Monitoring Patients in Hospital Beds Using Unobtrusive Depth Sensors

Tanvi Banerjee, *IEEE Student Member*, Moein Enayati, James M. Keller, *IEEE Life Fellow*, Marjorie Skubic, *IEEE Senior Member,* Mihail Popescu, *IEEE Senior Member*, and Marilyn Rantz

***Abstract*— We present an approach for patient activity recognition in hospital rooms using depth data collected using a Kinect sensor. Depth sensors such as the Kinect ensure that activity segmentation is possible during day time as well as night while addressing the privacy concerns of patients. It also provides a technique to remotely monitor patients in a non-intrusive manner. An existing fall detection algorithm is currently generating fall alerts in several rooms in the University of Missouri Hospital (MUH). In this paper we describe a technique to reduce false alerts such as pillows falling off the bed or equipment movement. We do so by detecting the presence of the patient in the bed for the times when the fall alert is generated. We test our algorithm on 96 hours obtained in two hospital rooms from MUH.**

INTRODUCTION

Falls occurring in hospital settings can cause severe emotional and physical injury to patients as well as increase healthcare costs for the hospitals as well as the patients and their families [4]. Studies such as [10] have indicated that any single intervention has proven insufficient to reduce falls in hospitals as well as care home facilities since the causes for the falls are so varied among the patients as well as the residents. Recent progress in using depth imagery to detect different activities makes it an efficient and unobtrusive technology for activity monitoring. This is of particular interest for remotely monitoring people at risk of falling or other hazards such as patients in hospital rooms. With multiple hospital rooms in each unit and limited staff for these units, it becomes imperative to have a monitoring system that can detect patient activity without infringing on their privacy. To that end, the Microsoft Kinect sensor offers a low cost depth sensor device that allows for a three dimensional (3D) representation of the sensor's field of view. The depth imagery provides the added advantage of functioning just as effectively under low illumination at night as in the day time which makes it suitable for around the clock monitoring. In this work, we discuss our algorithm to detect the presence of a patient in the bed as a means to reduce false alarms from an existing fall detection algorithm. This fall detection algorithm uses depth data collected using the Microsoft Kinect sensor mounted individually in each hospital room and generates real-time alerts for all the hospital rooms [11]. The work described in this paper complements the fall detection algorithm by eliminating some of the false alarms which occur when the patient is in the bed. This will strengthen the existing algorithm by improving its accuracy as well as prevent the need for the hospital staff to check on the patient when the alert is generated. The results are further described and analyzed in the rest of the paper.

RELATED WORK

There have been studies to monitor hospital rooms using depth images. In [6], Lea et al. used manually placed "markers" at specific locations such as the head of the patient, the ventilator, the staff computer. These selected points were then used to trigger specific activities related to the interaction of the hospital staff and the patient such as documentation observing, checking diagnostics, urinary catheter removal, and ventilator use. Along with this, the researchers also computed the orientation of the 3D data of the foreground at each image frame. These features were then input to classifiers to identify the activities. The classifiers implemented were the Support Vector Machine as well as the Decision Forest classifiers. The latter achieved a better performance with an average overall classification rate of 75%. Here, data were collected in an ICU Unit for a period of approximately 5 hours. In another approach Ni et al. [8] used a combination of Histogram of Oriented Gradients (HOG), Histogram of Optical Flow (HOF) and Motion History Images (MHI) to get a 48 dimensional feature vector for the manually selected bed region. This was then input to a Multiple Kernel Learning classifier to detect the event "patient gets up from the bed". The

Manuscript received April 15, 2011. This work was supported in part by the U.S. National Science Foundation under Grant 5R01HS018477.

T. Banerjee, M. Enayati, J. Keller, and M. Skubic are with the Electrical and Computer Engineering Department at the University of Missouri, Columbia, MO 65211 (e-mails: tsbycd@mizzou.edu, mektb@mizzou.edu, kellerj@missouri.edu, skubicm@missouri.edu).

M. Popescu is with Health Management and Informatics, University of Missouri Columbia, MO 65211 (email: popescum@missouri.edu)

M. Rantz is with School of Nursing, University of Missouri, Columbia, MO 65211 (email: rantzm@health.missouri.edu)

This work was supported in part by the Agency for Healthcare Research and Quality under grant R01-HS018477 (Rantz, PI).

results were tested on .5 hours data from a hospital room to achieve and overall accuracy rate of 98%. In bed detection algorithms, Kittipanya et al. [5] used Hough Transform to detect all lines in the color image from regular camera data after extracting the edges using the canny edge detected and then classified the lines into four categories (four sides of the bed) based on the angle and mid-point location constraints. The average of these groups of lines was then used to detect the bed location. The results were tested on around 1000 images randomly sampled from ten different datasets. The algorithm failed to detect some of the edges when there were nursing staff too close to the bed or when there were visitors which seemed to be the major drawback of this technique.

In this work, we do not propose to detect the bed directly since all the edges may not be visible in the field of view at all times; whether because the bed moved or due to the presence of other persons in the scene. Instead, we detect the largest non-ground horizontal surface present in the scene and estimate the gradients of the surface. This can help us obtain more information to determine the presence of a patient in the bed. The next section describes our method to detect the surface information.

ALGORITHM FOR FINDING PERSON IN BED

This section describes the techniques we employ to get more information about the surfaces in the image. As mentioned earlier, we collect depth data using the Microsoft Kinect sensors. In our application, we are not using the Microsoft skeleton tracker included with the Kinect OpenNI drivers. Under good conditions this tracking system will output a set of joints corresponding to positions of body parts such as the head, chest, and hands. However, owing to the limited performance range as well as its poor performance under occlusion, we do not use the skeletal tracking for our monitoring system. In the first step of our algorithm, we "clean" the image to get rid of noise using region filling techniques. The method we use is the image inpainting technique.

A. Image Inpainting

Most of our existing activity detection algorithms using vision sensors are centered on the analysis of foreground objects [1, 2]. This could be as simplistic as tracking the centroid positions to measure the walking speed or using shape descriptors to provide further insight into the activity state. This work proposes to use the cues provided by the scene itself to predict the activity of a person. However, in order to identify these cues, the image has to be filtered and features need to be extracted. The first step in this process is employing a region filling algorithm which removes the random noise generated by the depth sensor. This occurs when no depth values get returned from certain locations in the field of view. This happens when the infrared rays do not get reflected back to the CMOS sensor inside the Microsoft Kinect device. For the image inpainting algorithm, we use the technique describe in Telea et al. [12]. However, most region filling algorithms require some manual input to specify the affected regions which need to be identified. For our application, this is already provided since it is all the areas which do not return any depth values to the sensor. Those are the only regions modified by the inpainting algorithm. Figure 1 (a) shows the raw depth image of a hospital room obtained from the Kinect sensor and Figure 1 (b) shows the cleaner version after inpainting. Only the black regions in Figure 1(a) get modified in Figure 1 (b). We can see that some of the artifacts (noisy black regions) on the tray (left) and the wheelchair (right) disappear after the inpainting technique.

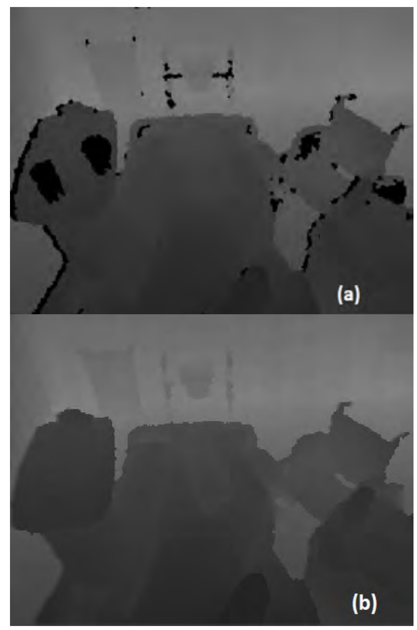

Figure 1. The raw depth image from a hospital room (a) and the inpainted result on the same image (b).

B. Dense SIFT Features

For the next step, we compute the Scale Invariant Feature Transform (SIFT) for all the pixels in the inpainted image. The SIFT feature was created by Lowe [7] and used to capture the local structural features of the image. These features are robust to noise as well as scale and can be used to compare area with similar features. These features are computed at key-points of interest such as corners detected in the image. An extension of this is the dense SIFT where the descriptors are computed for every pixel. For each pixel, the orientation is quantized into 8 bins for window size of 4×4. Then, the top three principal components are retained. The reason for three components is to be able to visualize the result by projecting it on RGB space. The result using this feature is shown in Figure 2. This is an image of a hospital room where the Kinect sensor was mounted on the television on the opposite wall compared to Figure 1. Here, the images with the same color have similar orientations. Hence, the hospital bed with tray (left) and the floor (bottom right) are colored the same color violet. These are both horizontal surfaces so they have similar gradient values. Similarly, the vertical surfaces facing the sensor are also similarly colored orange.

C. *Ground Plane Extraction*

The motion capture system for collecting the depth data automatically detects the ground plane using the RANSAC [9] plane fitting method approach described in Stone et al. [11].

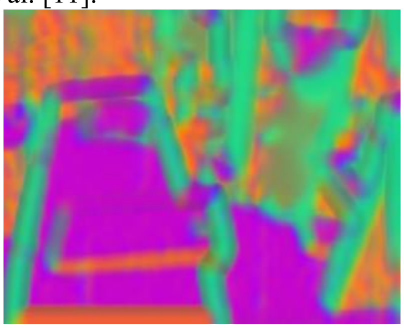

Figure 2. The gradient image obtained after implementing the dense SIFT on the depth image. Similar gradient regions are colored similarly. Hence, the hospital bed (left) has the same surface gradient as the ground (right).

Here, the assumption is that the sensor is tilted downwards so that the furthest plane from it is the ground plane. Figure 3 shows the detected floor plane highlighted in blue using this algorithm. The orange object is the detected foreground after implementing the background subtraction technique [11]. In this example, there is a pillow falling from the bed while the patient is in the bed. The falling pillow triggered the fall detection algorithm generating a false alarm. For a given video segment, we compute the aggregate ground over the entire segment by taking the union of the ground images in a time window of 30 seconds from the time stamp of the current frame. Heuristically, we found this time to provide enough information of the patient presence in bed. This gives a better description of the ground in case there are moving persons in the scene.

D. *Extracting the Horizontal Planes*

Once the dense SIFT features are extracted, we can infer the horizontal non-ground surfaces by removing the ground information from the set of horizontal surfaces detected. The result is further refined by using some image filtering steps to remove the smaller surfaces as well as fill gaps using image opening and closing operations [3]. Figure 4 (a) shows the depth image (left) and the obtained horizontal surfaces (right) of a patient lying in bed. The patient location is circled in yellow. As can be seen, there is a gap in middle of the bed (right) indicates the presence of a person. Another example is shown in Figure 4 (b) where the staff member is circled in green and the patient in yellow.

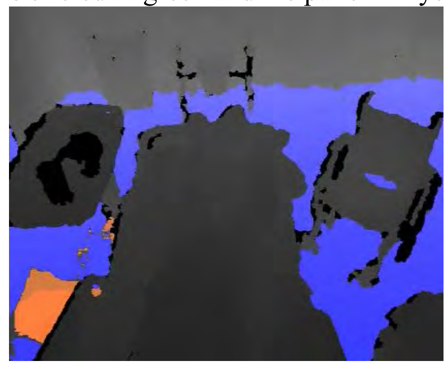

Figure 3. The ground plane (blue) and the moving object (orange) extracted from the depth image in a hospital room.

The patient is sitting on the edge of the bed so in the right image we can see that most of the bed is identified as a horizontal surface. There is another horizontal surface identified in Figure 4 (b) because of the presence of a tray in the field of view. However, for our calculation, only the surfaces at a predefined distance from each other are considered as part of the same surface.

Figure 4. The depth image and corresponding horizontal surfaces obtained after removing the ground plane. In Figure 4 (a), there is a patient lying in the bed circled in yellow and in Figure 4 (b), there is a patient and hospital staff member, circled in yellow and green respectively.

The bed region is identified as the largest horizontal surface identified in a time window of 10 minutes from the time stamp of the current image frame. If there are multiple flat surfaces within a distance of 30 pixels, the area is computed by the size of the smallest bounding box that encompasses these surfaces. If the size is below a predefined lower threshold, then the previous size is retained. If the size exceeds an upper threshold, then too the previous value is retained. Also, if there is a large difference in the location between the current centroid location of the bed and a new location of the bed, it relearns the new bed region from the current image frame. This takes care of the condition when the bed is moved around.

E. Computing the Area of the Person in Bed

Once the horizontal surfaces are computed for the images, we compute the area of the person in bed. We do this by dividing the area of the current horizontal surfaces by the area of the bed region as described above. The area of the person in the bed is then computed using equation 1. The results are further temporally smoothened with a moving average filter of window size 10.

$$\text{Area (person)} = 1 - \frac{\textit{Area of Current Surface}}{\textit{Area of Bed}} \tag{1}$$

EXPERIMENTAL RESULTS AND ANALYSIS

We evaluated our algorithm on five days of data collected from two different hospital rooms. This means that our data set comprised 96 hours of videos, though, in an effort to save memory space, when the depth frames have motion below a certain threshold, the frames are replaced by blank frames. Since the data collection occurred in real hospital rooms, the patient information and the number of patients in the rooms is unknown. For the five days of test data, there were four false alerts generated by the existing fall detection algorithm. An example of this is shown in Figure 3 where a pillow fell on the ground from the left side of the bed. During the same time period, the area of the person detected was greater than 0.6 as is shown in the Figure 5 below. The fall took place around frame 100 (x-axis). For three other false alerts, the area computed by our algorithm was above 0.5. However, for the fourth false alert, the patient was in the bathroom. The false alarm was generated when an object fell to the ground during housekeeping. However, the "fall" took place near the bed and the area computed was less than 0.3 so we could provide more contextual information that the patient was not in bed before the alert was generated. For each of videos for the five days, we used the raw depth data as the ground truth to detect the presence of the patient in bed.

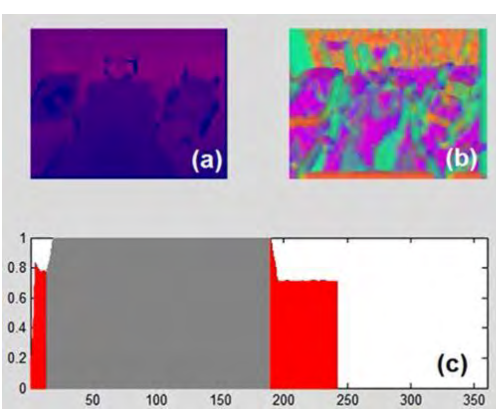

Figure 5. Result of the person detector on a false alert. Figure 5 (a) shows the depth image, (b) shows the dense SIFT image and (c) gives the area of the person during that time. The x-axis represents the depth frame number. The gray regions are when there is no depth data stored due to no movement.

CONCLUSION

A novel application to find surface information was used to complement an existing fall detection algorithm to provide more contextual information about patient activity while in bed. This technique can help eliminate false alarms caused by objects such as pillows and blankets dropping to the ground. Future work involves incorporating temporal information to get more information about patient movement by looking at the changes in the gradient values from the bed region. This can allow hospital staff to remotely monitor patient activity in a nonintrusive manner without the need to physically check each room. Ultimately, this technique will allow us to analyze falls and the events leading up to a fall to enable fall prevention strategies and improve patient safety in hospitals.

REFERENCES

[1] T. Banerjee, J. M. Keller & M. Skubic, "Resident Identification Using Kinect Depth Image Data and Fuzzy Clustering Techniques", *Proceedings, 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego CA, 5102-5105, 2012.

[2] T. Banerjee, M. Rantz, M. Li, M. Popescu, E. Stone & M. Skubic, "Monitoring Hospital Rooms for Safety Using Depth Images," Proceedings, *AAAI Fall Symposium Series - AI for Gerontechnology*, 2012.

[3] G. Gonzalez, R. Woods, "Digital Image Processing" (3rd Edition) Prentice Hall, 2007.

[4] S. K. Inouye, C. J. Brown, and M. E. Tinetti, "Medicare nonpayment, hospital falls, and unintended consequences." *New England Journal of Medicine*, 360 2390-2393, 2009.

[5] P. Kittipanya-Ngam, S. Ong, H. Eng, "Bed detection for monitoring system in hospital wards." *IEEE International Conference on Engineering in Medicine and Biology Society* (EMBC), Aug. 28 -Sept. 1, 5887-5890, 2012.

[6] C. Lea, J. Fackler, J. Hager, and R. Taylor, "Towards Automated Activity Recognition in an Intensive Care Unit." *MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions* (M2CAI), 2012.

[7] D.G. Lowe. "Object recognition from local scale-invariant features." *Proc. ICCV*, Kerkyra, Greece, 1150–1157, 1999.

[8] B. Ni, N. C. Dat, and P. Moulin, "RGBD-camera based get-up event detection for hospital fall prevention." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Kyoto, Japan, 1405-1408, March 2012.

[9] B. Oehler B., Stuckler, J., Welle, J., Schulz, D. and Behnke, S. "Efficient multi-resolution plane segmentation of 3D point clouds", *Proc. Of the International Conference on Intelligent Robotics and Applications* (ICIRA), Aachen, Germany, 145–156, 2011.

[10] D. Oliver, J. B. Connelly, C. R. Victor, "Strategies to prevent falls and fractures in hospitals and care homes and effect of cognitive impairment: systematic review and meta-analyses." *BMJ* 334, 82-82, 2007.

[11] E. Stone, and M. Skubic, "Fall Detection in Homes of Older Adults using the Microsoft Kinect", in press, *IEEE Journal of Biomedical and Health Informatics*, 2014.

[12] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method", *J. Graphics Tools*, 9, 23-34, 2004.