From Sensing to Visual Intelligence

Silvio Savarese
Sensing is the future
Everything can be a sensor...
Everything can be a sensor...
Everything can be a sensor...

- Night vision
- Thermal imaging
- Gravity detection
- Kinect
Sensing is not the hard problem

Intelligent understanding of the sensing data is the challenge!
What does it mean “intelligent understanding of the sensing data”? 
Is it about “where”?
Is this useful?
Is this sufficient?
What’s intelligent understanding of the sensing data?

• It’s not just about “where” things are

• … is it also about “what” are the things in the scene?
Computers can recognize objects better than humans! (*)

Is this sufficient?
Tea kettle!
Tea kettle!
NOT INTEGRATED OR FUNCTIONALLY USEFUL!
NOT INTEGRATED OR PREDICTIVE!

People waiting in a queue; about to get food

About to serve drink

People having a meal

Plates

Table
What’s intelligent understanding of the sensing data?

• It’s not just about “where”
• It’s not just about “what”

... it’s about functions, context, common sense reasoning, cue integration, generalization, prediction...
Not 'all that glitters is gold!

From DARPA 2015 Robotics Grand Challenge

July 13, 2016: Security robot accidentally attacks child!
Smart houses
Smart cities
From sensing to intelligence

- Integrating
- Contextualizing
- Predicting
From sensing to visual intelligence

- Integrating
- Contextualizing
- Predicting
Object understanding

- Integrating
- Contextualizing
- Predicting
Object understanding

Yu & Savarese, 2013-2015
Choy et al., 2015-2017
Objects become actionable!

Yu & Savarese, 2013-2015
Choy et al., 2015-2017
Smart cups!

Etc. Fener, Guibas, Savarese, 2017
Smart cups!
Scene understanding

- Integrating
- Contextualizing
- Predicting
Scene understanding

Choi, Chao, Pantofaru, Savarese, 2013-2015
Large-scale scene parsing

Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016
Large-scale scene parsing

Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016
Stanford Large-Scale Indoor dataset

Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016

6 buildings  ~500 rooms  ~6000m² area  ~6000 Building Elements
Stanford Large-Scale Indoor dataset

http://buildingparser.stanford.edu
Robot mobility

http://buildingparser.stanford.edu
James R. Croes Medal, 2013 (from the American Society of Civil of Engineers)
From sensing to visual intelligence

- Integrating
- Contextualizing
- Predicting
Activity understanding and prediction

- Integrating
- Contextualizing
- Predicting
Recognizing complex human activities

Choi, Pantofaru, Savarese, 2012-2014

FE = face each other;
SQ = standing after each other
SS = standing side by side
SR = standing in a row
How to “define” such rules and interrelations in the first place?
Human interrelations is a matter of social conventions
Social conventions and etiquette
Social conventions and etiquette
Social conventions and etiquette
Modeling and predicting social behavior

Alahi, Goel, Ramanathan, Robicquet, Fei-Fei, Savarese, 2016
Robicquet, Sadeghian, Alahi, Savarese, 2016
Ballan, Castaldo, Alahi, Palmieri, Savarese, 2016
A new experimental platform: The JackRabbit
A new experimental platform: The JackRabbit

- **Sensors**
  - Planar laser scanner
  - 3D stereo vision/laser scanner
  - Cameras
  - IMU+GPS
A new experimental platform: The JackRabbit

JackRabbit has been featured by BBC, MIT tech review, Financial Times...
A glimpse of the future
Thank you!