Application of Microsoft Kinect sensor for tracking construction workers

I.P. Tharindu WEERASINGHE¹, Janaka Y. RUWANPURA², Jeffrey E. BOYD³ and Ayman F. HABIB⁴

¹ Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada, T2N1N4; PH (403) 667-8313; email: ipweeras@ucalgary.ca
² Department of Civil Engineering, University of Calgary, Calgary, Alberta, Canada, T2N1N4; PH (403) 870-7503; email: janaka@ucalgary.ca
³ Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, T2N1N4; PH (403) 220-6038; email: boyd@cpsc.ucalgary.ca
⁴ Department of Geomatics Engineering, University of Calgary, Calgary, Alberta, Canada, T2N1N4; PH (403) 220-7105; email: ahabib@ucalgary.ca

ABSTRACT

The image processing based human recognition is yet a challenging task because of series of complications such as variations in pose, lighting conditions and complexity of background in the tracking environment. This study introduces a novel methodology to track construction workers using image processing techniques and depth information generated from the Microsoft Kinect sensor. Kinect is a new game controller technology introduced by Microsoft in November 2010. This automated real-time worker tracking system provides an opportunity to track the construction worker location and their movements in a specified indoor work area. The research study proposes a properly color coded “construction hardhat” as a key tracking object which can be used to differentiate site personnel (worker, supervisor, engineer, etc.).

The proposed method detects construction workers in three major stages which includes human recognition, hardhat recognition and 3D localization. The human recognition is done by analysing human body parts. 3D positions of body joints are accurately predicted from a single depth image. The construction hardhat detection is based on characteristics of the hardhat such as unique shape and color. Template based template matching is used as the pattern recognition technique.

Keywords: Microsoft Kinect sensor, worker tracking, worker performance, Image processing, construction productivity, project management.

INTRODUCTION

Kinect has been developed for Microsoft Xbox 360 game console and includes cameras that deliver depth information, color data. However, independent developers offer solutions for using Kinect separate from the game console and for the most common operating systems. CLNUI, OpenNI or Microsoft windows SDK enable applications to access and manipulate this data with all libraries needed for data processing. In this study, we used the OpenNI Platform to design the worker
tracking system. Figure 1 shows main components of the Kinect sensor which has been used as the main device in this research. A standard CMOS image sensor receives the projected structured infrared (IR) light pattern and processes the IR image and produces an accurate per-frame depth image of the scene (PrimeSense, 2011).

Figure 1: Main components of the Kinect sensor

Many construction projects are not systematically monitored due to difficulty of gathering reliable information to assess the worker tool time of ongoing construction work. The current method of productivity assessment requires manual data collection which is carried out by employing observers to collect worker tool time and performance information from construction sites. The quality of such manually collected data is significantly low because of human errors (biased working environment, non-standardized recognitions) and limitations in data availability. Implementing an automated data acquisition system in the construction field is recognized as the most suitable approach of extracting unbiased worker tool time and performance information. Several researchers have attempted to solve this problem by automating the monitoring of construction workers in various techniques such as systems using radio frequency identity (RFID) tags and receivers, image processing based systems, etc. However, the industry is still lack of a comprehensive solution which can facilitate productivity monitoring of a building construction project in its entirety. The major challenge has been the diversity of activities in building construction and the complexity of the construction process. This research attempts to address current data acquisition issues by developing an automated real-time system to track construction workers to assess worker tool time and performance. The worker tracking system provides an opportunity to differentiate site personnel (i.e. worker, supervisor, etc.), track construction worker location and their movements in a specified work area, which is necessary to measure the worker tool time and the efficiency. Tool time is defined as the time workers spend in producing a tangible output.
BACKGROUND

In the construction field, progress monitoring is an essential part of a project, which assists project managers in formulating strategies and making decisions in resource allocation in order to keep the project on track. There have been several techniques used to achieve the task of construction progress monitoring. Image processing based systems, 3D laser scanning methods, radio-frequency identification (RFID) tags, bar codes and embedded sensor systems are the leading technologies. The necessity of adding new tasks that need to be performed before, during or after the utilization of such technologies at a construction site is the main drawback in the application of laser scanners, RFID tags and embedded sensors (El-Omar & Moselhi, 2008; Kiziltas et al. 2008). RFID reduces performance when proximity of metals and tag size increases with increasing transmitting power. RFID sensors and bar codes need additional infrastructure to detect items and are time-consuming to set up. They are often costly and cannot be attached to many types of components. Laser scanners are also very costly, require operation expertise and may generate erroneous results within dynamic scenes. Peddi et al. (2009) developed a human pose recognition system to measure worker performance. However, this can be used only for selected construction activities which need to have unique human working poses to recognize the performance. The infrastructure cost, regular operational cost and range limitation are the common drawbacks of these methods.

SYSTEM INFORMATION

The tracking system is developed to detect the construction worker locations and their movements within a given work area. The software used to develop all image processing algorithms is MATLAB R2010a. All real-time RGB images and depth information are obtained from the kinect sensor. The worker tracking algorithm is based on the skeletonise figures, and characteristics of the hardhat (i.e. unique shape and colour of the hardhat). Furthermore, feature based template matching is used as the pattern recognition technique to detect hardhat shapes of the image. To reduce the complexity of the research study, two basic assumptions have been made about the site-end condition as follows;

1. All the site personnel use similar shaped hardhats.
2. All the site personnel follow correct color coded hardhats according to their job title. [Ex: yellow hardhats for labours, red hardhats for supervisors, etc]

Figure 2 shows the general equipment arrangement of the proposed methodology.
Technical overview of the kinect sensor

The range camera technology of the kinect device is developed by the PrimeSense. The PrimeSensor Reference Design performs more accurate sensory information by image registration and resultant is pixel-aligned images, which means that every pixel in the color image is aligned to a pixel in the depth image (PrimeSense, 2011). Technical specification of the kinect sensor is given in the Table 1 (PrimeSense, 2011).

Table 1: Technical specification of the kinect sensor

<table>
<thead>
<tr>
<th>Property</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field of View (Horizontal, Vertical, Diagonal)</td>
<td>58° H, 45° V, 70° D</td>
</tr>
<tr>
<td>Depth image size</td>
<td>VGA (640x480)</td>
</tr>
<tr>
<td>Spatial x/y resolution (@ 2m distance from sensor)</td>
<td>3mm</td>
</tr>
<tr>
<td>Depth z resolution (@ 2m distance from sensor)</td>
<td>1cm</td>
</tr>
<tr>
<td>Maximum image throughput (frame rate)</td>
<td>60fps</td>
</tr>
<tr>
<td>Operation range</td>
<td>0.8m - 3.5m</td>
</tr>
<tr>
<td>Audio: built-in microphones</td>
<td>Two microphones</td>
</tr>
<tr>
<td>Power consumption</td>
<td>2.25W</td>
</tr>
<tr>
<td>Operation environment (every lighting condition)</td>
<td>Indoor</td>
</tr>
</tbody>
</table>
TRACKING PROCEDURE

According to the assumptions mentioned above, all workers are expected to wear hardhats when they are in the site. Hence, hardhat can be used as the key tracking object to represent the worker in the field. In brief, the system recognizes human figures in the video by detecting the hardhat shape. Then the tracked human is differentiated based on their job title (ex. labour, supervisor, etc.) based on the color of the hardhat. Following diagram (Figure 3) illustrates the main tracking stages of this proposed study.

The proposed tracking system has four modules that work in sequence to increase robustness. The four modules are:
- Human recognition and skeletonising – to identify human figures and track skeleton image of people moving within the kinect field of view
- Colour filtration - to differentiate the predefined hardhat color patches
- Shape recognition – to recognize the reference model shapes (features of the hardhat) in filtered colour patches
- Object localization – to convert kinect local 3D coordinates to building coordinate system.

Skeleton image tracking

The system works on OpenNI platform to access and manipulate data with all libraries needed for data processing. We used matlab executable (mex) files which have been developed by Dirk-Jan Kroon (Kroon, 2011) and it is freely available in matlab central web site. This program is used to extract depth map and RGB image, and to track skeletons of human figures. All human figures need to be calibrated by standing on a specific pose at the first time to recognise the skeleton of each human figure. In the following illustration (Figure 4) shows that the depth map is aligned on the RGB image, and color range interprets the depth value of each pixel. As an example, areas in red color indicate objects that are farther from the camera and areas in blue color indicate objects close to the kinect device. In the kinect device, the range camera technology is developed by using infrared projector system. Further, human figure recognition is based on the depth information of the image and taking into consideration the proportions of body parts. Therefore, this human figure tracking can be used in low light conditions as well.
The skeleton of each figure consists with 15 number of human body joints. And the program generates horizontal and vertical coordinates of each body joint of each recognized human figure. This information can be used to determine the regions of interest (ROI) for the hardhat recognition process.

**Colour Filtration**

The primary objective of this human tracking element is to detect pre-identified four colour-coded hardhat shapes embedded in the image frame. In order to achieve the highest performance, at a relatively low computational cost, the image color feature extraction is identified. Therefore, colour based segmentation (CIELAB (CIE L*a*b*) colour space) is proposed to filter the image. The L*a*b* space consists of a luminosity 'L*' or brightness layer (L=0 black, 100 white), chromaticity layer 'a*' indicating where color falls along the red-green axis, and chromaticity layer 'b*' indicating where the color falls along the blue-yellow axis. The ROI for this color segmentation is selected by analysing the skeleton image. We used headzone: an effective radius centered from the coordinate of the head node of the skeleton in the same depth range as the ROI. The ROI selection process is illustrated in Figure 5.

**Pattern recognition**

Template matching is one of the popular techniques for finding objects of an image. Objects are directly compared with stored sample images or prototypes while taking into account all allowable poses (translation and rotation) and scale changes
Template matching can be classified into two categories: feature-based and template-based matching. In this study, template-based template matching is considered as the pattern recognition method since the template does not contain strong features and the whole image constitutes the matching image. Generally, two common matching algorithms are often used to measure the similarity between two images: minimum-distance measurement and the correlation measurement. Although the minimum-distance measurement technique calculates the distance rapidly, it is easily disturbed by noise (Chang et al. 2009). Hence, correlation measurement is used. In intensity-based correlation, the algorithm uses a certain similarity measurement to compare the gray value of the corresponding pixel points in the template and the image. The max-correlation measurement can match the target exactly even if there is noise in the image. In order to recognize hardhat shapes in the image, a comprehensive database of hardhat images and their characteristic features are developed (Weerasinghe and Ruwanpura 2010). The characteristic features will include both low-level image and shape features. Each image in the database contains a hardhat that is typically used in the construction site. The reference images are captured based on a spherical grid. The angle between two grid points at the centre is kept to 20 degrees and all the reference images are captured from a distance of 45 cm away from the camera. Figure 6 shows six raw images and the grid of the viewing sphere.

![Figure 6: Viewing sphere grid and raw images](image1.png)

Then these raw images are normalized into a common pose and scaled down to a standard size. In this stage, all raw images are converted into grayscale format and transformed into a standard image format which has the same image size and same orientation. The simplest and most widely accepted orientation is based on the principal axes of the object, which consists of its orientation and position with respect to an orthogonal frame or coordinate system. The image normalizing procedure is displayed in Figure 7.

![Figure 7: Image normalizing procedure](image2.png)
**Image similarity factor determination**

In this worker hardhat detection process, the system recognizes the best similar hardhat shapes embedded in the current image frame. The filtered color blob at each head node area is converted into gray scale format and normalized into a common pose and fit into a standard size same as the reference image preparation section. Standardization of images reduces the computational cost and the complexity of the similarity measurement process. To determine the maximum similarity between two images, the correlation function is applied as shown in below.

\[
\text{Corr} = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{i,j} \times y_{i,j})
\]

Where \(x_{i,j}\) and \(y_{i,j}\) are intensities of reference image and the filtered color blob (gray-scale format) with \((m, n)\) dimensions. And \(i, j\) represent row and column vectors. If correlation value exceeds the threshold level, the color blob is considered as a hardhat.

**Object localization**

The matlab executable files (Kroon, 2011) used to extract kinect sensory information are able to generate local 3D coordinates of a target with respect to the kinect device. Therefore, for a better utilization, these local coordinates need to be transformed into building coordinate system. First, the camera calibration procedure will be followed to determine interior and exterior parameters of the kinect camera (Brown, 1971). These parameters will be used to apply 3D transformation from the camera coordinate system to the building coordinate system. A set of ground control points onsite is used to calibrate the location and the rotation angles with respect to the building coordinate system. The mathematical model for the 3D transformation is illustrated below (Zeng, 2010);

\[
\vec{X}_G = \vec{X}_T + S R (\omega \varphi \theta) \vec{X}_i
\]

Where \(\vec{X}_G\) is ground coordinate vector, \(\vec{X}_T\) is translational (shift) vector, \(S\) scale factor, \(R (\omega \varphi \theta)\) is the rotational relationship between the image and the ground coordinate systems and \(\vec{X}_i\) is image coordinate vector. Once the above equation is applied, ground coordinates of the kinect device can be determined.

**RESULTS**

Figure 8 shows the results of the worker tracking system when multiple workers in different disciplines are on-site. As an example it recognizes people on-site and differentiates them according to their coloured hardhats. Then worker identity number and category are displayed in the middle of the person. In addition, transformed
building coordinates of tracked workers are displayed in a separate table which also includes block/area number currently they are working on.

<table>
<thead>
<tr>
<th>Category</th>
<th>X(mm)</th>
<th>Y(mm)</th>
<th>Z(mm)</th>
<th>Location</th>
<th>Duration(min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreman</td>
<td>400</td>
<td>5</td>
<td>3.127</td>
<td>Block A</td>
<td>12</td>
</tr>
<tr>
<td>Mover</td>
<td>-212</td>
<td>320</td>
<td>3.950</td>
<td>Block B</td>
<td>19</td>
</tr>
<tr>
<td>Inspector</td>
<td>+1000</td>
<td>1050</td>
<td>2.600</td>
<td>Block C</td>
<td>23</td>
</tr>
</tbody>
</table>

**Figure 8: Multiple workers in different disciplines**

**CONCLUSION**

This comprehensive research study outlines a conceptual framework for development of a real-time and fully automated system to measure worker movement patterns using image processing techniques. A Kinect device is used to detect humans in a construction site. A novel approach presented in this paper adds another dimension to the construction industry by providing an opportunity for more precise inputs to the tool time assessments and to automate the worker productivity measurements. The main drawback of this system is the distance limitation of the Kinect sensor. According to the camera specification, the maximum depth information can be generated only within the limit of 4m from the camera. Therefore, this project will focus only on close range of objects in indoor working environment. The main highlighted benefit of this proposed system is worker recognition and 3D positioning determined from a single viewpoint which reduces equipment cost while improving the practicability of implementation on-site. In addition, this system will be a better solution for image processing based human tracking under low lighting conditions, since the depth information is determined from an infrared structured light array system. Further, the system can effectively recognize human body parts including head and hands of a person in most of the poses. Therefore, worker recognition is achieved by analysing the proximity of hardhat and the head location. This information is used to eliminate false indications for nearly hardhat shapes and removed hardhats and elevates the robustness of the tracking process. Furthermore, real world 3D coordinates of tracked workers are determined using depth map generated from Kinect and the camera calibrated results including 3D location and rotation angles of the camera. In addition, information of coordinates of these body parts can be effectively utilized to analyze human poses related to construction...
activities in a future research. This study may lead to determine worker performance and assists project managers and planners as a planning tool in developing strategies for improving labour productivity and labour allocation.

ACKNOWLEDGEMENT

The authors wish to acknowledge the support and funding for this research project provided by Canadian Foundation for Innovation and CANA, PCL, Ellis Don, Graham, Ledcor, Revay & Associates, Stuart Olson, Canadian Construction Research Board, Calgary Construction Association, and the Natural Sciences Engineering Research Council under their Collaborative Research and Development Grant CRDPJ 341047 – 06.

REFERENCE


