

## Human Motion Analysis Using 3D Range Imaging Technology

Ray Gonsalves<sup>1</sup> and Dr. Jochen Teizer<sup>2</sup>

<sup>1</sup>RAPIDS Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Phone: +1-404-894-8269, Fax: +1-404-894-2278, e-mail: teizer@gatech.edu

### **Abstract**

Human motion analysis and tracking are a significant research area in the domain of computer vision. Existing systems of today focus on detection of human targets by analyzing their movements in order to recognize the different activities performed by them. Our research work mainly focuses on using detection and tracking of human targets using a 3D range image camera [1] for surveillance purposes in order to ensure the safety of construction workers and also to monitor their posture and movements for health related purposes in an active work zone. For this purpose the tracking algorithm proposed performs the segmentation of a human target (i.e. construction worker) from a range image video sequence and then models and tags them in order that their location can be continuously monitored. Unlike most other available systems, our system focuses on using the range or distance information since they indicate how far (in terms of meters) a human target is located away from the camera and more importantly because they are capable of generating a 3D perspective of the human target (i.e. by method of 3D point clouds). The range video sequence is obtained by using a special range image camera, which is an optical imaging system which offers real time 3D image data. Furthermore, the segmented human target is modeled by image skeletonization using a star skeleton structure [7]. This model in future research can be used in conjunction with HMM's (Hidden Markov Models) for human activity recognition. The system designed calculates the angles between different body parts to analyze the posture of a construction worker. It also incorporates the use of a particle filter [2] to trace the path of the construction worker in order to classify different work-related activities. Our system is also capable of detecting multiple people and tracking each of their paths separately in a given work environment.

**Keywords:** Range Image Processing, Real-Time, Safety, Detection, Tracking, Surveillance, Star-Skeleton Model, Construction Work Zone.

### **1. Introduction**

In the field of civil engineering, safety is becoming a growing concern on account of the numerous accidents occurring in construction work-sites. As per the Census of Fatal Occupational Injuries released by the U.S Department of Labor [13], out of the total of 5,448 fatal injuries which occurred in the year of 2007, 1,152 accidents occurred in the Construction industry while 1,423 occurred in the Transportation industry. This data clearly signifies the importance of increasing safety and detecting and avoiding fatal accidents before they occur in work-sites. In addition to safety, the health of construction workers is also an important issue, owing to the physical stress they are subject to while handling heavy equipment. Head injuries, spinal injuries, slip-discs, dislocations and fractures are some of the most common injuries which occur in construction sites. Most of these injuries can often result chronic health problems or in some cases even death if overlooked at construction sites.

With the help of computer vision technology we propose to device a real-time monitoring system to assist in providing safety of construction workers and monitoring their posture while working in order to determine whether they are subject to physical health related issues due to an incorrect posture. Managing safety in construction work-sites often require an in-depth analysis of real time visual data. Although, this can be done with the help of a human operator, the chances of error and wrong analysis increases manifold as the number of workers and the area of the site increases. With the advent of real-time intelligent systems, the complexity of analyzing safety in construction sites can greatly be reduced. It is important to note that the development of real-time monitoring systems is not to replace human operators (i.e. safety officers) but

instead to assist them to manage the safety of workers more efficiently and quickly with a lower rate of error especially in large construction sites with many workers performing various tasks simultaneously.

The system proposed is capable of detecting, tracking and monitoring construction workers using 3D Range Imaging technology [1]. The proposed system performs a series of steps which include background modeling, background subtraction, automatic threshold selection, noise removal [denoising], contour formation, centroid detection and tracking a human target using a particle filter [2]. The system then proceeds to form a star skeleton model which can be used to analyze the motion and posture of construction workers in a work zone site. The angles between the different segments of the star skeleton model are calculated for the purpose of motion analysis in order to detect the activity being performed by the construction worker.

## 2. Background Review and Range Imaging Technology

The primary purposes of tracking algorithms are to automate the process of segmenting and tagging objects from real-time video sequences. The objects which are to be tracked span a wide range including humans, vehicles, etc. Along with segmenting and tagging objects, many of the new generation tracking algorithms are also capable of analyzing the path traced by the detected objects. Our focus is mainly on developing and using tracking algorithms in civil engineering for monitoring the location of construction workers and interpreting the different activities performed them in active work zones for safety and health related issues.

Fujiyoshi and Lipton [7] developed a star-skeleton model in order to represent the human body structure. The star-skeleton model technique involves extracting the contour of the segmented object and then connecting the centroid of target object to extremities of the contour. Using this model they analyzed two motion cues namely cyclic motion and posture of the human star-skeleton model. Haritaoglu and Harwood [9] presented a method which used both shape analysis and tracking to locate humans and their body parts (head, hands, feet, and torso) and also created corresponding models in order so that they could be tracked even in presence of occlusions. The shape model used for their algorithm was the Cardboard Model [9], which represents the relative positions and size of body parts. However it is important to note that the cardboard model can be used only for upright people [straight posture]. Pfunder [3] is a real-time system for tracking a person which uses a multi-class statistical model of color and shape to segment a person from a background scene. It finds and tracks people's head and hands under a wide range of viewing condition. Yamota and Ohya [5] presented a human action recognition method based on a hidden Markov model (HMM). It is a feature-based bottom-up approach that is characterized by its learning capability and time-scale invariability. Yu and Aggarwal [8, 11] also incorporated the use of HMMs in their algorithm for detection of fence climbing from monocular video. To analyze the resulting time series, they built a block based discrete Hidden Markov Model (HMM) with predefined action classes {walk, climb up, cross over, drop down} as the state blocks. The detection was achieved by decoding the state sequence of the block based HMM. Many tracking algorithms try to solve problem in video analysis to track moving objects during a video sequence in presence of occlusions. Conte and Foggia [4] used a method of tracking objects through occlusions that exploits the wealth of information due to the spatial coherence between pixels, using a graph-based, multi-resolution representation of the moving regions. Bregler [12] uses many levels of representation based on mixture models, EM, and recursive Kalman and Markov estimation to learn and recognize human dynamics.

Range Imaging technology brings together a combination of obtaining the amplitude, intensity and the range (i.e. distance) information at every pixel of a two-dimensional sensor array of a range camera. Range Image cameras make use of ASP (Active Sensor Pixels) data of entire scene field-of-view in one frame. Range Image cameras operate on the TOF (Time of Flight) principle using phase shift measurement [1]. The camera uses conventional light emitting diodes (LED's) to actively illuminate a scene by emitting sinusoidal modulated (spatial or temporal) near-infrared light. The time the light needs to get to and to return from an impinged object in a scene back to the sensor is then measured using a practical synchronous sinusoidal demodulation. Focused through a lens and within the 3D range camera, a CMOS/CCD sensor chip is positioned to receive the incoming wave front. It then calculates the amplitude, intensity, and range values based on the TOF principle. Both the detection and the complete demodulation are performed in the

charge-domain using charge-coupled devices (CCD). This ensures almost noise free demodulation of the incoming light signal.

### 3. Methodology and Algorithm Design

The overall system architecture consists of seven modules namely Background Subtraction, Threshold Selection, De-Noising & Contour Formation, Centroid Locating, Tracking and Star-Skeleton Modeling [7] and Motion Analysis. The algorithm goes through five different phases which include Pre-processing, Denoising, Plotting, Tracking and Feature Recognition. The system initiates the algorithm by recording the range video sequence using a 3D range image camera. The raw data obtained by the camera is then processed to extract only the range data, while the intensity and amplitude information are discarded. Since the system works on the principle of static background modeling, initially the static background (200 frames at a rate of 25 frames/sec) of the indoor environment for the experiment is recorded. The system architecture along with a description of each of its modules involved in the algorithm is detailed below:

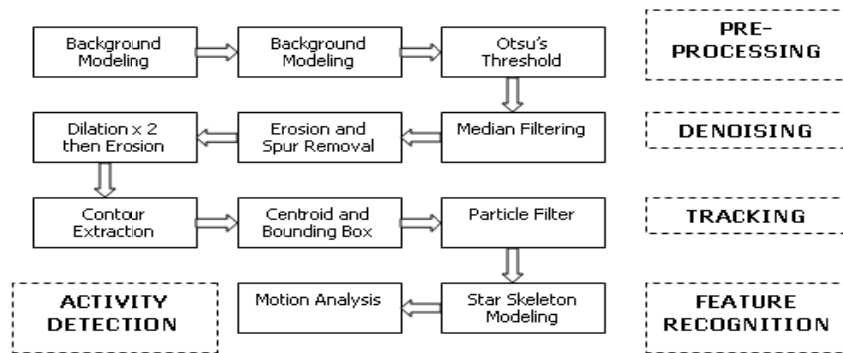


Figure 1: System Architecture

#### 3.1 Background Subtraction

In order to extract a human target from the range video sequence the algorithm uses the method of average background subtraction. For this purpose it initially records the range video sequence of only the background without any moving human target. It then models the background as a single frame, by calculating the time average of the background (i.e. the average value of the background over all the frames). This single modeled background is then treated as the reference background frame. The range video sequences for different experiments are then recorded. Segmentation of the human target is done by subtracting the modeled background from each frame in the range video sequence, and its absolute value is then taken. Thus the system then has an array containing the absolute difference between each frame and the background of the test video sequence.

#### 3.2 Threshold Selection

After obtaining the background subtracted image of the human target the algorithm then proceeds to threshold selection. In order to separate the foreground pixels from the background pixel (i.e. segmentation), it needs to threshold the absolute difference array. For this purpose, it uses *Nobuyuki Otsu's* method [6] of threshold selection from gray-level histograms. This method adaptively calculates the threshold based on the histogram of the difference array. The procedure utilizes only the zeroth and the first order cumulative moments of the gray level histogram to calculate the threshold. Note that the algorithm was used to calculate the threshold for each difference array of the test video sequence. Optimal sets of threshold were selected automatically and stably not based on the differentiation (i.e. a local property such as valley), but on the integration (i.e. a global property) of the histogram.

#### 3.3 Denoising and Contour Formation

Thresholding alone, however, is not sufficient to obtain clear foreground regions; it results in a significant level of noise, for example, due to illumination changes. For proper segmentation it is important to eliminate this noise. For this purpose a series of post-processing noise removal techniques are applied on the

thresholded array. Initially the thresholded array was subjected to median filtering. This was done in order to eliminate the salt-and-pepper noise present in the array. The array was then converted to a binary array (white representing the foreground pixels and black representing the background pixels). Erosion was then applied to the binary array. Erosion is one of the two basic operators in the area of mathematical morphology, the other being dilation. It is typically applied to binary arrays, but there are versions that work on grayscale arrays. The basic effect of the operator on a binary image is to erode away the boundaries of regions of foreground pixels (i.e. white pixels, typically). Then a fast binary connected component operator is applied to find the foreground regions and smaller regions are eliminated [9]. The largest blob (i.e. component) is then selected thus achieving 100% elimination of the noise pixels. After this process, all the small holes in the foreground region are filled. The position of the centroid and the dimensions of the bounding box are then calculated for the foreground regions (i.e. the segmented blob/human target). Boundary detection was applied on the segmented foreground pixels using *Canny-Edge* detection. This gives the resulting contour or boundary of the segmented blob/human target.

### 3.4 Centroid Location

Once the contour is formed, in order to achieve real time tracking of the blob/ human target in the range video sequence, the first step is to locate the centroid of the blob/ human target. The centroid is used to represent the human target moving in the range video sequence. The centroid calculated specifies the center of mass of the blob/human target. The centroid of the human target extracted is calculated by using the coordinates of the pixels on its boundary. Note that the first element of centroid is the horizontal coordinate (or x-coordinate) of the center of mass, and the second element is the vertical coordinate (or y-coordinate). The centroids are calculated for each person(s) in the range video sequence and are used as an input to the *Particle Filter*. This process serves as a primary step for tracking.

### 3.5 Tracking [Particle Filter]

In order to achieve tracking of the people in the range video sequence a particle filter was used. The basic idea of a particle filter [2] is to use a number of independent random variables called particles, sampled directly from the state space. In order to represent the posterior probability, and update the posterior by involving the new observations; the *particle system* is properly located, weighted, and propagated recursively according to the Bayesian rule. They are usually used to estimate Bayesian models and are the sequential ('on-line') analogue of Markov chain Monte Carlo (MCMC) batch methods and are often similar to importance sampling methods. If well designed, particle filters can be much faster than MCMC. They are often an alternative to the Extended Kalman filter (EKF) or Unscented Kalman filter (UKF) with the advantage that, with sufficient samples, they approach the Bayesian optimal estimate, so they can be made more accurate than the EKF or UKF. The centroids of the human target are extracted from each frame (using the process outlined in 5.4). The sequences of centroids over the entire sequence are then given as an input to the particle filter. The particle filter analyzes the shift in the position of the centroids over the entire video sequences and then accordingly tags and associates each blob/human target with a label. In this manner the algorithm can track when (i.e. at which frame sequence number) the human target enters and exits the range video sequence.

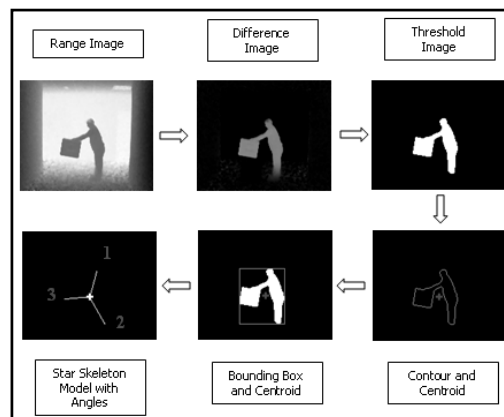


Figure 2: Algorithm Overview

### 3.6 Tracking Multiple People

The system designed has a provision for tracking multiple people in a given environment. For this purpose, the algorithm starts by segmenting the human targets from the range video sequence. The centroids are then detected by the algorithm for each of the human targets separately and later processed by the particle filter. The particle filter then uniquely tags and identifies each human target separately and traces its path of motion in the given range video sequence.

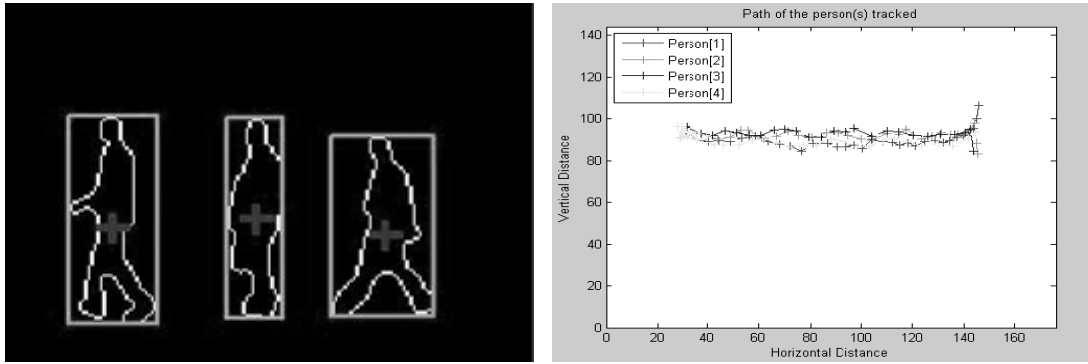


Figure 3: Particle filter output used for tracking four construction workers

### 3.7 Star Skeleton Modeling

In order to analyze the motion and activity of the human in the range video sequence, it is essential to extract and properly identify the extreme points lying on its boundary/contour. For this purpose the system uses the star skeleton model as proposed by Fujiyoshi. In his paper, Fujiyoshi describes a process of obtaining the star skeleton model from the contour of an extracted object, called as image skeletonization. The method proposed provides a simple, real-time, robust way of detecting extreme points on the boundary of the target to produce a star skeleton. The star skeleton consists of only the gross extremities of the target joined to its centroid in a star fashion. In order to obtain the star skeleton model the sequence of steps outlined below are followed [10]:

1. The centroid ( $x_c, y_c$ ) of the human target's boundary is determined using the points on the contour.
2. After this, the distance ( $d_i$ ) from each point on the boundary ( $x_i, y_i$ ) to the centroid ( $x_c, y_c$ ) is determined.
3. Next, low pass filtering of the signal ( $d_i$ ) in the frequency domain is performed to remove the noise present in the signal, becoming ( $d_{ic}$ ). Note that the signal  $d(i)$  is periodic with period  $N_b$  (number of pixels on the boundary of the human target)
4. The next step is to pick the maxima or peak points of the filtered signal. These points are then connected to the centroid of the target to form the star skeleton model.

The number of features (i.e. extreme points) to be detected can be controlled by the cut-off frequency ( $c$ ) of the low pass filter. The higher the value chosen for ( $c$ ), more extreme points can be detected in the filtered signal producing a more detailed star skeleton model. By default the value of  $c$  is kept as:  $c = 0.025 \times N_b$ . The main advantages of the star skeleton model are that they are computationally cheap, the scale sensitivity can be controlled and it does not depend on any apriori human model.

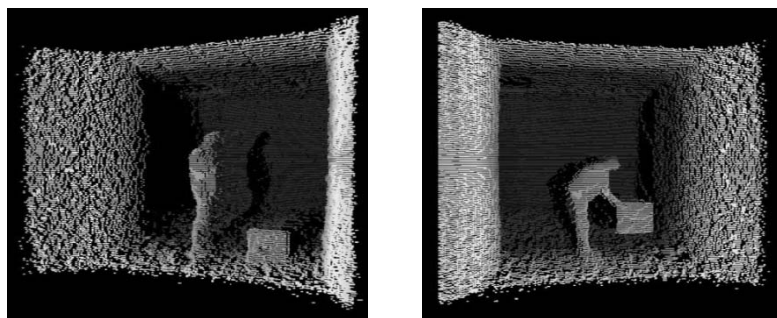


Figure 4: 3D Range Point Clouds of person picking a box

### 3.8 Angle Detection

Once the star-skeleton model is designed, angle detection between the various parts of the star-skeleton model is performed. The angles are calculated between the head with the left arm, the left arm with the left leg, the left leg with the right leg and the right leg with the right arm. The angle calculation is then repeated for each frame in the entire range video sequence. The angles calculated are stored in a data-structure for motion analysis. Motion analysis is a technique which is used to determine the type of motion activity which is performed by an object by analyzing the variation in the change of its path traced over a sequence of frames. The particle filter is then used, to trace the motion path of the person. However to get a more detailed view of the activity being performed by the worker, we incorporated the use of angles for motion analysis.

## 4. Experiments

The tracking algorithm was run on six different types of experiments which involved a worker performing various tasks in an indoor environment. The experiments included lifting a box, waving a flag to indicate warning, crawling, side-walking, sit-ups and normal straight walking. For each of these experiments, five different sets of range video sequences were recorded using a 3D range image camera. All of these experiments were recorded at a rate of 25 frames/sec. The pixel array resolution of the range image camera was 176 x 144 pixels. The output of the range camera contained the intensity, amplitude and range information of each of the pixels on the camera's sensor (176 x 144) on a per frame basis. The experiments were taken in an indoor environment due to the maximum distance constraints of the range camera.

## 5. Results and Analysis







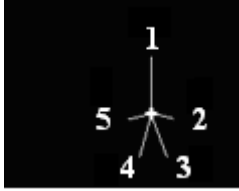
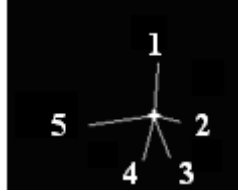

From the results in Section 7, we analyze the output of our algorithm for different test cases. Experiment 1 involves a worker performing a sidewalk experiment, in which the worker moves from left to right and vice versa. Experiment 2 involves a worker performing a crawling experiment on the ground again from left to right and vice versa. As seen above the worker has clearly been segmented from the scene and all associated noise has been cleaned from the input. The bounding box and centroid superimposed on the video sequence can be used to detect the real-time position of the worker. The star-skeleton model is in perfect conjunction with the position of the body parts of the worker as seen in the results. Once the star-skeleton model and centroid position of the worker is determined, then the motion path is traced with the particle filter designed. The plots of the tracked workers indicate the direction of motion as well as the change in the centroid position of the moving worker.

Experiment 3 involves a worker waving a flag up and down to indicate warning in a construction site. As seen in the results section, each of the body parts of the star-skeleton model are labeled successfully by our algorithm, starting with the head and then moving clockwise. From the calculations we observe that for this particular test case, the angles between [1-2], [2-3], [3-4] remain almost constant with a little variation. This is in accordance with our experiment since the relative position of the worker's head with the left arm, the left arm with the left leg and the left leg with the right leg remain almost constant in this experiment. The plot of the variation of angle between the different segments of the star-skeleton model is as shown in figure 5.

Experiment 4 taken involved a worker lifting a box. In this particular experiment all the angles between the segments of the star skeleton model vary as the worker bend down to life the box. Based on the angle of bending we can determine whether the worker is bending down in the right posture while lifting the box. The plot of the variation of angle between the different segments of the star-skeleton model is as shown in figure 6.

In experiments 3 and 4, we use the angle variations to determine if the worker is performing the activity in a correct posture. In experiment 3, where the worker is waving the flag to indicate a warning, the angle between the leg and the hand needs to reach at least 90 degrees to indicate a warning message. It needs to be observed that arm waving the flag is properly outstretched without bending. In case of experiment 4, if there is no angle between the back of the knees while bending down to lift the box, then it indicates incorrect posture. Another important observation to check is whether all the angles are varying while bending down. This is done to verify that the worker bends his back properly and does not maintain a straight back posture while picking the box

Table 1: Worker Waving a Flag

FRAME NO.	Frame 1	Frame 27	Frame 51
DENOISING AND CONDITIONING			
BOUNDING BOX AND CENTROID			
STAR SKELETON MODELING			
ANGLE DETECTION HEAD: 1 LEFT ARM : 2 LEFT LEG: 3 RIGHT LEG: 4 RIGHT ARM: 5	[4-5] = 61.05 [5-1] = 103.24 [1-2] = 104.03 [2-3] = 55.40 [3-4] = 36.26	[4-5] = 68.24 [5-1] = 102.52 [1-2] = 100.34 [2-3] = 54.09 [3-4] = 34.78	[4-5] = 99.33 [5-1] = 72.89 [1-2] = 98.88 [2-3] = 54.09 [3-4] = 34.78

We extended the angle calculation section of our algorithm to various other experiments performed and different trends were obtained for each of these cases. For example, in case of the sidewalk experiment, it was observed that the angle between the head and both arms remain constant, while the angles between the legs vary reach a minimum value of 0 degrees (i.e. when the worker joins his legs). For each experiment scenarios (walking, running, side-walking, crawling, etc.) we modeled the algorithm to analyze the angles and accordingly detected the corresponding activity performed by the worker.

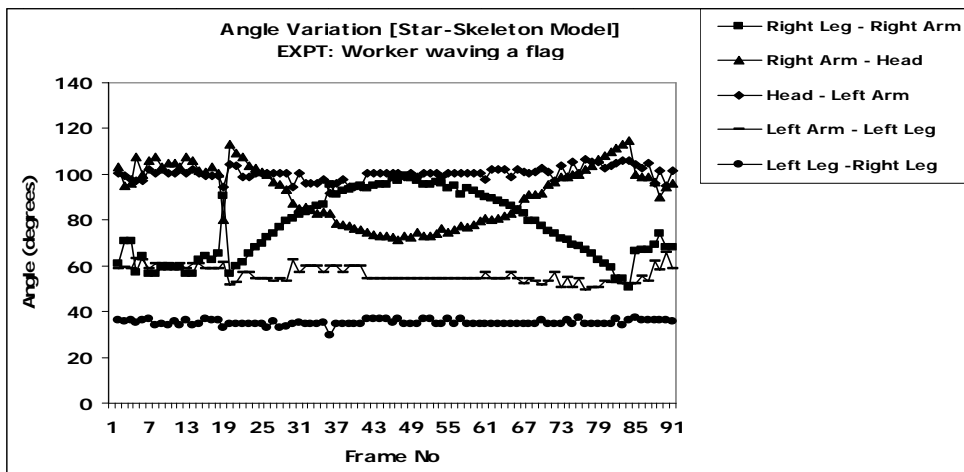






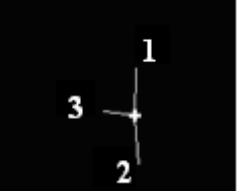

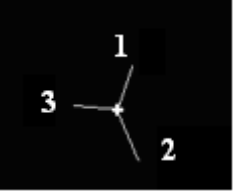


Figure 5: Plot of angle variation of the star-skeleton model for the flag experiment

The plot of the centroid variation with respect to the frame no. is as shown above for the case of the experiment of a construction worker lifting a box. From the graph we note that the worker performs the task in a periodic manner (i.e. same posture while bending each time). We also observe that the construction

worker does not bend properly since the minimum distance of the centroid from the ground is beyond the defined threshold value. If the construction worker had maintained a correct posture while bending down, then the centroid of the tracked target would have reached closer to the ground while lifting the box. This condition was not observed which indicates an incorrect posture. In a similar way, for the remaining experiments by setting a threshold parameter (e.g. angle between legs, angle between right arm and head, maximum centroid, etc), we can determine if the construction worker is maintaining a correct posture to avoid chronic physical stress related conditions.

Table 2: Worker Lifting a Box

FRAME NO.	Frame 23	Frame 35	Frame 60
DENOISING AND CONDITIONING			
BOUNDING BOX AND CENTROID			
STAR SKELETON MODELING			
ANGLE DETECTION HEAD: 1 LEFT LEG: 2 RIGHT LEG: 2 LEFT ARM : 3 RIGHT ARM: 3	[1-2] = 173.59 [2-3] = 84.20 [3-1] = 102.19	[1-2] = 174.49 [2-3] = 96.53 [3-1] = 88.96	[1-2] = 138.73 [2-3] = 103.24 [3-1] = 118.02

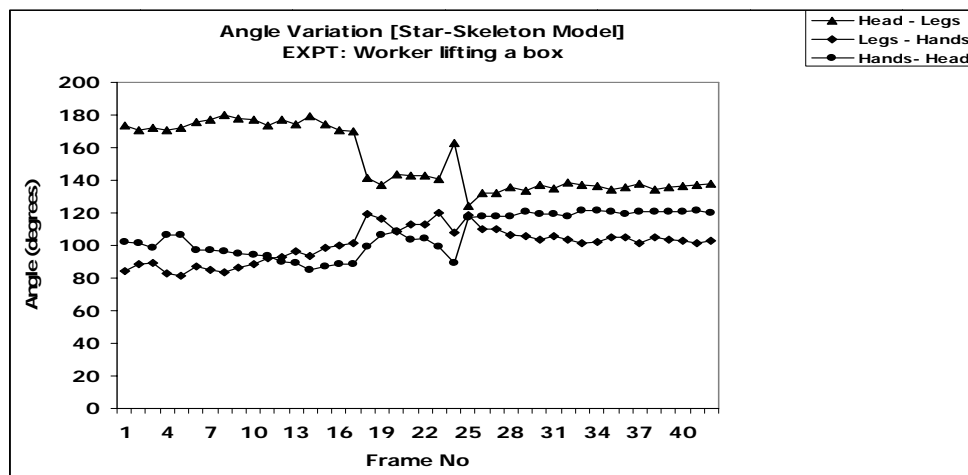


Figure 6: Plot of angle variation of the star-skeleton model for the box experiment



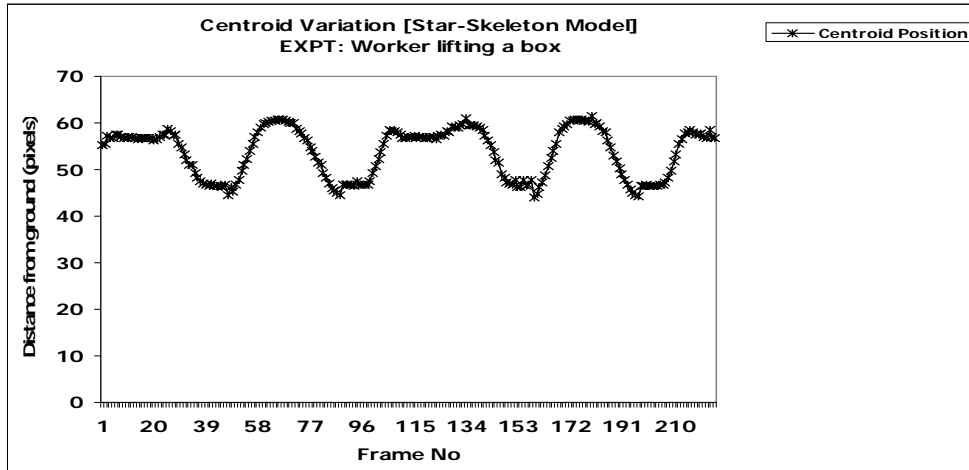


Figure 7: Plot of centroid variation of the star-skeleton model for the box experiment

## 6. Conclusions and Future Work

We present a real-time system for the tracking and surveillance of construction workers in a work-site for safety and health monitoring purposes. The safety of the worker is monitored continuously by tagging the target using a particle filter and then following it in the work-site. For extending the use of our system for health related issues, we modeled the worker using a star skeleton structure and then performed motion analysis to determine the variation in angles between the various segments of the model. On basis of this variation we can determine if the worker has performed the task maintaining a correct posture. Our system designed uses a 3D range image camera to segment a given target from a scene using distance based information thus providing segmentation in 3D space. We also were able to detect multiple workers and track and monitor each of them independently in the work-site. The algorithm is computationally inexpensive on resources making it a fast and robust real-time tracking system. Our system if used in conjunction with neural network training algorithms would certainly help to avoid fatal accidents as well as physical health related issues which are an important concern in the construction and transportation industry. Due to the maximum distance constraint imposed by the 3D range image camera we limited the scope of our experiments to an indoor work environment. In future we will extend our system to outdoor environments where several issues which include dynamic change in the environment and higher levels of noise will occur. Also we also plan to incorporate the use of an automated neural network system for making the training process of the system more robust.

## References

- [1] Teizer, Jochen, 3D range imaging camera sensing for active safety in construction, ITcon Vol. 13, Special Issue Sensors in Construction and Infrastructure Management, pg. 103-117, 2008.
- [2] J. C. Crocker and D. G. Grier, "Methods of digital video microscopy for colloidal studies," J. Colloid Interface Sci. 179, 298-310, 1996.
- [3] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, Pfnder: real-time tracking of the human body, Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96), p.51, Oct. 14-16, 1996.
- [4] Donatello Conte, Pasquale Foggia, Jean-Michel Jolion, Mario Vento, A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions, Pattern Recognition, v.39 n.4, p.562-572, April, 2006.
- [5] J. Yamato, J. Ohya and K. Ishii. "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 379--385, 1992.
- [6] N. Otsu, "A Threshold Selection Method from Gray Level Histograms," IEEE Trans. Systems, Man, and Cybernetics, vol. 9, pp. 62-66, 1979.

- [7] Hironobu Fujiyoshi, Alan J. Lipton, "Real-time Human Motion Analysis by Image Skeletonization," wacv, p.15, Fourth IEEE Workshop on Applications of Computer Vision (WACV'98), 1998.
- [8] Elden Yu, J.K. Aggarwal, "Detection of Fence Climbing from Monocular Video," icpr, pp. 375-378, 18th International Conference on Pattern Recognition (ICPR'06) Volume 1, 2006
- [9] I. Haritaoglu, D. Harwood, L.S. Davis, "W4: A Real Time System for Detecting and Tracking People," cvpr, p.962, 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
- [10] Hsuan-Sheng Chen , Hua-Tsung Chen , Yi-Wen Chen , Suh-Yin Lee, Human action recognition using star skeleton, Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, October 27-27, 2006, Santa Barbara, California, USA.
- [11] Elden Yu , J. K. Aggarwal, Detection of stable contacts for human motion analysis, Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, October 27-27, 2006, Santa Barbara, California, USA.
- [12] Christoph Bregler, Learning and Recognizing Human Dynamics in Video Sequences, Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR '97), p.568, June 17-19, 1997.
- [13] U.S Department of Labor, Bureau of Labor Statistics,  
<http://www.bls.gov/news.release/cfoi.t03.htm>