Computer vision class 2013

Object Representation II

Nov. 27. 2013
Bo Zheng
(zheng@cvl.iis.u-tokyo.ac.jp)

Outline

- 2D representation (for RGB image)
  - basics
  - research in the state of arts
- Sparse representation
  - basics
  - research in the state of arts
- 3D representation
  - basics
  - research in the state of arts
- Past & Future study on 3D vision

Last Class (Nov. 20): 2D representation, Sparse representation

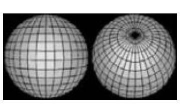
Today: 3D representation, Past & Future study on 3D vision

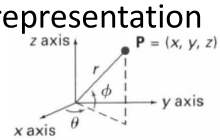
Basic techniques on 3D representation

Types

Continuity / Form		Discrete	Continuous
Parametric		Point cloud in polar coordinate…	Splines (piecewise polynomial),…
Nonparametric	Explicit	3D volumetric images, Polygon mesh,…	Explicit Polynomial…
	implicit	Signed Distance Field (SDF),…	Implicit Radial Basis Function & Algebraic surface…

Example: sphere representation





Parametric		$x = r\cos\phi\cos\theta,\quad -\pi/2 \le \phi \le \pi/2$ $y = r\cos\phi\sin\theta,\quad -\pi \le \theta \le \pi$ $z = r\sin\phi$
Nonparametric	Explicit	$z = f(x,y) = \pm\sqrt{r^2 - x^2 - y^2}$
	implicit	$f(x,y,z) = x^2 + y^2 + z^2 - r^2 = 0$

Examples of

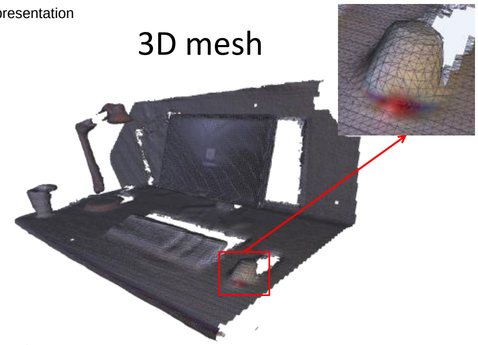
- Explicit representation
- Implicit representation
- Parametric representation

Explicit representation

3D mesh

Explicit representation

3D mesh



$G(V,E) \rightarrow V$: vertices, E: edges / triangles / polygons

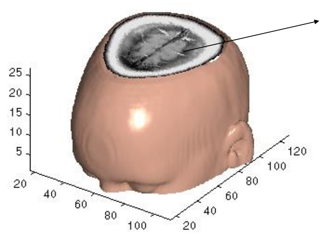
400,000 vertices and 130,000 triangles

Explicit representation

3D Volumetric Image

$I(x,y,z) = \text{intensity}$

Volumetric image



Brain MR image

Examples of

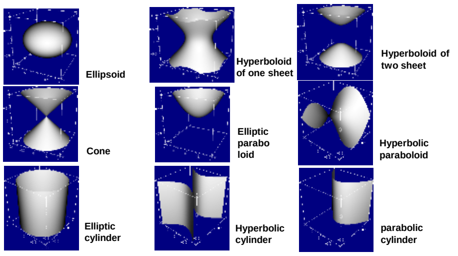
- Explicit representation
- Implicit representation
- Parametric representation

Implicit representation

Quadric Surface

- A polynomial of 2nd order

$ax^2 + by^2 + cz^2 + d + 2eyz + 2fzx + 2gxy + 2hx + 2iy + 2jz = 0$



Implicit Polynomial Surface with higher degree

formulation

$$f_n(x,y,z) = \sum_{i,j,k;i+j+k \le n} a_{ijk} x^i y^j z^k = 0$$

$x^3y + y^3z + z^3x = 0$ $\quad$ $x^4 + y^4 + z^4 - 200xyz = 0$

Explicit representation

3D 8-degree Polynomial

8-degree polynomial

Implicit Radial Basis Function (RBF)

$$f(\mathbf{x}) = v(\mathbf{x}) + \sum_{i=1}^{N} \lambda_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) = 0$$

Low degree polynomial

Radial basis (xi: control point)

e.g., $v(\mathbf{x})$

$$v(\mathbf{x}) = c_0 + c_1 x + c_2 y + c_3 z$$

e.g., $\phi(\mathbf{x})$

- Gaussian

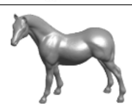
$$\phi(x) = e^{x^2/\sigma_i^2}$$

- Thin-plate radial basis

$$\phi(x) = x^2 \log(x)$$

Examples of RBF Surface

[Itoh, IEICE trans'06]

Results	Num. of points of Object	Num. of Con. Points
	35152	11717
	52251	17417
	64646	16161

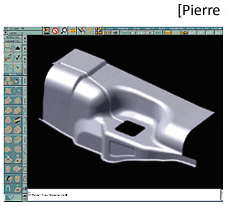
Examples of

- Explicit representation
- Implicit representation
- Parametric representation

Parametric curve/surface

- Manufactural design
- font（TrueType Font）
- Bézier curve/surface

[Pierre Bézier in 1966 for car design]

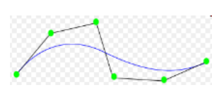


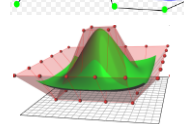
Non-uniform rational B-spline (NURBS)

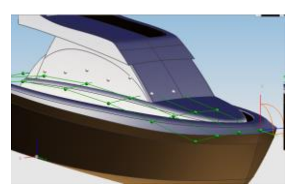
$$S(u,v) = \sum_i \sum_j B^{h}_{i,j} N_{i,k}(u) M_{i,l}(v)$$

Control points

Rational B-spline basis functions







computer-aided design (CAD)

A brief comparison

	Explicit	Parametric	Implicit
Data Interpolation	Difficult	Easy	Easiest
Smoothness	NO	Yes	Yes
Compact Representation	NO	Yes	Yes
Visualization	Easy	Easy	difficult
Surface Operations	Bad	Good	Very Good
Topology Preserving Deformation	Difficult	Easy	Easiest
Local Shape Control	Yes	Yes	No
Gradient Computation	Bad	Good	Good

Short survey

B. Zheng, J. Takamatsu and K. Ikeuchi (UT)
IEEE trans. on Pattern Recognition and Machine Intelligent (PAMI), 2010

Adaptively fitting implicit polynomials (IPs) to 2D/3D object shapes

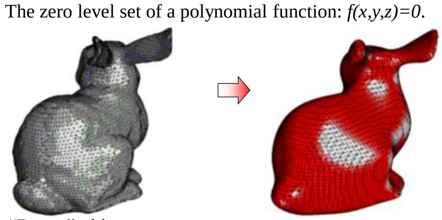
What is Implicit Polynomial (IP)?

3D Polynomial:

$$f_n(x,y,z) = \sum_{i,j,k;i+j+k\le n} a_{ijk} x^i y^j z^k = \mathbf{a}^{\mathrm{T}} \mathbf{m}_n(\mathbf{x})$$

IP surface:

The zero level set of a polynomial function: *f(x,y,z)=0*.

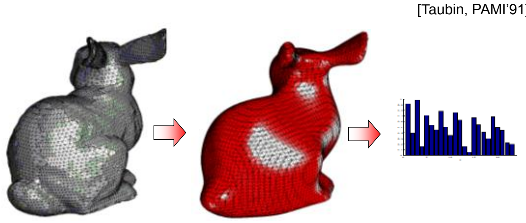


"Bunny" object

IP surface

Advantage1: Algebraic Invariants

[Taubin, PAMI'91]



Given Object: "Bunny"

IP representation

Invariants

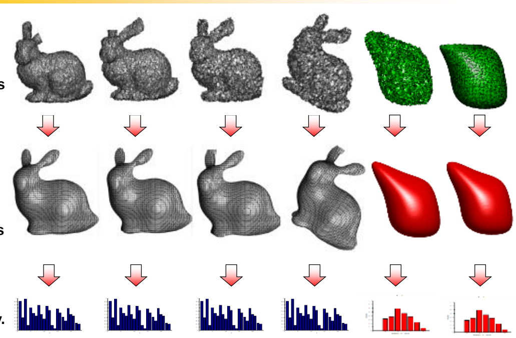
functions of the polynomial coefficients that do not change after the shape Euclidean transformed (rotated or translated).

Then what can we do?

objects

IPs

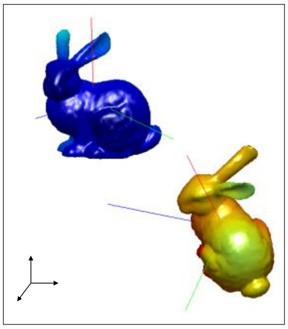
Inv.



Classify these objects

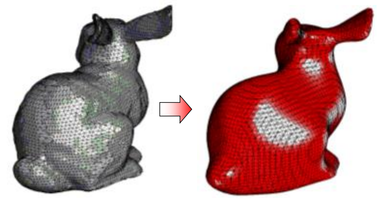
Advantage2: Pose estimation

[Taubin, PAMI'91]



Orientation (pose) of an object can be easily extracted.

How?



IP fitting method

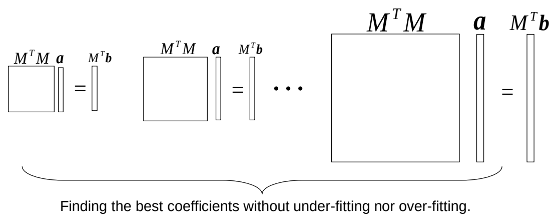
[Blane, PAMI'00]

$$\left.\begin{aligned} f_n(\boldsymbol{x}_1, a) &= b_1 \\ f_n(\boldsymbol{x}_2, a) &= b_2 \\ &\vdots \\ f_n(\boldsymbol{x}_N, a) &= b_N \end{aligned}\right\} \xrightarrow[\min_{a} \sum (\boldsymbol{m}(x_i)^T \boldsymbol{a} - b_i)^2]{\text{Linear LS Method}} M^T M \boldsymbol{a} = M^T \boldsymbol{b}$$

Given a polynomial for each point

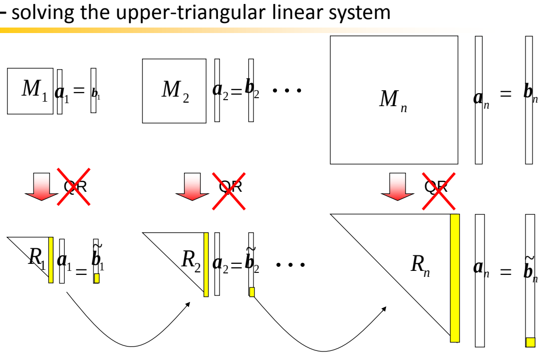
Obtain the coefficients $\boldsymbol{a}$ through solving this linear equations

A naive method for finding the moderate degree of IP

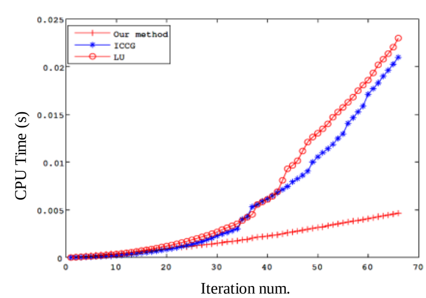


too time-consuming!

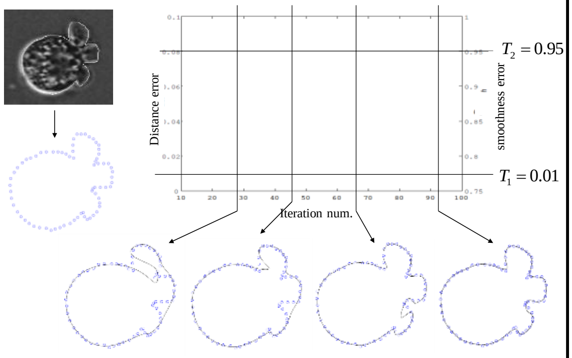
Incremental fitting
– solving the upper-triangular linear system



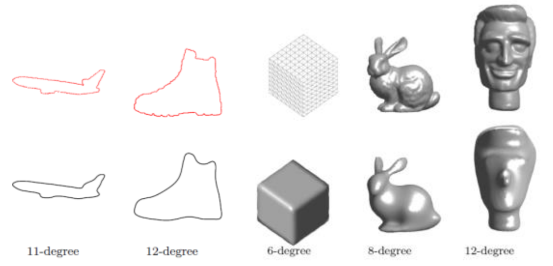
Computational efficiency



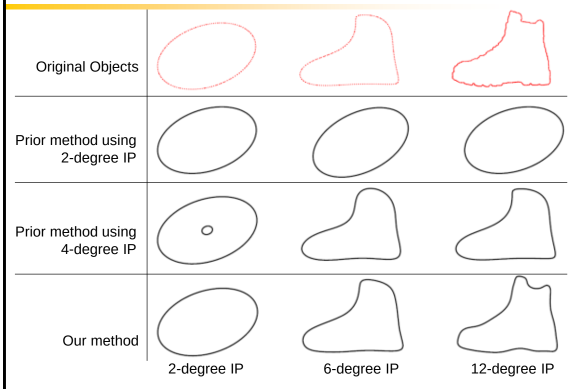
Example



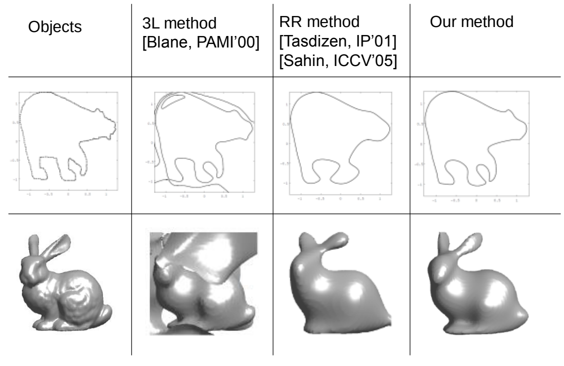
2D & 3D examples



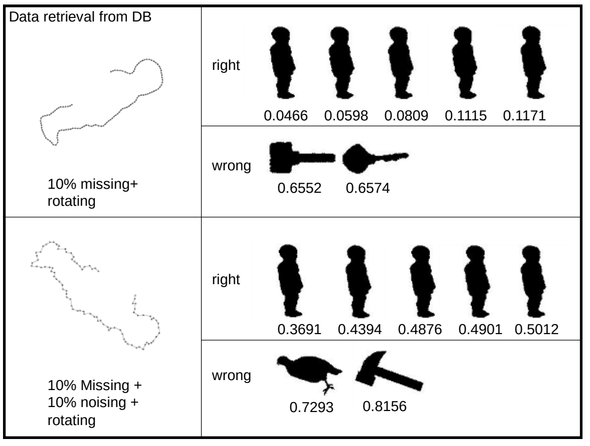
Comparison to degree-fixed method



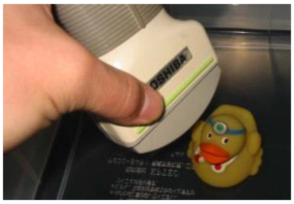
Comparison to prior methods



Applications



Example 2: ultrasound image

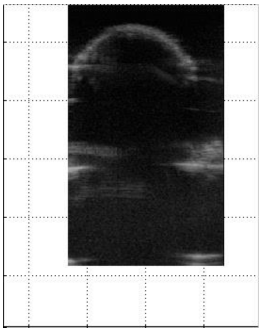


Ultrasound Scan

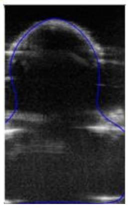
8-degree IP

37

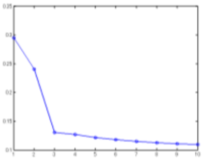
For one of the single frame



180 msec

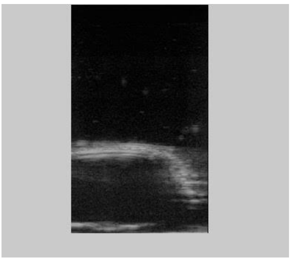


Cross section

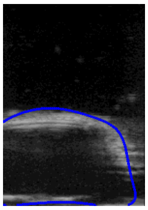


38

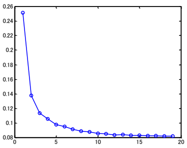
Another single frame



250 msec

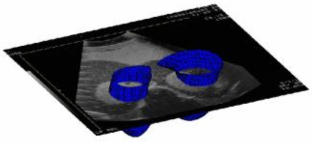


Cross section



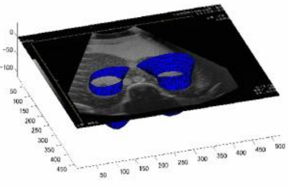
39

Pose Estimation



40

Tracking



41

Conclusion

- Adaptive IP fitting without under fitting nor over fitting.
- More globally stable and locally accurate

Outline

- 2D representation (for RGB image)
  – basics
  – research in the state of arts
- Sparse representation
  – basics
  – research in the state of arts

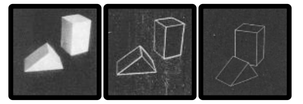
Last Class (Nov. 7)

- 3D representation
  – basics
  – research in the state of arts
- Past & Future study on 3D vision

Today

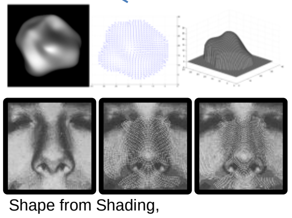
Past & future of 3D vision

1960s 1970s 1980s 1990s 2000s 2010s



Machine Perception of Three-Dimensional Solids,
Larry Roberts, PhD Thesis, MIT, 1963.

1960s 1970s 1980s 1990s 2000s 2010s



Shape from Shading,
Ikeuchi & Horn, MIT AI Memos 232, 1970.

1960s 1970s 1980s 1990s 2000s 2010s

Essential Matrix

$$X'_\mu Q_{\mu\nu} X_\nu = 0 \qquad (11)$$

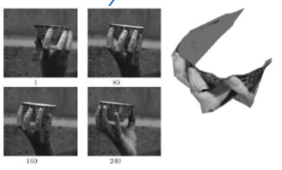
Dividing equation (11) by $X'_3 X_3$ we arrive at the desired relationship between the image coordinates:

$$x'_\mu Q_{\mu\nu} x_\nu = 0 \qquad (12)$$

3x3 Matrix mapping points to epipolar lines
- corresponding points x, x' satisfy x' Q x = 0
- camera matrices can be computed from Q

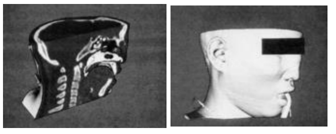
A computer algorithm for reconstructing a scene from two projections,
Longuet-Higgins, *Nature*, 1981.

1960s 1970s 1980s 1990s 2000s 2010s



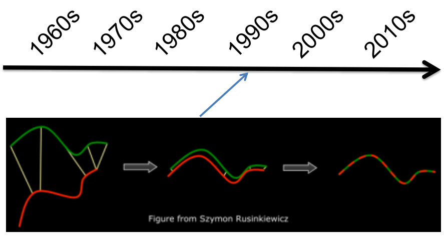
Structure-from-motion by factorization [Tomasi & Kanade, ICCV90]

1960s 1970s 1980s 1990s 2000s 2010s



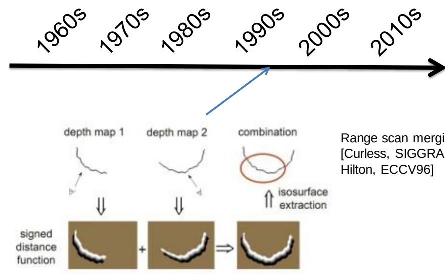
Marching Cubes
[Lorensen & Cline, SIGGRAPH' 87]

From Volume to Surface mesh
- Start at voxel containing surface
- Add polygon(s) based on configuration table
- earlier: 1970's Hummel & Zucker, 3D edge finding
- March to next voxel

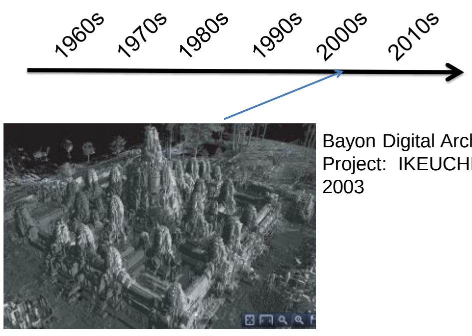


Iterative Closest Points (ICP)

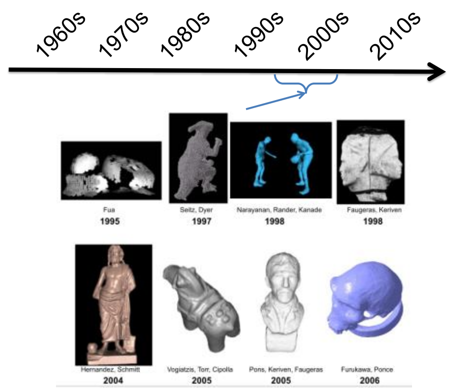
- Besl, McKay, "A Method for Registration of 3-D Shapes," PAMI 1992
- Chen, Medioni, "Object Modelling by Registration of Multiple Range Images," International Journal of Image and Vision Computing, 1992.
- Z. Zhang, Iterative point matching for registration of free-form curves, Research Report 1658, INRIA Sophia-Antipolis.
- T. Oishi, 3DIM 05

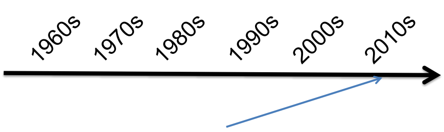


Range scan merging [Curless, SIGGRAPH96; Hilton, ECCV96]



Bayon Digital Archival Project: IKEUCHI Lab, 2003





City-scale SfM
- ~200K images, one day of processing
- 3 cities: Rome, Venice, Dubrovnik

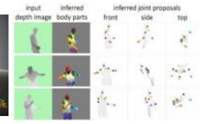
Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz and Richard Szeliski, "Building Rome in a Day," Communications of the ACM, Vol. 54, No. 10, Pages 105-112, October 2011.

Colosseum in Roma (2,106 images, 819,242 points)

Dubrovnik (4,619 images, 3,485,717 points)

1960s 1970s 1980s 1990s 2000s 2010s

2011: Kinect- Body pose from single depth image
• Fastest selling
Shotton, Fitzgibbon, Cook, Sharp, Finocchio, Moore, Kipman, Blake,
Real-Time Human Pose Recognition in Parts from a Single Depth Image, CVPR



1960s 1970s 1980s 1990s 2000s 2010s

Kinect Fusion: Microsoft research'2011
Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon, *KinectFusion: Real-Time Dense Surface Mapping and Tracking*, in *IEEE ISMAR*, IEEE, October 2011

1960s 1970s 1980s 1990s 2000s 2010s

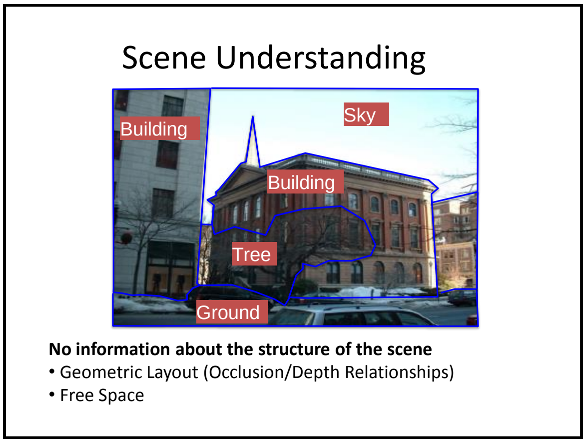
Blocks World Revisited:
[ECCV10 best paper awarded]
Abhinav Gupta, Alexei A. Efros and Martial Hebert, Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics, *European Conference on Computer Vision*, 2010.

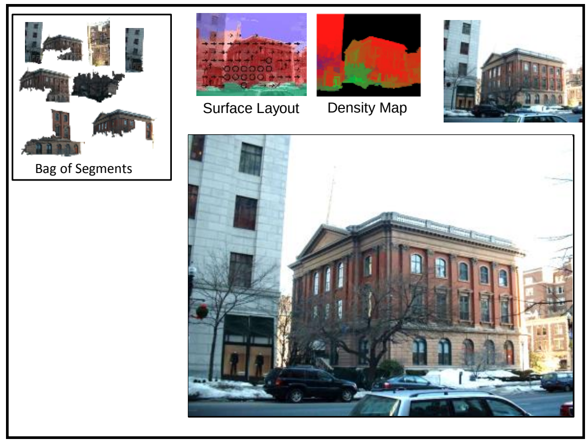
Goal

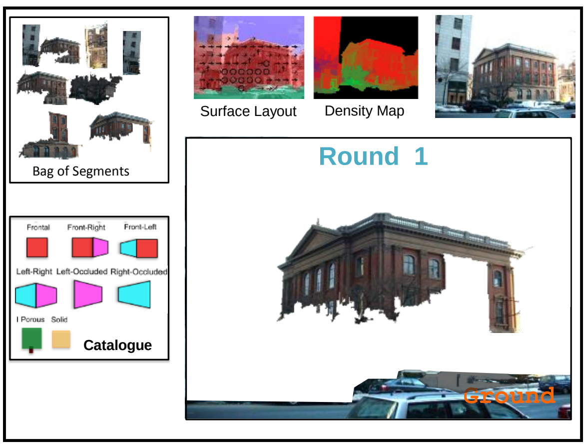
Hard Combinatorial Optimization

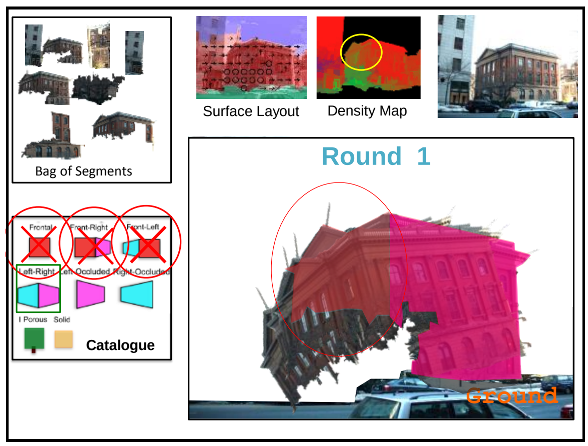
Blocks World Revisited:
Image Understanding Using Qualitative Geometry and Mechanics

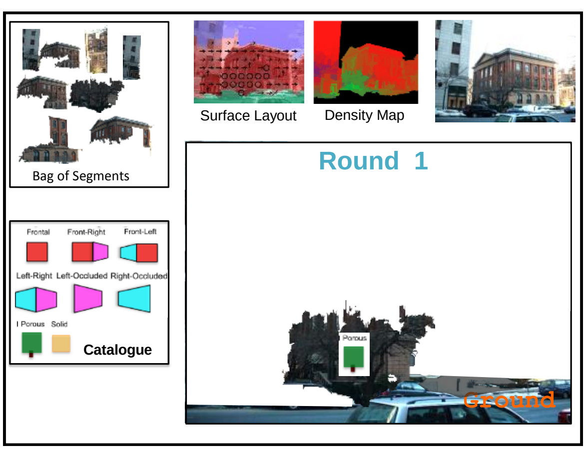
Abhinav Gupta, Alexei A. Efros, and Martial Hebert
Carnegie Mellon University

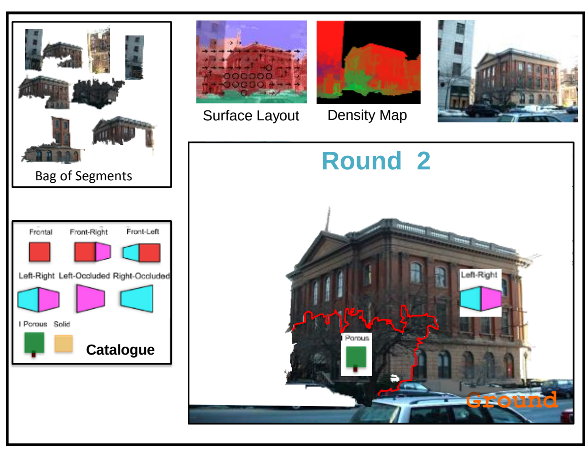

Scene Understanding
Building
Sky
Building
Tree
Ground
No information about the structure of the scene
• Geometric Layout (Occlusion/Depth Relationships)
• Free Space


Bag of Segments
Surface Layout
Density Map

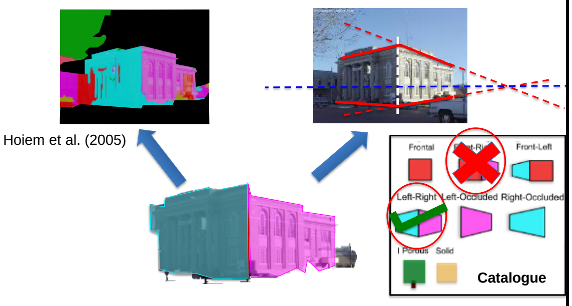

Bag of Segments
Surface Layout
Density Map
Round 1
Frontal
Front-Right
Front-Left
Left-Right
Left-Occluded
Right-Occluded
Porous
Solid
Catalogue
Ground


Bag of Segments
Surface Layout
Density Map
Round 1
Frontal
Front-Right
Front-Left
Left-Right
Left-Occluded
Right-Occluded
Porous
Solid
Catalogue
Ground


Bag of Segments
Surface Layout
Density Map
Round 1
Frontal
Front-Right
Front-Left
Left-Right
Left-Occluded
Right-Occluded
Porous
Solid
Catalogue
Porous
Ground


Bag of Segments
Surface Layout
Density Map
Round 2
Frontal
Front-Right
Front-Left
Left-Right
Left-Occluded
Right-Occluded
Porous
Solid
Catalogue
Left-Right
Porous
Ground

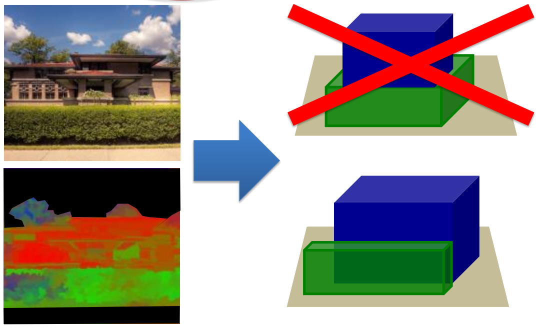
$$\mathcal{C}(\mathcal{B},\mathcal{S}) = \mathcal{F}_{geometry}() + \mathcal{F}_{contacts}() + \mathcal{F}_{intra}() + \mathcal{F}_{stability}() + \mathcal{F}_{depth}(),$$



$$\mathcal{C}(\mathcal{B},\mathcal{S}) = \mathcal{F}_{geometry}() + \mathcal{F}_{contacts}() + \mathcal{F}_{intra}() + \mathcal{F}_{stability}() + \mathcal{F}_{depth}(),$$

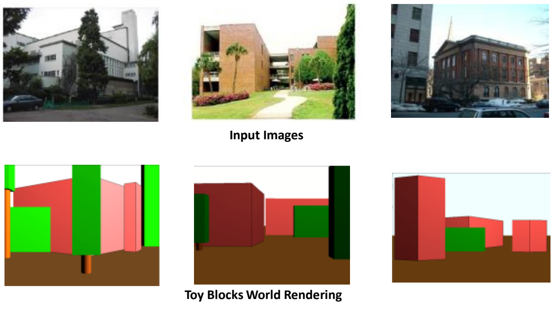
Static and Physically Stable World

$$\mathcal{C}(\mathcal{B},\mathcal{S}) = \mathcal{F}_{geometry}() + \mathcal{F}_{contacts}() + \mathcal{F}_{intra}() + \mathcal{F}_{stability}() + \mathcal{F}_{depth}(),$$

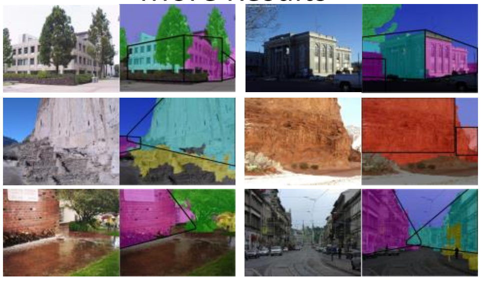


Fitting Cuboids

Building 3D Blocks World

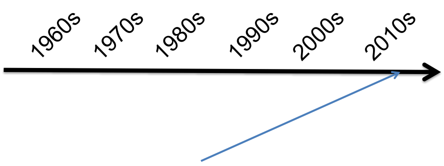


More Results



All results and preliminary version of code (Coming Soon):
http://www.cs.cmu.edu/~abhinavg/blocksworld

Past & future of 3D vision



Reconstructing the Museums:
[ECCV12 Best Student Paper Award]
Jianxiong Xiao and Yasutaka Furukawa
Reconstructing the World's Museums

The Goal

- Global texture-mapped 3D model
- Optimize for aerial viewing
- Enable effective indoor navigation

System Pipeline

1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints

System Pipeline

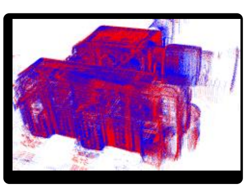
1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints

Art Project

www.GoogleArtProject.com

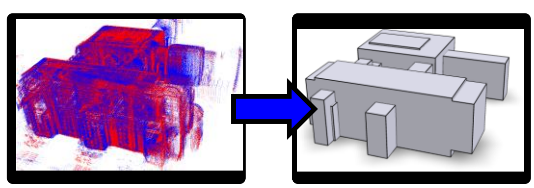
System Pipeline

1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints



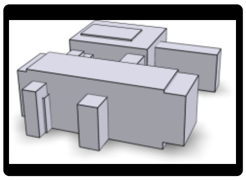
System Pipeline

1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints



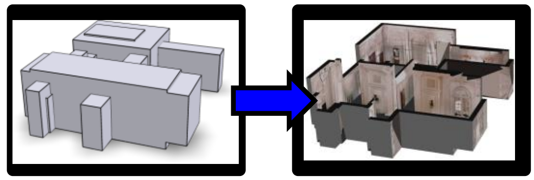
System Pipeline

1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints



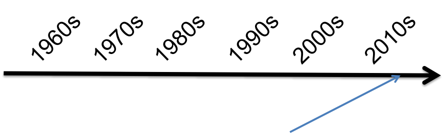
System Pipeline

1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints



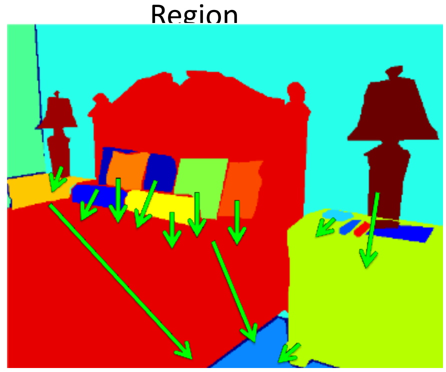
Results

Effective

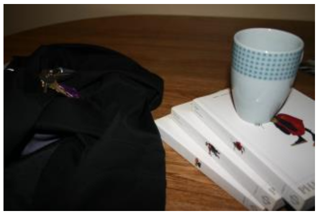


Physical relation:
[ECCV12 oral paper]
Nathan Silberman, Derek Hoiem, Pushmeet Kohli, Rob Fergus, "Indoor Segmentation and Support Inference from RGBD Images", ECCV 2012

Goal: Infer Support for Every Region

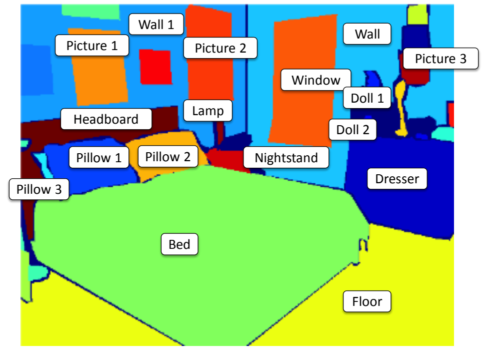


Why infer physical support?

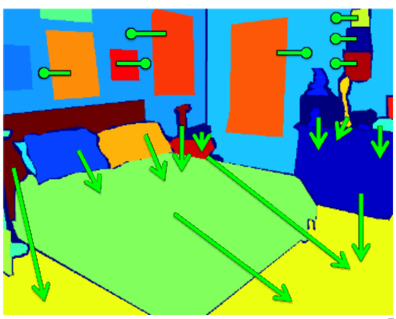


Interacting with objects may have physical consequences!

High Quality Semantic Labels



High Quality Support Labels

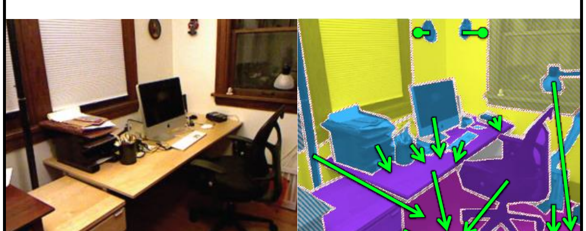


Support from below   Support from behind   Support from hidden

Experiments

Results

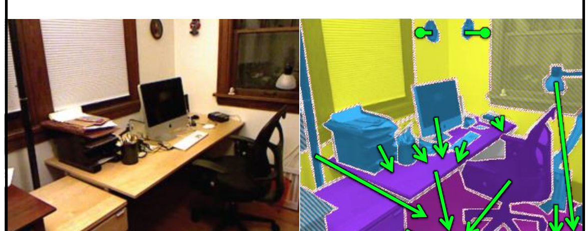
Ground Truth Regions



Floor   Prop
Furniture   Structure

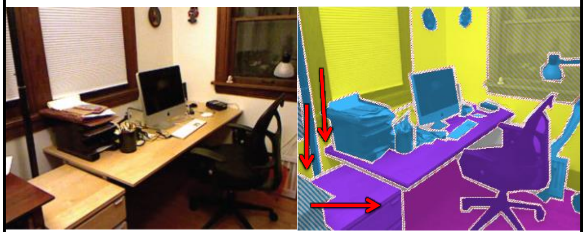
Results

Ground Truth Regions



Correct Prediction

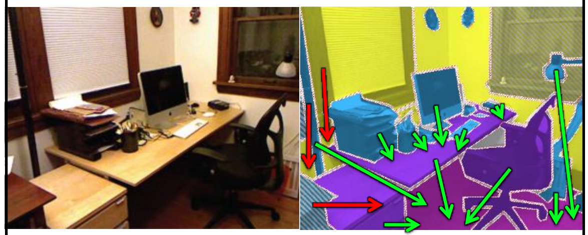
Results
Ground Truth Regions



Correct Prediction

Incorrect Prediction

Results
Ground Truth Regions

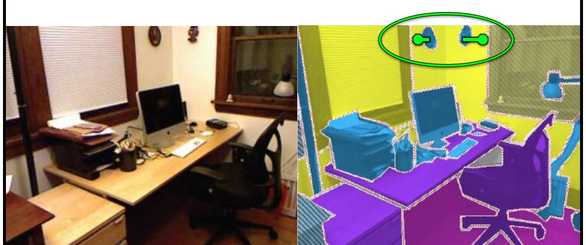


Correct Prediction

Incorrect Prediction

Support from below

Results
Ground Truth Regions



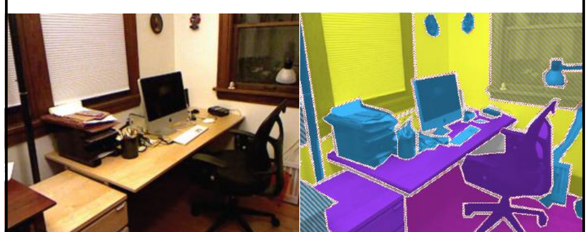
Correct Prediction

Incorrect Prediction

Support from below

Support from behind

Results
Ground Truth Regions



Correct Prediction

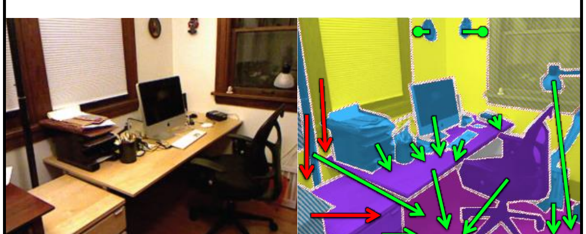
Incorrect Prediction

Support from below

Support from behind

Support from hidden region

Results
Ground Truth Regions



Correct Prediction

Incorrect Prediction

Support from below

Support from behind

Support from hidden region

Results
Ground Truth Regions

Correct Prediction

Incorrect Prediction

Support from below

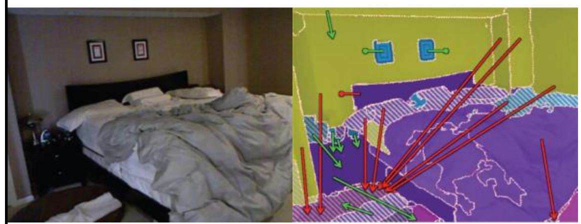
Support from behind

Support from hidden region

Results
Automatically Segmented Regions

Correct Prediction
Incorrect Prediction
Support from below
Support from behind
Support from hidden region

Results
Automatically Segmented Regions



Correct Prediction
Incorrect Prediction
Support from below
Support from behind
Support from hidden region

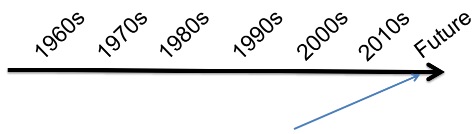
Conclusion

- Algorithm for inferring Physical Support
- Novel Integer Program Formulation
- 3D Cues for segmentation

Dataset:
- http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

Code:
- http://cs.nyu.edu/~silberman/projects/indoor_scene_seg_sup.html

Past & future of 3D vision



Dr. Steve Seitze: [Google talks, 2012]
2013: Digital Michelangelo from a few photos
2015: Models of everything
2020: Inverse CAD
2030: computer > human

3D vision

Signals (raw data) processing — *e.g. Denoising*
past ↓
Information — *e.g. Feature detection and description*
past ↓
Knowledge — *e.g. Examplar-based recognition*
past ↓
Cognition (future) — *e.g. Reasoning by various knowledge*

Beyond Point Clouds: Scene Understanding by Reasoning Geometry and Physics

B. Zheng[1], Y. Zhao[2], Joey. C. Yu[2], K. Ikeuchi[1], & S. –C. Zhu[2]

1) The University of Tokyo, Japan, 2) University of California, Los Angeles,U

Two observations

- The world can be represented by voxels (volumetric pixels).
- Mechanics is an important cue for reasoning the objects in a static scene.

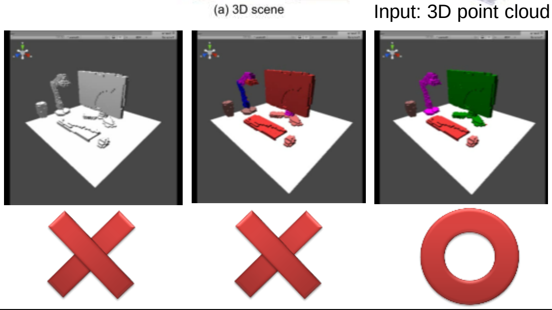
Gravity

- The useful information for scene understanding.

Our goal


(a) 3D scene


Input: 3D point cloud



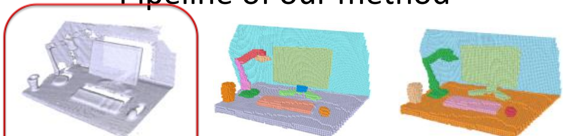
Related work

- Geometric methods
  - 3D segmentation [Attene, VC06]
  - Manhattan assumption [Furukawa, CVPR09]
- Physics reasoning
  - "Block world revisit" [Gupta, ECCV10]
  - Support relations inference [Silberman, ECC12]
- Cognitive science
  - Probabilistic representation [Hamrick, CogSc11]
- Physics engine?

Our contribution

- Geometric reasoning
  - Segmentation + volumetric completion
    (2.5D -> volumetric)
- Physical reasoning
  - novel model of intuitive physical stability
  - A novel stability optimization

Pipeline of our method

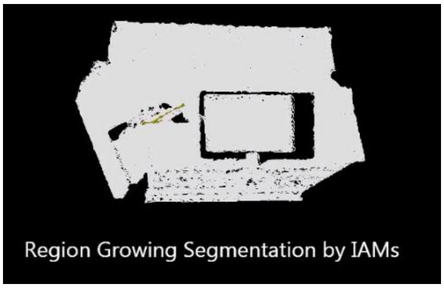


3D point cloud → Geometric reasoning → Physical reasoning

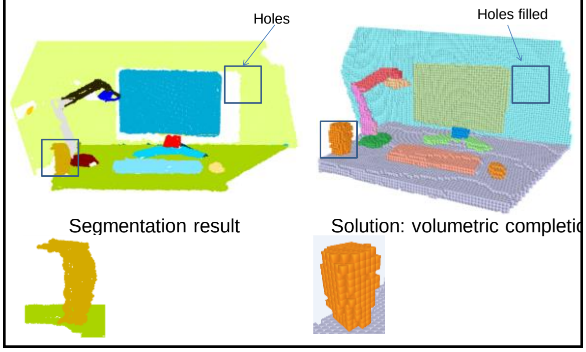
Geometric reasoning:
- 3D segmentation
- volumetric completion

Physical reasoning:
- stability maximization

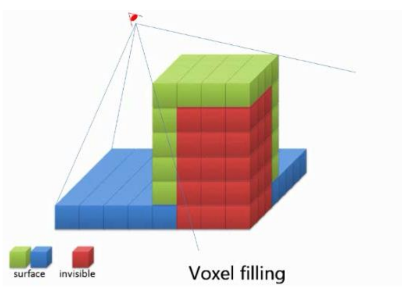
Region growing segmentation & convex connection merging



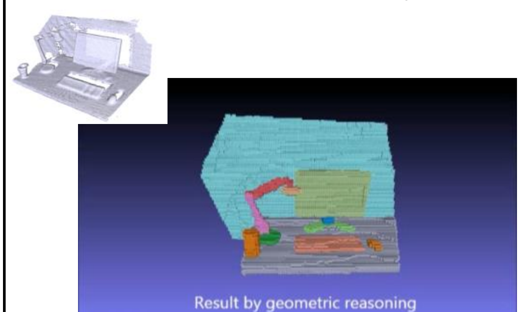
Current issue



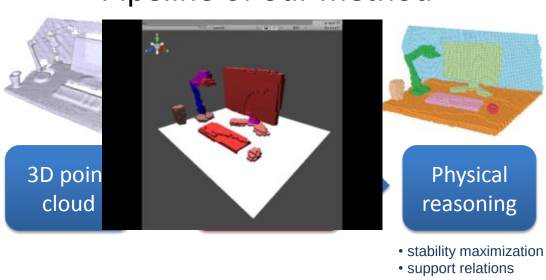
Volumetric completion



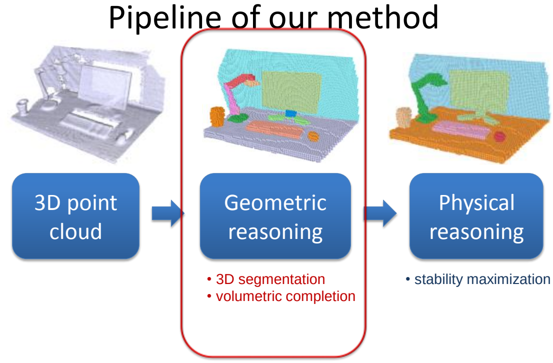
Result of volumetric completion



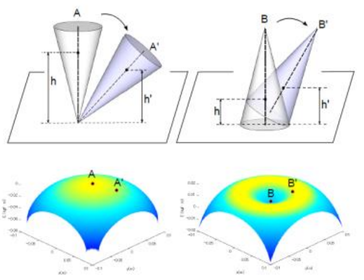
Pipeline of our method



Pipeline of our method

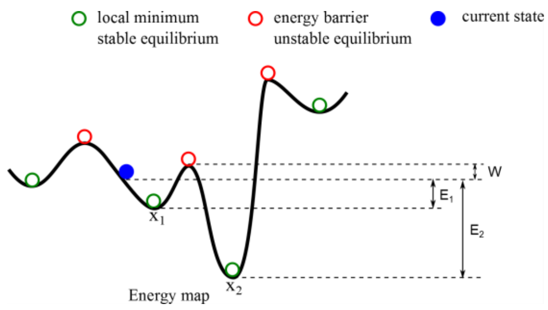


Object Stability



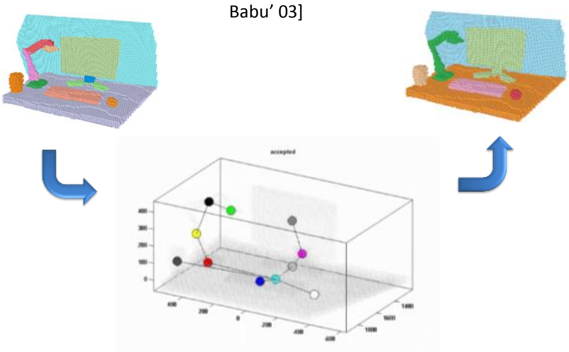
"unstable" "stable" – local minimum

Definition of stability



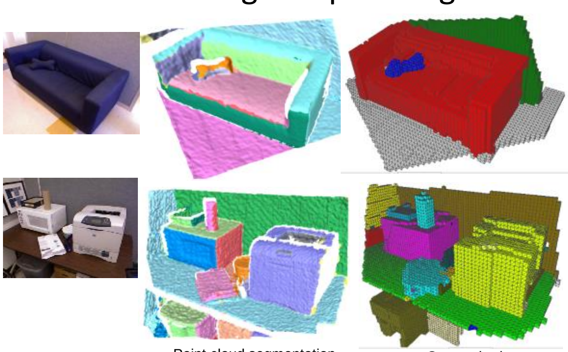
Given small energy, The less energy released, the stal

Swendsen-wang cut (SWC) [A. Babu' 03]



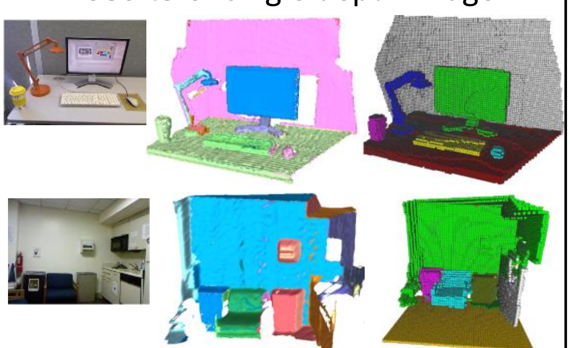
Experimental result

Results of single depth image



Point cloud segmentation Our method

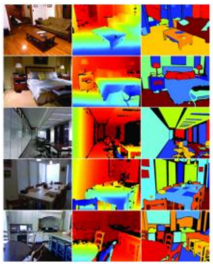
Results of single depth image

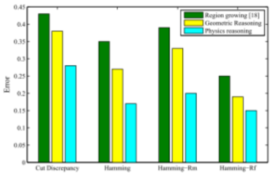


Point cloud segmentation Our method

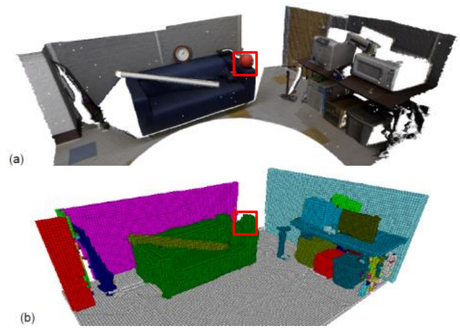
Segmentation comparison

- NYU dataset v2 (1449 labeled depth images)

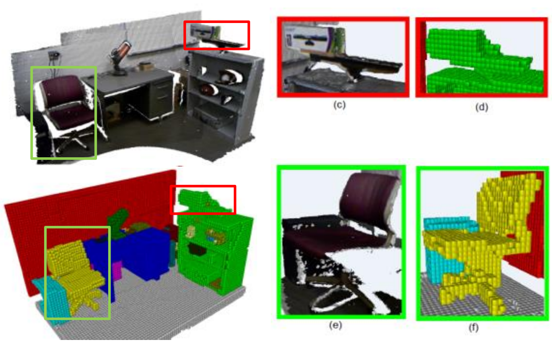




Large scale indoor scene

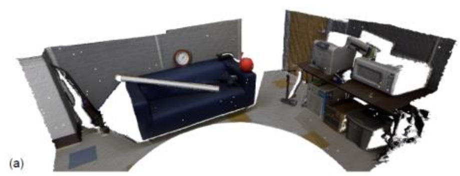


Large scale indoor scene



Precision of physical relation inference

- Dataset (15 labeled indoor scene data)

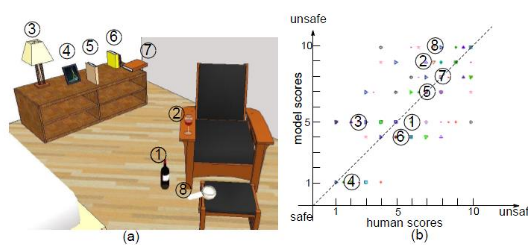


relations	Discriminative	Greedy	SWC
fixed joint	20.5%	66%	**81.8%**
support	42.2%	60.3%	**78.1%**

Summary

- Geometric reasoning
  - Segmentation + volumetric completion
    (2.5D -> volumetric)
- Physical reasoning
  - novel model of intuitive physical stability
  - A novel stability optimization

machine v.s. human



3D vision

Signals (raw data) processing — *e.g. Denoising*

past ↓

Information — *e.g. Feature detection and description*

past ↓

Knowledge — *e.g. Examplar-based recognition*

past ↓

Cognition (in future) — *e.g. Reasoning by various knowledge*

Machine > Human in 2030s?

Thank you for your listening!

Notice: Classroom change

Dec 4 (16:40-18:10) → E.Bld.2 Room #221

12月4日(5限) → 工2号館221号講義室