Tracking People with Multiple Kinects

University of St Andrews

Researcher: Chi-Jui Wu
Supervisor: Dr. David Harris-Birtill

sicsa* The Scottish Informatics & Computer Science Alliance

INTRODUCTION

People detection and tracking is the process of identifying and following people in an environment. Knowing where people are in realtime is important for human motion tracking and analysis. In particular, interactive systems often require people's spatial position and body features, such as the face and joints. For instance, gesture recognition relies on realtime information about a person's hand location and movement. Occlusion prevents tasks like this to completed.

The system resolves occlusion by leveraging the extended field of view with multiple Kinects. User studies show that the current system can reliably track people under occlusion within their personal space (about 15 cm) in various scenarios and Kinect placements. Future work could incorporate other algorithms or applications to enhance interactive experience.

Figure 1: An illustration of the occlusion problem. The skeletons are occluded by the red obstacle. In such situations, the current system uses multiple Kinects to actively track people under occlusion.

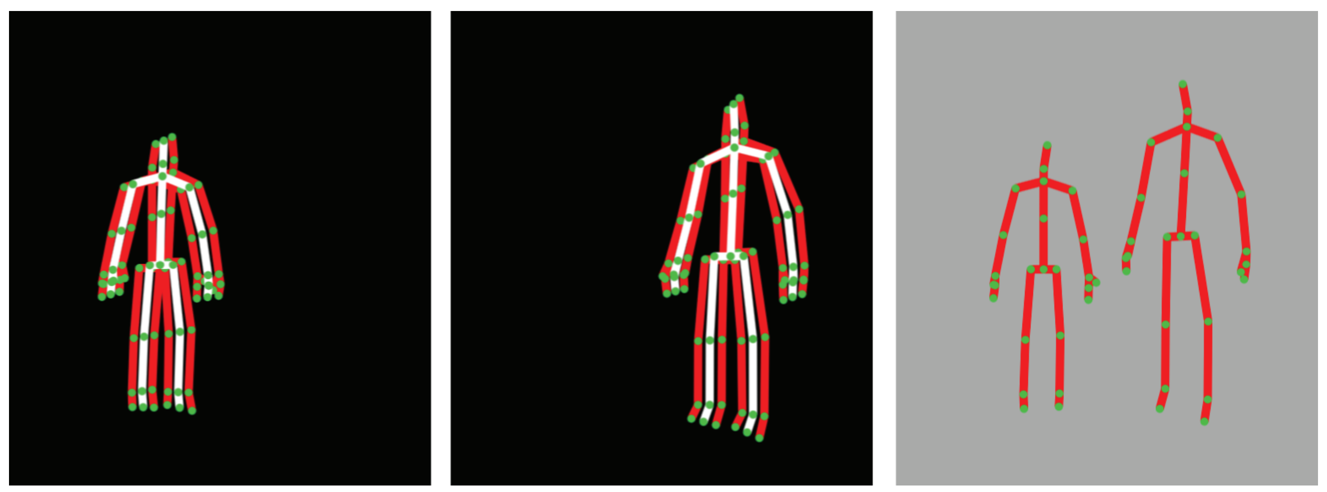

Figure 2: A demonstration of the tracking system. The left and middle images show the merged skeletons of the same person from the perspective of the front and 45° Kinect, respectively. The average skeleton is colored in white. The right image shows the skeletons in their original Kinect field of view.

CURRENT APPROACH

The tracking algorithm is based on 3D coordinate transformation around the y axis using unit quaternions. It consists of two parts: calibration and tracking.

Calibration measures each skeleton's initial center position and angle relative to each Kinect in the scene for a duration of 120 frames. The initial center position is the average of all joints coordinates, expressed as $C(X_C, Y_C, Z_C)$. The initial angle is the angle between the skeleton and Kinect in the last calibration frame, expressed as Θ. The following equation shows the transformation from the Kinect coordinate system to the World coordinate system:

$$K'_J(X_{K',J}, Y_{K',J}, Z_{K',J}) = (X_{K,J} - X_C,\ Y_{K,J} - Y_C,\ Z_{K,J} - Z_C)$$

$$W_J(X_{W,J}, Y_{W,J}, Z_{W,J}) = (X_{K',J}\cos\Theta - Z_{K',J}\sin\Theta,\ Y_{K',J},\ Z_{K',J}\cos\Theta - X_{K',J}\sin\Theta)$$

where X, Y, and Z are the coordinates, J is joint type, K is Kinect coordinate space point, and W is World coordinate space point.

Tracking is performed by updating the initial tracing result. Skeletons from multiple Kinects fields of view are matched based on their spatial proximity in the World coordinate system.

USER STUDIES

The efficacy of the tracking algorithm is measured by the differences in the x, y, and z coordinates between multiple skeletons of the same person from different Kinects fields of view. Each study involved 20 participants, except for the two person interaction task.

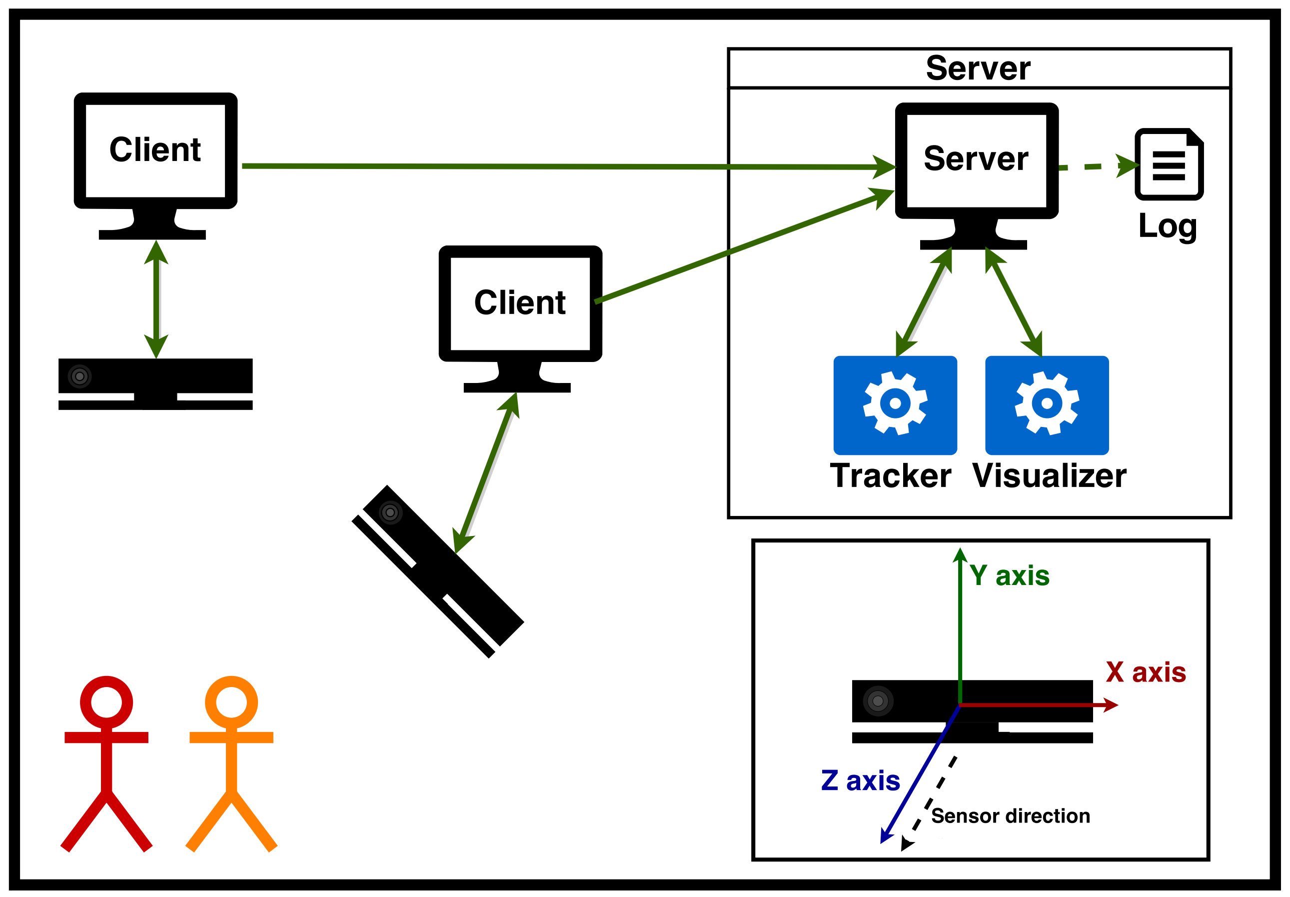

Figure 3: The system architecture. The clients stream skeleton data to the server, which performs the tracking and provides provides the visual feedback. The server also does logging during user studies.

Figure 4: The user studies setting

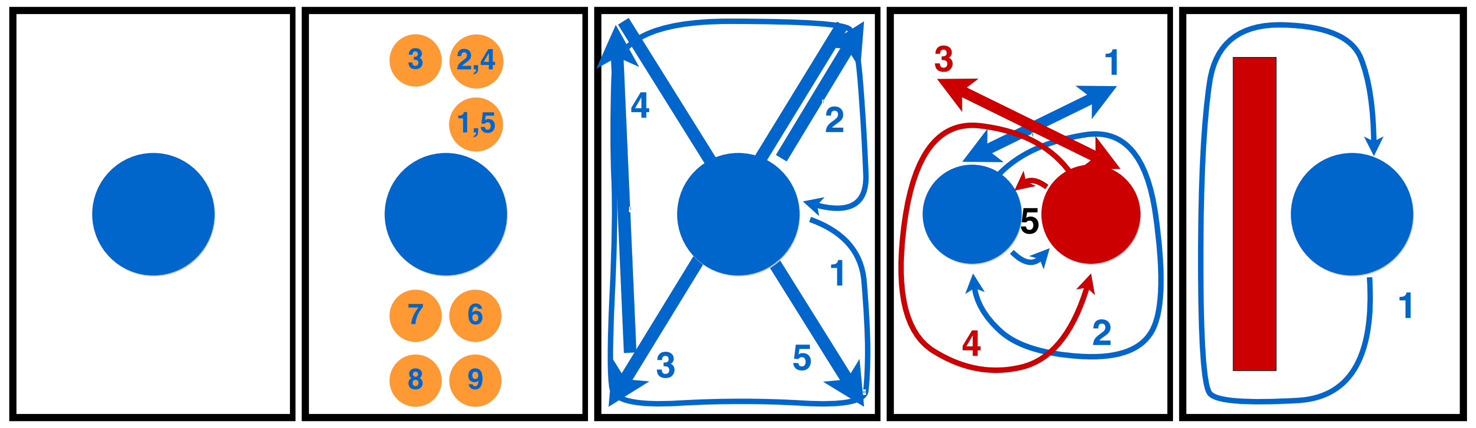

Figure 5: Tasks in the user studies. They are Stationary, Steps, Walk, Interactions and Obstacle, from left to right.

CONCLUSION

1. The system can persistently track people under occlusion
2. Average joint difference is within personal space (15 cm)
3. Average Δx, Δy, Δz, Δd increase with more difficult tasks
4. Average Δx, Δy, Δz, Δd increase with large angles between Kinects
5. The torso is more reliable than other body regions

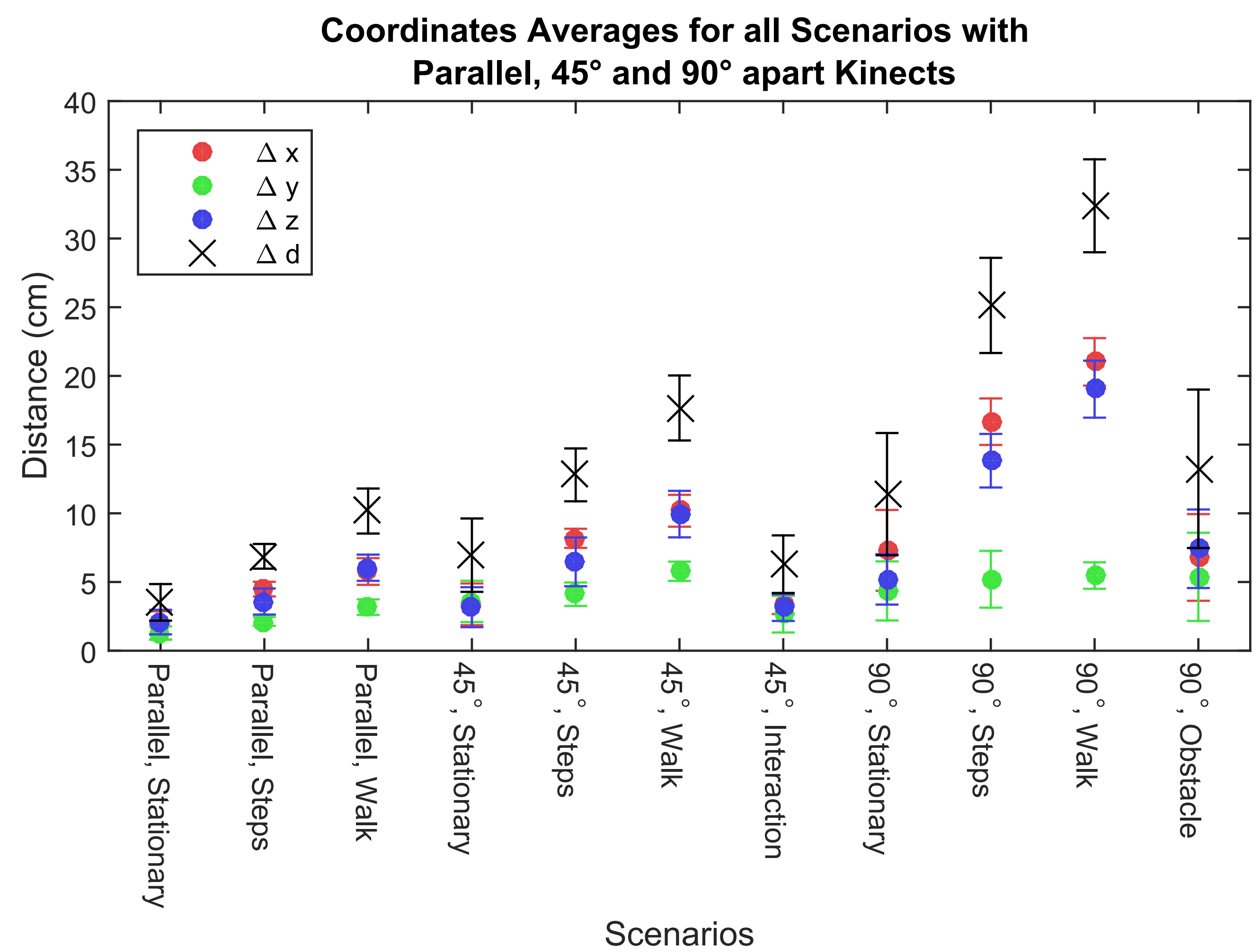

Figure 6: Plot showing the overall results in the user studies