# Motion Analysis For Supervision Of Medication Intake

ID: R296


**Christian Breiderhoff and Maria-Elena Algorri**

christian.breiderhoff@smail.th-koeln.de, elena.algorri@th-koeln.de

Technology Arts Sciences TH Köln


## Introduction

Medication non adherence is a major public health problem that has been called an "invisible epidemic" [1]. In the US non adherence to pharmacotherapy has been reported to range from 13% to 93%, with an average rate of 40% [2] and the problem encompasses all ages and ethnic groups. The 2003 report of the World-Health Organization (WHO) quotes Haynes et al. as saying "increasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatments" [3]. Among patients with chronic illness, approximately 50% do not take medications as prescribed [4][5]. This poor adherence to medication leads to increased morbidity and death and is estimated to incur costs of approximately $100 billion per year [6]. There are various forms of medication non adherence: Patients forget to take their medications or take them out of time, they run out of medication, take an incorrect dose or stop therapy before or after a deadline. It is therefore relevant to create a supervision system to help diminish the problem of medication non adherence in cases where the user forgets to take his medication or takes it out of schedule.

We wanted to implement an optical system that could take a decision of whether a person had taken her/his oral medications in a controlled fashion. In particular we implemented a system that detects whether a patient takes one of her/his hands to the mouth while simultaneously holding a cup of a particular color in that hand. We needed a system that could track a person in 3D space independently of the person's pose. We also wanted to track the arms and hands of the person to determine if the arms were flexed and if the hands were close enough to the head (using a distance threshold). The system is also capable of tracking an object of a particular color in 3D space independently of its pose. By fusing the information about the 3D tracking of both the hands and the object, the system takes a decision of whether the person has performed a drinking action (corresponding to oral medication intake) using the tracked object.

## Materials

The popular Kinect system consists of a low cost depth sensor (11 bits, 320 x 240) and an optical camera (8-bit 640x480 RGB) that allow the tasks of image segmentation and 3D tracking to be carried out as a synergetic process. The Kinect provides two video streams of a scene at a maximum rate of 30 fps: The depth video stream provides the 3D coordinates of the 320 x 240 pixels that it images and the camera video stream provides the color of the pixels. An advantage of the Kinect sensor is that open source SDKs are readily available for it (we use OpenNI/Nite).

## Methods

**Skeleton Tracking:** The skeleton tracking is done automatically by *OpenNI/Nite* from the depth sensor video stream of the Kinect. 14 body joints plus the head are tracked in 3D. We use *OpenGL* to visualize the 3D tracking of the joints and head of the skeleton as seen in Fig 1a,b.

**Object Tracking:** Object tracking requires the fusion of information from both the depth and the color video streams. Our process starts in the color video stream where we segment the contour of a red cup that we are interested in tracking. For all the steps of the segmentation we use OpenCV. The segmentation is based on the object's color and its contour.
In order to carry out the color-based segmentation, we perform a manual color calibration of the color to be segmented by placing the red cup in a scene and adjusting the HSV thresholds.
The actual segmentation is performed by finding the contours of the objects present in the color thresholded image. If several objects of the same color are present in a scene, a contour will be computed for each object. We eliminate contours whose area is smaller than a threshold and for each remaining contour we estimate its center of mass (average *x* and *y* coordinates), see Fig. 1c.
To perform the 3D tracking of the 2D segmented contours, we need to map the segmentation results from the color images into the depth images. To do this it is important to note that the color images and the depth images use different coordinate systems. The color images use *x, y* pixels that represent real world coordinates in millimeters while the depth images contain 3D projective coordinates.
Using built-in functions in OpenNI/Nite it is possible to map coordinates from one video stream into coordinates of the other video stream in order to fuse information from both video sources.
We do not map the coordinates of the whole contours (resulting from the color segmentation) into depth coordinates, instead, we only map the (*x,y)* coordinates of the centers of mass of the detected contours into depth coordinates. This simplified mapping of a single (*x,y)* coordinate per contour allows the system to measure the 3D distance from the centers of mass of the contours to the hand joints of the skeleton in real time.

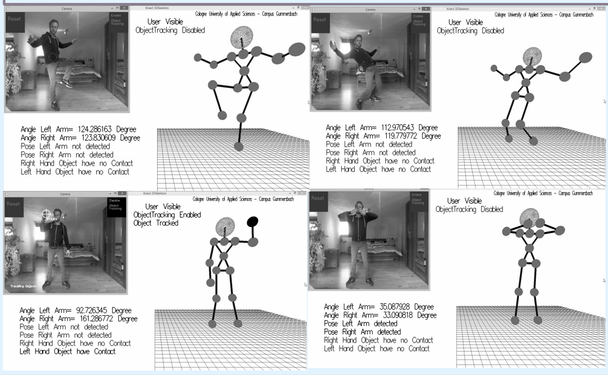

Fig 1:
(a) and (b): Show the user's skeleton in different poses.

(c): The user holding the tracked cup.

(d): Pose without cup detected.

**Automatic Detection of Drinking from a cup/glass.** We fuse the information about both trackers (skeleton and object) to take higher level decisions about the interaction of the skeleton with the object (cup/glass). To determine if the user is drinking from the cup, the system uses different criteria: The position of one hand should be near the head, the arm corresponding to that hand should be flexed and the position of the red cup should match the position of the hand near the head. Next we explain how the measurements are done.

Using the (x,y,z) positions of the shoulder, elbow and hand joints, we calculate four normalized 3D vectors for the right and left, upper and lower arms as follows:

$$v_{upperArm}[i] = |Elbow[i]\ (x,y,z) - Shoulder[i]\ (x,y,z)| \quad (1)$$

$$v_{lowerArm}[i] = |Hand\ [i]\ (x,y,z) - Shoulder[i]\ (x,y,z)|, \quad (2)$$

with: $i = 0,1$ for right and left joints. We calculate the angle between the vectors of the lower and upper arms using:

$$flexAngle_{rad}\ [i] = \arccos(v_{upperArm}\ [i] * v_{lowerArm}[i]) \quad (3)$$

$$flexAngle_{deg}\ [i] = 180 - (180\ /\ (pi * flexAngle_{rad}\ [i])\ ), \quad (4)$$

the system recognizes that an arm is flexed if $flexAngle_{deg}\ [i] < 50°$.
To determine if a hand joint is close to the mouth we do a simple 3D distance measurement from the hand joints to the head:

$$dist_{handHead}[i] = sqrt((Hand[i].x - Head[i].x)^2 + (Hand[i].y - Head[i].y)^2 + (Hand[i].z - Head[i].z)^2\ ), \quad (5)$$

the system recognizes that a hand is close to the head (and therefore the mouth) if $dist_{handHead}[i] < 100$ *pixel*. If both conditions: The hand is close to the head and the corresponding arm is bent, are satisfied, the system detects a successful pose for drinking. However, for the drinking action to be detected, the position of the red cup must match the position of the hand near the head. Fig. 1d shows two examples where a successful pose of the hands is determined (but still no drinking). Last, we compute the 3D distance of each center of mass of the segmented contours (from the color segmentation) to the hand joints:

$$dist_{handContour}[i][j] = sqrt((Hand[i].x - Contour[j].x)^2 + (Hand[i].y - Contour[j].y)^2 + (Hand[i].z - Contour[j].z)^2\ ), \quad (6)$$

for $j = 0..n$ where *n* is the number of segmented contours in the color images. The system takes a decision that the user has drank from the red cup if:

$$Detected_{drinking}\ [i][j] = (dist_{handContour}\ [i][j] < 100)\ \text{AND}\ (dist_{handHead}\ [i] < 100)\ \text{AND}\ (flexAngle_{deg}\ [i] < 50) \quad (7)$$

## Results

Fig. 2 shows a series of scenes where the drinking action, associated to oral medication intake, is detected. As long as the hands of the user and the red cup are visible to the Kinect sensor the detection of the drinking action is very robust. The drinking action is also detected even if the user is only partially visible. The system is able to detect the drinking actions under a variety of situations, for example, if various objects of the same color are present in the scene. In this case, all the objects are segmented and their centers of mass are 3D tracked in real time. As long as the objects are not used in a drinking action they do not alter the results of the detection.

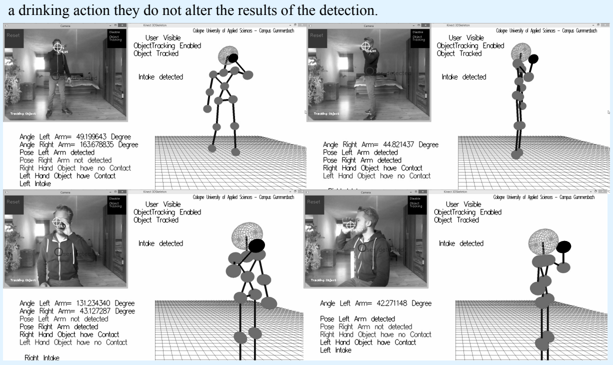

**Figure 2**: Detection of a drinking action: The user has the right pose and is holding a cup in his hand. The drinking action is detected under different poses.

The system also performs robustly when the user is wearing clothes of the same color as the drinking cup. Segmented clothes do not play an important role in the results because the center of mass of their segmentation is always distant from the center of mass of the hands of the user and from the head.

## Conclusions

In this paper we present a system for the detection of drinking actions that can be used to supervise the intake of oral medications by patients who take medications independently. Our system uses the depth and color cameras of a Kinect sensor to track the pose of a user and of a glass/cup of a predetermined color. The system requires a color calibration before the beginning of tracking and performs robustly under varying illumination conditions because computations are done in the HSV color space. One contribution of the paper is the extension of the Kinect skeleton tracking capabilities to tracking arbitrary objects in a scene. The tracking of objects is based on the color segmentation in the color images and the mapping of the centers of mass of the segmented objects to the coordinates in the depth images. We believe the main contribution is our proposed strategy to track objects in the depth images based only on their center of mass. With this strategy we are able to maintain the performance of the tracking system at 30 fps since only one point per object must be tracked. The strategy of only tracking the centers of mass of segmented objects also eliminates conflicts when several objects of the same color are segmented in a scene, since only relatively small objects whose centers of mass coincide with the position of a hand trigger a drinking detection. The main limitation of the system is that it only senses a scene from one point of view. The drinking glass/cup and at least one hand should be visible to the camera for drinking to be detected.

### Bibliography


[1] Nichols-English G., Poirier S., Optimizing Adherence to Pharmaceutical Care Plans, J. of the American Pharm. Ass., 40(4), 2000.

[2] Bond W.S., Hussar D.A., Detection methods and strategies for improving medication compliance. Am J Hosp Pharm., 48, pp. 1978 -- 88, 1991

[3] Brown, M. T., and Bussell J. K., Medication Adherence: WHO Cares?, Mayo Clinic Proceedures, 86(4): 304--314, Apr 2011

[4] Sabaté E, editor. , ed. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.

[5]Lee JK, Grace KA, Taylor AJ. Effect of a pharmacy care program on medication adherence and persistence, blood pressure, and low-density lipoprotein cholesterol: a randomized controlled trial. JAMA, 2006

[6] Osterberg L, Blaschke T. Adherence to medication. N Engl J Med., 353(5), pp.487--497, 2005