

Image Understanding and Computer Vision Research at MSR Redmond and Cambridge

Microsoft Research
Faculty Summit
2015
July 8-9, 2015



Image Understanding and Computer Vision Research at MSR Redmond and Cambridge

Microsoft Research

Faculty Summit
2015

Zhengyou Zhang

Research Manager/Principal Researcher
Microsoft Research, Redmond, WA

Computer Vision has made
tremendous progress!

Microsoft researchers say their newest deep learning system beats humans — and Google



Image Credit: Robert Scoble/Flickr

Jian Sun's team
at MSRA

Convolutional Neural Network Demo

mtgranite-03 - Remote Desktop Connection

0 1 2 3

4 5 6 7

8 9 10 11

Perf

16 Images / Second

WCS 1.0 Server
(CPU Only)

Obvious mistakes made by computer



- GT: letter opener**
1: drumstick
2: candle
3: wooden spoon
4: spatula
5: ladle



- GT: letter opener**
1: Band Aid
2: ruler
3: rubber eraser
4: pencil box
5: wallet



- GT: letter opener**
1: fountain pen
2: ballpoint
3: hammer
4: can opener
5: ruler

Computer Vision
still falls far short of what
Human Vision can do

Image Ingredients

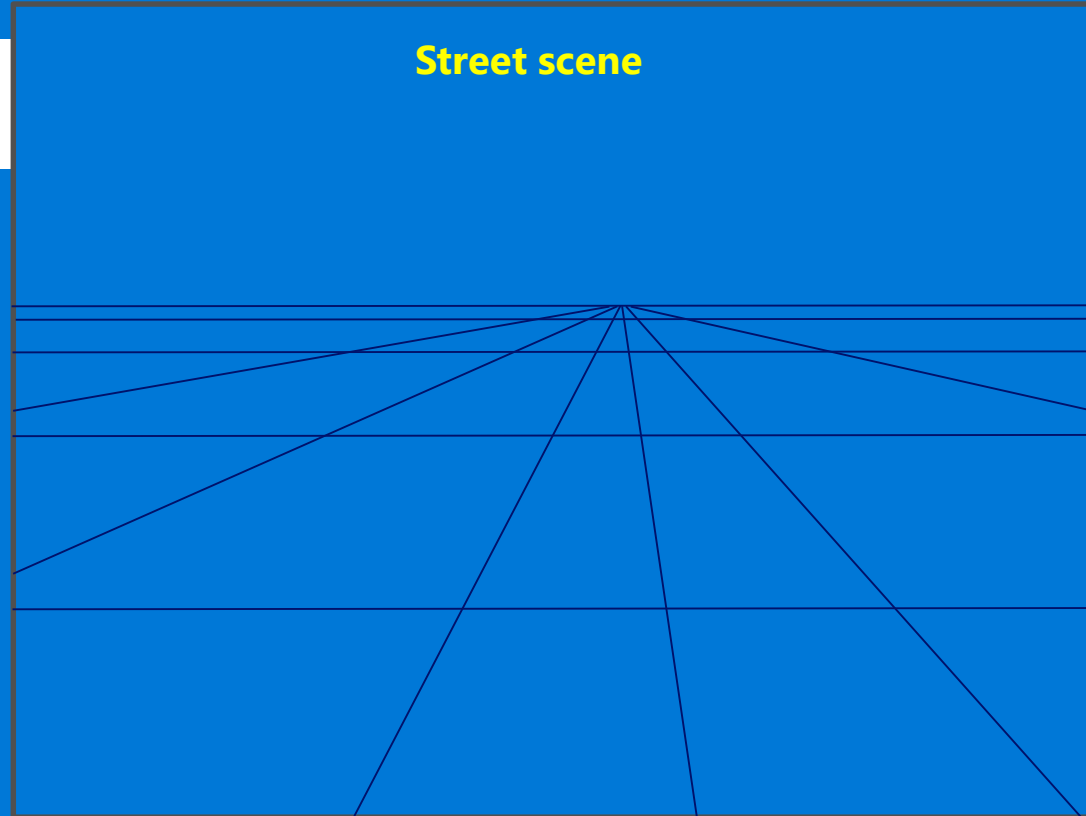
Very many
sources of
variability



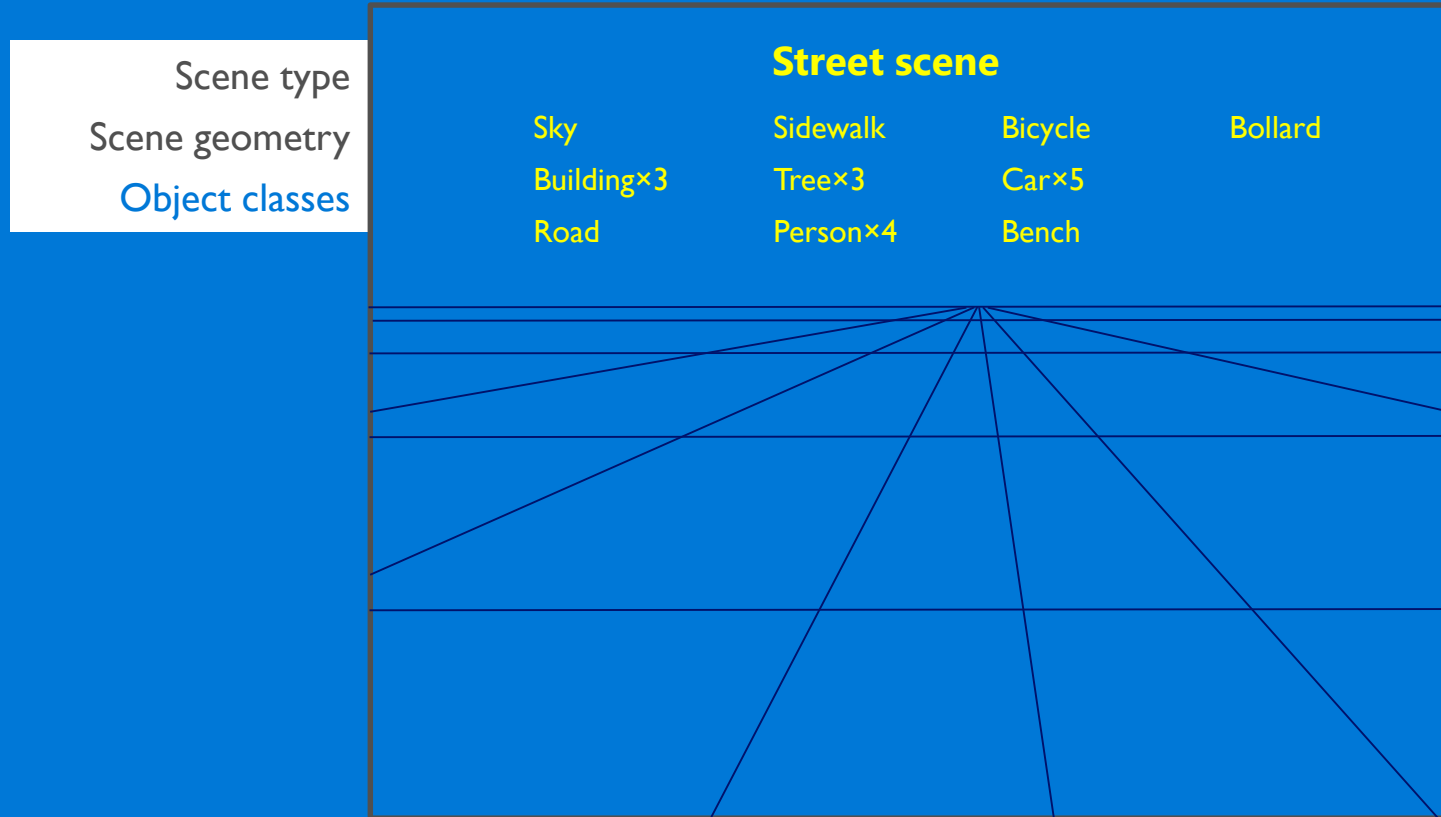
Image

Sources of image variability

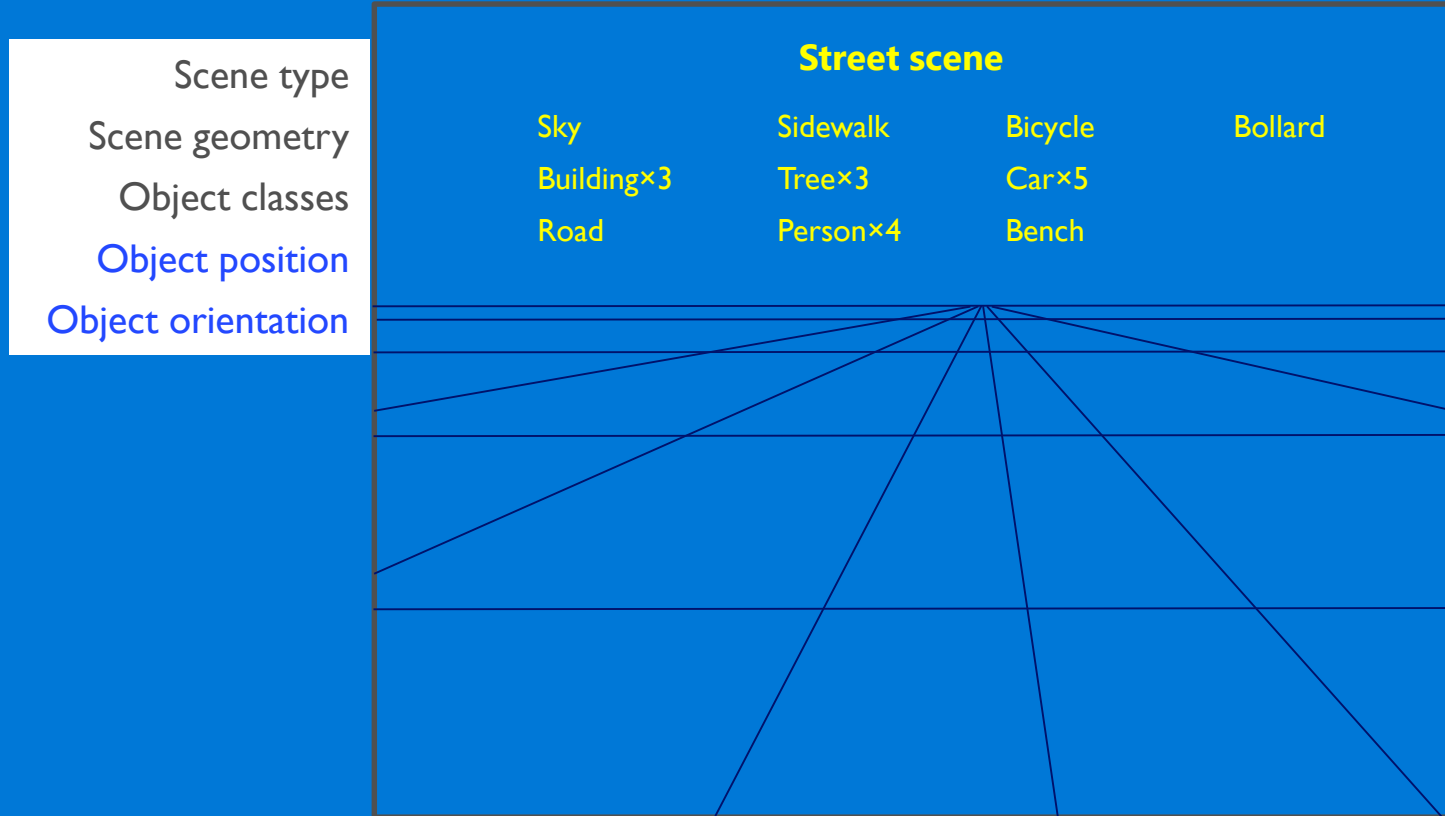
Scene type
Scene geometry



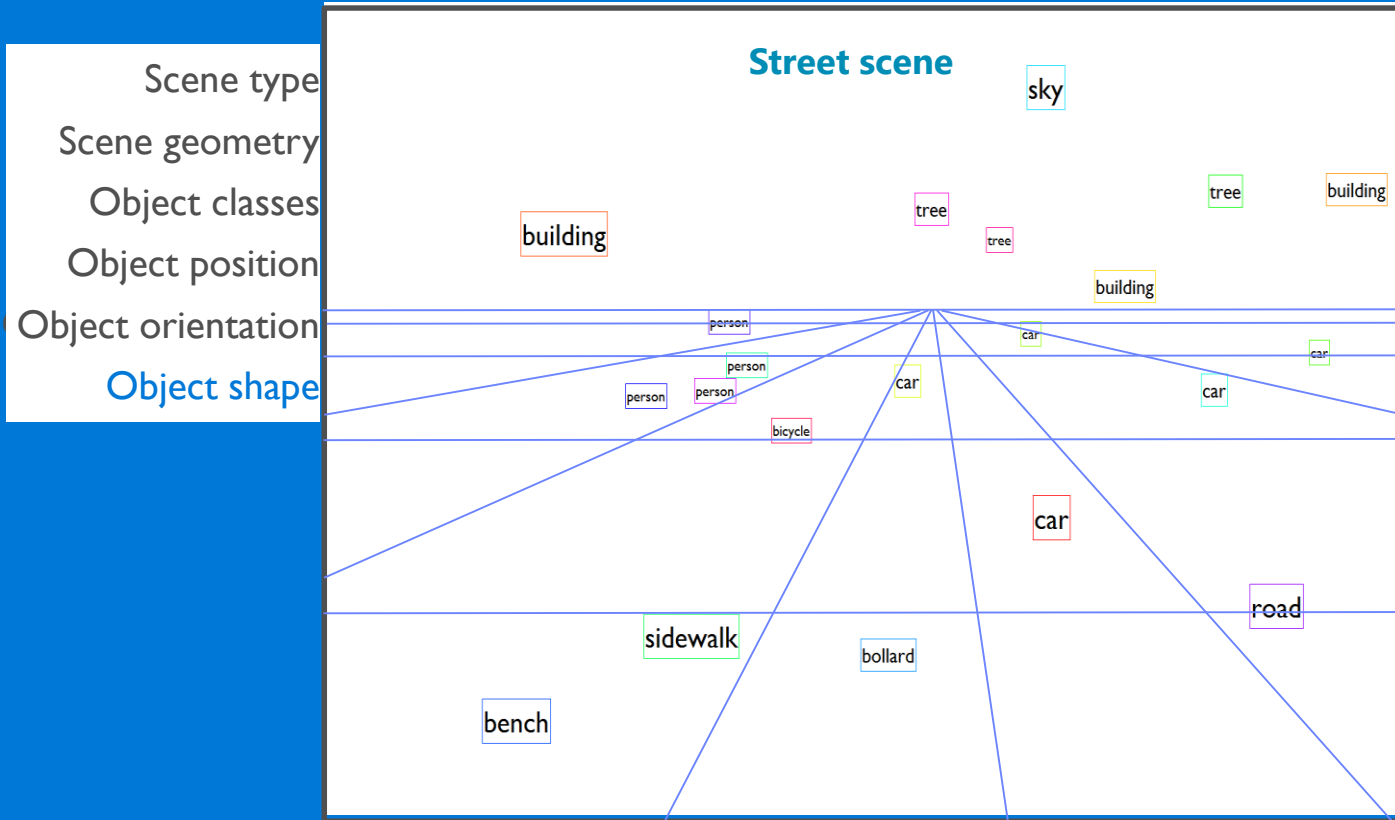
Sources of image variability



Sources of image variability

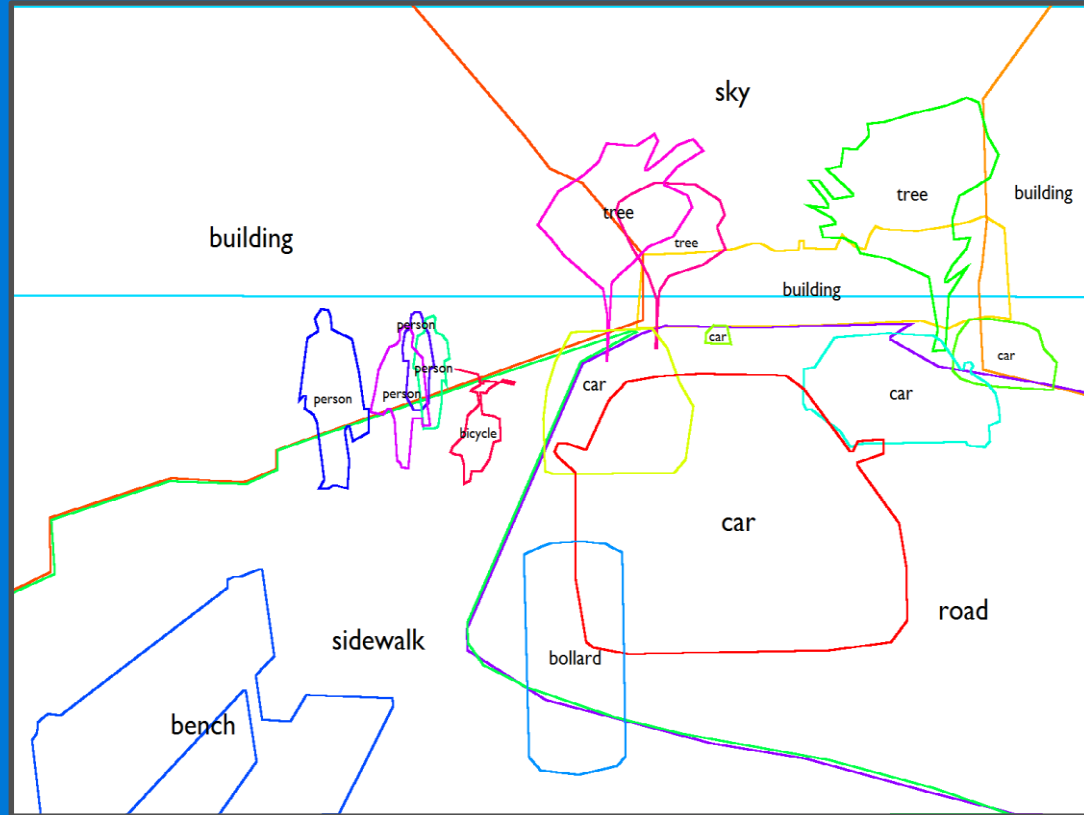


Sources of image variability



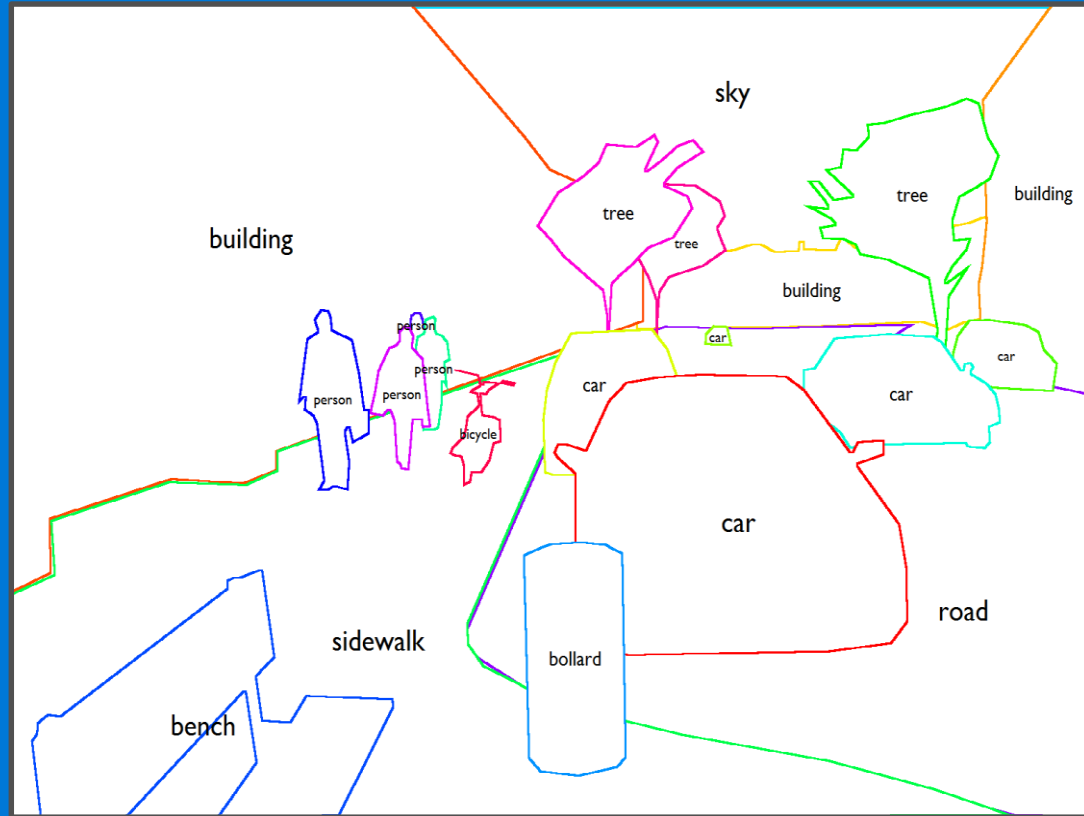
Sources of image variability

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions



Sources of image variability

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance



Sources of image variability

Scene type
Scene geometry
Object classes
Object position
Object orientation
Object shape
Depth/occlusions
Object appearance
Illumination
Shadows



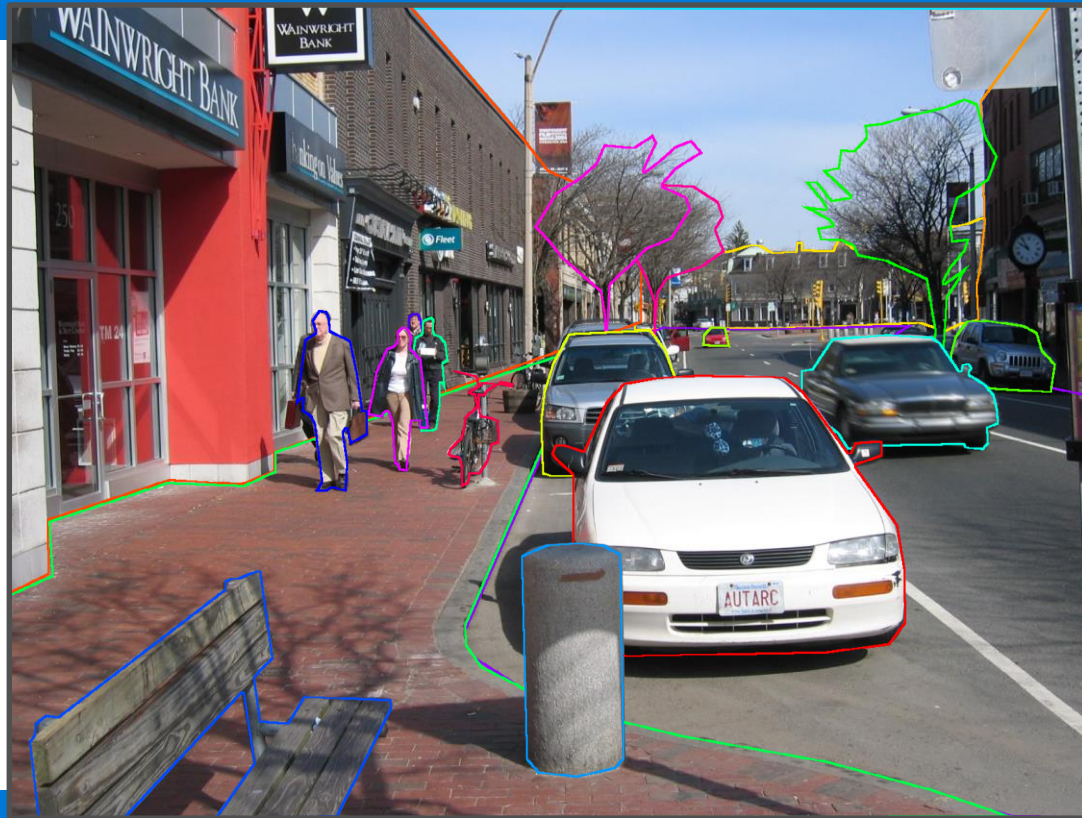
Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination**
- Shadows**



Sources of image variability

- Scene type
- Scene geometry
- Object classes
- Object position
- Object orientation
- Object shape
- Depth/occlusions
- Object appearance
- Illumination
- Shadows
- Motion blur
- Camera effects



Now.. The Good News

Interesting problems in vision can be solved if they are sufficiently constrained.

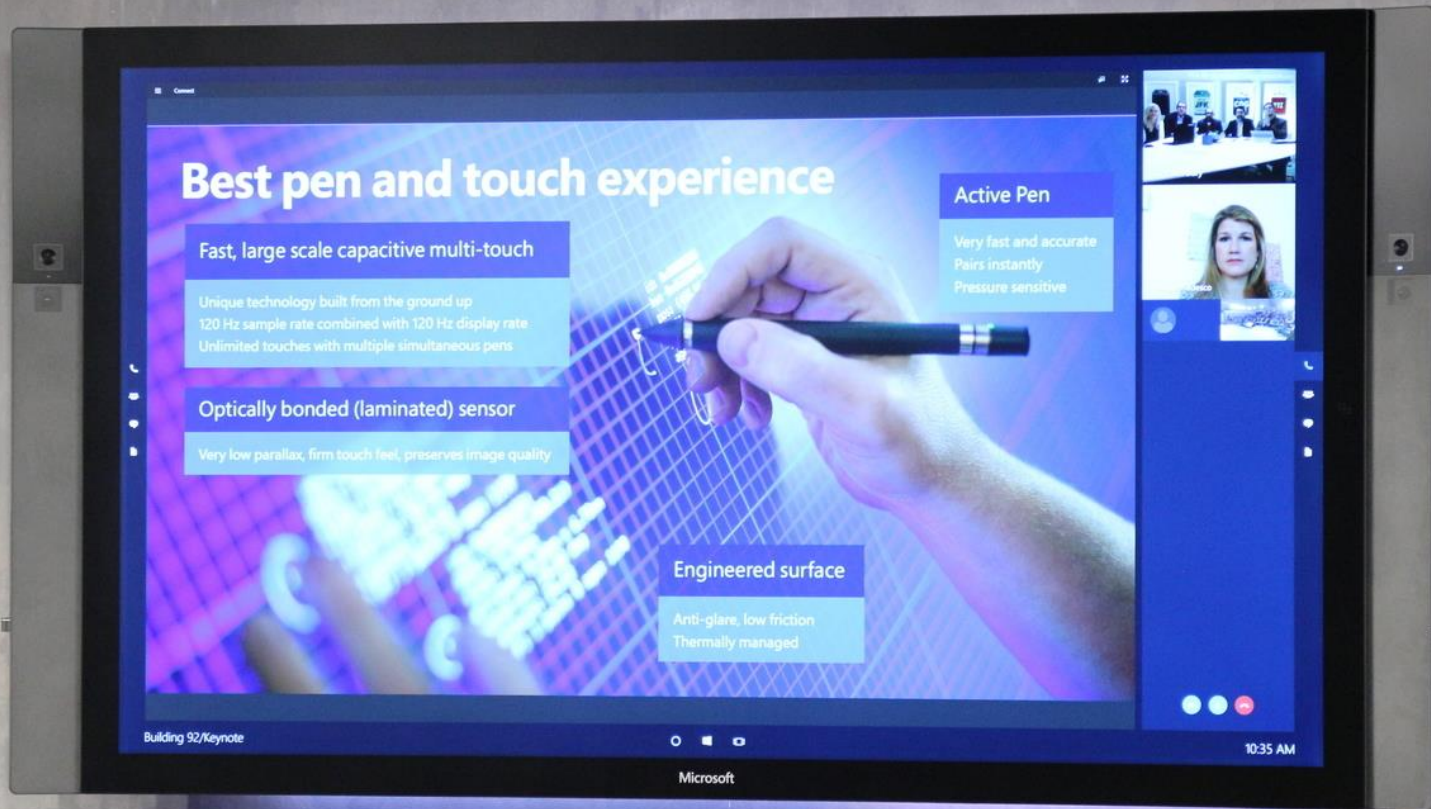
Collaborative Office Space



Big Boards & Massive Office Screens



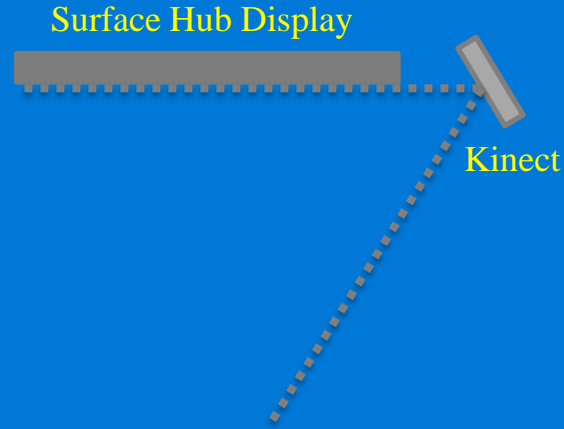
Microsoft Surface Hub



ViiBoard:

Vision-enhanced Immersive Interaction with Touch Board

Experimental Setup



Big Touch Board (Surface Hub) + RGB-D Sensor (Kinect)
leads to more natural and immersive interaction with touch boards

ViiBoard:

Vision-enhanced Immersive Interaction with Touch Board

VTouch

Natural and Rich Interaction Beyond Touch with important cues from RGB-D sensors

- Position and proximity w.r.t. Touch Board
- Person ID
- Hand ID
- Gesture ID
- Intention

ImmerseBoard

Immersive Remote Collaboration *as if writing on a physical whiteboard side-by-side*

- Seeing the **reference point**
- Sharing the **same space**
- Being aware of **gaze**
- Predicting **intention**

Vision-enhanced Touch Experience

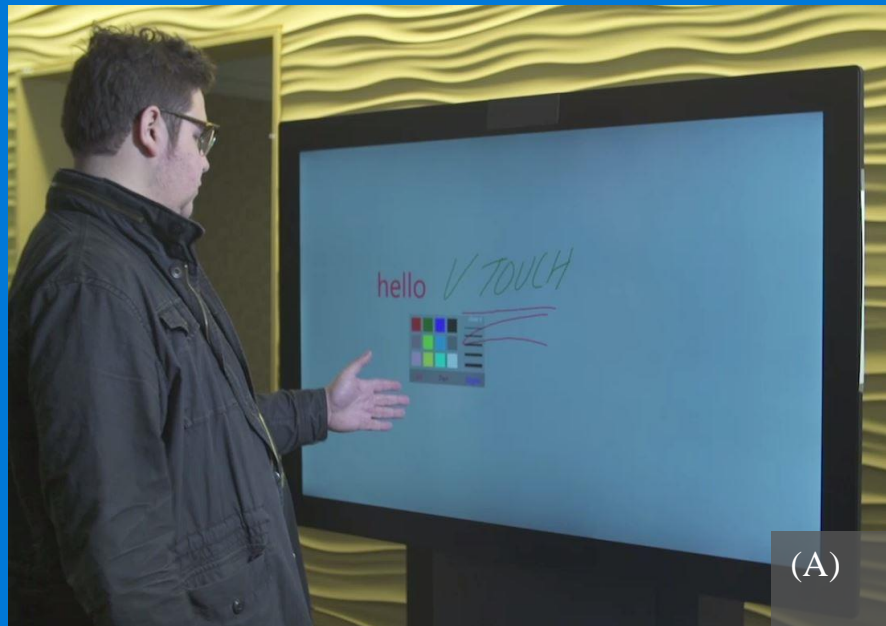
VTouch

Key Vision Technologies

- Sensor-Display Calibration
- Human Skeletal Tracking
- Hand Gesture Recognition
- Person Recognition

VTouch: Sample applications

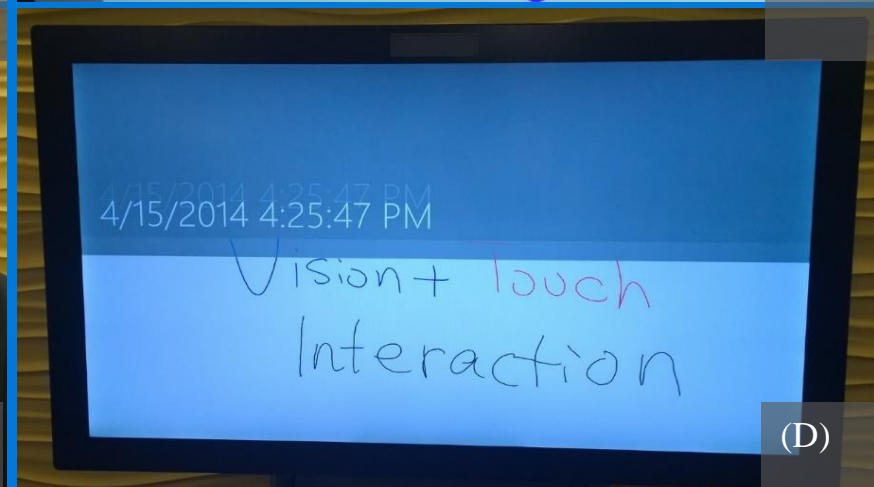
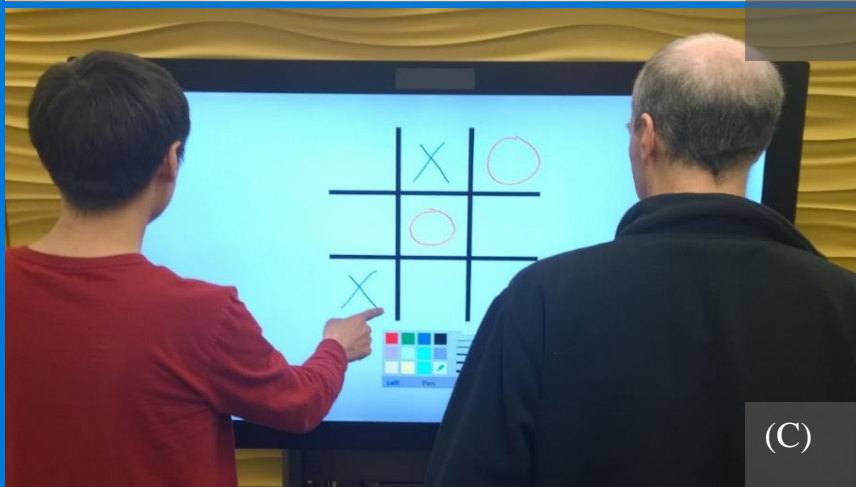
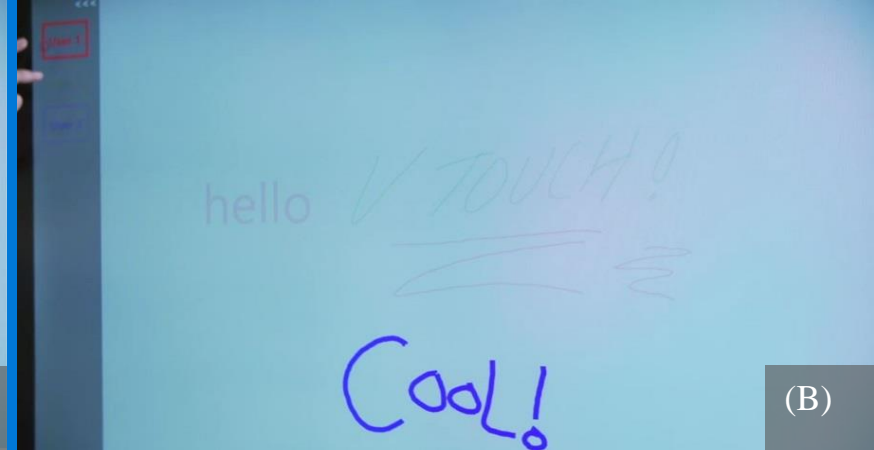
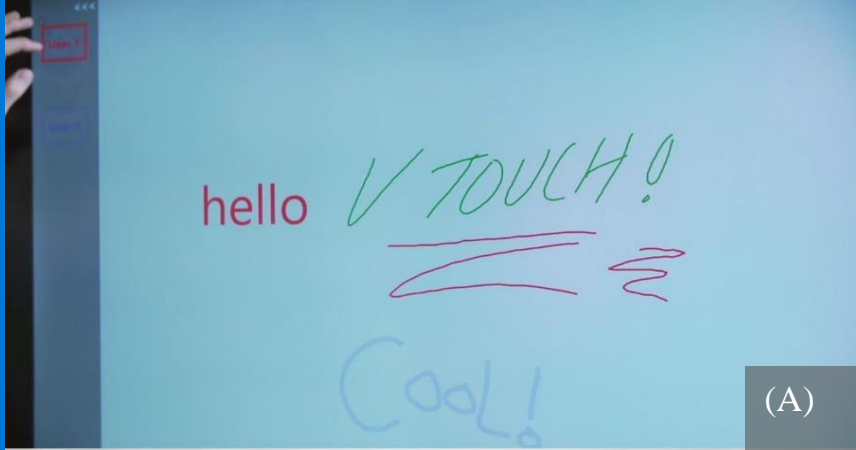
- Bring up menu without touch
- Display menu where you are
- Augment touch with
 - HandID, PersonID, GestureID
- Hover
- Pointing
- Auto lock of the display
- Auto unlock only with meeting participants



(A)

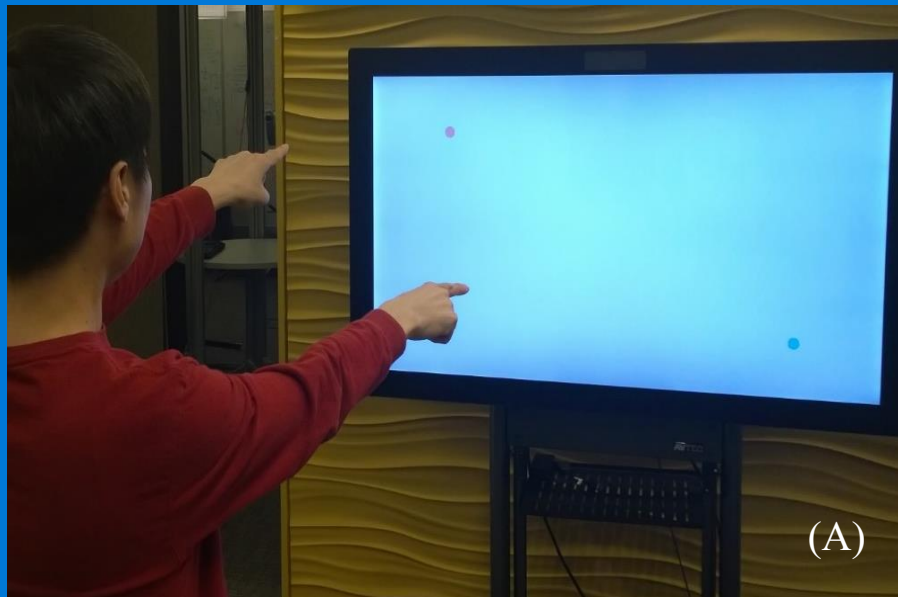


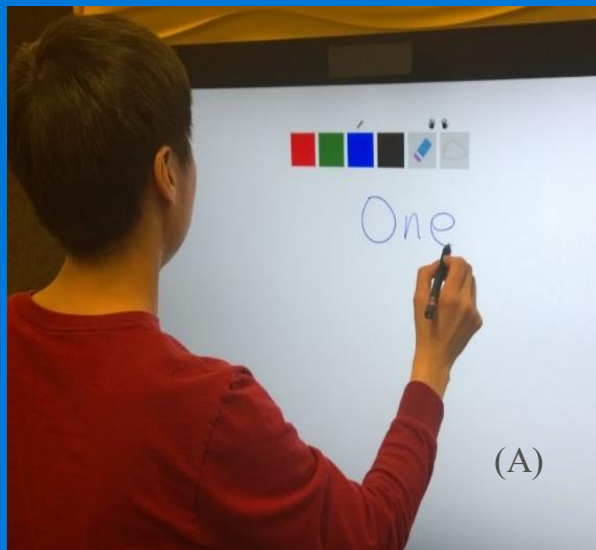
(B)



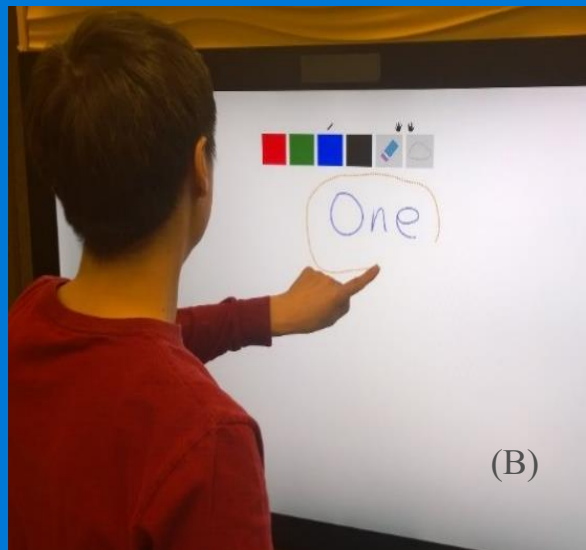
4/15/2014 4:25:47 PM
4/15/2014 4:25:47 PM

Vision + Touch
Interaction

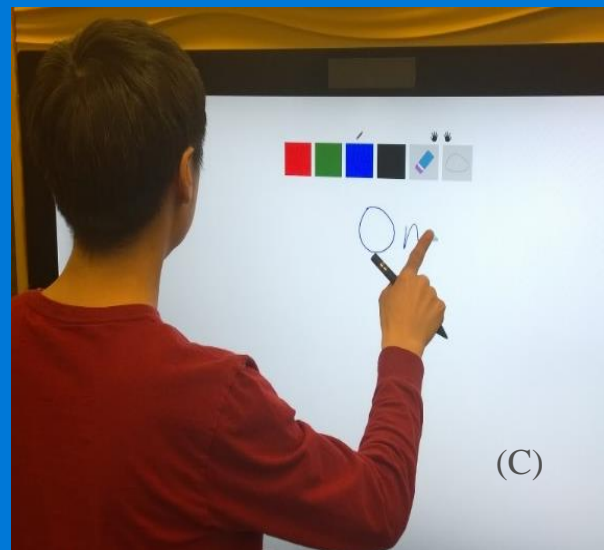




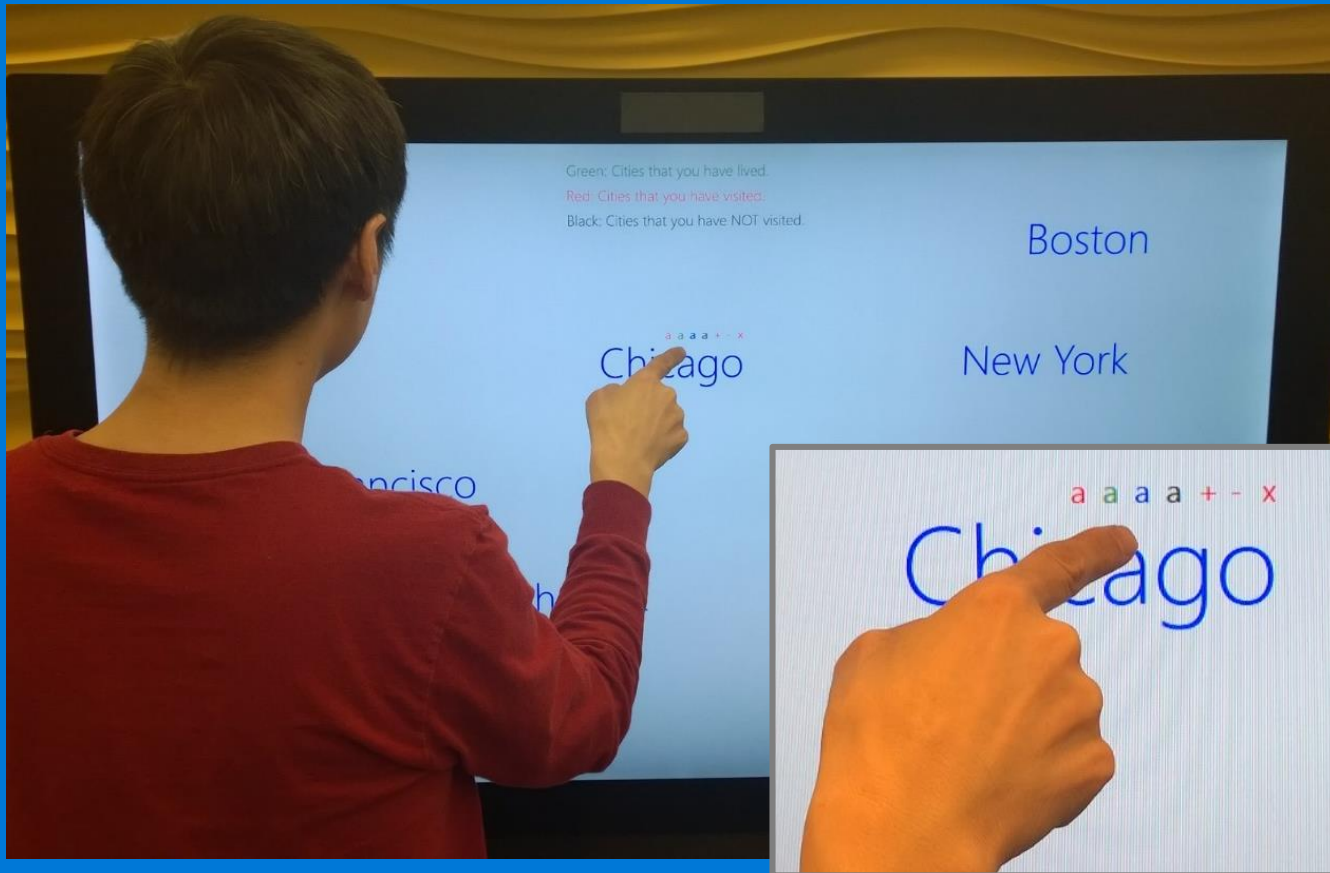
(A)



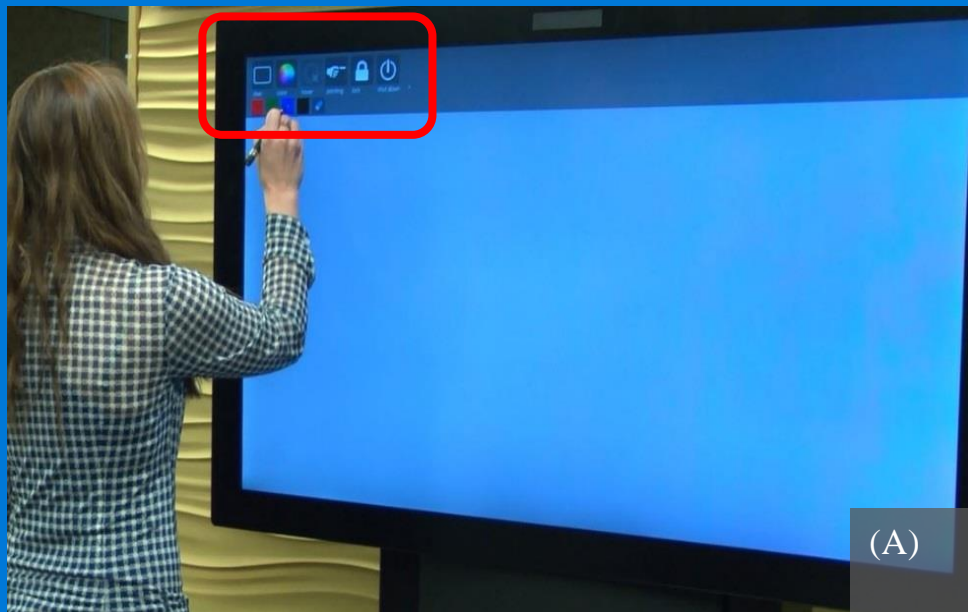
(B)



(C)

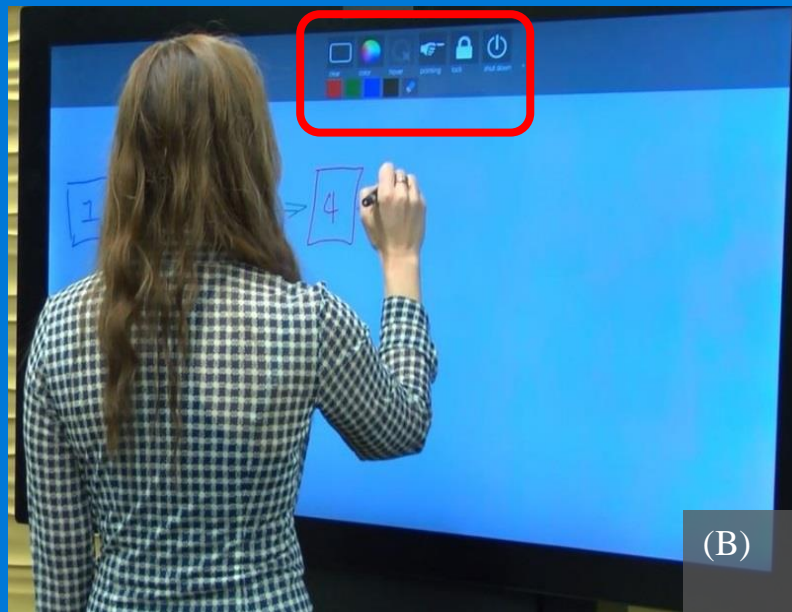


Menu Buttons



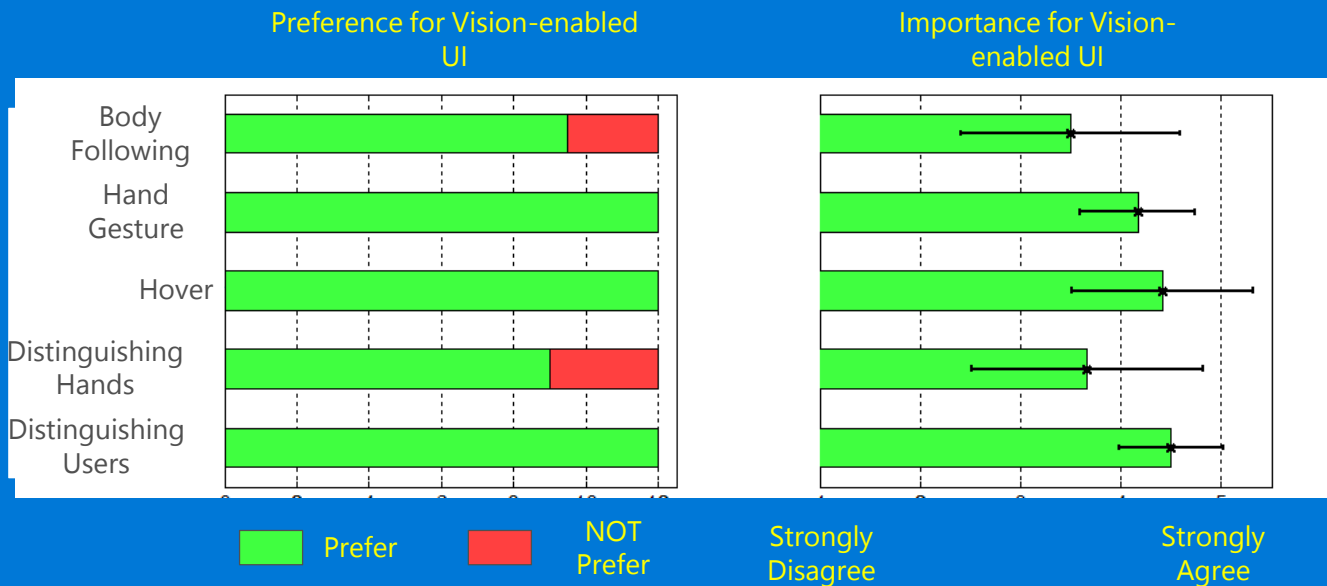
(A)

Menu Buttons

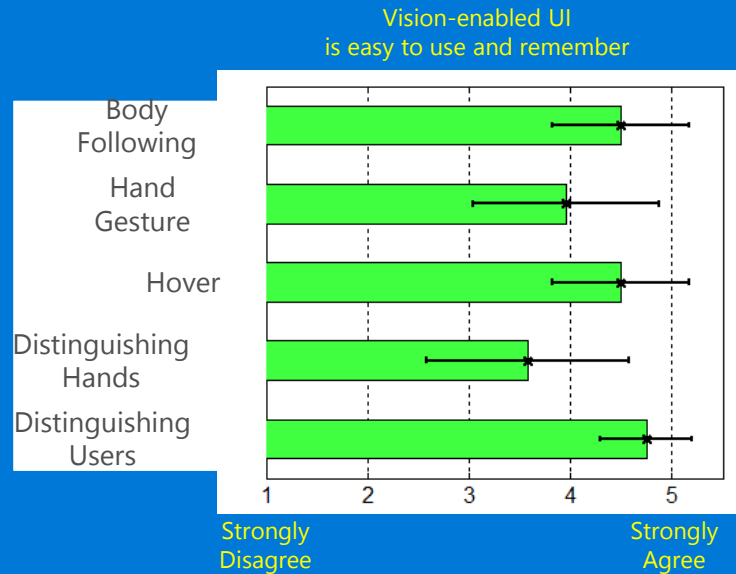


(B)

User Study



User Study

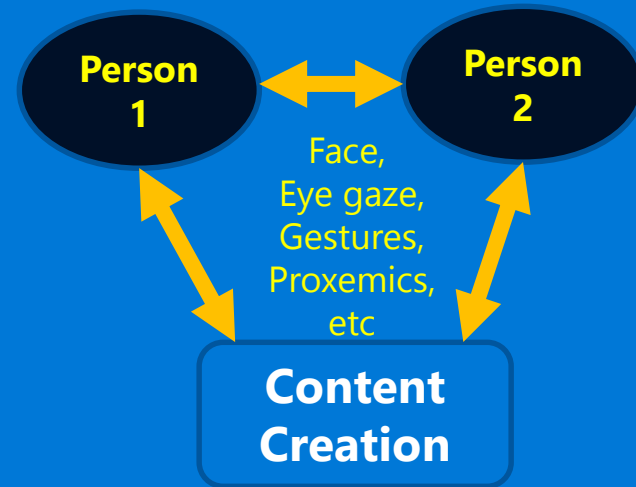
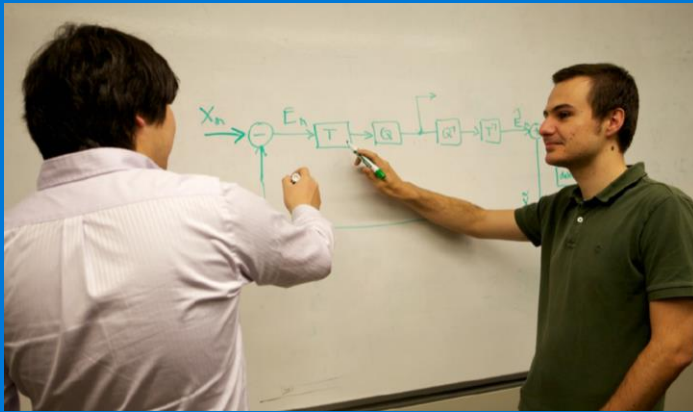


Vision-enhanced immersive remote collaboration

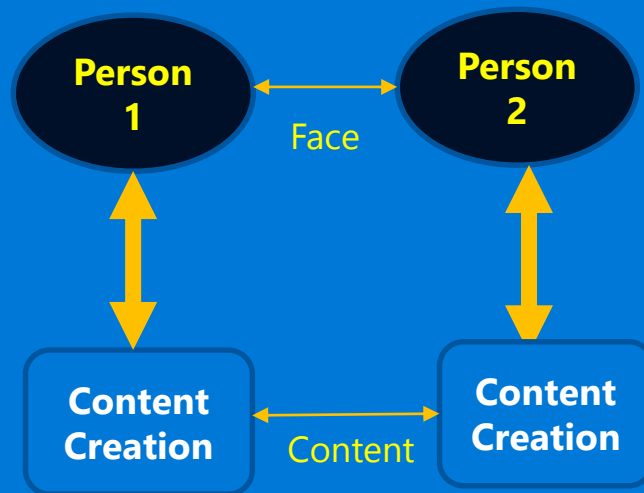
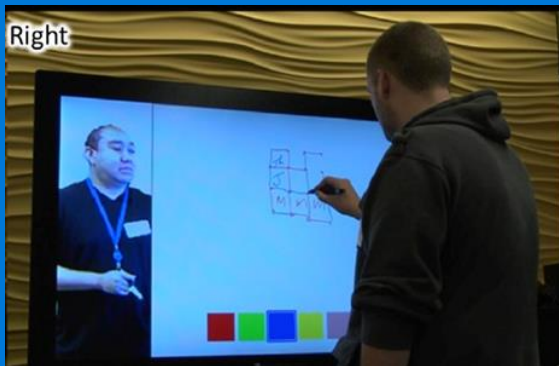
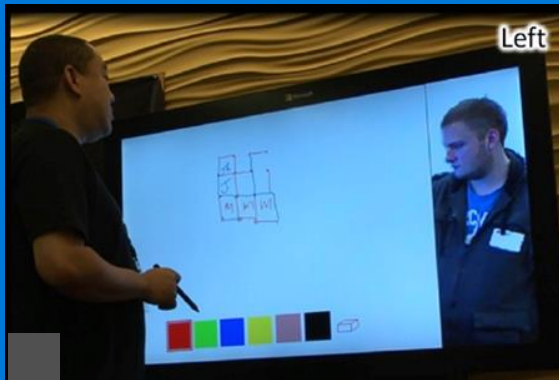
ImmerseBoard

Co-located Collaboration

- Physical whiteboard



Remote Collaboration: Now



RGBD Sensor (Kinect) + Touch Board (Surface Hub)
= Immersive Remote Collaboration

as if writing on a physical whiteboard side-by-side

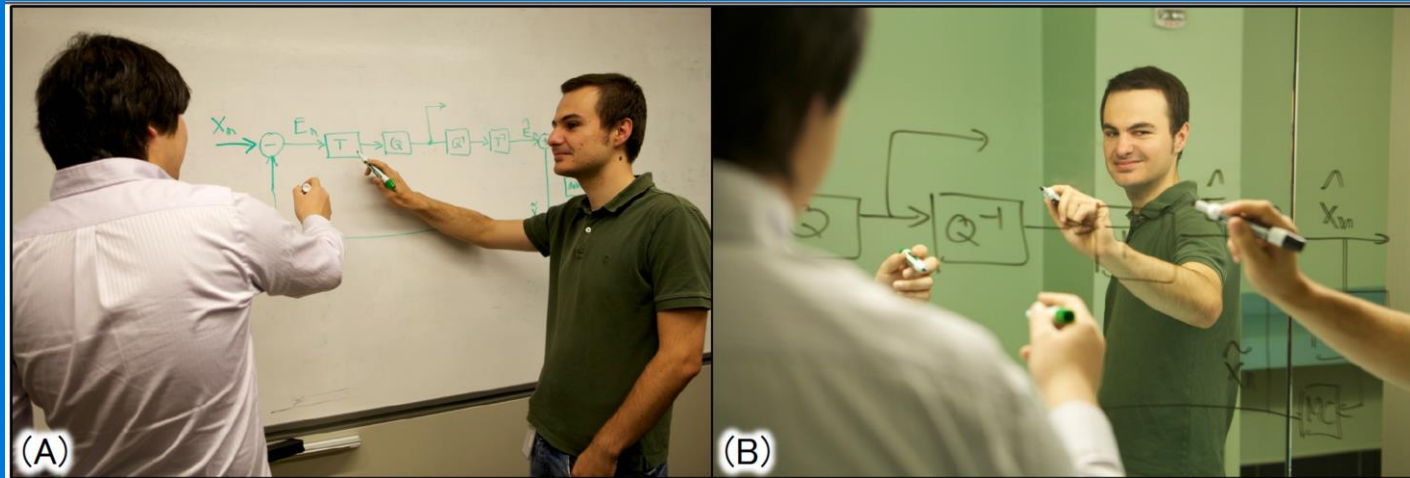
- Seeing the **reference point**
- Sharing the **same space**
- Being aware of **gaze**
- Predicting **intention**

Two Metaphors

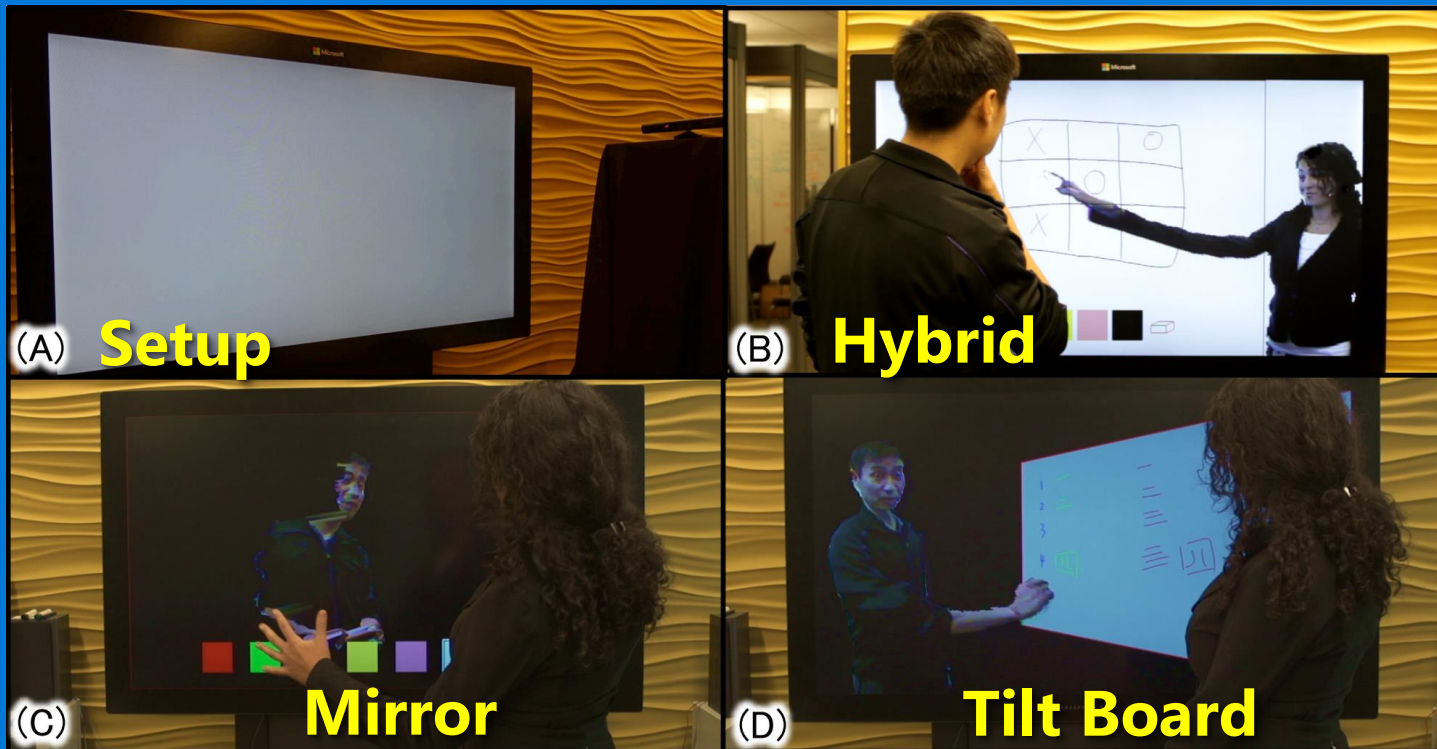
Side-by-side writing

on a whiteboard

on a mirror



ImmerseBoard: Implemented Conditions



ImmerseBoard

Natural Remote Collaboration

ViiBoard:

Vision-enhanced Immersive Interaction with Touch Board

Big Touch Board (Surface Hub) + RGB-D Sensor (Kinect)

leads to more natural and immersive interaction with touch boards

VTouch

Natural and Rich Interaction Beyond Touch
with important cues from RGB-D sensors

- Position and proximity
w.r.t. Touch Board
- Person ID
- Hand ID
- Gesture ID
- Intention

ImmerseBoard

Immersive Remote Collaboration
*as if writing on a physical whiteboard
side-by-side*

- Seeing the **reference point**
- Sharing the **same space**
- Being aware of **gaze**
- Predicting **intention**

Additional Projects from MSRC

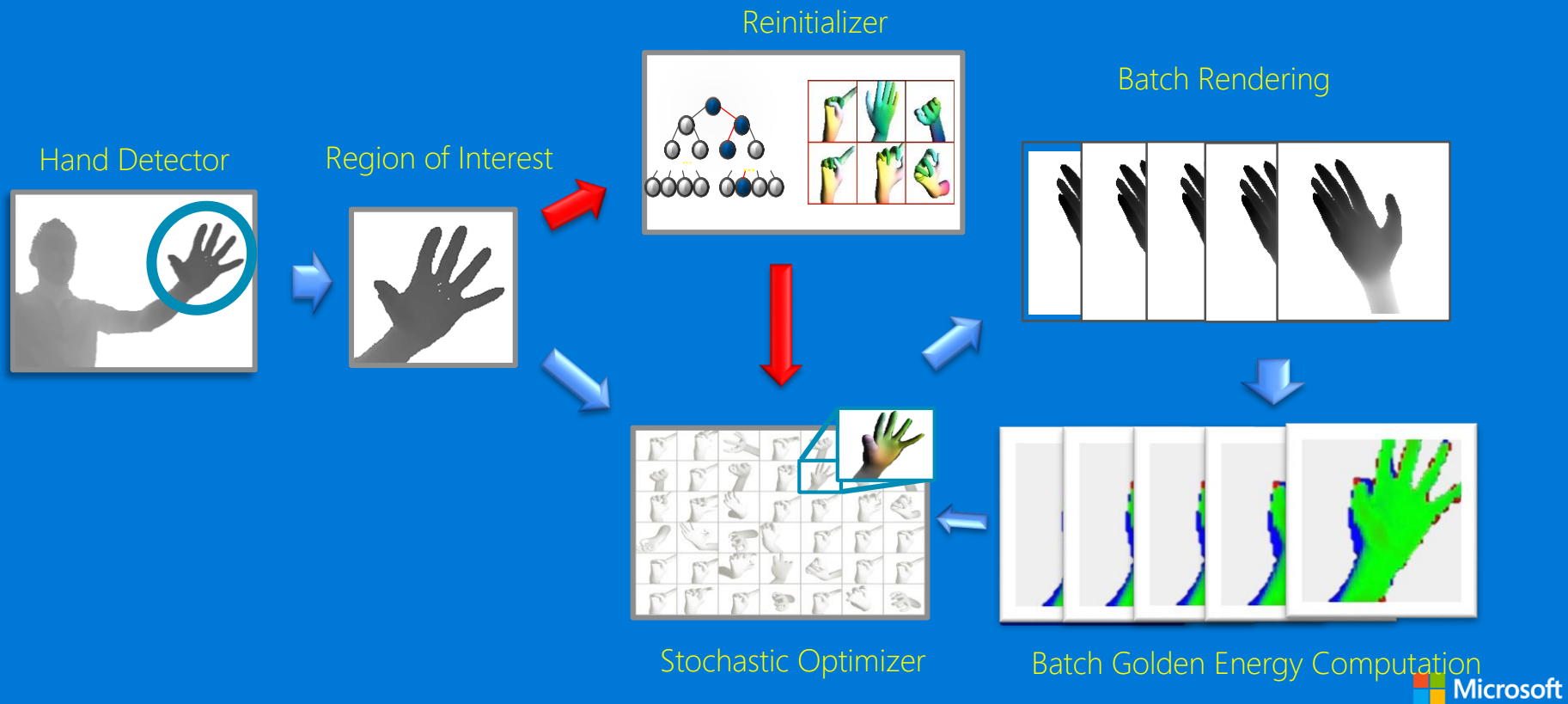
Effective (Hand) Shape And Pose Inference from Depth Images

Varun Ramakrishna
Aaron Hertzmann
Toby Sharp
Cem Keskin
Duncan Robertson
Jonathan Taylor
Jamie Shotton
Ido Leichter
Alon Vinnikov



Richard Stebbing
Sameh Khamis
David Kim
Christopher Rhemann
Yichen Wei
Daniel Freedman
Eyal Krupka
Andrew Fitzgibbon
Shahram Izadi

Architecture





Understanding Reality for Generating Credible Augmentations

Pushmeet Kohli

Microsoft Research

1: Labelling Point Clouds

[With Shapovalov et al. CVPR '12]



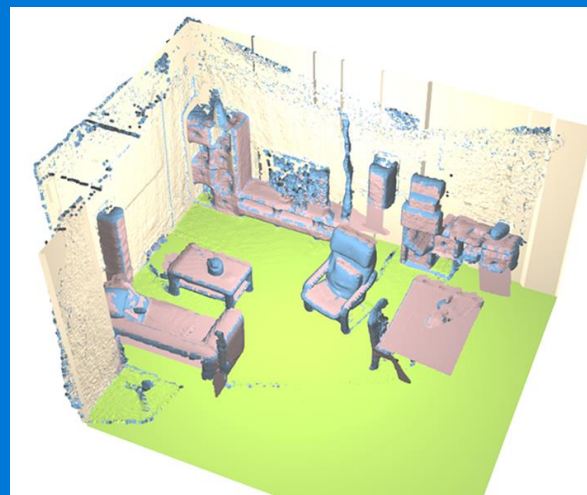
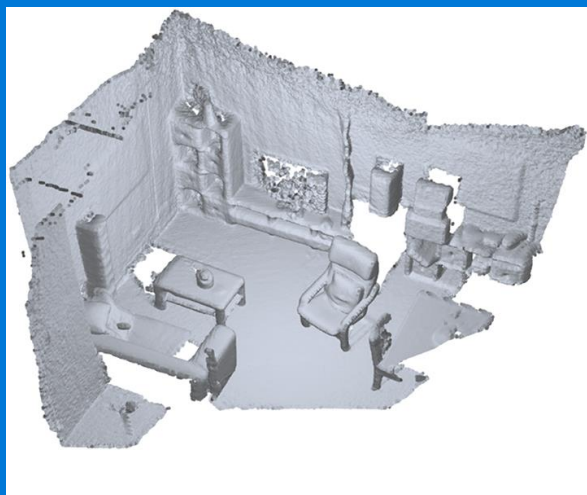
Inference Machine
=
Extension of
Random Forests



Colours represent different
object categories

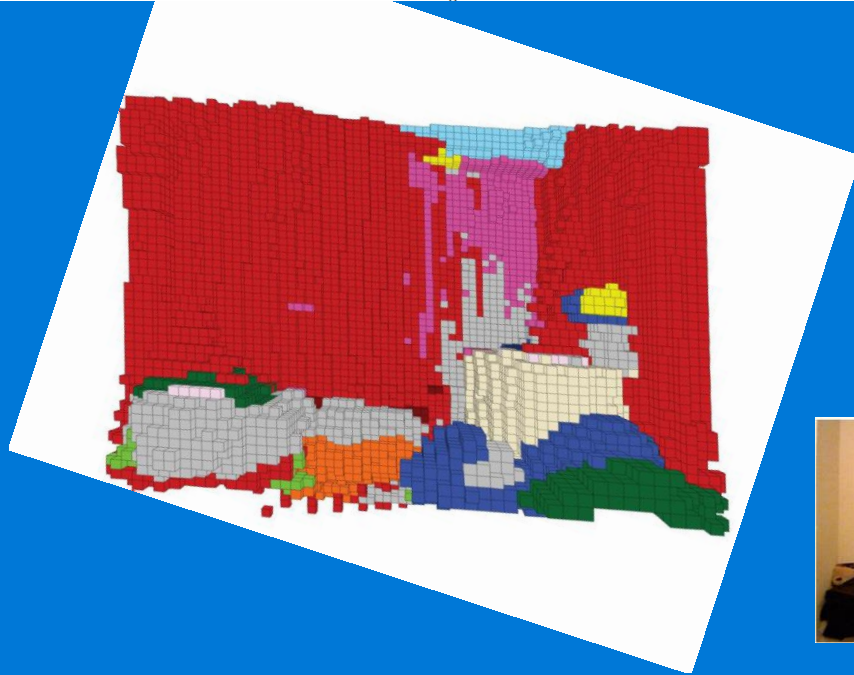
2. Scene Completion

[Silberman, Shapira, Gal, Kohli, ECCV 2014]



3. Semantic Labelling through Voxel CRF

[Kim, Kohli, Saverese, ICCV 2013]



4. Inferring Support Relationships

[Silberman, Hoiem, Kohli, Fergus, ECCV 2012]

**Interacting with objects
requires understanding of
support relationships!!**



Can I move the book?

5. Dynamic Capture and Labelling

[With Oxford Brookes, Shahram Izadi, TOG 2015]

Video

