Enhancements to 3D Capture of Room-sized Dynamic Scenes with Pan-Tilt-Zoom Cameras

Asad Ullah Naweed, Lu Chen, Mingsong Dou, Adrian Ilie, Henry Fuchs

Overview

We present a method to improve the geometry and texture quality of 3D reconstructions generated from fixed off-the-shelf color + depth cameras (Microsoft Kinects) [Dou14] by using Pan-Tilt-Zoom cameras to provide more details in areas of interest , such as textured backgrounds and human faces. We apply our method to both static scenes and moving objects.

Improving the 3D Reconstruction of Static Scenes

We improve the offline scanning results of [Dou14] with high resolution textures generated by scanning using multiple PTZ cameras at a high zoom level and multiple pan-tilt angles. See Figures 1 and 2.

Steps for each camera image:

- Estimate camera pose by matching SIFT features in the image with features in the Kinect RGB-D frames from the 3D reconstruction
- Iteratively refine the pose estimate using the optical flow from nearby images with known camera poses
- Reproject the PTZ image as a texture onto the surface from [Dou14]

Results:

- 6x improvement in quality, measured as color samples per unit area
- 64% success rate in camera pose estimation
- Mean 1.25 pixels (1.65 mm in our scene) of reprojection error
- 64% of matches as inliers in the estimated model

Improving the 3D Reconstruction of Dynamic Objects

We use three PTZ cameras to improve the reconstruction and texture quality of the online reconstruction of [Dou14]. See Figure 3.

Steps for each iteration:

- Zoom in cameras to obtain high-resolution images of a dynamic object as it moves and deforms.
- Use Semi-Global Block-Matching for pair-wise stereo reconstruction
- Merge the three resulting point clouds using the Iterative Closest Point algorithm
- Remove noise using a Laplacian smoothing transform
- Triangulate the result using Poisson triangulation
- Estimate the camera poses by matching SIFT features in the static background
- Fuse our result with the result of [Dou14] using the Iterative Closest Point algorithm

Results:

- A much denser mesh than that the one generated using Kinect data
- Camera pose estimation algorithm always converged in less than 200 iterations
- Mean reprojection error of 7.8 mm in the background feature points

Future Work

- Experiment using other types of features for better and more robust camera pose estimation
- Incorporate automatic methods to track moving objects in real time using the cameras
- Investigate the possibility of a real-time implementation using GPUs

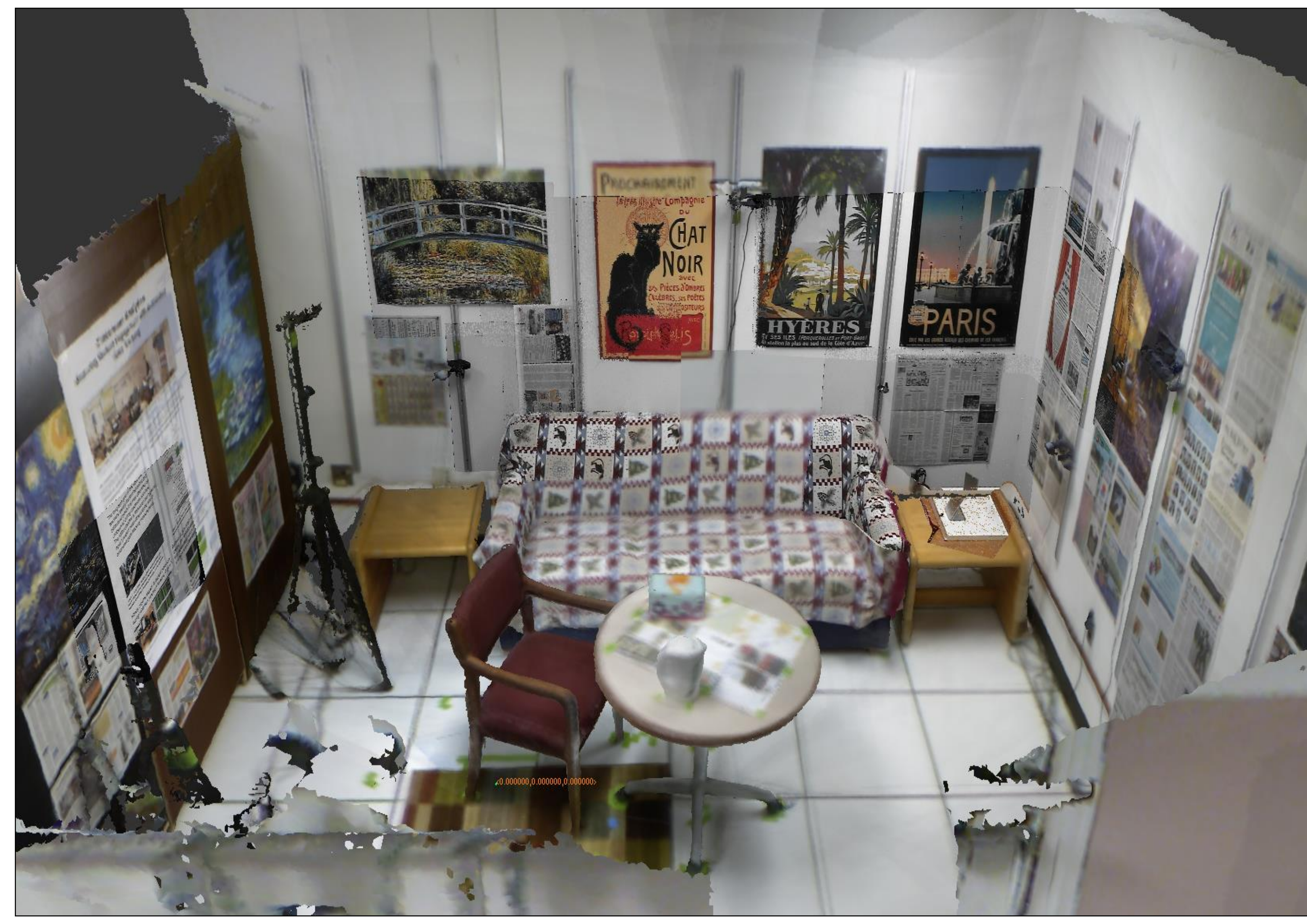

***Figure 1** – Regions with improved texture combined with regions with low resolution texture: overall view.*

***Figure 2** – Regions with improved texture combined with regions with low resolution texture: zoomed-in views.*

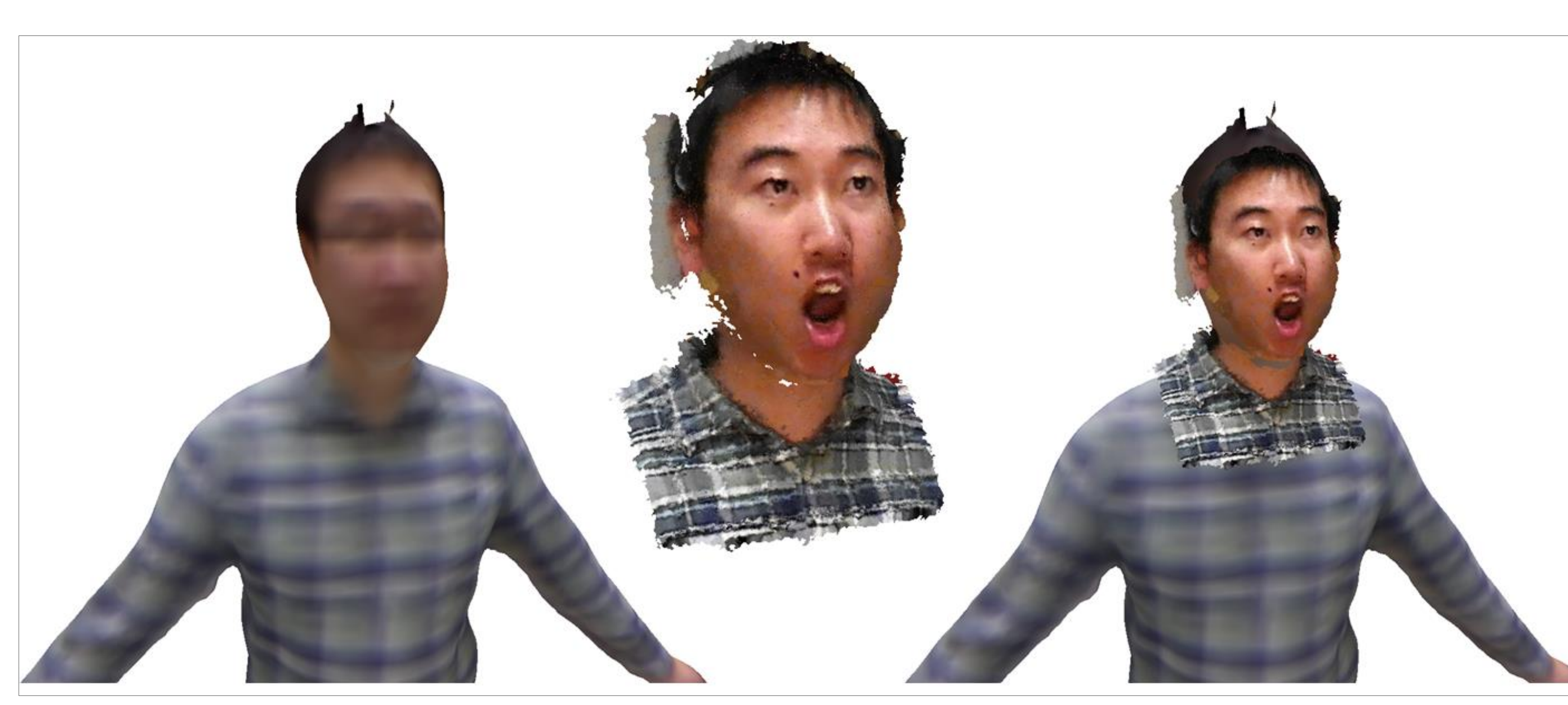

***Figure 3** – The surface from multiple Kinects, from* [Dou14] *(left), the stereo reconstruction from PTZ cameras (middle) and our fused result (right)*

References:

- [Naw14] Asad Ullah Naweed, Lu Chen, Mingsong Dou, Henry Fuchs, *"Enhancement of 3D Capture of Room-sized Dynamic scenes with Pan-Tilt-Zoom Cameras,"* International Symposium on Visual Computing, Las Vegas, 2014
- [Dou14] M. Dou and H. Fuchs, *"Temporally Enhanced 3D Capture of Room-sized Dynamic Scenes with Commodity Depth Cameras,"* IEEE VR, Minneapolis, 2014. Award: Best short paper

UNC DEPARTMENT OF COMPUTER SCIENCE

This research was supported in part by Cisco Systems and by the BeingThere Centre, a collaboration among ETH Zürich, NTU Singapore, and UNC Chapel Hill, supported by ETH, NTU, UNC, and the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative

CV²R Consortium for Vision and Virtual Reality