SIBGRAPI 2012

Workshop on Industry Applications

# Computer Vision in the Kinect for Windows

Alisson Sol
Principal Software Design Engineer Lead
Kinect for Windows Team
IEB – Interactive Entertainment Business Division
Microsoft

KINECT™
for Windows®

Microsoft

# Thank You…

- SIBGRAPI 2012 Organization
- WGARI Committee
- CNPq, CAPES, Brazilian taxpayers…
- Microsoft
- Family and Friends…

# Before I Forget…

- Lots of people helped to created the content of this presentation…
  - Not only from K4W, but also DPE, Xbox, MSR, etc.

# Diversity for WGARI 2012

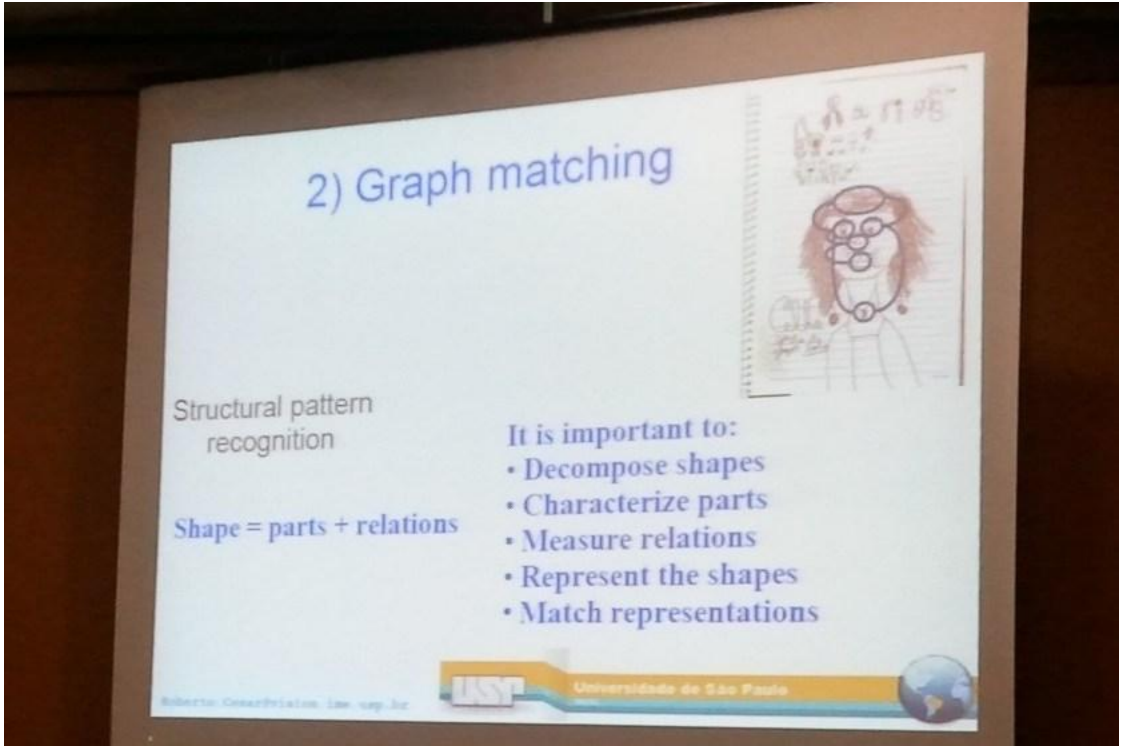
- 3D reconstruction
- Agriculture
- Augmented Reality
- Geology
- Iris recognition
- Morphing
- Video navigation

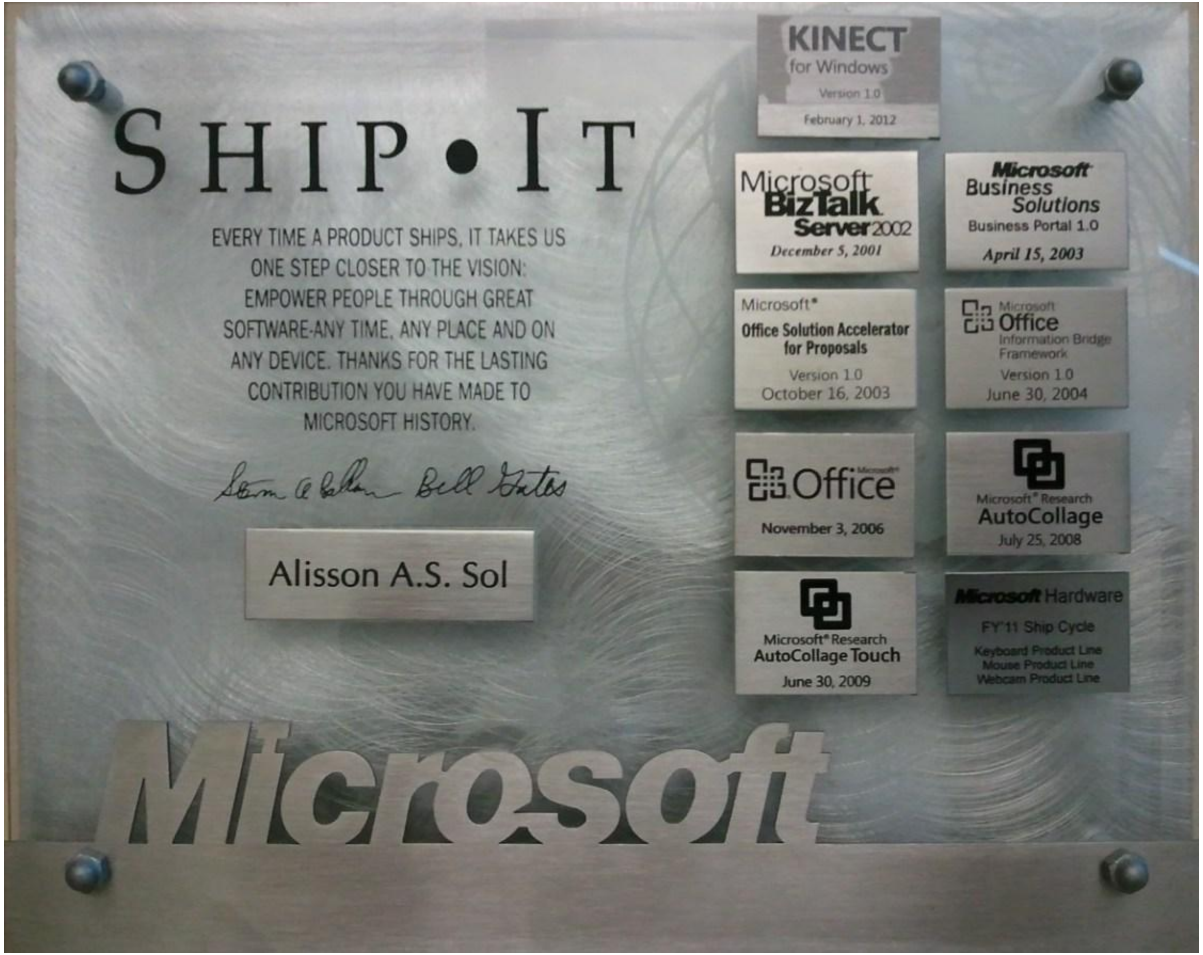
- C, C++
- iOS, LAMP, Windows
- LibSVM
- Matlab
- OpenCV
- OpenGL
- Point cloud

# From Software Top 100 - 2011

| # | Company | Software Revenue ($ million) | Software Revenue Growth | Total Revenue ($ million) | Software Revenue Share |
|---|---|---|---|---|---|
| 1 | Microsoft | 54,270 | 10.6% | 67,383 | 80.5% |
| 2 | IBM | 22,485 | 5.1% | 99,870 | 22.5% |
| 3 | Oracle | 20,958 | 12.8% | 30,180 | 69.4% |
| 4 | SAP | 12,558 | 10.5% | 16,654 | 75.4% |
| 5 | Ericsson | 7,274 | -4.2% | 30,307 | 24.0% |
| 6 | HP | 6,669 | 7.9% | 126,562 | 5.3% |
| 7 | Symantec | 5,636 | 1.3% | 6,013 | 93.7% |
| 8 | Nintendo | 5,456 | -19.8% | 13,766 | 39.6% |
| 9 | Activision Blizzard | 4,447 | 3.9% | 4,447 | 100.0% |
| 10 | EMC | 4,356 | 10.0% | 17,015 | 25.6% |
| - | - | - | - | - | - |
| 28 | Apple | 1,358 | 11.5% | 75,660 | 1.8% |
| 57 | Intel | 751 | 54.8% | 43,623 | 1.7% |
| 79 | Google Inc. | 543 | 42.5% | 29,321 | 1.9% |
| 84 | Totvs | 478 | 11.4% | 681 | 70.2% |

# Earlier Today…





SHIP • IT
EVERY TIME A PRODUCT SHIPS, IT TAKES US ONE STEP CLOSER TO THE VISION: EMPOWER PEOPLE THROUGH GREAT SOFTWARE-ANY TIME, ANY PLACE AND ON ANY DEVICE. THANKS FOR THE LASTING CONTRIBUTION YOU HAVE MADE TO MICROSOFT HISTORY.
Alisson A.S. Sol
KINECT for Windows Version 1.0 February 1, 2012
Microsoft BizTalk Server 2002 December 5, 2001
Microsoft Business Solutions Business Portal 1.0 April 15, 2003
Microsoft Office Solution Accelerator for Proposals Version 1.0 October 16, 2003
Microsoft Office Information Bridge Framework Version 1.0 June 30, 2004
Office November 3, 2006
Microsoft Research AutoCollage July 25, 2008
Microsoft Research AutoCollage Touch June 30, 2009
Microsoft Hardware FY'11 Ship Cycle
Keyboard Product Line
Mouse Product Line
Webcam Product Line
Microsoft


Microsoft
Windows 7,
精彩尽在指尖
Microsoft
TOUCHMOUSE
微软多点触控鼠标
Windows 7
Windows 7

# Wizard of Oz and the Scarecrow

- *"…, everybody can have a brain. That's a very mediocre commodity. … Back where I come from, we have universities, seats of great learning, where men go to become great thinkers. And when they come out, they think deep thoughts and with no more brains than you have. But they have one thing you haven't got: a* diploma*."*

# Dangerous Generalization…

## Physics

- Theory Unification
  + String Theory
- Collaboration
- Publications

## Computer Science

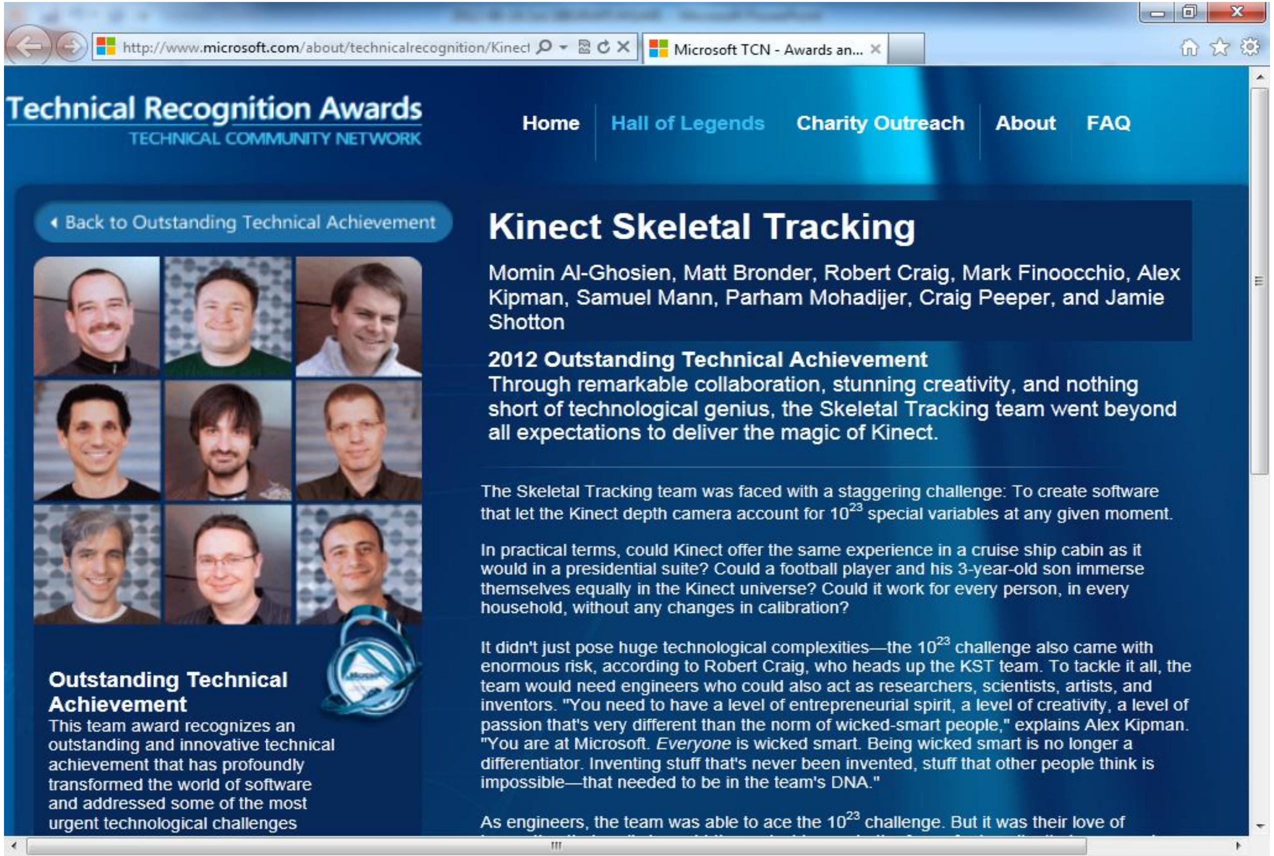
- “Natural User Interface”
  + Cloud Computing
- Competition
- Patents


Remington


1
ABC 2
DEF 3
GHI 4
JKL 5
MNO 6
PRS 7
TUV 8
WXY 9
0

# The Kinect… for Windows

# The Computer Vision Task v1.0



# Environment Differences…

## Xbox

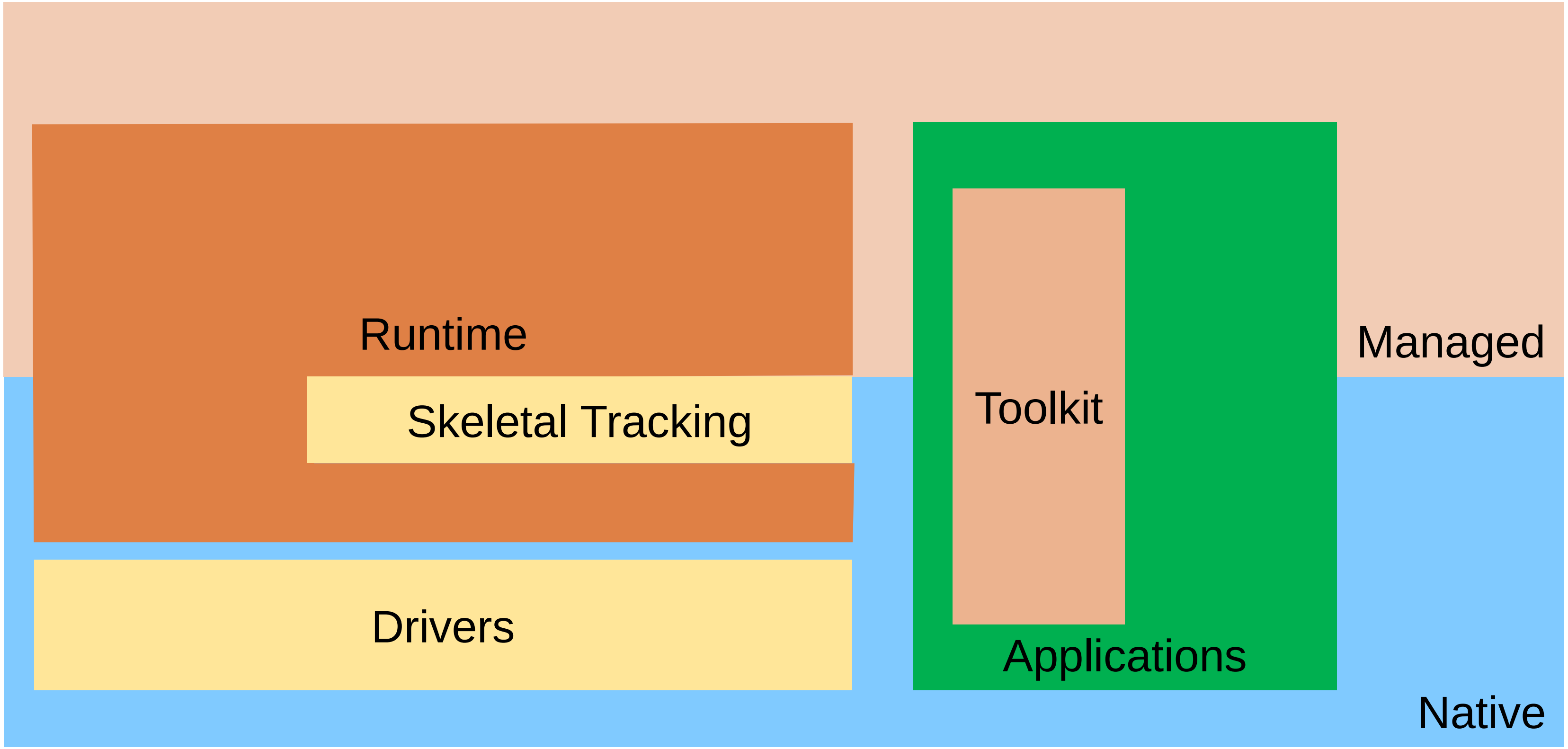
- Known CPU: PPC
- Known bus
- 1 device per machine
  - Only 1 supported
- Known architecture
- Known GPU
- Selected audience
- Mainly games

## Windows

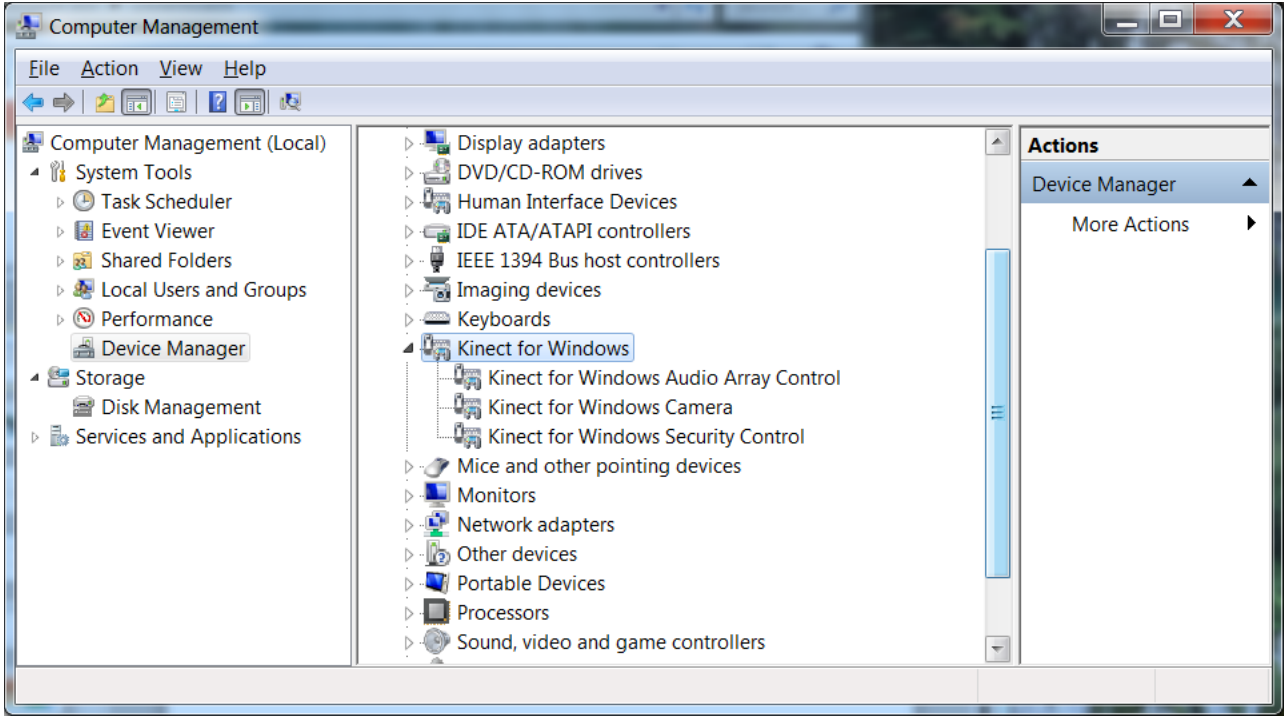
- Intel, AMD, …
- USB 2.0+
- Multiple devices
  - 1 per USB controller
- Win32, x64
- V1: no GPU requirement
- General audience
- Unbounded scenarios…

# Demo

# K4W SDK Block Diagram



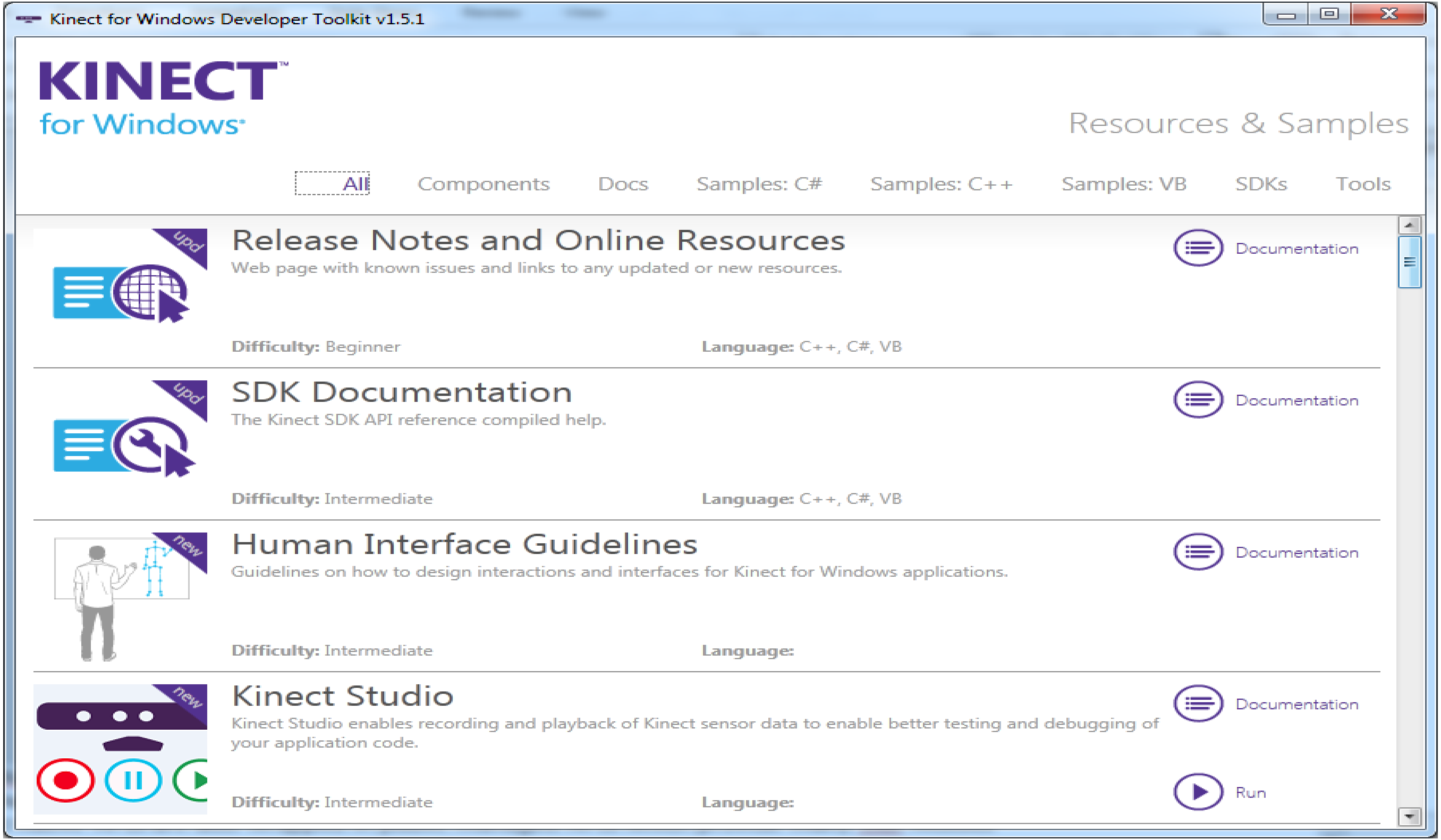
# Drivers



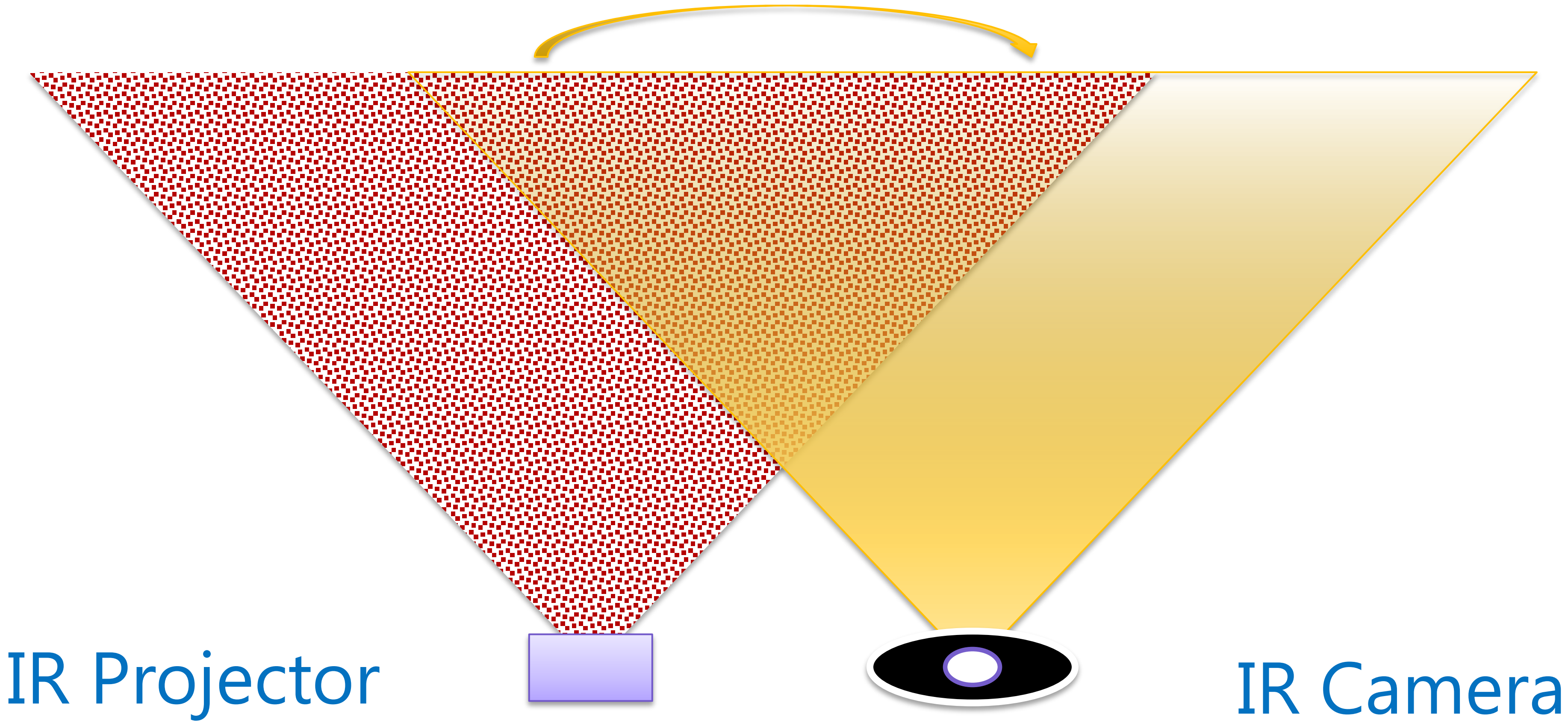
Computer Management
File Action View Help
Computer Management (Local)
System Tools
Task Scheduler
Event Viewer
Shared Folders
Local Users and Groups
Performance
Device Manager
Storage
Disk Management
Services and Applications
Display adapters
DVD/CD-ROM drives
Human Interface Devices
IDE ATA/ATAPI controllers
IEEE 1394 Bus host controllers
Imaging devices
Keyboards
Kinect for Windows
Kinect for Windows Audio Array Control
Kinect for Windows Camera
Kinect for Windows Security Control
Mice and other pointing devices
Monitors
Network adapters
Other devices
Portable Devices
Processors
Sound, video and game controllers
Actions
Device Manager
More Actions

# Runtime

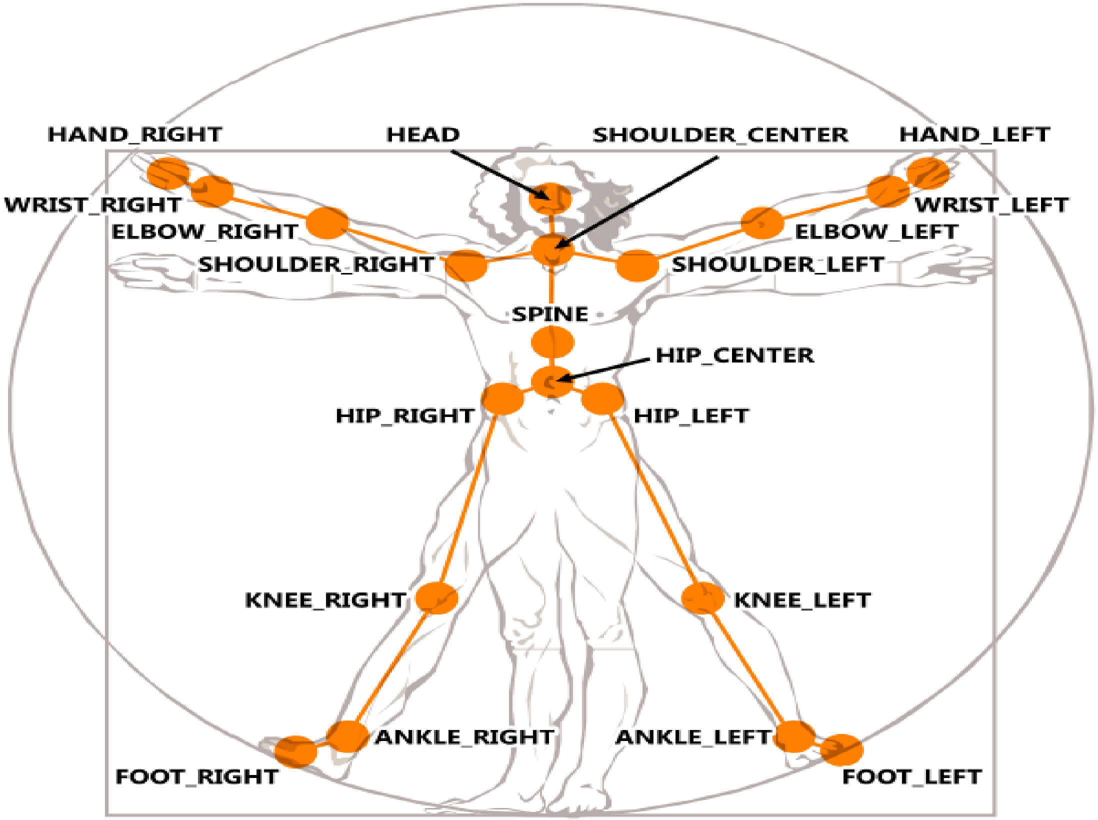
- Sensor discovery, initialization and notification
- Frame delivery supports event-based notification and polling models
- Emphasis on low-latency, low per-frame allocations
- Supports virtual sensors, including test tools and Kinect Studio


Kinect for Windows Developer Toolkit v1.5.1
KINECT for Windows
Resources & Samples
All
Components
Docs
Samples: C#
Samples: C++
Samples: VB
SDKs
Tools
upd
Release Notes and Online Resources
Web page with known issues and links to any updated or new resources.
Documentation
Difficulty: Beginner
Language: C++, C#, VB
upd
SDK Documentation
The Kinect SDK API reference compiled help.
Documentation
Difficulty: Intermediate
Language: C++, C#, VB
new
Human Interface Guidelines
Guidelines on how to design interactions and interfaces for Kinect for Windows applications.
Documentation
Difficulty: Intermediate
Language:
new
Kinect Studio
Kinect Studio enables recording and playback of Kinect sensor data to enable better testing and debugging of your application code.
Documentation
Run
Difficulty: Intermediate
Language:

# Skeletal Tracking – Source Input



# Skeletal Tracking - Output



# K4W Skeletal Tracking Pipeline

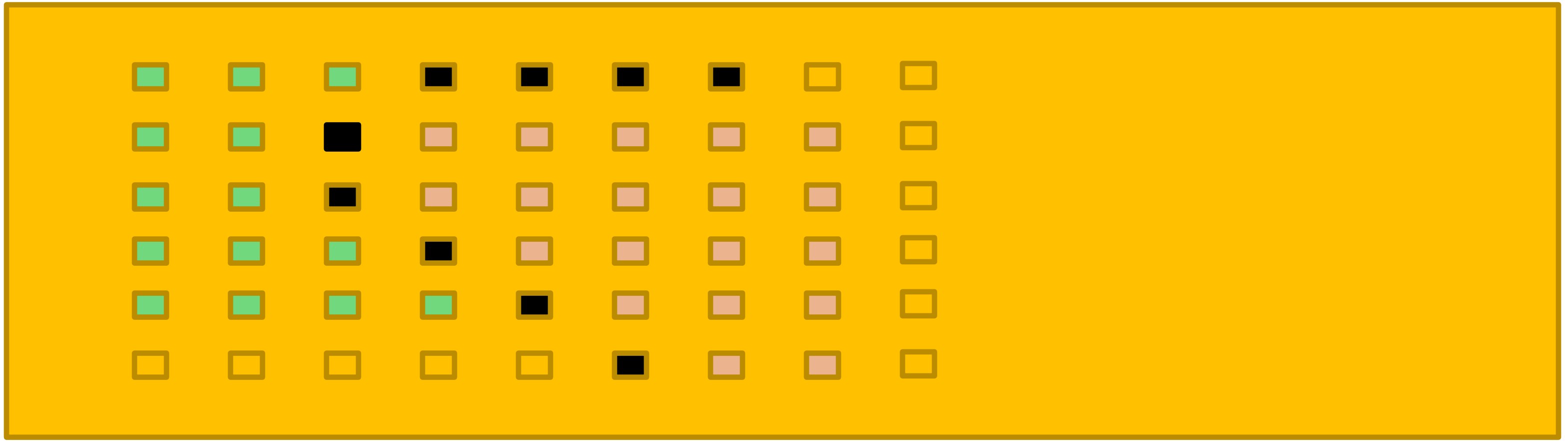
- Stages
  - Depth Source
  - BGR
  - Exemplar
  - Centroids
  - Model fitting
- Data Objects
  - Depth frame
  - Player Mask
  - Classification Map
  - Centroids Tree
  - Skeleton Data

# Depth Computation

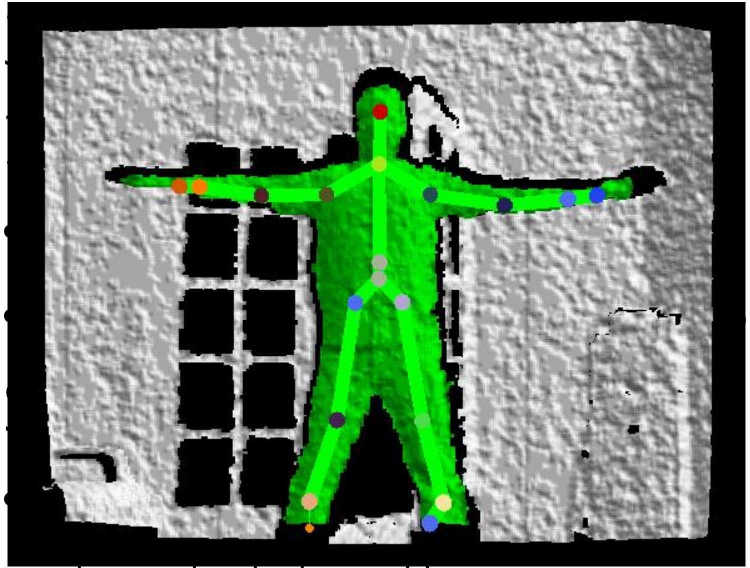
# Depth Map

# Generating Islands

- Voxels without any connections to either left or top neighbor define boundaries between islands.



# Output: Depth and Segmentation Map



- …m: …
- …n m…
- 1 – skeleton 0
- 2 – skeleton 1
- …

# Real-Time Human Pose Recognition in Parts from Single Depth Images

Jamie Shotton Andrew Fitzgibbon Mat Cook Toby Sharp Mark Finocchio
Richard Moore Alex Kipman Andrew Blake
Microsoft Research Cambridge & Xbox Incubation

## Abstract

*We propose a new method to quickly and accurately predict 3D positions of body joints from a single depth image, using no temporal information. We take an object recognition approach, designing an intermediate body parts representation that maps the difficult pose estimation problem into a simpler per-pixel classification problem. Our large and highly varied training dataset allows the classifier to estimate body parts invariant to pose, body shape, clothing, etc. Finally we generate confidence-scored 3D proposals of several body joints by reprojecting the classification result and finding local modes.*

*The system runs at 200 frames per second on consumer hardware. Our evaluation shows high accuracy on both synthetic and real test sets, and investigates the effect of several training parameters. We achieve state of the art accuracy in our comparison with related work and demonstrate improved generalization over exact whole-skeleton nearest neighbor matching.*

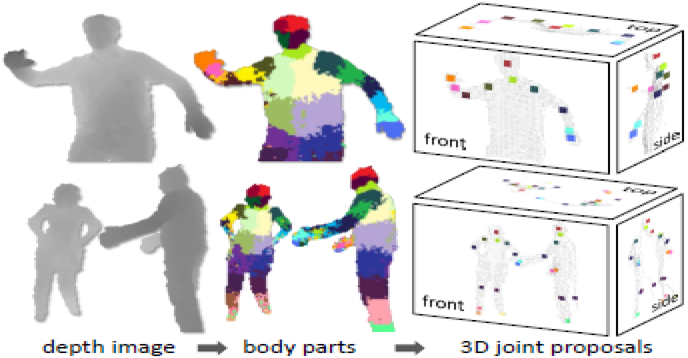



Figure 1. **Overview.** From an single input depth image, a per-pixel body part distribution is inferred. (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals). Local modes of this signal are estimated to give high-quality proposals for the 3D locations of body joints, even for multiple users.

joints of interest. Reprojecting the inferred parts into world

# Machine Learning Classification

- Input
  - Point cloud: pixels in 3-D space
    - Focus on player masks
- Classification problem
  - Which body part each point belongs to?
- Features?
  - Arrangement of body parts in space

# Training and Features


Figure 2. **Synthetic and real data**. Pairs of depth image and ground truth body parts. Note wide variety in pose, shape, clothing, and crop.

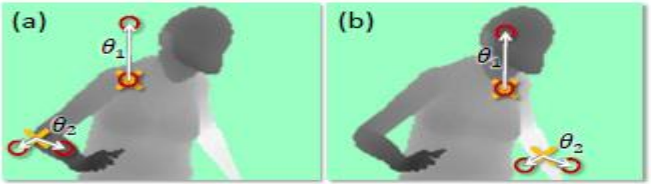


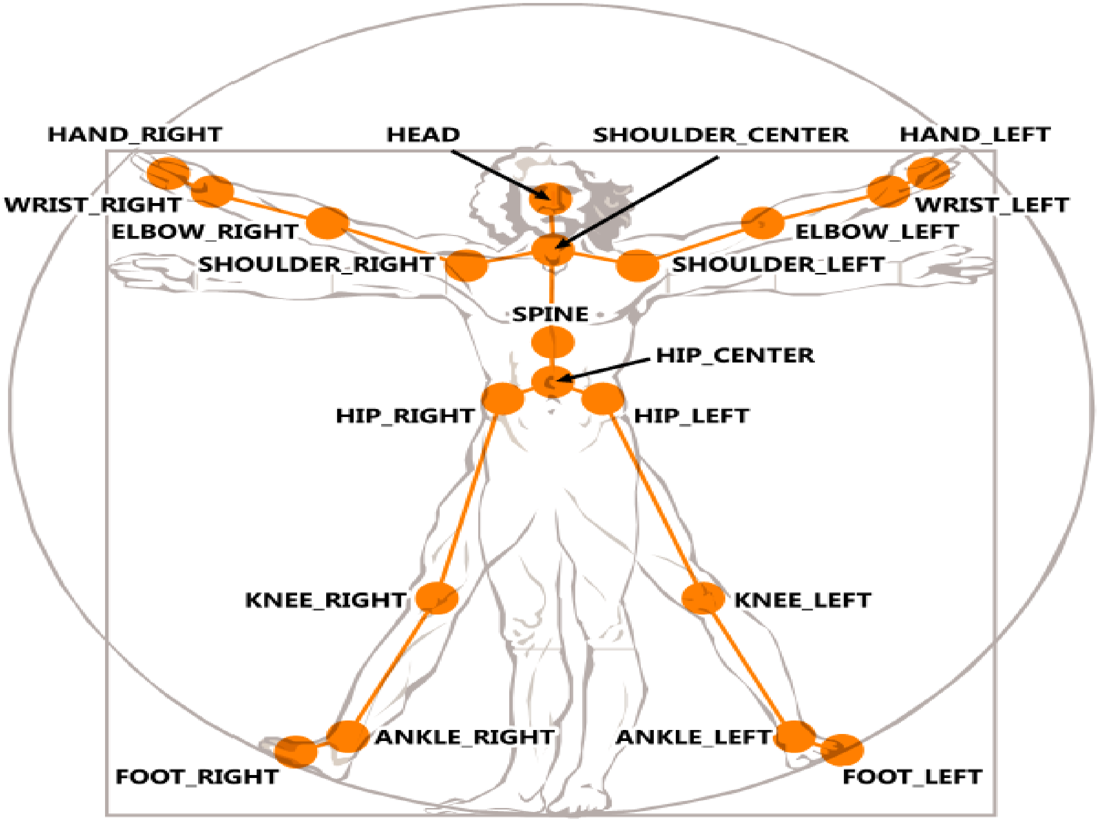

Figure 3. **Depth image features.** The yellow crosses indicates the pixel $\mathbf{x}$ being classified. The red circles indicate the offset pixels as defined in Eq. 1. In (a), the two example features give a large depth difference response. In (b), the same two features at new image locations give a much smaller response.

# Let's See It Working…

- Video from CVPR 2011 paper
  - *Real-Time Human Pose Recognition in Parts from Single Depth Images*
    Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake
    *Microsoft Research Cambridge & Xbox Incubation*

IEEE Computer Vision and Pattern Recognition

# Skeletal Tracking - Output



# Known Issues (Future Work)

# Conclusions

- Doing software for living is <10% coding
- Successful software results from knowledge of several areas
- Developing software is far easier than developing software teams