relative to the vertical plane. This angle can be read by accessing the inbuilt gyroscope of the mobile device.

After drinking the water, patient brings the cup away from the mouth and here we find this action by tracking lost of intersection between the cup and the face. After this action, instructions are given to the patient to
remain in front of the camera for another thirty seconds. Here, the face detection algorithm runs continuously to verify the patient’s presence. Finally, that video is saved and uploaded to the hospital system for observation. Figure 10 shows the video observation menu in the hospital system.

**Limitations**

One limitation of the study is that the algorithm was developed to detect feature rich surfaces rather than plain objects. Therefore, a feature rich cup holder needs to be used which allows the system to be adapted to any shape or size of cups.

This study proposed a system to automate the recording process of the patients taking the medication. Similar to VOT, a healthcare worker will still need to manually verify from the recordings if the patient has taken their medication.

**Conclusion and Future Work**

The automated video observation system is implemented on a mobile platform and integrated with a medication reminder system to enhance the capabilities of existing virtual observed therapy approach. This delivers a new experience of medication monitoring to the patient as the system ‘talks’ to the patient and gives instructions from start to end of the medication taking process. Currently, we track the cup and face intersection in XY plane. This can be further enhanced by implementing depth intersection in XZ plane. We are also planning to conduct a study with 8-10 users to evaluate the efficiency, effectiveness and usability of this system in the near future.

**Acknowledgements**

This research is supported by the Singapore National Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

**References**


