Privacy Preserving Dynamic Room Layout Mapping

Xinyu Li[1], Yanyi Zhang [1], Ivan Marsic [1], and Randall S. Burd [2]

1 Department of Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ, USA

2 Division of Trauma and Burns, Children's National Medical Center, Washington, D.C., USA

Abstract:

We present a novel and efficient room layout mapping strategy that does not reveal people's identity. The system uses only a Kinect depth sensor instead of RGB cameras or a high-resolution depth sensor. The users' facial details will neither be captured nor recognized by the system. The system recognizes and localizes 3D objects in an indoor environment, that includes the furniture and equipment, and generates a 2D map of room layout. We evaluated this system in two challenging real-world application scenarios: a laboratory room with four people present and a trauma room with up to 10 people during actual trauma resuscitations. The system achieved 80% object recognition accuracy with 9.25 cm average layout mapping error for the laboratory furniture scenario and 82% object recognition accuracy for the trauma resuscitation scenario during six actual trauma cases.

Major Contribution:

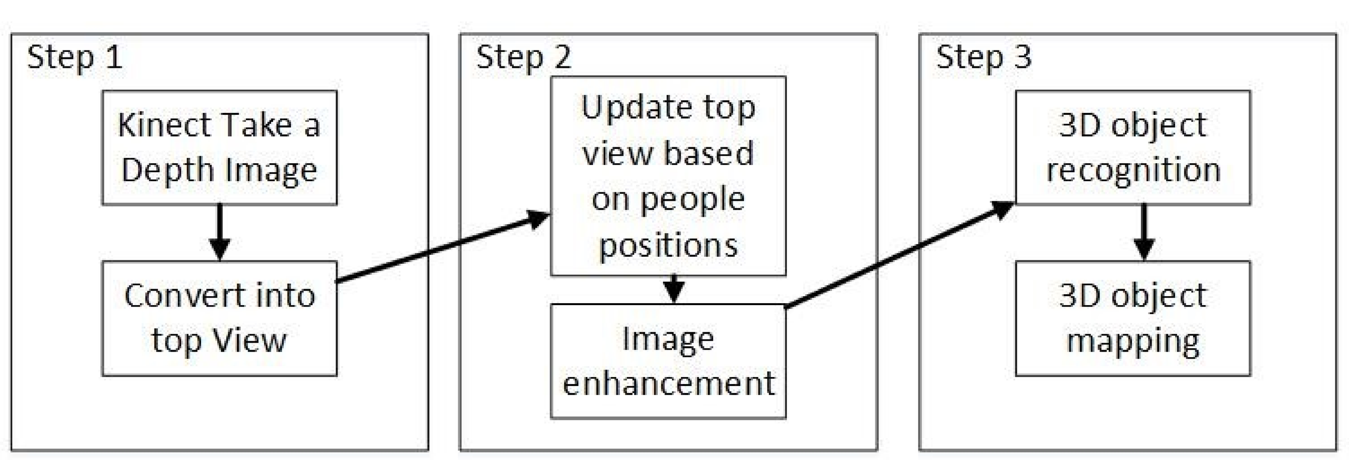
- A novel dynamic, privacy-preserving room layout mapping strategy using only a commercial depth sensor.
- A strategy to restore the missing information in Kinect depth-maps caused by random noise or view occlusion by moving people.
- The implementation and evaluation of the system in two challenging real-world applications.

System Structure:

Step1: Generate a point-cloud map of the 3D environment based on the depth sensor and converts it into top-view image.

Step2: Process the top-view image by first restoring the part of view occluded by people in the room and then enhancing the top-view image to eliminate the undefined pixels caused by Kinect sensor's random noise.

Step3: Recognize 3D objects (equipment, furniture, etc.) based on shape and height using 2D template matching and then layout mapping based on recognition results.



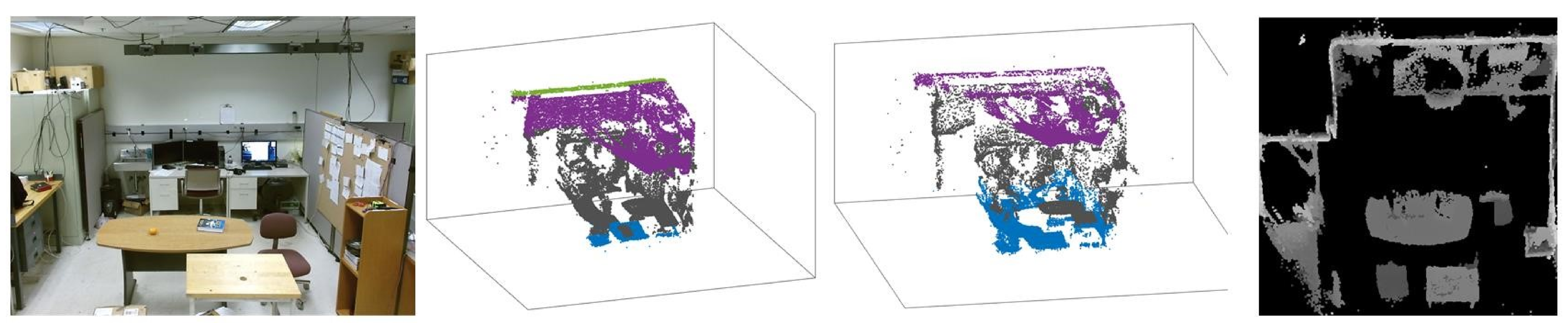
Methodology:

Step 1: From Point-Cloud to Top View

To have a clear view of the room, we mounted the Kinect H meters above the ground with a tilt angle α so that people and objects in the room are more likely to be seen in camera view (in our application, H = 2.5 m and α = 7'). Before converting the 3D point-cloud into a 2D top view, each point in the point-cloud needs to be adjusted for the tilt angle α so that the camera space of the Kinect is aligned with the actual setting using rotation matrix:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} 0 \\ H \\ 0 \end{bmatrix}$$

Where (X, Y, Z) denotes each point in the point-cloud in camera space and (X', Y', Z') denotes the corresponding point in the room, given the tilt angle α and height of Kinect sensor H.



(a) (b) (c) (d)

(a) The picture of the laboratory. (b) The point cloud of the room before tilt angle adjustment. (c) The point cloud of the room after tilt angle adjustment. (d) The top view image of the room.

Step 2: Top-View Image Enhancement

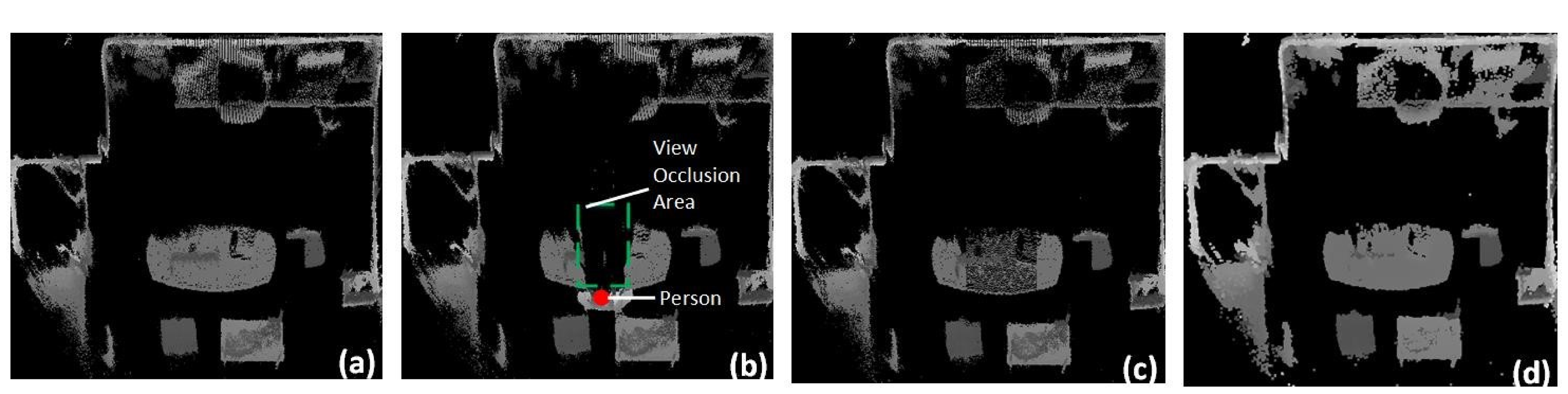
View occlusion leads to information loss and makes it difficult to recognize 3D objects in top-view images. The system restore the missing information in top-view images with two methods:

○ **View Occlusion Compensation:**

- Start system when room is empty. (fig. a).
- Capture a depth image and convert into top-view image every 100ms. (fig. b).
- Update the top-view image if the room remains empty of people.
- If a person is detected by the Kinect, updates the pixels outside the occluded area based on the top-view image and keeps the pixels in the occluded area unchanged from the previous top-view image. (fig. c).

○ **Random Noise Compensation:**

- Find undefined pixels in current top view image, and assign the value of that pixels using the average value of same pixel in n previous top View images.
- Image dilation. (fig. d)



(a) Top view image with no people. (b) View occlusion caused by people in the room. (c) Compensating for the information loss. (d) Top view image after image enhancement.

Step 3: Object Recognition and Room Layout Mapping

We generated templates for objects by manually selecting examples of each object from 10 top-view images and averaging them for the template. The system will take the area that has maximum matching score with certain template as the location of object.

Experimental Results:

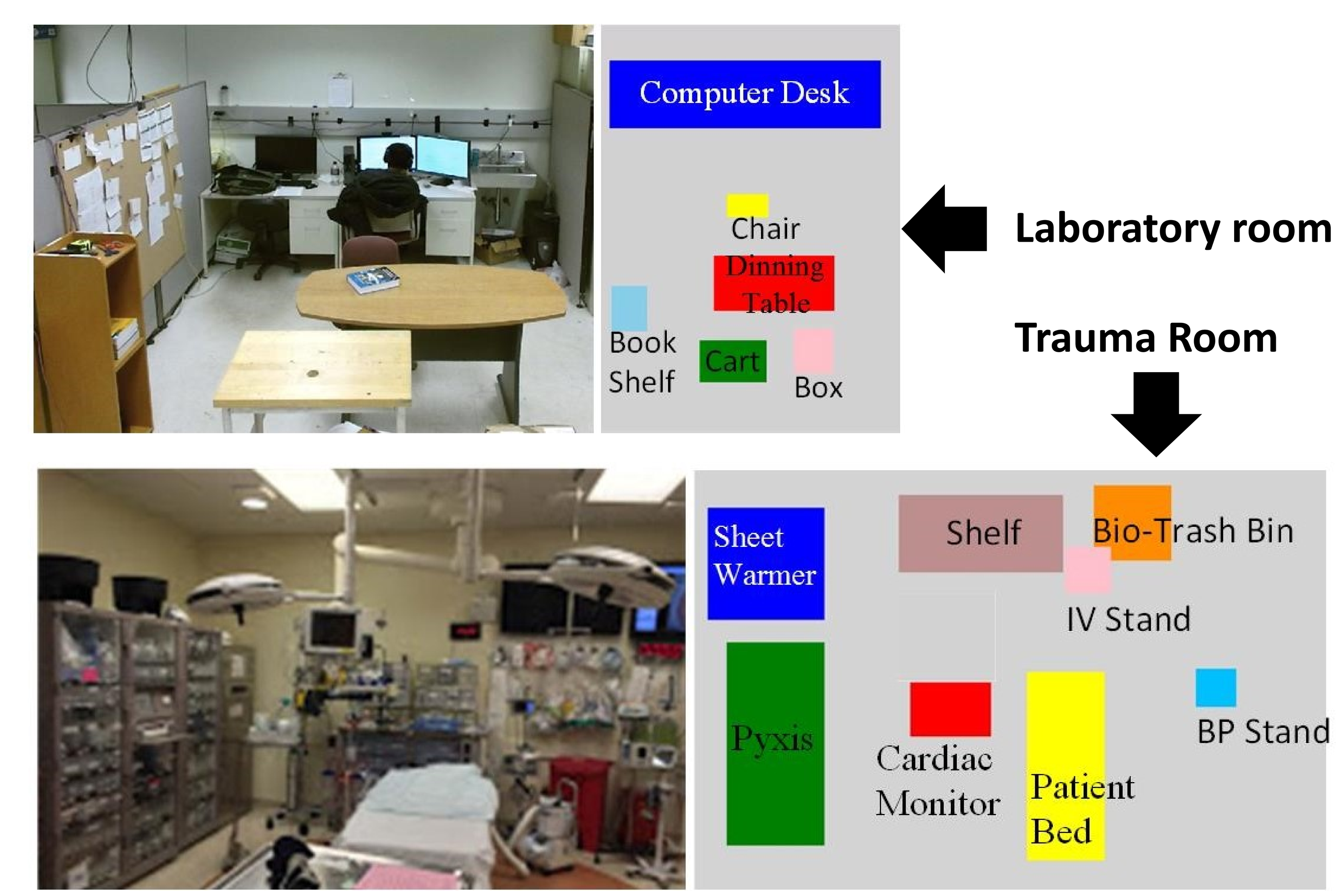
Table 1. Objects used for experiments in laboratory room and trauma room with their dimensions and recognition accuracy. The Gray shaded objects were relocated during the experiments.

Objects(lab)	Size (inch)	Accuracy (0/4 people)	Objects (Trauma room)	Size (inch)	Accuracy (up to 10 people)
Dinning Table	58×28×30	100%/100%	Patient Bed	31×32×80	92.3%
Cart	28×18×36	100%/87.5%	Shelf	57×42×20	100%
Chair	H=31", irregular shape	100%/42.8%	IV Stand	H=68", irregular shape	67.8%
Box	15×13×19	100%/67.1%	BP Stand	H=43", irregular shape	46.1%
Book Shelf	12×14×16	98.0%/89.1%	Cardiac Monitor Adapter	30×28×17	62.5%
Desk	91×30×29	99.1%/90.2%	Bio-Trash Bin	19×32×16	93.2%
			Pyxis	82×27×78	100%
			Sheet Warmer	30×29×72	99.8%

Objects	Dinning table	Cart	Chair	Box	Book Shelf	Desk
Error in X-axis (cm)	13	9.7	20.0	13.0	11.6	11.6
Error in Z-axis (cm)	2.7	13.2	5.5	1.6	2.7	6.4

Table 2. Room layout mapping error in X-axis and Z-axis in centimeter.

Room Layout Mapping (Screenshot)



Acknowledgement

This research was supported by the National Library Of Medicine of the National Institutes of Health under Award Number R01LM011834