My goal is to build intelligent visual machines that can perceive and understand the 3D world.
Humans perceive the world in 3D

Biederman, Mezzanotte and Rabinowitz, 1982
Objects are constrained by the 3D space

The 3D space is shaped by its objects

Modeling this interplay is critical for 3D perception!
Visual processing in the brain

where pathway (dorsal stream)

what pathway (ventral stream)
Visual processing in the brain

where pathway (dorsal stream)

Pre-frontal cortex

what pathway (ventral stream)

V1
Current state of computer vision

3D Reconstruction
- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

2D Recognition
- Object detection
- Texture classification
- Target tracking
- Activity recognition
Current state of computer vision

3D Reconstruction

- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

References:
- Lucas & Kanade, 81
- Chen & Medioni, 92
- Debevec et al., 96
- Levoy & Hanrahan, 96
- Fitzgibbon & Zisserman, 98
- Triggs et al., 99
- Pollefeys et al., 99
- Kutulakos & Seitz, 99
- Levoy et al., 00
- Hartley & Zisserman, 00
- Dellaert et al., 00
- Rusinkiewic et al., 02
- Nistér, 04
- Brown & Lowe, 04
- Schindler et al., 04
- Lourakis & Argyros, 04
- Colombo et al., 05
- Golparvar-Fard, et al. JAEI 10
- Pandey et al. IFAC , 2010
- Pandey et al. ICRA 2011
- Savarese et al. IJCV 05
- Savarese et al. IJCV 06
- Microsoft’s PhotoSynth
- Snavely et al., 06-08
- Schindler et al., 08
- Agarwal et al., 09
- Frahm et al., 10
Current state of computer vision

3D Reconstruction

- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

Lucas & Kanade, 81
Chen & Medioni, 92
Debevec et al., 96
Levoy & Hanrahan, 96
Fitzgibbon & Zisserman, 98
Triggs et al., 99
Pollefeys et al., 99
Kutulakos & Seitz, 99
Levoy et al., 00
Hartley & Zisserman, 00
Dellaert et al., 00
Rusinkiewic et al., 02
Nistér, 04
Brown & Lowe, 04
Schindler et al, 04
Lourakis & Argyros, 04
Colombo et al. 05
Golparvar-Fard, et al. JAEI 10
Pandey et al. IFAC, 2010
Pandey et al. ICRA 2011
Savarese et al. IJCV 05
Savarese et al. IJCV 06
Microsoft’s PhotoSynth
Snavely et al., 06-08
Schindler et al., 08
Agarwal et al., 09
Frahm et al., 10
Current state of computer vision

2D Recognition

- Object detection
- Texture classification
- Target tracking
- Activity recognition

Turk & Pentland, 91
Poggio et al., 93
Belhumeur et al., 97
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Ramanan & Forsyth, 03
Weber et al., 00
Vidal-Naquet & Ullman 02
Fergus et al., 03
Torralba et al., 03
Vogel & Schiele, 03
Barnard et al., 03
Fei-Fei et al., 04
Kumar & Hebert ’04
He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09
Current state of computer vision

2D Recognition

- Object detection
- Texture classification
- Target tracking
- Activity recognition

Turk & Pentland, 91
Poggio et al., 93
Belhumeur et al., 97
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Ramanan & Forsyth, 03
Weber et al., 00
Vidal-Naquet & Ullman 02
Fergus et al., 03
Torralba et al., 03
Vogel & Schiele, 03
Barnard et al., 03
Fei-Fei et al., 04
Kumar & Hebert 04
He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09
Current state of computer vision

Perceiving the World in 3D

- 3D Object detection
- 3D Scene Understanding
Outline

• 3D Object detection
• 3D Scene Understanding
Modeling objects and their 3D properties
State of the art object detection

“Car” model

Too many missed detections!
Some false alarms

Turk & Pentland, 91
Poggio et al., 93
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Vasconcelos ’00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Weber et al., 00
Fergus et al., 03
Torralba et al., 03
Fei-Fei et al., 04
Leibe et al., 04
Dalal & Triggs, 05
Savarese et al., CVPR 06
Felzenszwalb et al., 08
Lampert et al., 09
State of the art object detection

"Car" model

mixture model

Too many false alarms
State of the art object detection

“Car” model

mixture model

Too many false alarms
Poor scalability
No 3D properties
3D object detection

“Car” model

- Few missed detection and false alarms
- Estimate 3D pose & distance from camera

Savarese et al., ICCV 07
Su et al., ICCV 2009
Sun, et al., CVPR 2009
Yu & Savarese, CVPR 2012

- Thomas et al. ’06-09
- Yan et al., ’07
- Kushal et al., ’07
- Hoiem et al., 07
- Chiu et al ’07
- Liebelt et al 08, 10
- Xiao et al 08
- Arie-Nachimson & Barsi ’09
- Sandhu et al ’09
- Farhadi ’09
- Zhu et al. ’09
- Ozuysal et al. ‘10
- Stark et al.’10
- Payet & Todorovic, 11
- Glasner et al., ’11
- Zia et al. 11
- Pepik et al. ‘12
3D object representation

- Object is represented by a collection of parts
- Parts relationship are learnt from training images
- Inference by a novel algorithm based on variational EM
- Part configuration is modeled as a 3D conditional random fields

Savarese et al., ICCV 2007
Su et al., ICCV 2009
Sun, et al., CVPR 2009
Yu & Savarese, CVPR 2012
Results

<table>
<thead>
<tr>
<th>Object</th>
<th>a</th>
<th>e</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR</td>
<td>330</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>MOUSE</td>
<td>300</td>
<td>45</td>
<td>23</td>
</tr>
<tr>
<td>SHOE</td>
<td>240</td>
<td>45</td>
<td>11</td>
</tr>
</tbody>
</table>
Results

ImageNet dataset [Deng et al. 2010]

CHAIR  a=0  e=30  d=7

TABLE  a=60  e=15  d=2

SOFA   a=345  e=15  d=3.5
       a=60  e=30  d=2.5

BED    a=30  e=15  d=2.5
Results
Examples of failure (wrong category)

3D object dataset [Savarese & Fei-Fei 07]

This can’t be a shoe!
Outline

• 3D Object detection
• 3D Scene Understanding
Scene understanding is an interplay between objects and space
3D space is shaped by its objects
Objects are placed into 3D space
Interplay between objects and space

Interactions between:
- Objects-space
- Object-object

Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007
Desai et al, 2009
Sadeghi & Farhardi, 2011
Li et al, 2012
Hoiem et al, 2006
Herdau et al., 2009
Gupta et al, 2010
Fouhey et al, 2012
A 3DGP encodes **geometric** and **semantic** relationships between groups of objects and space elements which frequently co-occur in **spatially consistent configurations**.
3D Geometric Phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13

- W/o annotations
- Compact
- View-invariant

Using Max-Margin learning w/ novel Latent Completion algorithm
Results

Sofa, Coffee Table, Chair, Bed, Dining Table, Side Table

Estimated Layout

3D Geometric Phrases
Results

Sofa, Coffee Table, Chair, Bed, Dining Table, Side Table

Estimated Layout

3D Geometric Phrases
Scene understanding from images

- Interaction between object-space
- Interaction among objects
Scene understanding from images

- Interaction between object-space
- Interaction among objects
- Transfer semantics across views

Bao & S. Savarese, CVPR 2011
Bao, Bagra, Savarese. CORP – ICCV 2011 (Best student paper award!)
Bao, Bagra, Chao, Savarese, CVPR 2012
Bao, Xiang, Savarese, ECCV 2012
Results

Input images
Results

Input images
Results

Input images

From the office dataset [Bao et al., 11]
Results

Input images

From the office dataset [Bao et al., 11]
From objects to activities

Choi & Savarese, ECCV 2010
Choi, Pantofaru, Savarese, CORP 2011
Choi, Pantofaru, Savarese, PAMI 2013
Choi et al., VSWS 09
Choi et al., CVPR 11
Choi & Savarese, ECCV 2012 (oral)
X: Crossing, S: Waiting, Q: Queuing, W: Walking, T: Talking, D: Dancing
Conclusions

- From images to the 3D physical world
- Interplay between space and semantics
Applications

- Vision for the blinds
- Mobile vision
- Construction monitoring
- Safe driving

Diagram showing:
- Objects understanding
- Space understanding
- Activity understanding
New intelligent digital interfaces between us and the 3D world
New generation of autonomous agents that can operate safely alongside humans in dynamic environments
Automatizing large scale information and environmental management tasks
Computer vision meets civil engineering
It came in:

- At 5 times the estimated original cost!
- 6 years behind schedule
Volume of construction industry in US

Loss: ~10 billions USD/year!

900 billions USD/year

[Census Bureau, www.census.gov, 2007]
It’s not a surprise!

- Manual
- Time consuming
- Non-systematic
- Error prone

“An improvement of as little as 1% can lead to up to 900 millions USD in savings in construction business”

[Census Bureau, www.census.gov, 2007]
Opportunity to modernize age-old process in a profound way freeing up critical human resources
Towards modernity

- Barcodes & RFID tags 😞
Towards modernity

- Barcodes & RFID tags
- 3D laser scans

Still time consuming and expensive!
Our solution

Golparvar-Fard, Peña-Mora & Savarese, 2008-2012

- Thousands of images
- Computer vision
Our solution

Images are cheap!
Our solution
Our solution
Our solution
<table>
<thead>
<tr>
<th>Task</th>
<th>Planned</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete Excavate Upper Footings Area C</td>
<td>15SEP06A</td>
<td>14NOV06</td>
</tr>
<tr>
<td>Concrete Excavate Footings Area D</td>
<td>25SEP06A</td>
<td>27SEP06A</td>
</tr>
<tr>
<td>Concrete Pour Footings Area D</td>
<td>27SEP06A</td>
<td>14NOV06</td>
</tr>
<tr>
<td>Concrete Form Walls Area B</td>
<td>30SEP06A</td>
<td>16NOV06</td>
</tr>
<tr>
<td>Concrete Pour Upper Footings Area D</td>
<td>02OCT06A</td>
<td>03OCT06A</td>
</tr>
<tr>
<td>Concrete Excavate Column Pads Area A</td>
<td>10OCT06A</td>
<td>14NOV06</td>
</tr>
<tr>
<td>Concrete Pour Column Pads Area A</td>
<td>10OCT06A</td>
<td>14NOV06</td>
</tr>
<tr>
<td>Concrete Waterproof First Lift for Drain</td>
<td>10OCT06A</td>
<td>06FEB07</td>
</tr>
<tr>
<td>Tile</td>
<td>11OCT06A</td>
<td>17NOV06</td>
</tr>
<tr>
<td>Concrete Form/Pour Upper Walls Area C</td>
<td>11OCT06A</td>
<td>16NOV06</td>
</tr>
<tr>
<td>Exterior Perimeter Drain Area A</td>
<td>11OCT06A</td>
<td>17NOV06</td>
</tr>
<tr>
<td>Concrete Pour Walls Area B</td>
<td>19OCT06A</td>
<td>17NOV06</td>
</tr>
<tr>
<td>Concrete Form/Walls Area D</td>
<td>08NOV06A</td>
<td>29NOV06</td>
</tr>
<tr>
<td>Concrete Mock-Up</td>
<td>09NOV06A</td>
<td>08NOV06A</td>
</tr>
</tbody>
</table>

**Ahead of Schedule**

- **On Schedule**
- **Behind Schedule**
Summary

• Automate communication of performance deviations
• Reduction in delivery time
• Potential to identify unsafe locations/components
• Large impact in the civil engineering community

• James R. Croes Medal, October 2013 (from the American Society of Civil of Engineers)
• Best paper award from journal of CEM, 2011
• Best paper award at AEC/FM 2010
• Best paper award at Construction Research Congress 2009
Thank you!
Performance analysis of construction projects

12/02/2006; 1:13:00 PM (As-built)

Project: College of Business Instructional Facility, Photograph Courtesy of Facilities & Services, UIUC.
Performance analysis of construction projects

- Manual progress monitoring:
  - Time consuming
  - Non-systematic

- Improvement of as little as 1% can lead to up to 900 millions USD in savings in construction business

[Census Bureau, www.census.gov, 2007]
Performance analysis of construction projects

- Golparvar-Fard, Pena-Mora, Savarese, 2008-2012

- **James R. Croes Medal 2013**
- Best paper award from journal of CEM, 2011
- Best paper award at **AEC/FM 2010**
- Best paper award at Construction Research Congress 2009

- Large impact in the civil engineering community
- Opportunity to modernize age-old labor-intensive process in a profound way, freeing up critical human resources
What’s ahead

- Representation
- Learning
- Computational demands
Thank you!
## Results

### Average precision in localizing objects in the 3D space

<table>
<thead>
<tr>
<th></th>
<th>Hoiem et al. 2011</th>
<th>SSFM no int.</th>
<th>SSFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORD CAMPUS</td>
<td>21.4%</td>
<td>32.7%</td>
<td><strong>43.1%</strong></td>
</tr>
<tr>
<td>OFFICE</td>
<td>15.5%</td>
<td>20.2%</td>
<td><strong>21.6%</strong></td>
</tr>
</tbody>
</table>

### Average precision in detecting objects in the 2D image

<table>
<thead>
<tr>
<th></th>
<th>DPM [1]</th>
<th>SSFM 2 views no int.</th>
<th>SSFM 2 views</th>
<th>SSFM 4 views</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORD CAMPUS</td>
<td>54.5%</td>
<td>61.3%</td>
<td>62.8%</td>
<td><strong>66.5%</strong></td>
</tr>
<tr>
<td>OFFICE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*FORD CAMPUS dataset [Pandey et al., 09]*

*Office dataset [Bao et al., 11]*

*[1] Felzenszwalb et al. 2008*
## Results

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Camera translation error</th>
<th>Camera rotation error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SFM</td>
<td>SSFM no int.</td>
</tr>
<tr>
<td>FORD CAMPUS</td>
<td>26.5°</td>
<td>19.9°</td>
</tr>
<tr>
<td>OFFICE</td>
<td>8.5°</td>
<td>4.7°</td>
</tr>
<tr>
<td>STREET</td>
<td>27.1°</td>
<td>17.6°</td>
</tr>
</tbody>
</table>

FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]
Street dataset [Bao et al., 11]
Applications

Vision for the blinds

Mobile vision

Construction monitoring

Safe driving

Objects understanding

Space understanding

Activity understanding