3D Capture of Room-sized Dynamic Scenes with Commodity Depth Cameras

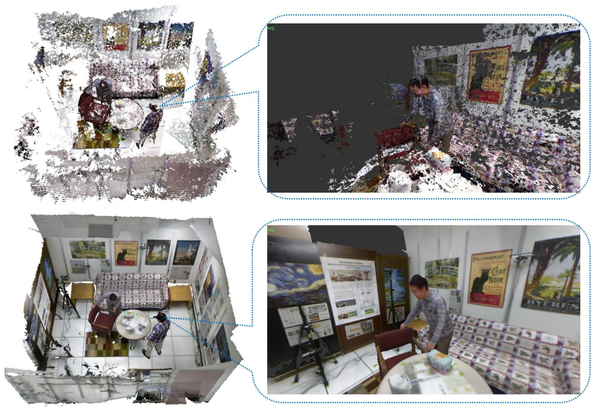
UNC DEPARTMENT OF COMPUTER SCIENCE

Mingsong Dou YoungWoon Cha Rohan Chabra Federico Menozzi Eric Wallen, MD Henry Fuchs

CV²R Consortium for Vision and Virtual Reality

Overview

We present a serious of our works for 3D capture of a room-sized dynamic scene with commodity depth and high-resolution cameras



The original input from 10 Kinects (top row); Enhanced results (bottom row). [3)]

The scene is partitioned into three categories:

1. ***Static Background***. Never move, such as walls, floor.
2. ***Semi-static Objects***. Move infrequently and do not deform, such as chairs.
3. ***Dynamic Objects***. Constantly move and deform, such as human beings.

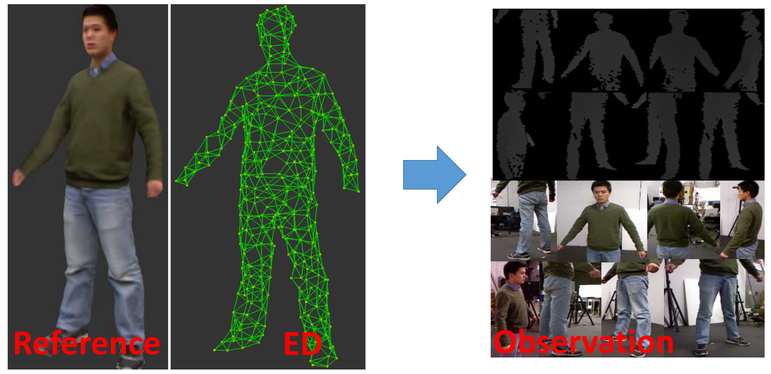
The static parts of the room are pre-scanned, and the movements of semi-static objects are tracked. Dynamic parts are accumulated over time using ***Non-rigid Alignment***.

Room Scanning with One Hand-held Kinect

(Left) Point cloud from 10 Kinects; (Right) Pre-scanned Result.

3D points are aligned using pair-wise image matching and ***Bundle Adjustment.***

Non-rigid Alignment for Dynamic Objects



Align a reference surface (left) with a newly observed data (right).

Non-rigid movement parameterization via Embedded Deformation Model (EDM): [1)]

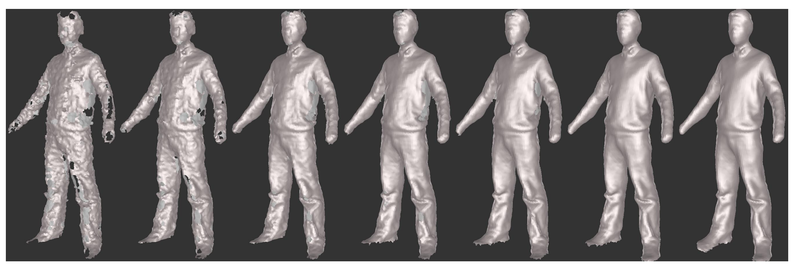
Randomly and uniformly sample ED nodes from the reference mesh; each node is associated with a local affine transformation $\langle \boldsymbol{A}, \boldsymbol{t} \rangle$ around its position $\boldsymbol{g}$. Thus, motion parameters are $\boldsymbol{G} = \{\langle \boldsymbol{A_k}, \boldsymbol{t_k} \rangle\}_{\boldsymbol{k=1}}^{\boldsymbol{K}}$. To deform a vertex on the reference:

$$\tilde{v} := ED(v; G) = \sum_{k \in \mathcal{N}} w_k [A_k(v - g_k) + g_k + t_k]$$

Non-rigid Alignment by Energy Minimization: [4)]

MIN: $w_{rot}\boldsymbol{E_{rot}} + w_{reg}\boldsymbol{E_{reg}} + w_{dns\_pts}\boldsymbol{E_{dns\_pts}} + w_{clr}\boldsymbol{E_{clr}}$

$\boldsymbol{E_{rot}}$, $\boldsymbol{E_{reg}}$: *regularization term on the EDM;* $\boldsymbol{E_{dns\_pts}}$: *dense point cloud alignment term;* $\boldsymbol{E_{clr}}$: *color consistency term*



Accumulated surfaces when fusing more and more frames. The newly observed data is transformed back to the reference pose and fused at the reference using a volumetric method. [2)]

October, 2015

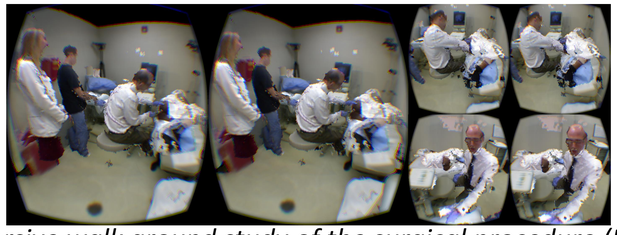
Texture Enhancement using High-Resolution Cameras

The improved textures are mapped from the calibrated high-resolution cameras.

Per-vertex colors from Kinect (left) Textures transferred from High-res images (right)

Immersive 3D Experience with HMD

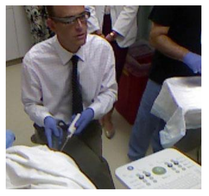
An entire surgical experience is captured for multiple Kinect cameras. The surgical sequence can be reconstructed for education training.



Immersive walk around study of the surgical procedure (Image from Oculus Rift DK2)

Future Work

Improve reconstruction quality with head-worn cameras. Immersive annotation and playback controls for students.



1) *RW Sumner, J Schmid, M Pauly, "Embedded Deformation for Shape Manipulation," SIGGRAPH 2007*
2) *M Dou, H Fuchs, JM Frahm, "Scanning and Tracking Dynamic Objects with Commodity Depth Cameras," ISMAR 2013*
3) *M Dou, H Fuchs, "Temporally Enhanced 3D Capture of Room-sized Dynamic Scene with Commodity Depth Cameras," VR 2014*
4) *M Dou, J Taylor, H Fuchs, A Fitzgibbon, "3D Scanning Deformable Objects with a Single RGBD Sensor," CVPR 2015*

Acknowledgement: This research was supported in part by Cisco Systems, by NSF Award II-1405847: Seeing the Future: Ubiquitous Computing in EyeGlasses, 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.