A Brief Introduction to 3D Capture Technology

Mingsong Dou

Google

12/05/2017
My experiences with 3D Capture

Autostereoscopic Display + 3D Data Acquisition

It starts with the Telepresence Project...
My experiences with 3D Capture

Gaze-correction for room-sized telepresence, IEEE VR 2012

Room scanning with bundle adjustment of points and planes, ACCV 2012

Non-rigid Surface Scanning with dense nonrigid bundle adjustment, CVPR 2015
Scanning and Tracking Dynamic Objects, ISMAR 2013

Fusion4D
SIGGRAPH’16

Room-sized Dynamic Scence Capture,
IEEE VR 2014

Holoportation
UIST’16

Motion2Fusion
SIGGRAPH Asia, 2017
Outline

• 3D Capture Sensors and Depth Estimation
  • Stereo
  • Structured Light
  • Time-of-Flight
  • Multive-view capture

• World Reconstruction
  • SLAM
  • Kinect Fusion

• People Reconstruction
  • Parametric Tracking
  • Non-Rigid Tracking and Fusion

• Applications in Mixed/Virtual Reality
  • Holoportation
Camera Pinhole Model

• Camera model:
  • Map a 3D point $X = [x; y; z]$ in the world to a 2D pixel position on the image $x = [u; v]$
  • Intrinsics: focus length, pixel size, lens distortion, etc. Usually represented as a 3x3 matrix $K$.
  • Extrinsics: camera position and orientation, represented by a rotation matrix $R$ and translation vector $t$. 
3D Capture/Reconstruction

• Reverse Rendering Problem
  • From 2d image point $x$ to the corresponding 3D world point $X$
  • Triangulation
Stereo

• 1D search problem:
  • search along the epipolar line

**Epipolar Constraints:** Potential matches for $x$ have to lie on the corresponding epipolar line $l'$. 
Stereo

• 1D search problem:
  • search along the epipolar line
Stereo

• 1D search problem:
  • search along the epipolar line

• A trick to boost the performance
  • Image rectification
Stereo

- 1D search problem:
  - search along the epipolar line

- A trick to boost the performance
  - Image rectification

- Disparity map
  - Disparity is inversely proportional to depth

\[
\text{disparity} = x - x' = \frac{B \cdot f}{z}
\]
Stereo

- 1D search problem:
  - search along the epipolar line

- A trick to boost the performance
  - Image rectification

- Disparity map
  - Disparity is inversely proportional to depth
Active Stereo
Active Stereo
Active Stereo

• Other commercial sensors
  • Microsoft Kinect, Intel RealSense, ...

• IR: doesn’t work well at outdoor environment
Other 3D capture techniques

• Structured Light
Other 3D capture techniques

- Structured Light
- Time-of-flight

Microsoft Kinect V2
Other 3D capture techniques

• Structured Light
• Time-of-flight
• LIDAR
Shape-from-silhouette

Intersection of foreground cones
Shape-from-silhouette

Offline, Controlled environment
Geometry quality is low, but sharp edges
Shape-from-silhouette

Offline, Controlled environment
Geometry quality is low, but sharp edges
Microsoft Free Viewpoint Video (H-Cap)
Reconstruct the world

• Simultaneous localization and mapping (SLAM)
  • One moving camera

Reconstruct the world

- Simultaneous localization and mapping (SLAM)
  - One moving camera
- Google ARCore/Apple ARKit
  - mobile camera + IMU
Reconstruct the world

• Simultaneous localization and mapping (SLAM)
  • One moving camera
• Google ARCore/Apple ARKit
  • mobile camera + IMU
• Bundle Adjustment
  • a technique to simultaneously optimize both geometry and camera poses

Reconstruct the world

• Simultaneous localization and mapping (SLAM)
  • One moving camera

• Google ARCore/Apple ARKit
  • mobile camera + IMU

• Bundle Adjustment
  • a technique to simultaneously optimize both geometry and camera poses

• KinectFusion
  • One depth camera, detailed geometry

"KinectFusion: Real-time dense surface mapping and tracking," Newcombe, Richard A., Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. ISMAR 2011
Bundle Adjustment

\[ \min \{ \{ X \downarrow k \}, \{ R \downarrow i, t \downarrow i, K \downarrow i \} \} \sum_{i,k} \| x \downarrow i,k - P \downarrow \text{proj}(X \downarrow k; R \downarrow i, t \downarrow i, K \downarrow i) \|^2 \]
SLAM

- detect sparse feature points, eg. *SIFT*
- Initial the system with two-view-geometry
- Estimate camera poses for later frames by matching 2D features with 3D points

“Parallel Tracking and Mapping for Small AR Workspaces”, Georg Klein and David Murray, ISMAR'07
KinectFusion

- Data accumulation in a TSDF grid
- Camera pose tracking with **Iterative Closest Point (ICP)**
People Reconstruction
The need for temporal consistency

Live Data

Temporally Consistent Model
The need for temporal consistency

Multiple Point Clouds (Bilaterally-Smoothed)  Fused Live Data (Kinect Fusion)  Temporally consistent model
Reconstruct People

• Articulated Body tracking
  • Human Body tracking

“Marker-less Motion Capture of Skinned Models in a Four Camera Set-up using Optical Flow and Silhouettes”, L. Ballan and G. M. Cortelazzo, 3DPVT 2008
Reconstruct People

• Articulated Body tracking
  • Human Body tracking
  • Hand Tracking

“Articulated Distance Fields for Ultra-Fast Tracking of Hands Interacting”, Jonathan Taylor, Vladimir Tankovich, Danhang Tang, Cem Keskin, David Kim, Philip Davidson, Adarsh Kowdle, Shahram Izadi. SIGGRAPH Asia 2017
Reconstruct People

• Articulated Body tracking
  • Human Body tracking
  • Hand Tracking

“Articulated Distance Fields for Ultra-Fast Tracking of Hands Interacting”, Jonathan Taylor, Vladimir Tankovich, Danhang Tang, Cem Keskin, David Kim, Philip Davidson, Adarsh Kowdle, Shahram Izadi. SIGGRAPH Asia 2017
Reconstruct People

• Articulated Body tracking
  • Human Body tracking
  • Hand Tracking

• Discriminative method
  • Kinect body tracker
  • deep learning

Reconstruct People

• Articulated Body tracking
  • Human Body tracking
  • Hand Tracking
• Discriminative method
  • Kinect body tracker
  • deep learning
• Face Tracking

Digital Emily Project, Paul Debevec
Less infra-structure, no green screen.
No pre-scan, works for general surface
Non-rigid Alignment

Parameterize motion field as **Embedded Deformation Graphs**
- One affine transform per node
- Enforce smoothness over edges.

Skinning: attach vertices to neighboring nodes
- Based on **geodesic distance**
- Linear blending

\[ v = \sum_{k \in \mathcal{N}} w_k [A(qk)(v-gk)+gk+t_k] \]

- geometry alignment
- learned correspondences
- texture/color consistency
- free-space penalization

as rigid as possible: minimum distortion

Reference

Deformation Graph

Data

Non-rigid Alignment

Non-linear least square problem

- **Custom/efficient** sparse Levenberg-Marquardt solver
- PC GPU ~1ms per iteration

- solve the normal equation at each step:

\[
(J^T J + \delta I) h = -J^T f
\]

A sparse block matrix, each pixel represents a 7x7 matrix.
Data Term

• Alignment residual between model and data

\[ E_{\text{data}}(G) = \sum_{m=1}^{M} \sum_{n=1}^{N} (\Psi(v_{\downarrow}m; D_{\downarrow}n))^2 \]

\[ \Psi(v_{\downarrow}m; D_{\downarrow}n) = \delta_{mn} (v_{\downarrow}m - \Gamma_{\downarrow}n(v_{\downarrow}m)) \]

\( \Gamma_{\downarrow}n(v_{\downarrow}m) \): **projective correspondence** of \( v_{\downarrow}m \) in depth map \( D_{\downarrow}n \)

\[ \delta_{mn} = \begin{cases} 1 & \text{if } v_{\downarrow}m \text{ is visible to } D_{\downarrow}n \\ 0 & \text{otherwise} \end{cases} \]
Visual Hull Constraint

Visual Hull Constraint

\[ E_{\text{hull}}(G) = \sum_{m \in M} \| \mathcal{H} (v \downarrow m) \|_2 \]

\( \mathcal{H} \): the Distance Transform to the visual hull (0 inside visual hull)

Measure the difference between the observed color images \( \{ \mathcal{I}_k \} \) \( \downarrow k = 1 \uparrow K \) and the reconstructed image by projecting the per-vertex-colored mesh:

\[
E_{\text{clr}} = \sum_{i} \sum_{k=1}^{K} \delta_{ik} \| \mathcal{I}_k (\Pi \downarrow k (v_{\downarrow i})) - c_{\downarrow i} \|_2
\]
Single Sensor Results
High Speed 360 Capture
Holoportation
Normal environment
24 cameras
Real-time
Holoportation
Holoportation
Results
Non-human examples

real-time multi-view reconstruction
Usage Scenarios
Challenges for the future: FoV
Challenges: FoV
Challenges: Infrastructure
Challenges: Headset removal
Compression

• Raw Data from Cameras = 23 Gb/s
• Current HoloPort Compression = 1 Gb/s
• HD video = 10 Mb/s
• Exploit temporally consistency and texture atlas
Thank you!