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
- 3D vision: Beyond the “what is where”

Last Class

Today

Object Representation II

Oct. 28, 2014
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
• 3D vision: Beyond the “what is where”

Today

Basic techniques on 3D representation

Types

<table>
<thead>
<tr>
<th>Form</th>
<th>Continuity</th>
<th>Discrete</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric</td>
<td></td>
<td>Point cloud in polar coordinate...</td>
<td>Splines (piecewise polynomial),...</td>
</tr>
<tr>
<td>Nonparametric</td>
<td>Explicit</td>
<td>3D volumetric images, Polygon mesh,...</td>
<td>Explicit Polynomial,...</td>
</tr>
<tr>
<td></td>
<td>implicit</td>
<td>Signed Distance Field (SDF),...</td>
<td>Implicit Radial Basis Function &amp; Algebraic surface...</td>
</tr>
</tbody>
</table>

Example: sphere representation

Example 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

Brain MR image

Examples of

• Explicit representation
• Implicit representation
• Parametric representation

Explicit representation

3D Volumetric Image

Explicit representation

Quadric Surface

• A polynomial of 2nd order

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

Implicit representation

Quadric Surface

Implicit Polynomial Surface with higher degree

$$f_q(x, y, z) = \sum_{i,j,k} a_{ijk} x^i y^j z^k = 0$$

$${x}^3 {y} + {y}^3 {z} + {z}^3 {x} = 0 \quad {x}^4 + {y}^4 + {z}^4 - 200xyz = 0$$
Explicit representation

3D 8-degree Polynomial

\[ f(x) = v(x) + \sum_{i=1}^{N} \lambda_i \phi\left(\|x - x_i\|\right) = 0 \]

Low degree polynomial

Radial basis (xi: control point)

\[ \phi(x) = e^{-\|x-x_i\|^2/\sigma^2} \]

Examples of

- Explicit representation
- Implicit representation
- Parametric representation

Shape representation - using RBF basis

[Car et al. (SIGGRAPH 01)]

544,000 point cloud

8000 control points

Parametric curve/surface

- Manufacturing design
- Font (TrueType Font)
- Bézier curve/surface

[Car et al. (SIGGRAPH 01]

Non-uniform rational B-spline (NURBS)

\[ S(u, v) = \sum_i \sum_j B_{ij}^h N_{i,k}(u) M_{j,l}(v) \]

Control points
Rational B-spline basis functions

computer-aided design (CAD)
A brief comparison

<table>
<thead>
<tr>
<th></th>
<th>Explicit</th>
<th>Parametric</th>
<th>Implicit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Interpolation</td>
<td>Difficult</td>
<td>Easy</td>
<td>Easiest</td>
</tr>
<tr>
<td>Smoothness</td>
<td>NO</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Compact Representation</td>
<td>NO</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Visualization</td>
<td>Easy</td>
<td>Easy</td>
<td>difficult</td>
</tr>
<tr>
<td>Surface Operations</td>
<td>Bad</td>
<td>Good</td>
<td>Very Good</td>
</tr>
<tr>
<td>Topology Preserving Deformation</td>
<td>Difficult</td>
<td>Easy</td>
<td>Easiest</td>
</tr>
<tr>
<td>Local Shape Control</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Gradient Computation</td>
<td>Bad</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>

A Brief Introduction on Implicit Polynomial (IP)

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:

\[ f_x(x,y,z) = \sum_{i,j,k} a_{ijk} x^i y^j z^k = a^T m(x) \]

IP surface:

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

“Bunny” object

IP surface

Then what can we do?

Advantage 1: 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).
Orientation (pose) of an object can be easily extracted.

How?

Given an Object

IP representation

Advantage 2: Pose estimation

[Taubin, PAMI’91]

IP fitting method

[Blanco, PAMI’00]

A naive method for finding the moderate degree of IP

Finding the best coefficients without under-fitting nor over-fitting.

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

Conclusion

Source code: http://www.cvl.iis.u-tokyo.ac.jp/~zheng

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

→ 3D vision: brief introduction

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

Last Class (Nov. 7)

Today
Past & future of 3D vision

Input
Gradient
Output


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

Essential Matrix

\[
X^t Q X = 0
\]


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

From Volume to Surface mesh
- Start at voxel containing surface
- Add polygon(s) based on configuration table
- March to next voxel

Marching Cubes [Lorensen & Cline, SIGGRAPH 87]

Iterative Closest Points (ICP)
- 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

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

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

Blocks World Revisited:

Scene Understanding
- No information about the structure of the scene
  - Geometric Layout (Occlusion/Depth Relationships)
  - Free Space
$C(B,S) = F_{\text{geometry}} + F_{\text{contacts}} + F_{\text{intra}}$

$+ F_{\text{stability}} + F_{\text{depth}},$

Static and Physically Stable World

$C(B,S) = F_{\text{geometry}} + F_{\text{contacts}} + F_{\text{intra}}$

$+ F_{\text{stability}} + F_{\text{depth}},$

Fitting Cuboids

Building 3D Blocks World

Input Images

Toy Blocks World Rendering

More Results

Past & future of 3D vision

Reconstructing the Museums:
[ECCV'12 Best Student Paper Award]
Jianxiong Xiao and Yasutaka Furukawa
Reconstructing the World’s Museums

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

www.GoogleArtProject.com
System Pipeline
1. Take pictures inside the rooms
2. Reconstruct the 3D shape
3. Render from aerial viewpoints

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

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

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

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

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

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

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

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

Past & future of 3D vision


CVPR workshop on Vision meets Cognition (FPIC2014): Beyond the “what is where”

Y.-B. Zhao (UCLA), Craig Yu (MIT), B. Zheng (U. Tokyo), Tao (MIT), Peter (MIT)

Beyond “what is where”

Beyond “what is where”

CVPR2013

Beyond Point Clouds: Scene Understanding by Reasoning Geometry and Physics

Beyond Point Clouds: Scene Understanding by Reasoning Geometry and Physics

B. Zheng\textsuperscript{1}, Y. Zhao\textsuperscript{2}, Joey. C. Yu\textsuperscript{2}, K. Ikeuchi\textsuperscript{1}, & S. –C. Zhu\textsuperscript{2}

Two observations

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

Gravity

- The useful information for scene understanding.

Our goal

Our goal

(a) 3D scene

Input: 3D point cloud

- The useful information for scene understanding.
### 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

1. 3D point cloud
2. Geometric reasoning
   - 3D segmentation
   - Volumetric completion
3. Physical reasoning
   - Stability maximization

### Region growing segmentation & convex connection merging

### Current issue

- Segmentation result
- Solution: volumetric completion

### Volumetric completion

- Voxel filling
Result of volumetric completion

Geometric reasoning

Physical reasoning

Pipeline of our method

3D point cloud ➔ Geometric reasoning ➔ Physical reasoning
- 3D segmentation
- volumetric completion
- stability maximization

Object Stability

"unstable"  "stable" – local minimum

Definition of stability

Given small energy, the less energy released, the more stable

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

Experimental result
Results of single depth image

Segmentation comparison
• NYU dataset v2 (1449 labeled depth images)

Large scale indoor scene

Precision of physical relation inference
• Dataset (15 labeled indoor scene data)
Summary

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

Scene Understanding:
Potential Falling Risk for Objects
by Inferring Human Action and Natural Disturbance

Goal
understand the potential falling
objects

“Oh, it’s dangerous!”

Motivations

- safety surveillance robot,
  - children, elders and people with disabilities
- Robotics – rescue
  - DARPA robotics Challenge
- Human can imagine but machine cannot.
- Doing the serious physical simulation?
  - various collisions
  - large number of objects
  - huge variation in size, shape, material
- Time consuming!

Observation – causality of the falling risk

- “Cause” – the physical disturbance (energy absorbed)
- “Result” – much uncontrolled energy released
1. Physical reasoning

2. Disturbance field

Pipeline

Risk evaluation

Physical reasoning

Physical risk definition

Example

Disturbance field by human activities
Other disturbances

- human activity
- wind
- earthquake

Discussion: Human v.s. Machine?

- There is no ground truth
- People have big variance on safety understanding

3D vision

Signals (raw data)

processing

past

Information

past

Knowledge

past

Cognition

(in future)

e.g. Denoising

e.g. Feature detection and description

e.g. Examplar-based recognition

e.g. Reasoning by various knowledge

Machine > Human in 2030s?