Using Computer Vision and Depth Sensing to Measure Healthcare Worker-Patient Contacts and Personal Protective Equipment Adherence Within Hospital Rooms

Junyang Chen,[1] James F. Cremer,[1] Kasra Zarei,[1] Alberto M. Segre,[1] and Philip M. Polgreen[2]

[1]Department of Computer Science and [2]Departments of Internal Medicine and Epidemiology, University of Iowa, Iowa City

Background. We determined the feasibility of using computer vision and depth sensing to detect healthcare worker (HCW)-patient contacts to estimate both hand hygiene (HH) opportunities and personal protective equipment (PPE) adherence.

Methods. We used multiple Microsoft Kinects to track the 3-dimensional movement of HCWs and their hands within hospital rooms. We applied computer vision techniques to recognize and determine the position of fiducial markers attached to the patient's bed to determine the location of the HCW's hands with respect to the bed.

To measure our system's ability to detect HCW-patient contacts, we counted each time a HCW's hands entered a virtual rectangular box aligned with a patient bed. To measure PPE adherence, we identified the hands, torso, and face of each HCW on room entry, determined the color of each body area, and compared it with the color of gloves, gowns, and face masks. We independently examined a ground truth video recording and compared it with our system's results.

Results. Overall, for touch detection, the sensitivity was 99.7%, with a positive predictive value of 98.7%. For gowned entrances, sensitivity was 100.0% and specificity was 98.15%. For masked entrances, sensitivity was 100.0% and specificity was 98.75%; for gloved entrances, the sensitivity was 86.21% and specificity was 98.28%.

Conclusions. Using computer vision and depth sensing, we can estimate potential HH opportunities at the bedside and also estimate adherence to PPE. Our fine-grained estimates of how and how often HCWs interact directly with patients can inform a wide range of patient-safety research.

Keywords. computer vision; gloves; gowns; hand hygiene; personal protective equipment.

In recent studies, several approaches have been developed to electronically monitor hand hygiene adherence [1–3]. It is relatively straightforward to determine whether a healthcare worker has practiced hand hygiene on entering or leaving a room. However, it is much more difficult to monitor what is happening inside a patient room at the bedside, where most of the opportunities to spread pathogens from the hands of healthcare workers to patients occur. Accordingly, much infection control attention has been focused on observing such opportunities (eg, the World Health Organization [WHO] 5 moments [4]).

If we could determine, in an automated fashion, when a healthcare worker's hands come close enough to touch a patient or the patient's immediate environment, we could produce more informed estimates of hand hygiene opportunities and ultimately adherence. Moreover, if we could analyze color information related to healthcare workers' hands, torsos, and faces, we could determine whether they were wearing gloves, masks, and gowns, respectively. Thus, we could measure personal protective equipment (PPE) adherence, a common intervention designed to prevent the spread of healthcare-associated infections [5, 6].

The purpose of this study is to determine the feasibility of using computer vision and depth sensing to measure not only direct patient contacts, but also compliance with recommendations for healthcare workers to wear gloves, gowns, and masks.

METHODS

To measure healthcare worker movement we used a Kinect placed on either side of a patient's bed. A Kinect is a motion-sensing device developed by Microsoft to enable the use of body movements for controlling video games. However, the Kinect can be also be used for other purposes via custom software that controls and communicates with the Kinect through its software development kit (SDK) [7, 8].

Each Kinect includes a color video camera combined with a depth sensor, enabling analysis of both color and distance of objects. The Kinect's proprietary software processes these fine-grained range and color data to provide high-level SDK functions for tracking 3-dimensional (3D) movements of individuals. Because the Kinect was designed for multiple

simultaneous users, it can recognize up to 6 axial skeletons. At any moment, 2 skeletons can be tracked in full detail, meaning the Kinect provides 3D positions of 20 predefined skeletal points on each skeleton. The Kinect tracks only a representative "center" position of any additional skeletons in view. In addition to the SDK-provided functionality, we apply standard computer vision techniques to raw Kinect data to recognize and accurately determine the position of fiducial markers attached to key locations or equipment; this enables our software to locate the Kinect relative to room features of interest.

To protect patient privacy, our software does not capture any human interpretable image or video data for more than a few milliseconds before deleting it. The only data saved are in the form of Cartesian (x, y, and z) coordinates from which we can compute distances and recreate skeletal paths and the locations of fiducial markers. None of the saved data could be used to reconstruct recognizable images of faces or anatomic detail above the level of a stick figure with 20 skeletal points.

For this work, we focus on tracking the hand positions of axial skeletons as they move near the patient, saving only 3D trajectories of hand center points during contact events. For PPE compliance, we determine and record only binary decision information—eg, gloves vs no gloves, mask vs no mask, and gown vs no gown—based on video and depth data that are analyzed in real time and discarded immediately. The University of Iowa's Institutional Review Board determined our approach to be nonhuman-subjects research.

Tracking Potential Healthcare Worker-Patient Contacts

We developed custom software that uses the Kinect SDK to obtain data that allow us to do the following: (1) identify axial skeletons of up to 6 people within view of each Kinect; (2) track detailed 3D locations of 20 defined "joint" positions on any 2 of the skeletons; (3) track 3D location of the "center" of any additional skeletons; (4) recognize fiducial markers placed at the foot of a hospital bed and compute the relative position between these markers and the Kinect devices; (5) compute bed position and orientation based on known position of the Kinect with respect to fiducial markers; and (6) track position of the skeletal hands with respect to the patient bed.

We placed the 2 Kinects on opposite sides of a hospital bed, each near the head facing diagonally across the bed toward the foot end (see Figure 1). Thus, given that we know the position of the Kinects in relation to the bed, the size of the bed, the
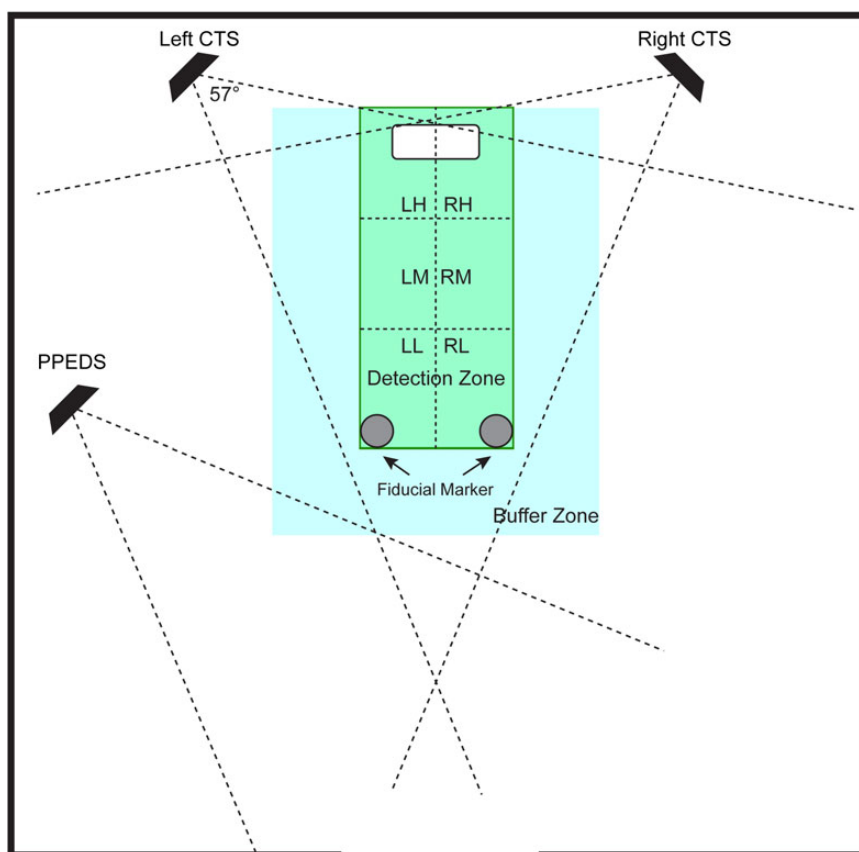
Figure 1. Two Kinects (Left CTS and Right CTS) comprise the Contact Tracking System that tracks healthcare worker hands near the patient. A third Kinect (PPEDS) gathers data for the Patient Protective Equipment Detection System that monitors personal protective equipment compliance. The detection zone refers to the bed itself. Bed areas include left head (LH), right head (RH), left middle (LM), right middle (RM), left lower (LL), and right lower (RL). The buffer zone refers to the area that is within touching range of the bed.
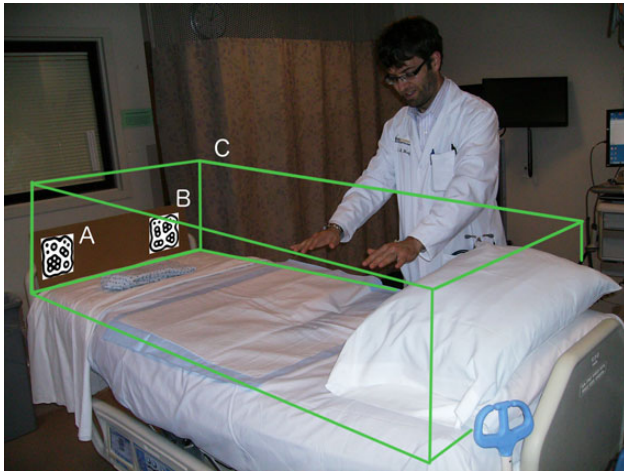
Figure 2. Here, a healthcare worker places both of his hands into the virtual rectangular box above the patient's bed. The box is superimposed on the photo for illustrative purposes. Although we track the motion of healthcare worker hands in the view of each Kinect, we only count when hands enter the virtual box as a potential healthcare worker-patient contact. Note that at the foot of the bed, there are fiducial markers that allow us to "find the location of the patient's bed" in order to determine the location of healthcare worker hands with respect to the bed.

crosses the boundary of a virtual rectangular box aligned with the patient's bed (ie, having the same length and width as the bed and a height reaching 0.8 m up from the bed surface. See Figure 2).
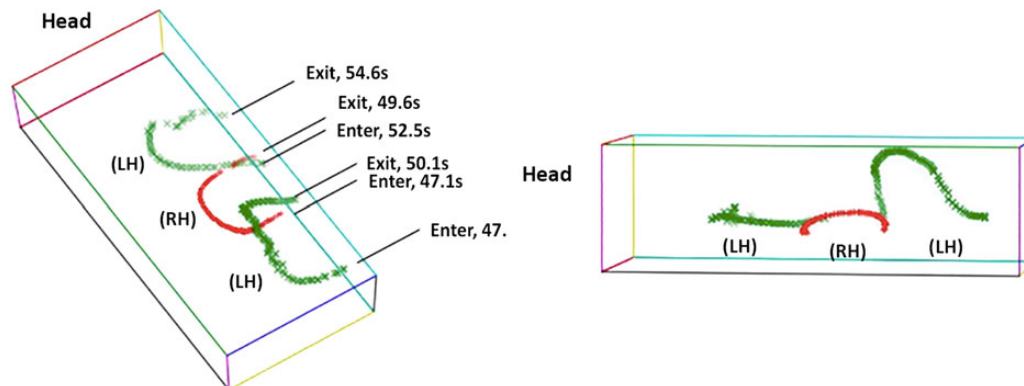
Tracking Personal Protective Equipment Compliance

Using the video and depth sensor data from the Kinect, our software analyzes color and movement information to determine whether healthcare workers are compliant with use of PPE.

Specifically, to detect compliance with gloves, gowns, and masks, we place a third tripod-mounted Kinect within a hospital room facing the entrance (see Figure 3) and proceed as follows. First, the software determines whether a healthcare worker is entering or leaving a room by analyzing the motion of an axial skeleton. The Kinect provides skeletal point position information at 30 Hz, which allows us to track point positions relative to the room entrance over a sufficient period of time to determine whether a skeleton near the doorway is moving into or out of the room.

Next, the software samples pixels in regions of the video image corresponding to the following: the skeleton's hands, the skeleton's torso, and the mouth and nose of the skeleton. The software then determines whether the sampled colors match the color of the gloves, gowns, and face masks, respectively. Thus, if the pixel colors sampled in location of the hands matches the color of the gloves, we record that the healthcare worker is wearing gloves; if the colors do not match closely enough, we record that the healthcare worker is not wearing gloves. We record gown and face mask compliance in a similar fashion.

It is important to note that we implemented custom tracking software in C#, making use of both the Kinect SDK for Windows and the reacTIVision image processing library for tracking fiducial markers. (http://reactivision.sourceforge.net). Finally, we also wrote postcollection data processing and visualization software in Python.

location of the bed, and the location of skeletal hands, we can estimate when and for how long a healthcare worker's hands are within touching range of a patient. It is important to note that the software continually monitors the position of the fiducial markers, enabling the system to continue to function correctly even if the bed or the Kinect is moved.

To determine whether a healthcare worker touched a patient, we make 2 simplifying assumptions. First, when the hand of a healthcare worker comes close enough to touch a patient, we deem the event a contact. Second, we only consider contacts that would occur while a patient is in the hospital bed. Thus, a contact is defined to occur when the hand of a healthcare worker

Figure 3. Two views of sample trajectories of healthcare worker hands in a virtual rectangular box above the patient bed. In each figure, the rightmost green trajectory (LH) and red trajectory correspond to an ~3-second long 2-handed patient contact in which the healthcare worker's left and right hands entered the region at time $t = 47$ seconds and exited the region at near $t = 50$ seconds. The leftmost green trajectory corresponds to a patient contact with just the healthcare worker's left hand, beginning at time $t = 52.5$ seconds and ending at time $t = 54.6$ seconds. Abbreviations: LH, left hand; RH, right hand.

Evaluation of Accuracy

To determine whether our system can detect when and for how long healthcare workers touch a patient or are close enough to touch a patient, we tested our system in an empty hospital room. We also attached a video camera to the room's ceiling directly above the hospital bed. Note that this additional video camera is not part of our hand tracking system; it was used solely to capture "ground truth" to support the evaluation of our system.

Next, we recruited a number of test subjects to approach the hospital bed and touch or not touch different areas of the bed repeatedly, as if they were examining a patient. The different scenarios described below were designed to test whether our system could detect "contacts" both short and long in all areas of the virtual box. In addition, we wanted to test situations mimicking common touch encounters typical of clinical practice (eg, doing a clinical exam, accessing the patient, washing or cleaning the patient). It is important to note that the data produced by our Kinect-based tracking system represent a set of detailed contact events over time (eg, hand enters the virtual rectangular box aligned with the patient bed at location a and time $t1$ then tracks the hand trajectory until it exits the virtual rectangular box at position b and time $t2$). We independently and visually examined the ground truth video recording, producing a manual count and characterization of each contact event. We compared the tracking system's results with the manual results to compute sensitivity and precision.

To determine whether we can accurately detect whether subjects were wearing gloves, gowns, or facemasks when entering a hospital room, we had subjects repeatedly enter and exit either wearing or not wearing gloves, gowns, and facemasks. We again compared the tracking system's results with manual counts from ground truth video data.

Statistical Analysis

We computed touch-detection sensitivity and also calculated the precision of our system (positive predictive value). For this application, specificity is not well defined because there is no fully satisfactory way to quantify the number of opportunities for false positives (times when the healthcare workers are not touching anything within the rectangular box). To evaluate PPE compliance, we computed the sensitivity and specificity for detecting gloves, gowns, and facemasks.

RESULTS

Tracking Healthcare Worker Hands

In 4 separate scenarios, we instructed 1 or more subjects to approach a patient's bed and place their hands on the bed in various positions: Scenario 1 comprised 1 subject, both sides of bed, 60 seconds; Scenario 2 comprised 1 subject 1 side of bed, 270 seconds; Scenario 3 comprised 3 subjects, both sides of bed, 135 seconds; Scenario 4 comprised 2 subjects, both sides of bed, 480 seconds.

In Scenario 1, 1 subject moved from the head of the left of the bed toward the foot, placing his hands in and out of the bed zone in various ways. The subject then moved to the right side of the bed, working from foot to head and then back to the foot. Finally, the subject returned to left side and moved foot to head once more. During the 60-second scenario, 10 in-range bed-touching events occurred on the left side and 12 on the right side. All 22 events were detected by our system.

In Scenario 2, 1 subject, a physician, moved back and forth along the right side placing hands in and out of the bed in various ways. During the 270-second scenario, 97 bed-touching events occurred and all 97 were detected by our system. There were 4 false-positive detections that seemed to be due to window reflections in the background (these 4 spurious events correspond to noticeably erratic data that could easily be discarded with appropriate postprocessing).

In Scenario 3, 3 subjects moved back and forth along both sides of the bed, sometimes on opposite sides, sometimes all on the same side. During the 135-second scenario, 68 bed-touching events occurred. On the left side, there were 8 actual touch events and all were detected by our system. On the right side, there were 60 actual touches, many of them by 2 subjects at once, with 61 touches detected by our system. The extraneous touch detection occurred during a short sequence of 4 touches involving 2 subjects who were entering/leaving the bedside, where we falsely detected parts of 2 separate touches as an additional touch.

In Scenario 4, 2 subjects moved back and forth along both sides of the bed, usually only 1 at the bedside at a time. During the 408-second scenario, 181 bed-touching events occurred. On the right side, 45 actual touches occurred and all were detected by the system. On the left side, there were 136 touching events. Our system correctly distinguished 134 of them. Two touches in the same area of the bed with a 1 second gap between them were reported as 1 continuous touch. These events occurred in the extreme lower left region of the bed farthest from the observing Kinect.

Overall, for touch detection the sensitivity was 99.7%, and the precision (positive predictive value) was 98.7%.

Assessment of Touch Event Time and Trajectory Accuracy

Most touches lasted between 2 and 7 seconds, with the average being ~3 seconds. Approximately 90% of the touches, both the touch start and touch end times determined by the system, matched those determined by a viewer watching the recorded overhead video to within 0.25 seconds, with trajectories tracking well both in position and time. For ~8% of the touches, the system experienced 1 or more short gaps (usually .25–.5 seconds) in position tracking. For these touches, the total touch time remains accurate, but position data are less detailed. For ~2% of the touches, the system reported either too-early-start or too-late-end times that were incorrect by between .25 and 2 seconds.

Tracking Personal Protective Equipment Compliance

We instructed 3 different test subjects to enter and leave a patient room in rapid succession. There were 122 actual entries. The system missed 6 entries and detected 2 extraneous entries, reporting a total of 118 entries. The 2 "false-positive" entries resulted from other people in the hall who started to enter the room but did go far enough to qualify an entry. In the remainder of the section, "entries" refers to entries among the 116 correctly detected actual entries. For gowned entrances, sensitivity was 100.0% and specificity was 98.15%. For masked entrances, sensitivity was 100.0% and specificity was 98.75%; for gloved entrances, the sensitivity was 86.21% and specificity was 98.28%.

DISCUSSION

Our results demonstrate that we can use computer vision and rangefinding to better understand hand hygiene opportunities at the patient's bedside. Although we cannot yet determine when a healthcare worker is compliant with hand hygiene according to the specific WHO 5 moments [4], we can estimate how often and for how long a healthcare worker's hands are within touching distance of a patient in a bed, and we can do this without the healthcare worker wearing any specialized badge or equipment. We can use the same computer vision approach to determine whether healthcare workers are entering or exiting a room and whether or not they are wearing gloves, gowns, or facemasks and thus determine compliance with PPE guidelines.

Historically, hand hygiene data have been both sparse and coarse. Human observations, although considered the gold standard [9], suffer from multiple limitations especially in busier healthcare settings [10]. In most studies, the number of observations performed by human observers is quite small in comparison to the number of true opportunities [11]. Moreover, even where and when observations occur affect the number and diversity of opportunities observed [11]. Finally, although a host of new automated monitoring approaches will help inform future hand hygiene-research efforts, most of these investigations will likely focus on in- and out-of-room-based measurements.

Computer vision approaches have a number of advantages. Although they are based on "imaging", they need not capture any recognizable images beyond the few milliseconds required for data extraction, and no saved data can be used to recognize patients or healthcare workers. Furthermore, because our approach is automated, we do not need humans to monitor the data collected or review images, as video-based hand hygiene approaches have previously required [12, 13]. Because so little is known about what happens within patient rooms, our approach could help inform hand hygiene research even without widespread deployment to answer central questions related to hand hygiene monitoring. For example, on average, how many times does a healthcare worker touch a patient while in a room for 15 seconds vs 30 seconds vs 60 seconds? How does the number of healthcare worker-patient direct contact times differ in different healthcare units (ie, intensive care units, general medical wards)?

Answers to these questions will also help inform current observer-based monitoring approaches. For example, human observers often count an opportunity as any entrance into a room regardless of how long a healthcare worker spends in the room: should all in-room dwell times be considered equally? Other infection control monitoring is also likely to profit from the kind of fine-grained information produced by our system. Room "heat maps" highlighting high-touch surfaces could be used to better understand the order of operations in daily patient care as well as to inform cleaning policies and procedures.

Another benefit from computer vision approaches to compliance monitoring is that we can also track adherence to other important infection control measures, specifically the use of PPE. The wide-ranging preparedness efforts on the part of hospitals after the Ebola outbreak in Western Africa [14, 15] highlights the importance of efforts to improve PPE compliance. Data regarding compliance with PPE are sparse relative to data regarding hand hygiene adherence. However, existing reports show suboptimal adherence [16], and healthcare workers may need PPE technique training [17]. Furthermore, glove use may actually contribute to suboptimal hand hygiene [18, 19]. Likewise, although masks and respirators are important to help prevent the spread of droplet and airborne infections, respectively [5], there are few if any ways to monitor compliance with these measures other than by direct human observations.

The fine-grain movement data collected from our system has several patient safety applications beyond estimating hand hygiene opportunities, monitoring PPE compliance, or even determining the cleaning of high-touch surfaces around patient beds. First, data from our approach can better inform infection control modeling. For example, incorporating granular measurements of contact patterns into models affects results [20]. Similarly, modeling efforts that incorporate the probability of a healthcare worker touching a patient may yield dramatically different results than simply considering whether the healthcare worker entered a room. Thus, our approach could inform such models by providing probability estimates of touches based upon how long a healthcare worker spent in a room. Second, we train our system to detect patterns of movement that might indicate agitation or levels of sedation for a ventilated patient. It may also be possible to train a system to detect risks for falls or developing decubitus ulcers.

Our work has several limitations. First, we cannot currently differentiate between healthcare workers and visitors. We could address this limitation if healthcare workers wore identifiably different apparel or fiducial markers. Second, the Kinect can only detect healthcare workers within its direct field of view; if another healthcare worker occludes the Kinect's field of view, our ability to monitor will be limited as long as the

obstruction persists. This could be partially overcome by hanging the Kinects from the ceiling. Third, our work focused on the zone around the patient's bed. Some hand hygiene opportunities occur away from patient's bed; these events would be missed unless we added additional Kinects. Fourth, we cannot recognize whether healthcare workers are performing specific sterile or clean procedures, although with suitable training data, it should be possible to apply machine learning methods to identify specific healthcare worker events (eg, central line placement, intubation, bathing a patient). Because some of these procedures are "clean", it may be possible to differentiate between specific sterile or clean procedures to better inform hand hygiene opportunities. Fifth, the current version of Kinect used here can only track 2 healthcare workers at one time (although a new version addresses this limitation). Sixth, our ability to measure some colors depends upon lighting; although blue was easy to detect, it was more difficult to differentiate between yellow and skin colors in low-light situations. Future versions could automatically adjust for lighting conditions or give ranges of confidence of estimates depending upon lighting conditions. Furthermore, we cannot detect the use of skin tone-colored gloves.

A final limitation to our system is it was designed for an intensive care unit. In units where patients spend more time out of bed, our current software will not always be able to differentiate between the hands of healthcare workers and those of patients. The Kinect was designed to find people with an upright posture with most of their body visible. Thus, the potential for patient hands being confused with those of healthcare workers is low when patients are lying down. We evaluated our system with both manikins and real people inside our virtual box (data not shown). The Kinect can detect patients' hands if patients are sitting up with their arms exposed, and the hands of patients could be confused with the hands of healthcare workers as patients transition from being in bed to being out of bed or sitting on the side of the bed. Additional processing would need to occur to control or adjust for such situations.

CONCLUSIONS

Despite our limitations, we demonstrate that we can perform estimates of hand hygiene opportunities and PPE adherence using computer vision. Future studies deploying this approach will help provide fine-grained estimates of how and how often healthcare workers interact directly with patients. Such data will not only help inform hand hygiene-monitoring efforts but will also have other patient safety applications.

References

1. Boyce JM. Measuring healthcare worker hand hygiene activity: current practices and emerging technologies. Infect Control Hosp Epidemiol 2011; 32:1016–28.
2. Polgreen PM, Hlady CS, Severson MA, et al. Method for automated monitoring of hand hygiene adherence without radio-frequency identification. Infect Control Hosp Epidemiol 2010; 31:1294–7.
3. Marra AR, Edmond MB. New technologies to monitor healthcare worker hand hygiene. Clin Microbiol Infect 2014; 20:29–33.
4. Sax H, Allegranzi B, Chraïti MN, et al. The World Health Organization hand hygiene observation method. Am J Infect Control 2009; 37:827–34.
5. Weber DJ, Rutala WA, Schaffner W. Lessons learned: protection of healthcare workers from infectious disease risks. Crit Care Med 2010; 38:S306–14.
6. Williams CK, Carnahan H. Development and validation of tools for assessing use of personal protective equipment in health care. Am J Infect Control 2013; 41:28–32.
7. Zhang Z. Microsoft kinect sensor and its effect. MultiMedia, IEEE 2012; 19:4–10.
8. Han J, Shao L, Xu D, Shotton J. Enhanced computer vision with Microsoft Kinect sensor: a review. IEEE Trans Cybern 2013; 43:1318–34.
9. Boyce JM. Hand hygiene compliance monitoring: current perspectives from the USA. J Hosp Infect 2008; 70:2–7.
10. Sharma D, Thomas GW, Foster ED, et al. The precision of human-generated hand-hygiene observations: a comparison of human observation with an automated monitoring system. Infect Control Hosp Epidemiol 2012; 33:1259–61.
11. Fries J, Segre AM, Thomas G, et al. Monitoring hand hygiene via human observers: how should we be sampling? Infect Control Hosp Epidemiol 2012; 33:689–95.
12. Armellino D, Trivedi M, Law I, et al. Replicating changes in hand hygiene in a surgical intensive care unit with remote video auditing and feedback. Am J Infect Control 2013; 41:925–7.
13. Armellino D, Hussain E, Schilling ME, et al. Using high-technology to enforce low-technology safety measures: the use of third-party remote video auditing and real-time feedback in healthcare. Clin Infect Dis 2012; 54:1–7.
14. Association for Professionals in Infection Control and Epidemiology (APIC). Ebola preparedness one year later: a poll of APIC members 2015. Available at: http://www.apic.org/Resource_/TinyMceFileManager/Topic-specific/APIC_Ebola_Survey_Results_November_2015.pdf. Accessed 27 November 2015.
15. Polgreen PM, Santibanez S, Koonin LM, et al. Infectious disease physician assessment of hospital preparedness for Ebola virus disease. Open Forum Infect Dis 2015; 2:ofv087.
16. Clock SA, Cohen B, Behta M, et al. Contact precautions for multidrug-resistant organisms: current recommendations and actual practice. Am J Infect Control 2010; 38:105–11.
17. Beam EL, Gibbs SG, Boulter KC, et al. A method for evaluating health care workers' personal protective equipment technique. Am J Infect Control 2011; 39:415–20.
18. Girou E, Chai SH, Oppein F, et al. Misuse of gloves: the foundation for poor compliance with hand hygiene and potential for microbial transmission? J Hosp Infect 2004; 57:162–9.
19. Fuller C, Savage J, Besser S, et al. "The dirty hand in the latex glove": a study of hand hygiene compliance when gloves are worn. Infect Control Hosp Epidemiol 2011; 32:1194–9.
20. Hornbeck T, Naylor D, Segre AM, et al. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. J Infect Dis 2012; 206:1549–57.