Computer Vision Systems Programming VO

3D Vision Applications

Christopher Pramerdorfer

Computer Vision Lab, Vienna University of Technology

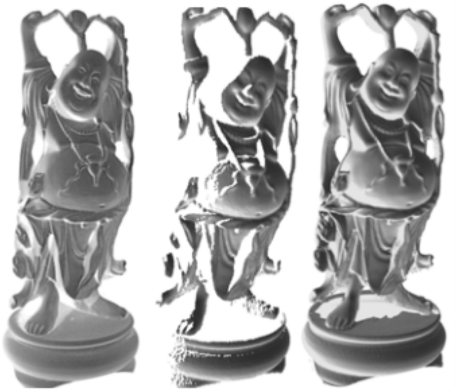
Topics

CV applications utilizing scene geometry (3D data)

- Focus on those based on Kinect

Images by Ryuzo Okada, Shotton et al. 2011, Newcombe et al. 2011

3D Reconstruction



Images from Curless and Levoy 1996

3D Reconstruction

Construction of accurate 3D models from range data

- Usually involves combing multiple point clouds

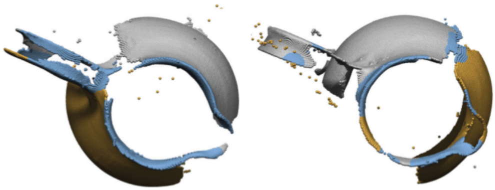
Accomplished in two steps

- Align range data (map to common coordinate system)
- Merge range data in a way that minimizes errors

Often followed by surface reconstruction

Popular method for aligning two point clouds $\{\mathbf{r}\}$, $\{\mathbf{s}\}$

- Goal is to find parameters $\boldsymbol{\theta}$ of some transformation $\mathcal{T}$
- Usually assuming a rigid transformation



Images from Aiger, Mitra, and Cohen-Or 2008

Algorithm iterates between

- Finding point correspondences based on distance, $\{(r_n, s_n)\}_n$
- Finding the $\boldsymbol{\theta}$ that minimizes $\sum_n \|\mathbf{r}_{r_n} - \mathcal{T}(\mathbf{s}_{s_n}; \boldsymbol{\theta})\|_2^2$

Converges towards a local minimum

- Requires good initial estimate of $\boldsymbol{\theta}$

`https://www.youtube.com/watch?v=ii2vHBwlmo8`

Truncated signed distance functions (TSDFs)

- Similar to distance transforms in 3D ($0 =$ surface)
- But distances are signed, measured along view rays

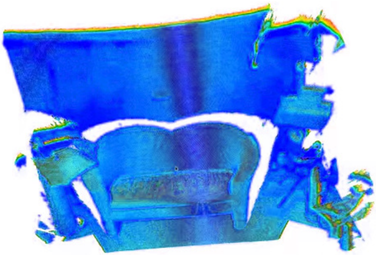
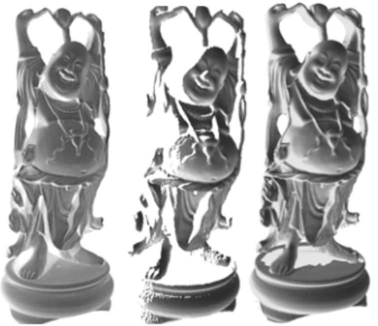


Image from https://www.youtube.com/watch?v=AjjSZufyprU

Merged data = weighted average over aligned TSDF voxels

- Weights based on e.g. object distance, angle

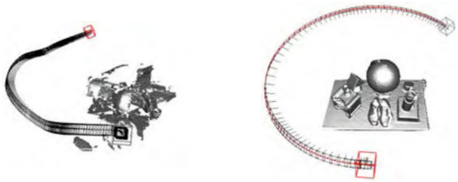


Images from Curless and Levoy 1996

Temporal fusion of Kinect depth maps

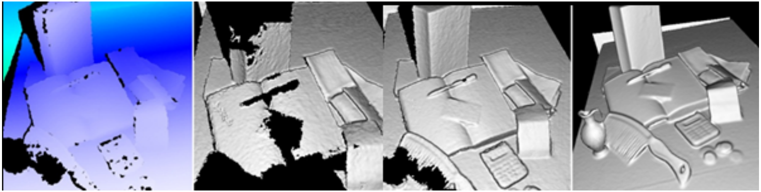
Based on the above methods (ICP & TSDF fusion)

- But $\{\mathbf{r}\}$ is synthesized from merged model
- Suppresses alignment error accumulation



Images from Newcombe et al. 2011

https://www.youtube.com/watch?v=quGhaggn3cQ



Images from microsoft.com

Reconstruction of surface mesh from point cloud

- Results in a (locally) watertight 3D model
- Allows for further processing (e.g. texturing)

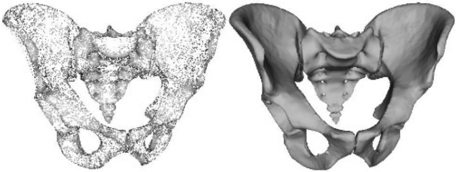


Image from Kazhdan 2005

Correction of point cloud errors

- Noise, outliers, alignment errors, missing data

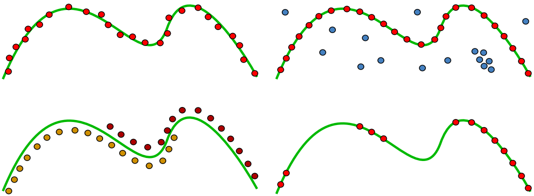
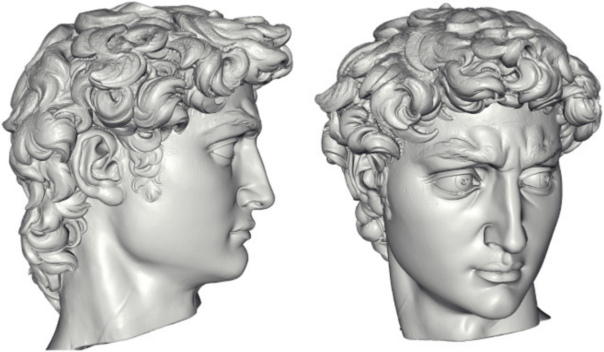


Image adapted from Berger et al. 2014

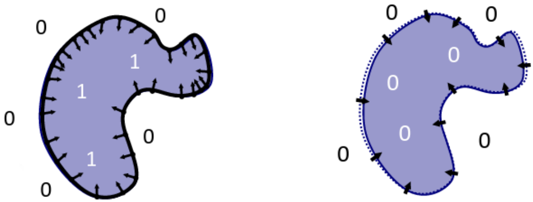
3D Reconstruction
Surface Reconstruction – Poisson Surface Reconstruction



Images from Kazhdan, Bolitho, and Hoppe 2006

Define $\chi(\mathbf{x}) = 1$ if $\mathbf{x}$ inside the object, 0 otherwise

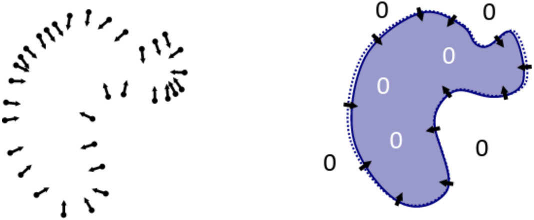
- Surface is at $\chi(\cdot) = 0.5$
- $\nabla\chi$ equals surface normal near surface, 0 otherwise



Images from Gotsman & Kazhdan's slides

Regard oriented points $\{(\mathbf{x}, \mathbf{n})\}$ as samples from $\nabla\chi$, $\nabla\chi(\mathbf{x}) = \mathbf{n}$

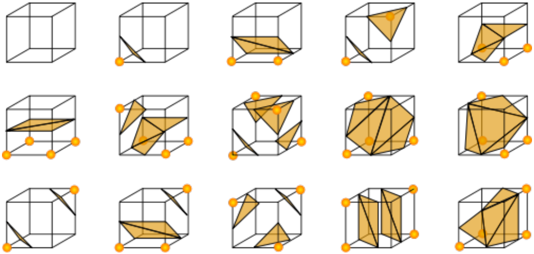
- Points define vector field $\mathcal{V}$ that corresponds to $\nabla\chi$
- Sought χ minimizes $\|\nabla\chi - \mathcal{V}\|$



Images from Gotsman & Kazhdan's slides

Once χ is known, the isosurface $\chi(\cdot) = 0.5$ can be extracted

- Using marching cubes, for example



Images from wikipedia.org

C++ open-source library for point cloud processing

Includes implementations of the above methods

Image from `pointclouds.org`

Preservation of physical artifacts

Image from Levoy et al. 2000

Project Tango

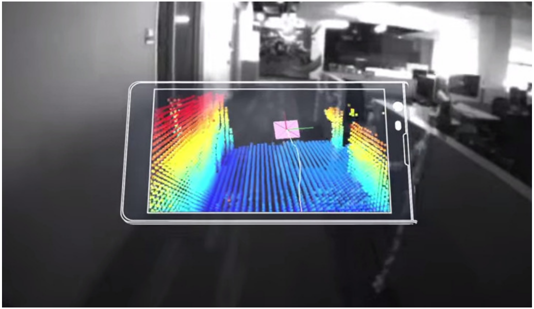
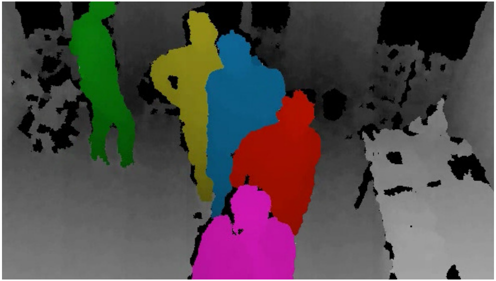


Image from https://www.youtube.com/watch?v=Qe10ExwzCqk

Person Detection

3D data enables reliable person detection

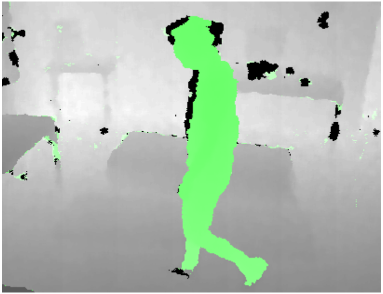
- Robust motion detection
- Distinctive, invariant features



Person Detection
Motion Detection

Reliable motion detection via background subtraction

- Measurements represent object distances
- Not affected by illumination, clothing, shadows



Person Detection
Features

Scene geometry allows for distinctive, invariant features

- Object size, extent, volume, shape, ...

More on object detection later

Person Detection
Applications – Breaking Assistance

https://www.youtube.com/watch?v=oU4XQvxO10k

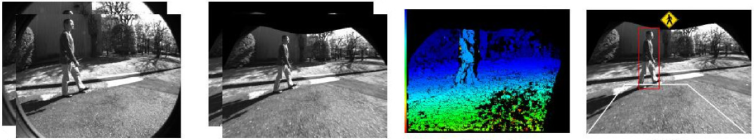


Image from Ryuzo Okada, Toyota

Person Detection
Applications – Interactive Art Installations

Image from ortlos.com

Fall detection system developed at CVL

- Uses data from a single Kinect sensor
- Detects falls by tracking the height of persons

Person Detection
Applications – Entertainment (Kinect Player Pose Estimation)

https://www.youtube.com/watch?v=p2qlHoxPioM

Kinect Player Pose Estimation

Let's take a look at how this works

Assuming we have already detected the person

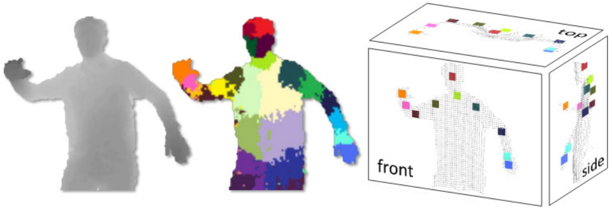


Image from Shotton et al. 2011

Estimate body part of each pixel independently

Perform clustering to obtain joint position proposals

Fit skeleton model to joint proposals

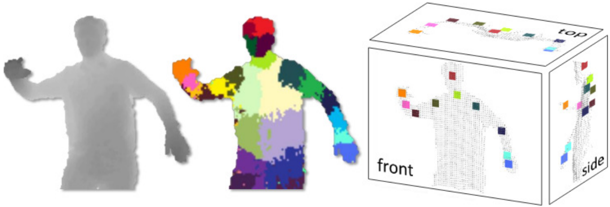


Image from Shotton et al. 2011

Kinect Player Pose Estimation
Pixel Classification

For each pixel $\mathbf{x}$ with depth $d(\mathbf{x})$ compute $\Pr(w|\mathbf{x})$

- With w representing the body part, $w \in \{0, \ldots, 30\}$
- Note that this is a discriminative model

Image from Shotton et al. 2011

Classification using simple depth offset features $f_{\boldsymbol{\theta}}$,

$$f_{\boldsymbol{\theta}=(\mathbf{u},\mathbf{v})}(\mathbf{x}) = d\left(\mathbf{x} + \frac{\mathbf{u}}{d(\mathbf{x})}\right) - d\left(\mathbf{x} + \frac{\mathbf{v}}{d(\mathbf{x})}\right)$$

Image adapted from Shotton et al. 2011

In isolation, one such feature $f_{\boldsymbol{\theta}_1}$ is weak

But a strong classifier can be built by combining several features

Method uses a forest of T **randomized decision trees**

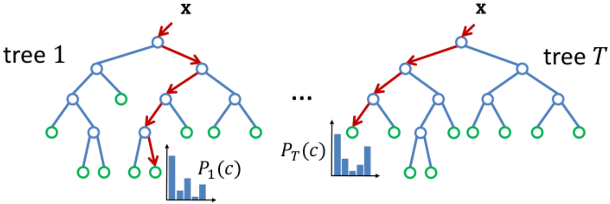


Image adapted from Shotton et al. 2011

Each tree t consist of split and leaf nodes

Each split node consists of a feature $f_{\boldsymbol{\theta}}$ and a threshold τ

- $\mathbf{x}$ branches down based on $f_{\boldsymbol{\theta}_k} > \tau_k$
- Until a leaf node is reached, which stores $\mathrm{Pr}_t(w|\mathbf{x})$

Tree trained from training samples $(\mathbf{x}, w)$

- $\boldsymbol{\theta}_k, \tau_k$ selected randomly (hence random tree)

All trees contribute to result, $P(w|\mathbf{x}) = \sum_{t=1}^{T} \mathrm{Pr}_t(w|\mathbf{x})$

Bibliography I

Aiger, Dror, Niloy J Mitra, and Daniel Cohen-Or (2008). **4-points congruent sets for robust pairwise surface registration**. ACM TOG.

Berger, Matthew et al. (2014). **State of the Art in Surface Reconstruction from Point Clouds**. Eurographics.

Curless, Brian and Marc Levoy (1996). **A volumetric method for building complex models from range images**. CGIT.

Kazhdan, Michael (2005). **Reconstruction of solid models from oriented point sets**. Eurographics.

Kazhdan, Michael, Matthew Bolitho, and Hugues Hoppe (2006). **Poisson surface reconstruction**. Eurographics.

Bibliography II

Levoy, Marc et al. (2000). **The digital Michelangelo project: 3D scanning of large statues**. CGIT.

Newcombe, Richard A et al. (2011). **KinectFusion: Real-time dense surface mapping and tracking**. ISMAR.

Shotton, Jamie et al. (2011). **Real-Time Human Pose Recognition in Parts from a Single Depth Image**. CVPR.