Do patients take their medications? An optical based supervision system

Christian Breiderhoff, Maria-Elena Algorri, and Daniel Gaida

Robotics Laboratory, Steinmllerallee 1, 51643 Gummersbach, Germany
{elena.algorri,daniel.gaida}@fh-koeln.de

Abstract. Studies show that 35 to 50% of all prescription medication in Germany are not taken correctly or at all. We use motion capture techniques to build a system that can supervise if a person takes his medications. Our system detects when a patient uses a glass (of a particular color) to take his medications orally by fusing color, contour and depth information extracted from the video streams of a Kinect sensor. The system is able to 3D track arbitrary objects that have been color segmented from the video stream without compromising the speed of the Kinect skeleton tracking. The drinking action is detected by analyzing the pose of a user and the 3D position of the segmented glass. We present the methodology used for color calibration, color segmentation, 3D tracking and information fusion and show that the system performs robustly under different illumination and ambient conditions.

Keywords: Motion Analysis, 3D Tracking, Color Segmentation

1 Introduction

Medication nonadherence is a major public health problem that has been called an "invisible epidemic" [1][2]. In the US nonadherence to pharmacotherapy has been reported to range from 13% to 93%, with an average rate of 40% [3] 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" [4]. Among patients with chronic illness, approximately 50% do not take medications as prescribed [5][6]. This poor adherence to medication leads to increased morbidity and death and is estimated to incur costs of approximately $100 billion per year [7]. There are various forms of medication nonadherence: 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 use motion capture techniques to create a supervision system to help diminish the problem of medication nonadherence in cases where the user forgets to take his medication or takes it out of schedule.

1.1 Optical Tracking

In recent years, optical based motion analysis has been increasingly used in medical applications [9]. Common optical systems record marker positions in order to extract patient motion information [10]. Since camera technology has become less expensive, its use in medical motion analysis is very popular [8]. Lately, 3D depth sensors have also become inexpensive and multi-sensor systems like the affordable Microsoft Kinect (includes an optical camera and a 3D depth sensor), are an attractive choice for motion detection and analysis. Optical motion analysis is used in various application fields such as joint angle derivation [11], evaluation of patient activity [13], patient posture analysis [15], robotics [14] and gesture recognition [12]. In this paper we propose an application that uses the Kinect cameras to supervise if a patient takes his medications orally. We propose a simple but ingenious way to track arbitrary objects for robust scene analysis. Our system is placed so that it can sense the location where a patient usually takes his medications (bathroom, kitchen) and can trigger an alarm if the patient does not adhere to a predetermined schedule.

1.2 Problem Description

We wanted to implement an optical system that could take a decision of whether a person had performed a specific action using a given object. In particular we wanted to implement a system that could detect wether a patient took one of his hands to his mouth while simultaneously holding a red cup 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 closer to the head than a given threshold. Our system must also be capable of tracking an object in 3D space independently of its pose. By fusing the information about the 3D tracking of both the hands and the object, the system should take a decision of whether the person had performed a drinking action using the tracked object.

1.3 System Proposal

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). From the depth sensor video stream, the available SDKs detect persons inside a scene as 3D skeletons containing 15 to 20 joints and continuously track the 3D positions of the joints. Using the RGB video stream we extend the skeleton tracking capabilities to track arbitrary objects inside the scene. We fuse the information about both trackers (skeleton and object), to take a higher level decision about the interaction of the skeleton with the object.

2 Methodology

The two contributions of this paper are: the implementation of a tracker based on color segmentation and an algorithm to fuse information from the object and the skeleton trackers to take robust decisions about the interaction of the skeleton with the object. Both implementations are described next.

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 the 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. The segmentation starts by converting each video frame from the RGB camera to the HSV color space. The HSV representation of the color is more robust against illumination variations than the RGB one because it takes brightness and saturation into account (which are the variations introduced by changing illumination) instead of only intensity levels of red, green and blue (RGB model). 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 as shown in Fig. 1.

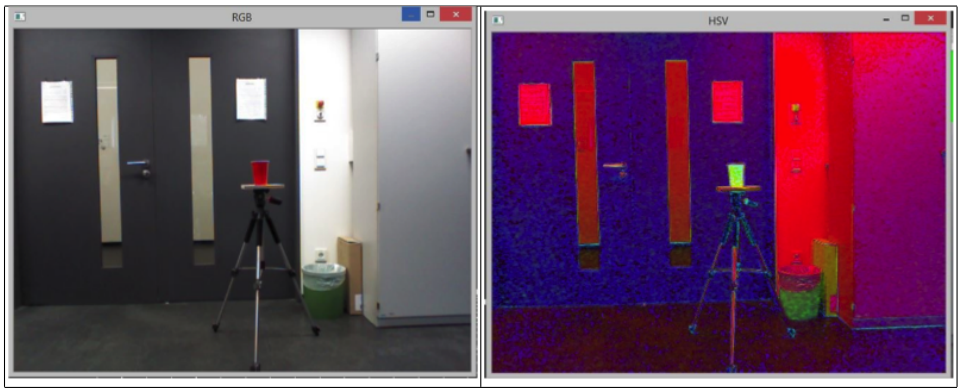

Fig. 1. a)The red cup to segment is used to calibrate the color that will be thresholded from the image. b) Before color calibration the video stream is mapped to the HSV color space.

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.2.

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

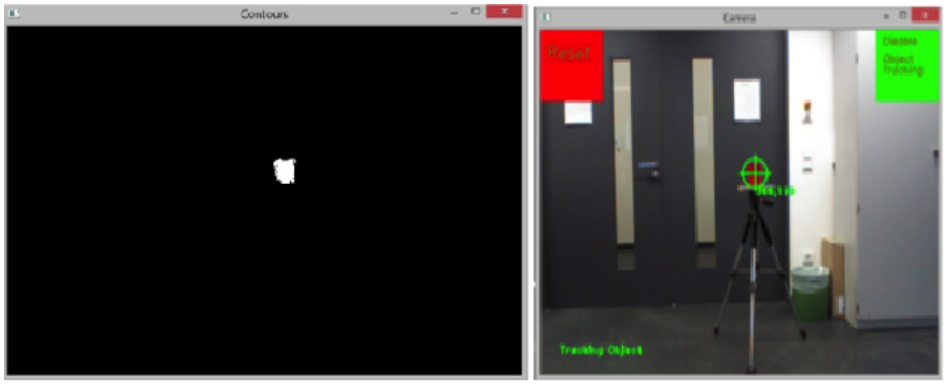

Fig. 2. a) Filled contour extracted from the thresholded image. b) Center of mass of the segmented contour shown in the original image as a green circle.

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 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.

Skeleton Tracking The skeleton tracking is done automatically by OpenNI/Nite. The tracking is done over 14 body joints plus the head. We use OpenGL to visualize the 3D tracking of the joints and head of the skeleton as seen in Fig.3.

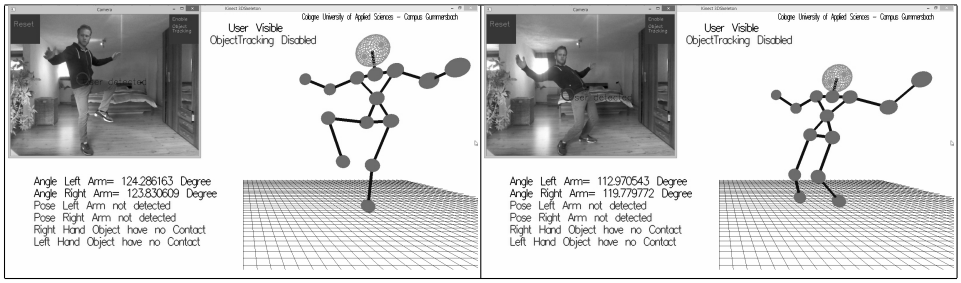

Fig. 3. Skeleton tracking (15 joints) using OpenNI/Nite.

Automatic Detection of Drinking from the Cup For the system to determine if the user is drinking from the cup, different criteria should be met: 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.

From the 3D positions of the shoulder, elbow and hand joints we calculate four normalized 3D vectors for the right and left, upper and lower arms :

$$up\_arm[i] = [elb[i].x - shoul[i].x, elb[i].y - shoul[i].y, elb[i].z - shoul[i].z] \quad (1)$$

$$low\_arm[i] = [hand[i].x - elb[i].x, hand[i].y - elb[i].y, hand[i].z - elb[i].z] \quad (2)$$

$$\hat{up\_arm}[i] = \frac{up\_arm[i]}{||up\_arm[i]||} \quad (3)$$

$$\hat{low\_arm}[i] = \frac{low\_arm[i]}{||low\_arm[i]||} \quad (4)$$

where in Eqs. 1 - 4: $i = 0, 1$ for right and left joints, *elb* = elbow, *shoul* = shoulder. We calculate the angle between the vectors of the lower and upper arms using:

$$\begin{aligned} flex\_angle[i] = arcos[&up\_arm[i].x * low\_arm[i].x + up\_arm[i].y* \\ &low\_arm[i].y + up\_arm[i].z * low\_arm[i].z] \end{aligned} \quad (5)$$

$$flex\_angle[i] = 180 - \frac{180}{\pi * flex\_angle[i]} \quad (6)$$

The system recognizes that an arm is flexed if $flex\_angle[i] < 50°$.

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:

$$\begin{aligned} dist\_hand\_head[i] = sqrt((&hand[i].x - head.x)^2 + (hand[i].y - head.y)^2 + \\ &(hand[i].z - head.z)^2) \end{aligned} \quad (7)$$

The system recognizes a hand is close to the head (and therefore the mouth) if $dist\_hand\_head[i] < 100$ pixels. If the conditions that a 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. 4 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 to the hand joints.

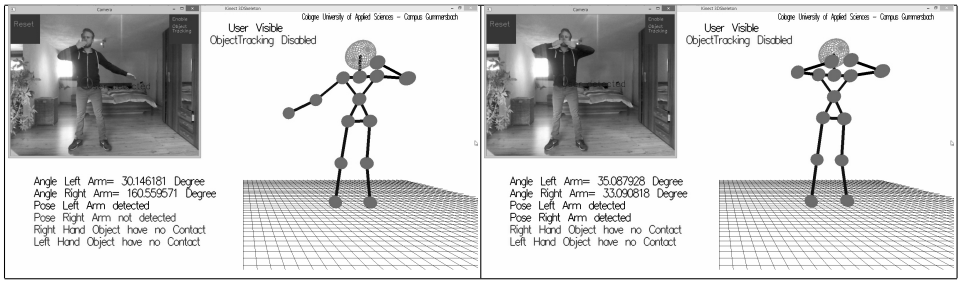

Fig. 4. Detection of drinking pose of the hands without the cup.

$$dist\_hand\_contour[i][j] = sqrt((hand[i].x - contour[j].x)^2 + (hand[i].y - contour[j].y)^2 + (hand[i].z - contour[j].z)^2) \quad (8)$$

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:

$$drinking\_detected = (dist\_hand\_contour[i][j] < 100)AND (dist\_hand\_head[i] < 100)AND(flex\_angle[i] < 50); \quad (9)$$

3 Results

Fig. 5 shows two scenes where the user is holding the cup in his hands, but because the drinking pose is not correct, no drinking action is detected. Fig. 6 show a series of scenes where the drinking action 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. However, as long as the objects are not used in a drinking action they do not alter the results of the detection.

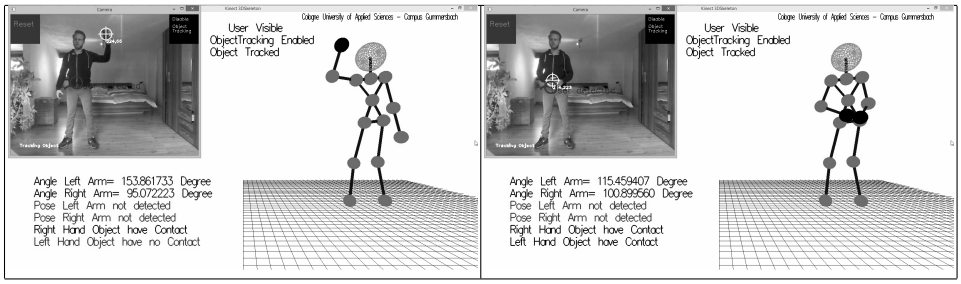

Fig. 5. The system detects that the user is holding the cup in his hand but is not drinking from it.

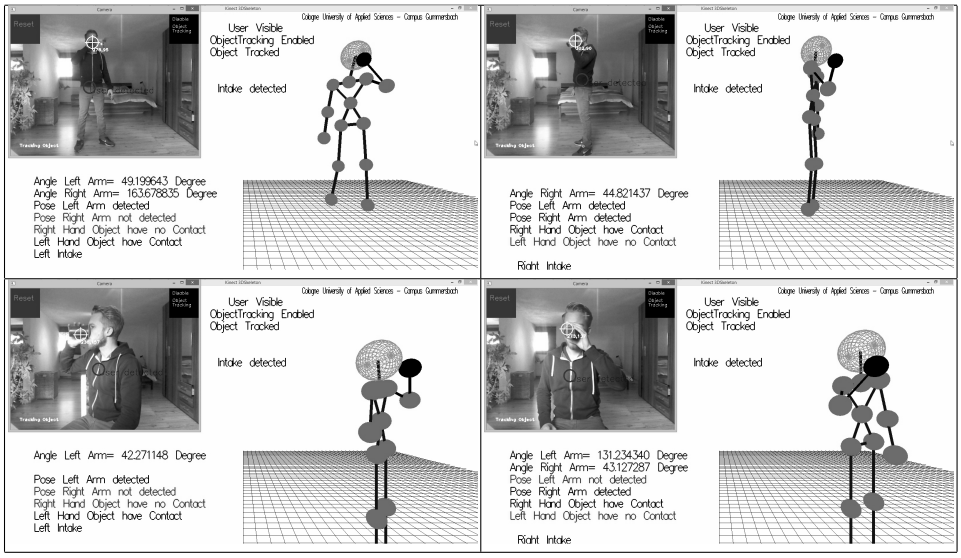

Fig. 6. Detection of a drinking action: The user has the right pose and is holding a cup in his hand. The drinking action is detected under different poses.

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.

4 Discussion

In this paper we present a system for the detection of drinking actions that can be used to supervise patients that need to take medications independently. Our system uses the depth and color cameras of a Kinect sensor to track the pose of a user as well as an object 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 object and at least one hand should be visible to the camera for drinking to be detected. In the current scenario, if a user is standing with his back to the camera no drinking actions will be detected. Our system needs to be expanded with at least a second Kinect sensor to have better visibility of the user independently of his pose. The methodology proposed in this paper can be used for a variety of purposes where a user interacts with an object in a scene.

References

1. Smith, M.C,. Predicting and detecting noncompliance. In: Smith MC, Wertheimer AI, eds. Social and Behavioral Aspects of Pharmaceutical Care. New York, NY: Pharmaceutical Products Press, Inc; 1996.
2. Nichols-English G., Poirier S., Optimizing Adherence to Pharmaceutical Care Plans, Journal of the American Pharmaceutical Association, 40(4), 2000
3. Bond W.S., Hussar D.A., Detection methods and strategies for improving medication compliance. Am J Hosp Pharm., 48, pp. 1978 – 88, 1991
4. Brown, M. T., and Bussell J. K., Medication Adherence: WHO Cares?, Mayo Clinic Proceedures, 86(4): 304–314, Apr 2011
5. Sabat E, editor. , ed. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
6. 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
7. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med., 353(5), pp.487–497, 2005
8. Kalkbrenner, C., Hacker S., Algorri, M.E., Blechschmidt-Trapp, R., Motion capturing with IMU and Kinect, Tracking of limb movement using optical and orientation information, Biodevices 2014, France
9. Claasen, G., Martin, P., and Picard, F, Highbandwidth low-latency tracking using optical and inertial sensors, Automation, Robotics and Applications (ICARA), 5th International Conference on, pages 366 – 371, 2011
10. Liguo, H., Yanfeng, Z., and Lingyun, Z, Body motion recognition based on acceleration sensor, Electronic Measurement Instruments (ICEMI), 10th International Conference on, volume 1, pages 142 – 145, 2011
11. Bo, A., Hayashibe, M., and Poignet, P, Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect. In Engineering in Med. and Biol. Soc., EMBC, Annual International Conference of the IEEE, pages 3479 – 3483, 2011
12. Patsadu, O., Nukoolkit, C., and Watanapa, B., Human gesture recognition using kinect camera, Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on, pages 28 – 32, 2012
13. Cordella, F., Di Corato, F., Zollo, L., Siciliano, B., and Van Der Smagt, P., Patient performance evaluation using Kinect and Monte Carlo-based finger tracking, Biomedical Robotics and Biomechatronics (BioRob), 4th IEEE RAS EMBS International Conference on, pages 1967 – 1972, 2012
14. El-laithy, R., Huang, J., and Yeh, M., Study on the use of microsoft kinect for robotics applications, Position Location and Navigation Symposium (PLANS), IEEE/ION, pages 1280 – 1288, 2012
15. Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., Pavel, M., Accuracy and robustness of kinect pose estimation in the coaching of elderly population, Eng. Med. and Bio. Soc. (EMBC), Int. Conf. IEEE, pp 1188 – 1193, 2012