



# Audio and Video Research in Kinect

Alex Acero  
Research Area Manager  
Microsoft Research



# Introducing Kinect for Xbox 360

**You are the controller. No gadgets, no gizmos, just you!**

Kinect brings games and entertainment to life in extraordinary new ways without using a controller. Imagine controlling movies and music with the wave of a hand or the sound of your voice. With Kinect, technology evaporates, letting the natural magic in all of us shine.

## You are the controller

**Everything you need to know to get started with Kinect.**

### **You're Ready to Play**

Easy to use and instantly fun, Kinect gets everyone off the couch and moving, laughing and cheering. See a ball? Kick it. Want to join a friend in the fun? Simply jump in.







# ***FASTEST-SELLING CONSUMER ELECTRONICS DEVICE***

8M units / First 60 days on sale (11/4/2010 to  
1/3/2011)

= 133,333 units per day



# Kinect Sensor



RGB  
camera

infra-red  
camera

infra-red  
projector



Microphones  
Motor  
USB

**For only \$150 !**

# Outline

- Kinect in the PC
- Audio processing
- Depth Sensor
- Skeletal Tracking
- Head pose & facial expression tracking

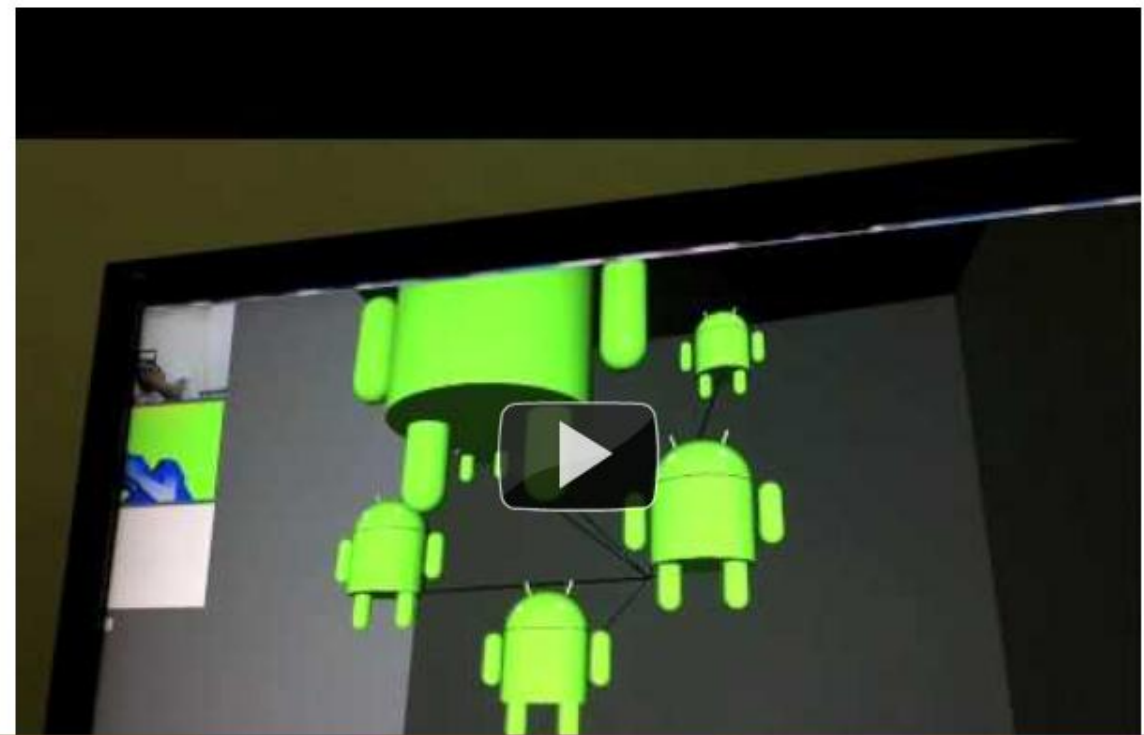
# KINECTHACKS

- HOME
- FORUMS
- FAQ
- GUIDES
- TOOLS AND RESOURCES
- ABOUT

## Crazy Head Tracking Androids

December 4th, 2010 Madhav K 4 Comments and 6 Reactions

Sittiphol Phanvilai @ Hua Lampong Co.,Ltd has implemented head tracking using Kinect and creates crazy 3D effects with android dolls. This is a modification of their earlier Kinect VR project which had spheres instead of androids.



Search



### FORUMS



### FACEBOOK

[Sign Up](#)

Create an account or [log in](#) to see what your friends like.

[Kinect Hacking on Facebook](#)

765 people like Kinect Hacking

Dan

Graham

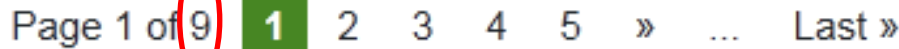
Eric

Stephan

Lawrence

# ~ 90 Projects as of 12/04/2010

58 pages as of 5/18/2011



Page 1 of 9 1 2 3 4 5 » ... Last »

Every few hours new applications are emerging for the Kinect and creating new phenomenon that is nothing short of revolutionary.

- Quote from KinectHacks.net



# 3D Video Capture

# Music Video

# Navigational Aids for the Visually Impaired

# NUI: Natural User Interaction

- **Testing the Kinect as a Navigation Device**
  - [YvesQuemener](#) is using Kinect as a navigation device to navigate a 3D environment.
- **Controlling Google Maps With Kinect**
  - [wafflejock](#) is controlling Google Maps with hand gestures with the Kinect. The program uses the kinect running on an Ubuntu box and forwards the depth/rgb data across the network to a windows machine running the Flex/Flash Builder client.
- **TUIOKinect**
  - Tuiokinect tracks simple hand gestures using the Kinect controller and sends control data based on the [TUIO](#) protocol. This allows the rapid creation of gesture enabled applications with any platform or environment that supports TUIO. Check out this extensive list of TUIO enabled software for further information: <http://www.tuio.org/?software>
- **Convert Monitor or TV into a Touchscreen**
  - Kinect tracks the finger and detects its proximity to the monitor or TV and registers the touch. This program shows a laptop connected to a TV and used with a Kinect to make the TV a touchscreen.
- **Kinect as mouse in Windows 7**
  - Liam Nichols, more commonly known as [l14m333](#), the creator of the [TV-touchscreen demo](#) we covered earlier, has released the source code of a very promising project. It's a Kinect mouse driver for Windows 7 – it allows Windows users to control their mouse movements and clicks with a Kinect.
- **Full Multitouch implemented on Windows 7 using kinect**
  - Wolfgang Herfurtner, the CEO of Evolve, a technology driven company, that offers multitouch and gesture computing solutions has implemented full multitouch on Windows 7 using kinect. Check out the awesome video.
- **Two more Kinect web browsing videos**
  - [Dustin O'Connor](#) has released a video showing him browsing the web using gestures and voice recognition.

# Virtual/Augmented Reality

- **Virtual Hurling With Xbox Kinect**

- Diarmuid Wrenne shows how he can extract the rotation of objects seen by the kinect and use that rotation to change the orientation of virtual objects within the Box2d space to create a virtual bat out of a real one!!.

- **3D Object Manipulation using Kinect**

- Dominick D'Aniello has created a program to manipulate 3D objects using his hands with a Kinect. He will reach out to the 3D object with his hands, grab the object, turn, rotate and play with the object as if it were a real solid object in his hands. This could really have great applications in games & 3D modelling.

- **Kinect Green Screen**

- Green Screen is used in Movies where actors shoot in front of the screen and the green is replaced by a different background and make it look like the actors are somewhere else.

- **Playing basketball with a hacked Kinect**

- Dustin O'Connor has been pumping out new demos in these last days. Here he is playing basketball in a self made game. It was made with the KinectTools for Quartz Composer plugin and the Unity3D game engine

- **Kinect 3D Banana Slapping Game**

- Dennis Ippel has created a homemade Kinect 3D banana slapping game. The accuracy is still not as good as retail games but it looks like it can be a fun game when fully developed. Enjoy the video...



# Digital Visual Art

- **Kinect Art – Hand Printing**

- OpenKinect group in Korea have created this Kinect art video where they are doing hand printing.

- **Hollow Man**

- [TakayukiFukatsu](#) has created Hollow Man, the Optical Camouflage program built with Kinect and Openframeworks. How exactly it works is unknown but the effect is cool.

- **DaVinci Goes Touchless With Kinect**

- Razorfish has ported the [DaVinci: Microsoft Surface Physics Illustrator](#) to be used with the Kinect. The prototype they have created is stunning. This seems like one of the best Kinect apps till date. The video is spectacular.

# Digital Audio/Music

- **Audio Modulated Point Clouds**

- Kinect is not only about visual recognition but audio is also a great part of the Kinect experience which brings computing and gaming to the human senses. Here is a demonstration where the software has calibrated the Kinect to create ripples in the 3D point clouds based on the tone of music.

- **Chinese music using Kinect**

- Tim88588585, a fellow member of [our forum](#) (are you registered yet?), have modified the MIDI controller code from Bextan [we wrote about](#) a couple of days ago and made it sound way more beautiful. His is playing classical Chinese music and he's actually quite skilled.

- **OpenKinect Piano**

- Developers from all over the world are working their asses off hacking their Kinects. This video shows a man playing a virtual piano which he can see on his pc, but is using his feet to create music with it. The newer versions of the openKinect drivers are showing a very good amount of accuracy which is making these kinds of applications work very well and the future seems very promising.

- **Playing GarageBand instruments with Kinect**

- ptone805, the creator of the [Kinect controlled Christmas lights](#) demo, has uploaded a new video. This time he is playing instruments with the Mac OS X iLife application Garageband. Kinect is connected to Garageband like it was a normal MIDI device.

# Human Activity Analysis

- **Crazy Head Tracking Androids**

- **Sittiphol Phanvilai @ Hua Lampong Co.,Ltd** has implemented head tracking using Kinect and creates crazy 3D effects with android dolls. This is a modification of their earlier Kinect VR project which had spheres instead of androids.

- **Real-time People detection and tracking with multiple Kinect cameras**

- **alexutubeutube** has created a program for realtime people detection and tracking with multiple Kinects. Cameras are connected to a single laptop. People are detected given the noisy depth measurements from multiple kinects. Stationary and moving people are detected in day or night time or with highly varying lighting conditions.

- **Kinect Skeleton Test (new)**

- CADET ( Center for Advances in Digital Entertainment Technologies) have been working to write an open source sdk for skeleton tracking. Here are some preliminary tests they have conducted and uploaded the video.

# Other Applications

- **Object Recognition using Kinect on the PC**

- Solid objects are held in front of Kinect and they are recognized. The names of the objects have been taught to the Kinect before only. The process of teaching Kinect to recognize objects is not very difficult. They have used OpenKinect(drivers) + OpenCV(image processing and recognition) + FestVox (speech synthesis).

- **3D object scanning using Kinect**

- An application to scan and create 3D objects by capturing in space real world solid objects. In this video he uses his bike and creates 3D bike on his computer with great detail...

- **Interactive Puppet Prototype with Kinect**

- Emily Gobeille & Theo Watson of design-io.com have created an interactive puppet by using skeleton tracking on the arm and determining where the shoulder, elbow, and wrist is, using it to control the movement and posture of the giant funky bird!

- **Hacked Kinect on a mobile robot does 3D mapping**

- Philipp Robbel of Personal Robots Group of MIT has installed a hacked kinect on a robot and used it to create 3D reconstruction of the environment by making the robot move around.
- Also the robot uses gestures for taking directions. Its uses kinect for 3D mapping. Kinect streams the depth and color images to a remote host for SLAM and 3D map processing.

- **Crazy Mashup of Kinect Ideas**

- [jvcleave](#) has made a mashup of many great ideas for Kinect and made this awesome video. After watching the video you can download the [source code and get more details here](#).



## Non-commercial Windows software development kit

### Access to deep Kinect system information

- Depth data
- Audio
- Direct control of the Kinect sensor
- System API



# Outline

- Kinect in the PC
- Audio processing
- Depth Sensor
- Skeletal Tracking
- Head pose & facial expression tracking

# Kinect: Voice Control



*Harry Potter and the Sorcerer's Stone* available on Zune Marketplace

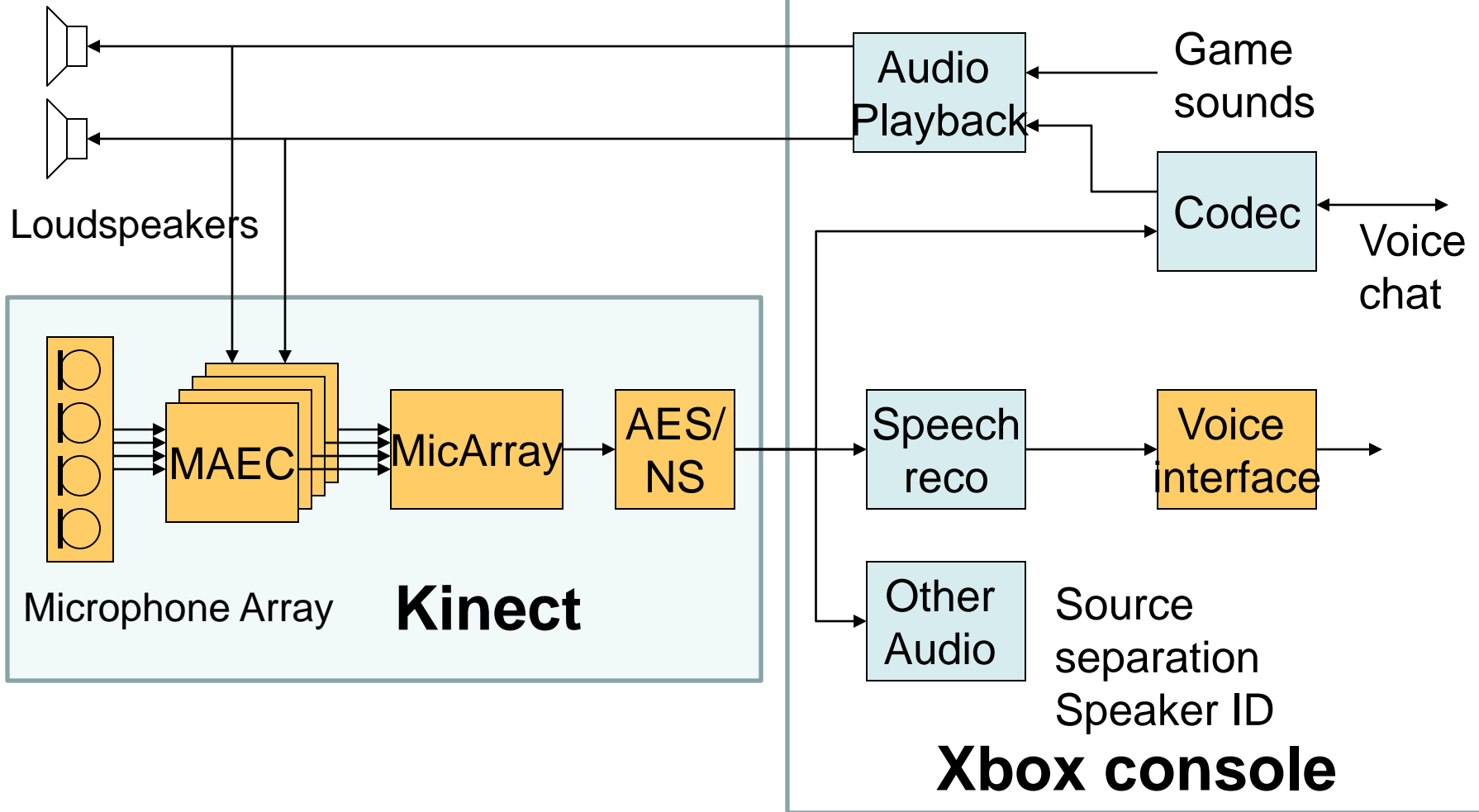
All titles available on Zune Marketplace  
HARRY POTTER characters, names and related indicia are trademarks of and © Warner Bros. Entertainment Inc. Harry  
Potter Publishing Rights © JKR, Harry Potter, and the Sorcerer's Stone © 2001 Warner Bros. Entertainment Inc. All Rights Reserved.

# Kinect: Speech recognition

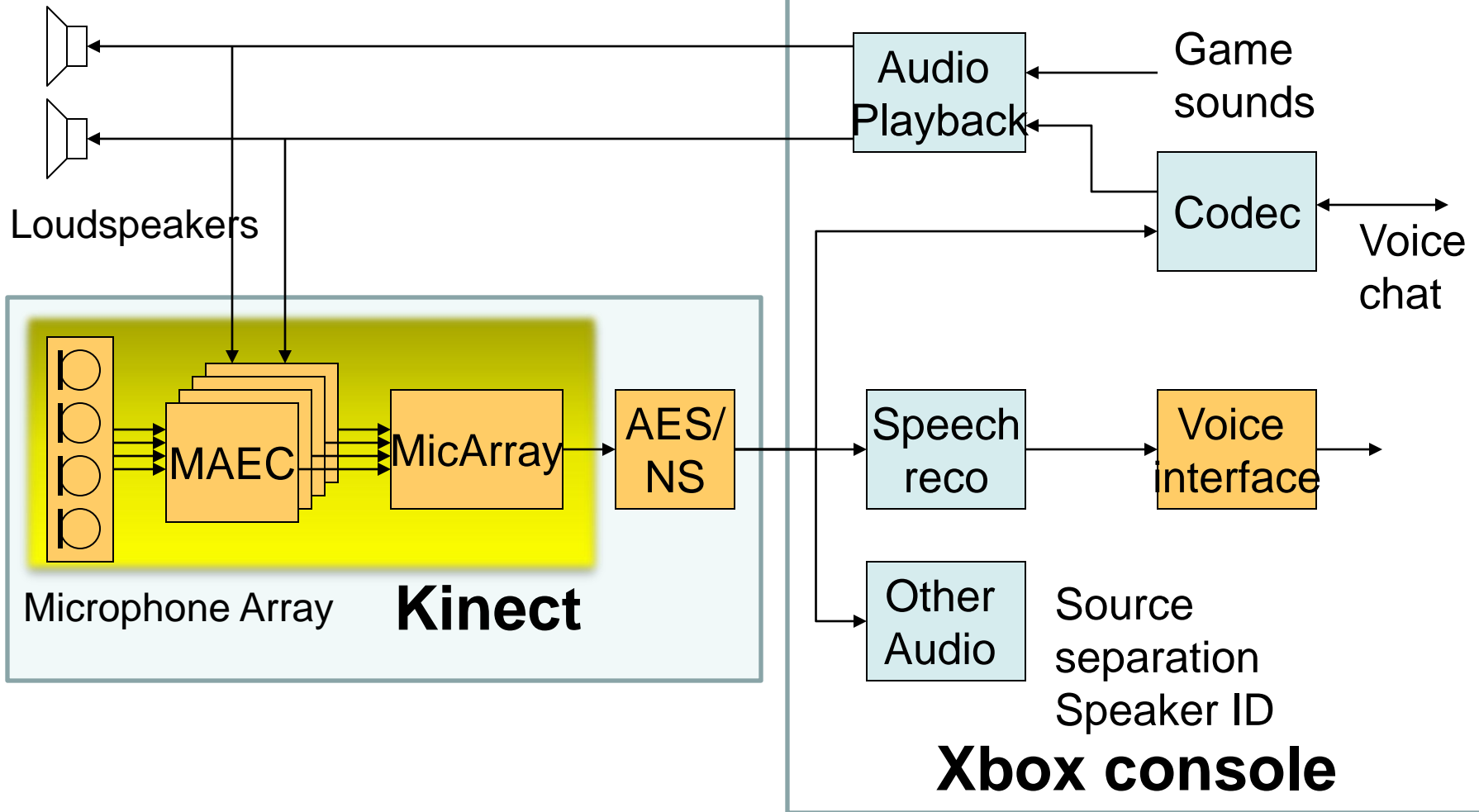
- Speech recognition
  - Complementary to gesture
  - Want to talk to your animal
  - Voice control without on-screen buttons
  - Access long lists
- From headsets to hands free
  - Needs relatively good quality audio!
  - Loud gaming sounds from Xbox
  - Noise and reverberation in the room



# Audio Stack



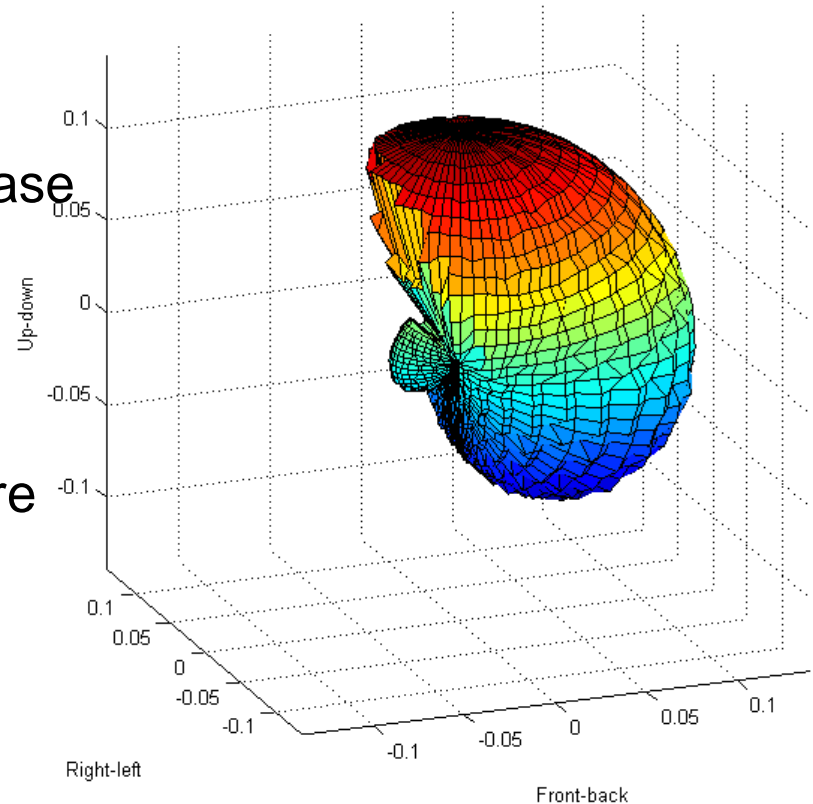
# Audio Stack





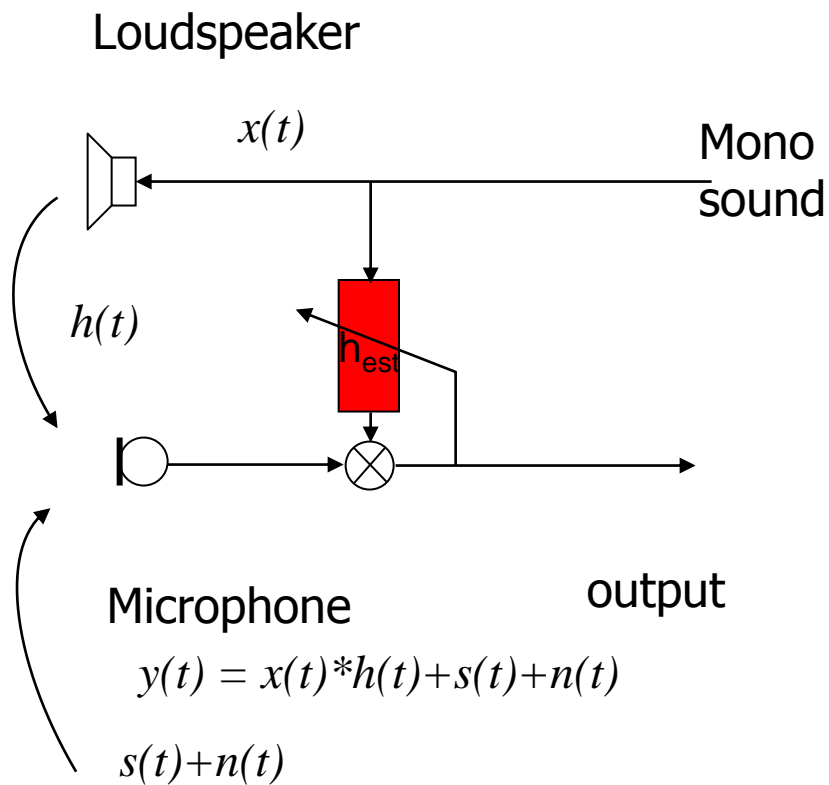
# Directional Microphones

- Acoustical design
  - Using the enclosure shape to increase the microphones directivity
- Optimized microphone array geometry
  - Non-equal spacing, covers the entire bandwidth

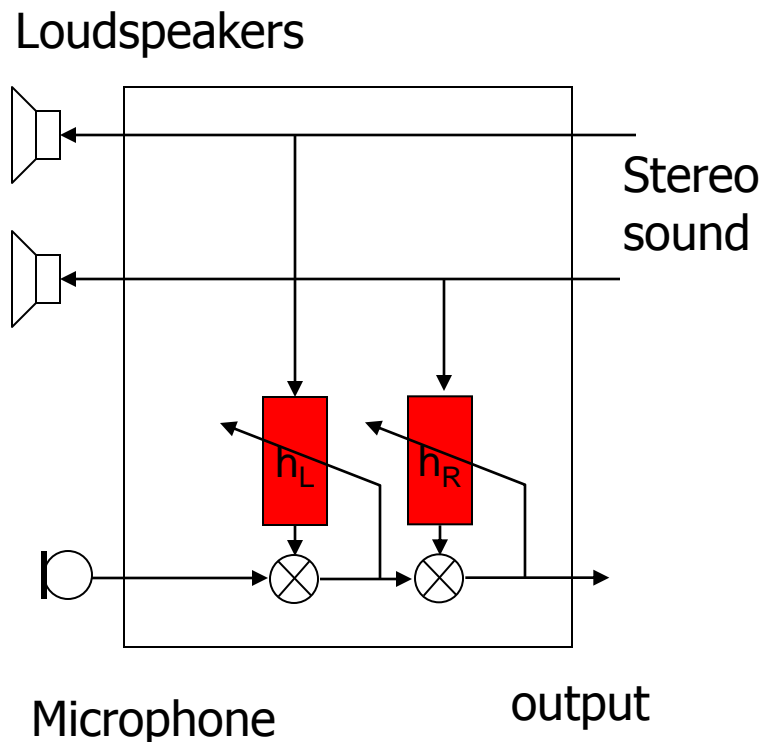


# Mono Acoustic Echo Cancellation

- Acoustic echo cancellation
  - Mono AEC – part of each speakerphone



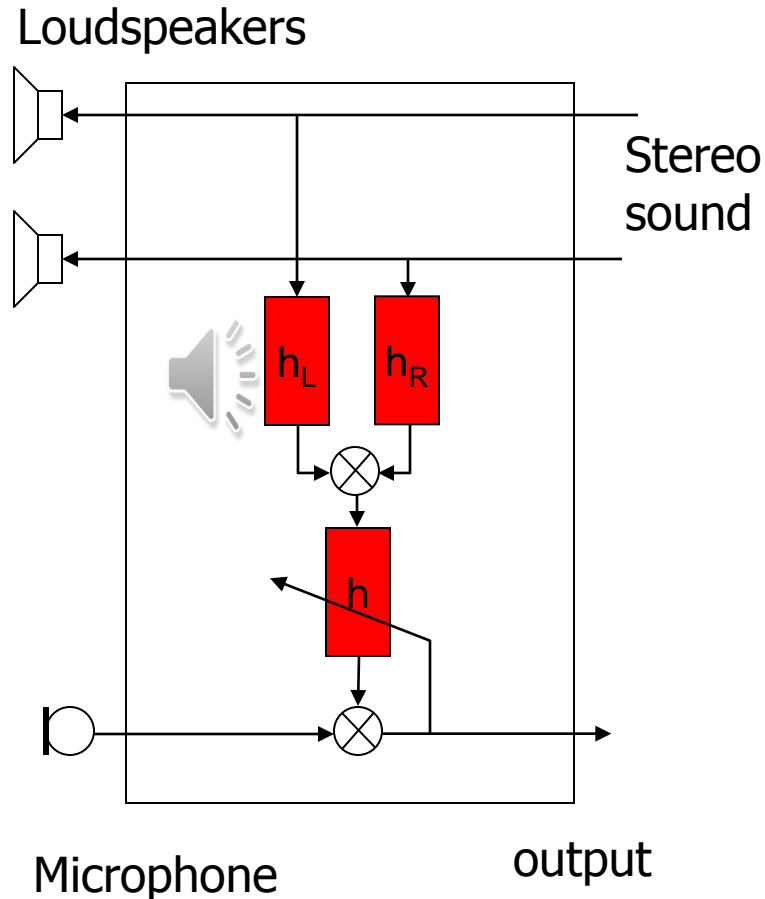
# Multichannel Acoustic Echo Cancellation



- Acoustic echo cancellation
  - “Stereo AEC has a non-uniqueness problem that presents a fundamental limitation” (Sondhi et al. Bell Labs, 1995)

# Multichannel Acoustic Echo Cancellation

Ivan Tashev 2008

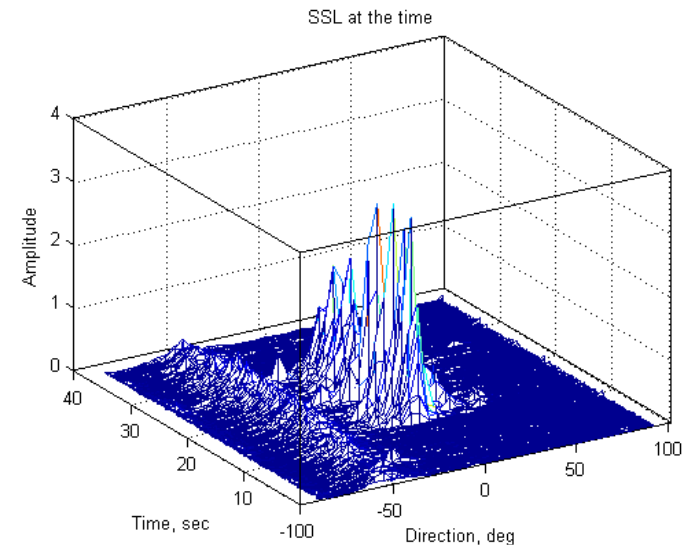
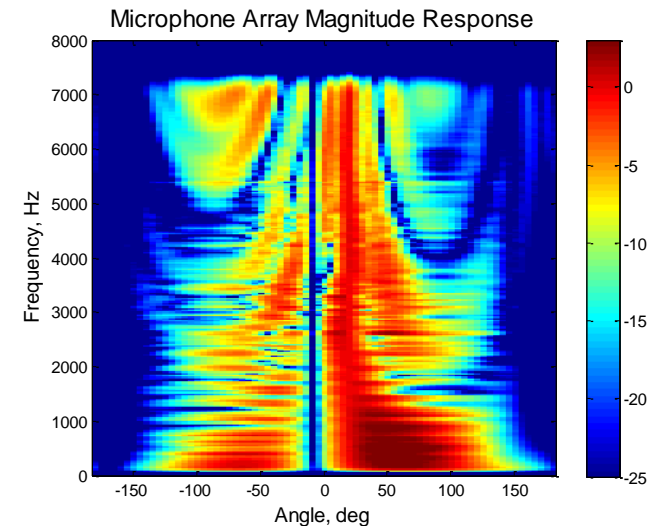


- Acoustic echo cancellation
  - “Stereo AEC has a non-uniqueness problem that presents a fundamental limitation” (Sondhi et al. Bell Labs, 1995)
- Multichannel AEC
  - Use calibration pulses, lock mixing filters, use one adaptive filter
  - Reduces 15-20 dB echo
  - Entire audio pipeline: ~35 dB

# Microphone array processing

Ivan Tashev 2008

- Adaptive beamformer
  - Acts as a steerable directional microphone
  - Can suppress interferers as well
  - Reduces 3-6 dB noise
- Spatial filtering
  - Sound source localization per frequency bin
  - Suppresses sounds outside desired direction range
  - Suppresses 6-12 dB noise



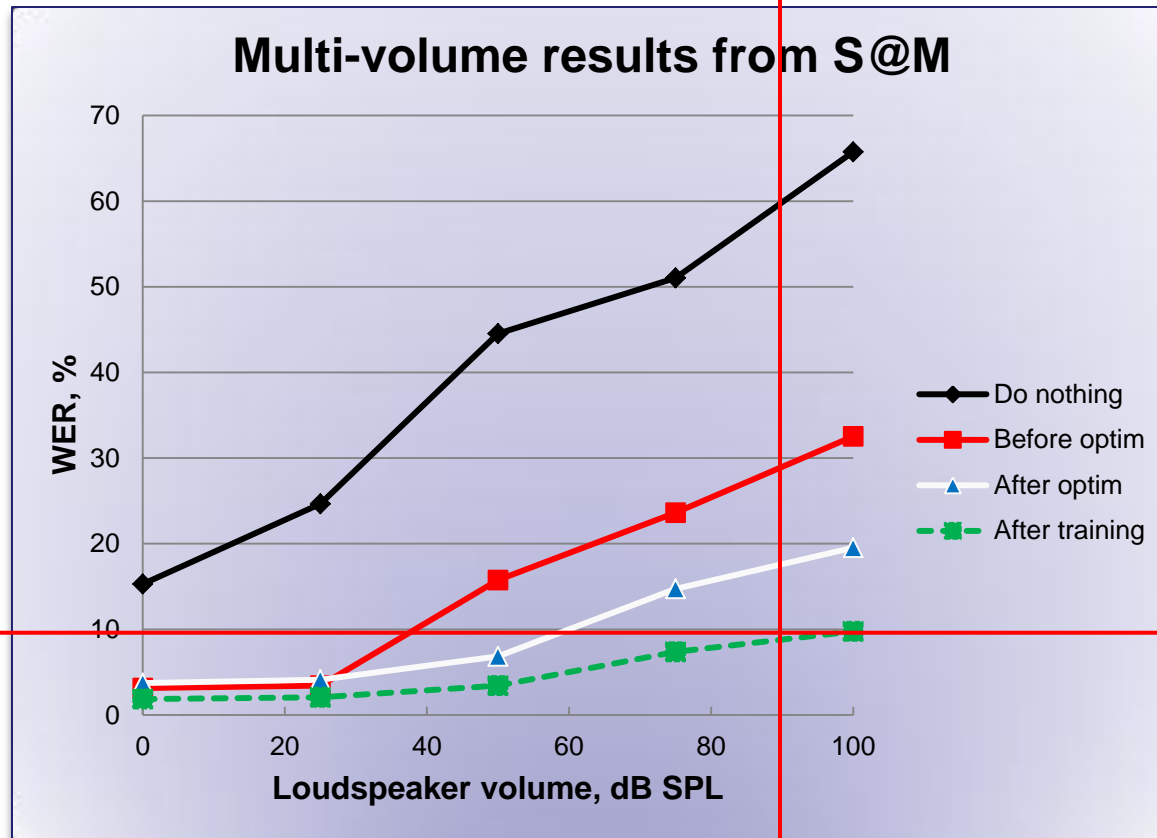


# End-to-end optimization

Ivan Tashev 2008

- A chain of optimal processing blocks is suboptimal
- Optimization criterion:
  - Perceptual Evaluation of Sound Quality (PESQ)
- 25 parameters for optimization
  - Time constants, thresholds
- Parallelized processing on cluster
  - Large data corpus
- Results with speech recognizer

# End-to-end optimization

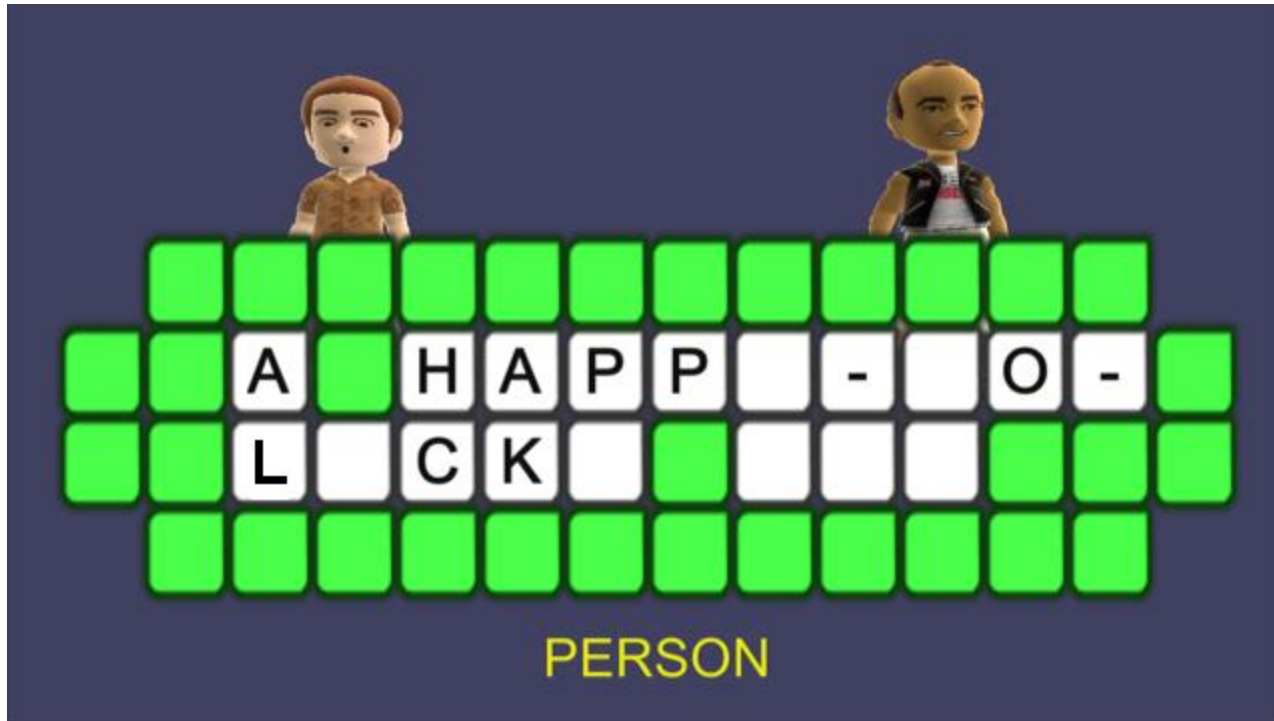


Supported levels  
up to here

Speech NU  
good  
up to here

# Utterance Verification in Games

YC Ju 2010



- Engines will typically assign similar scores to “A Happy Go Lucky **Guy**” and “A Happy Go Lucky **Man**”
- **Word-dependent** utterance verification

# Outline

- Kinect in the PC
- Audio processing
- Depth Sensor
- Skeletal Tracking
- Head pose & facial expression tracking

# Motion capture



[Vicon]

- ✓ very accurate
- ✓ high frame rate

- ✗ suit / sensors
- ✗ expensive



[Xsens]

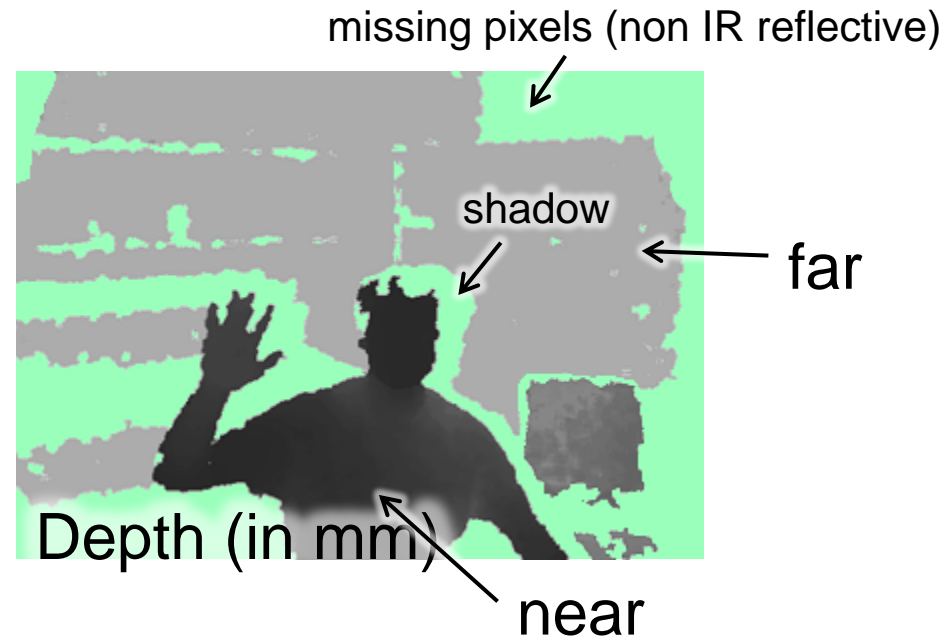
- ✗ large space
- ✗ calibration



# Depth cameras



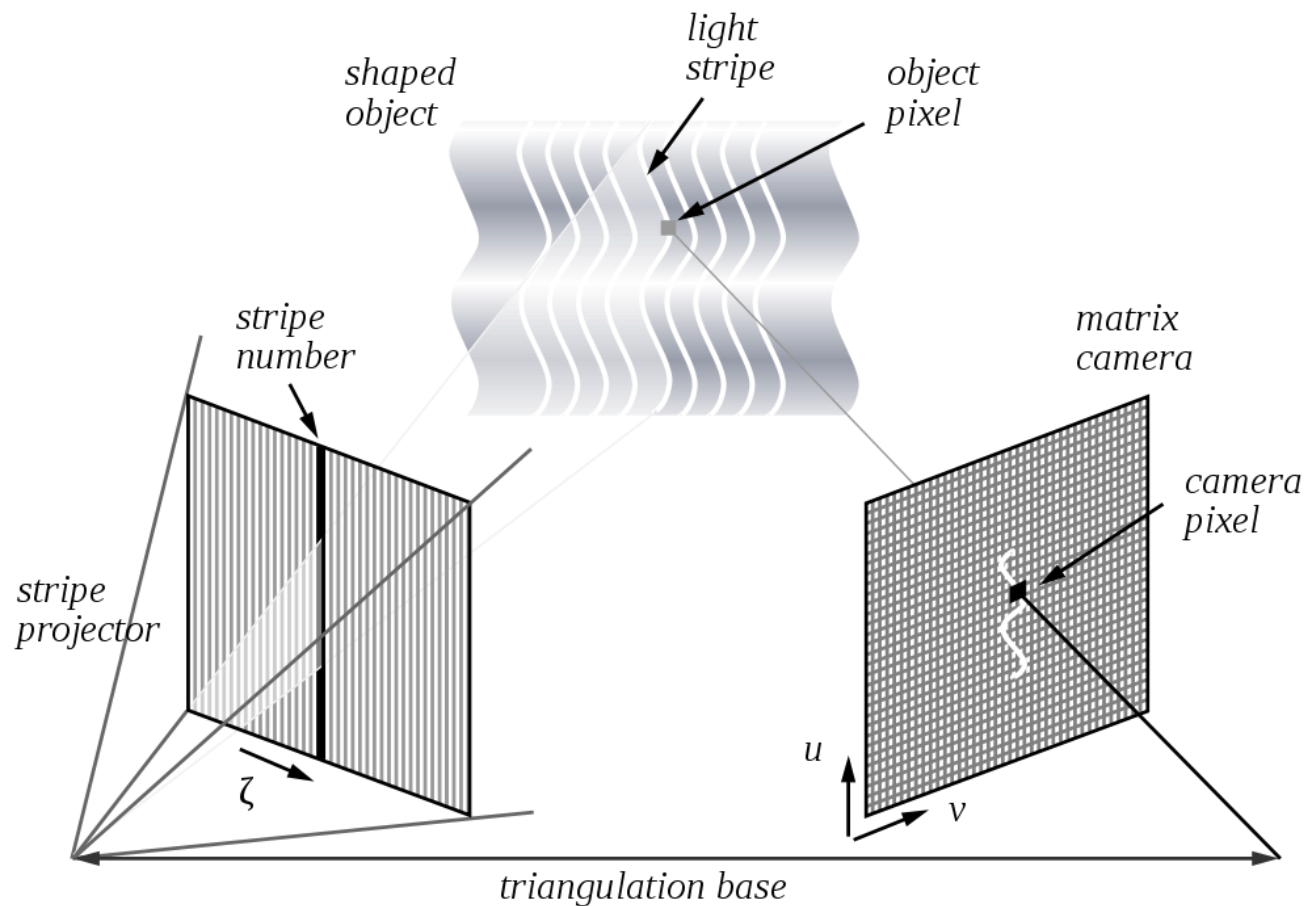
- Technology
  - structured IR light



- ✓ cheap, fast, accurate    ✗ missing pixels, shadows

# How it works?

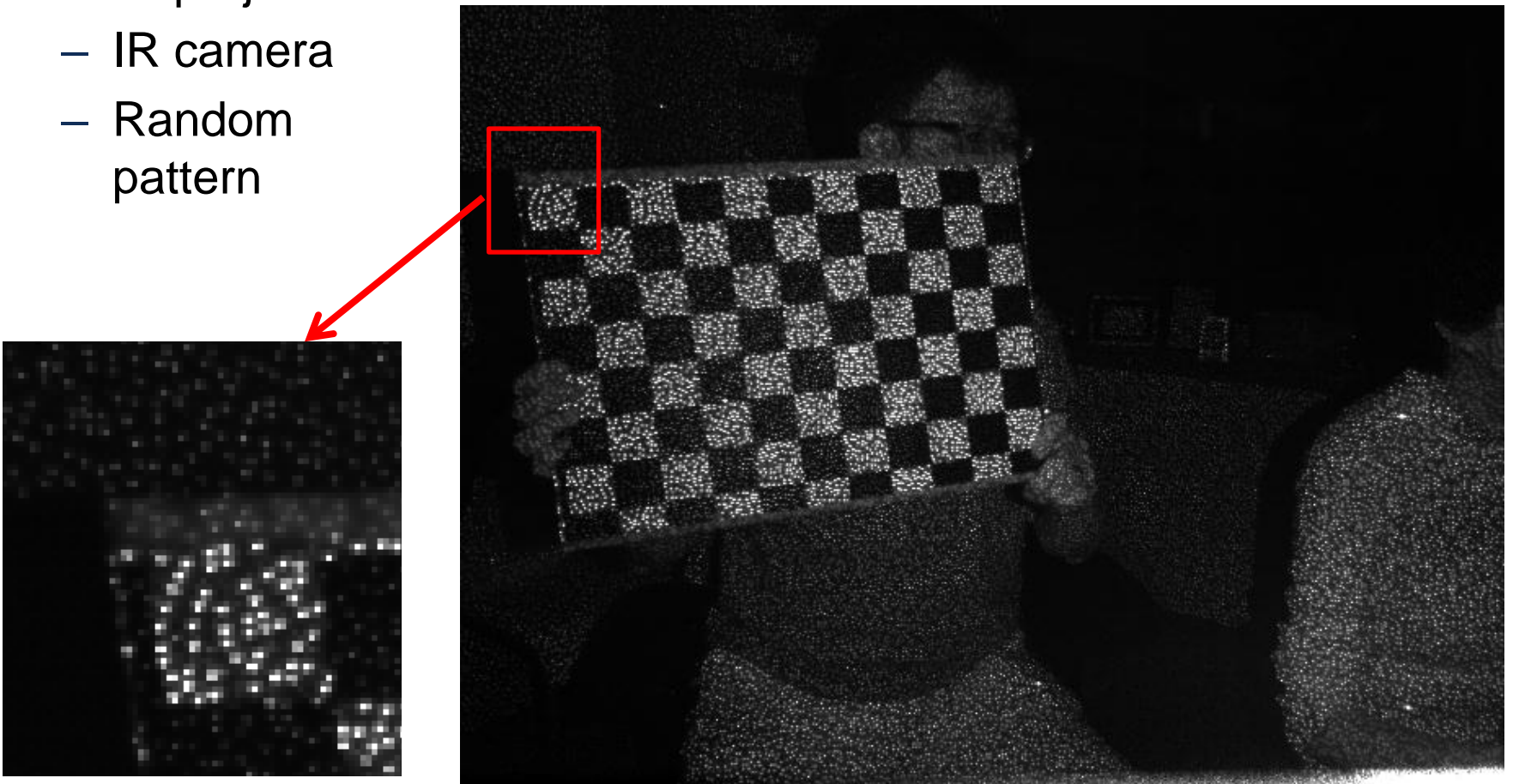
- **Structured light 3D scanner**





# How it works? Kinect Sensor

- Modified structured light 3D scanner
  - IR projector
  - IR camera
  - Random pattern



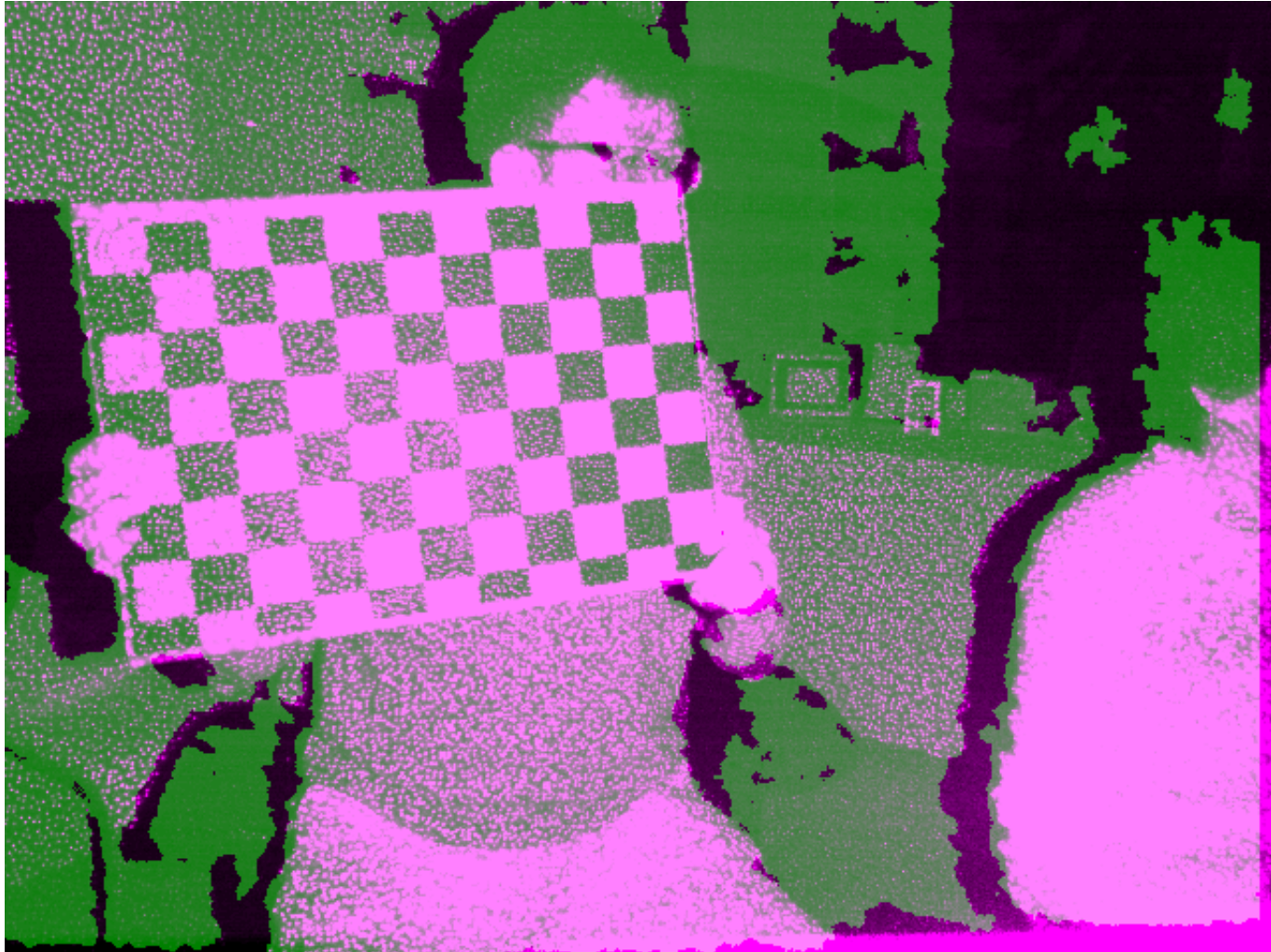


# Matching & Depth Map

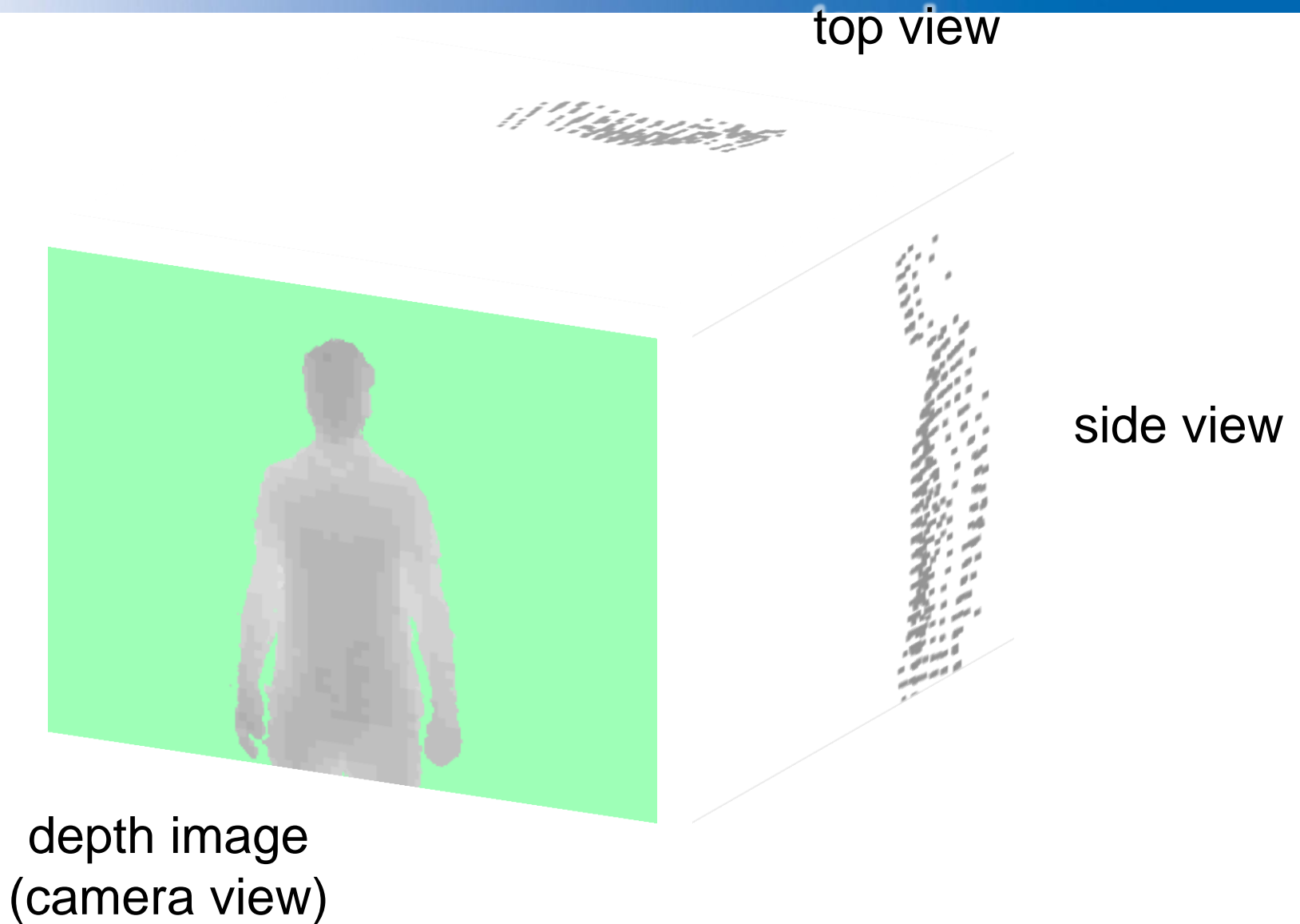
- Correlation



# Overlay of Depth Map on IR Image



# Depth cameras



# RGB vs depth for pose estimation

## RGB

- ✗ Only works well lit
- ✗ Background clutter
- ✗ Scale unknown
- ✗ Clothing, skin colour

much  
easier  
with  
depth!



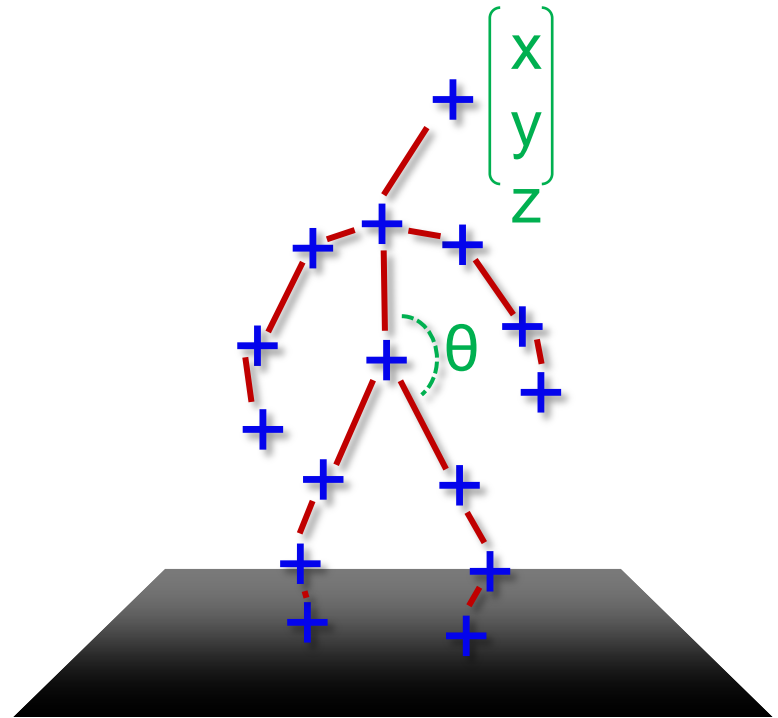
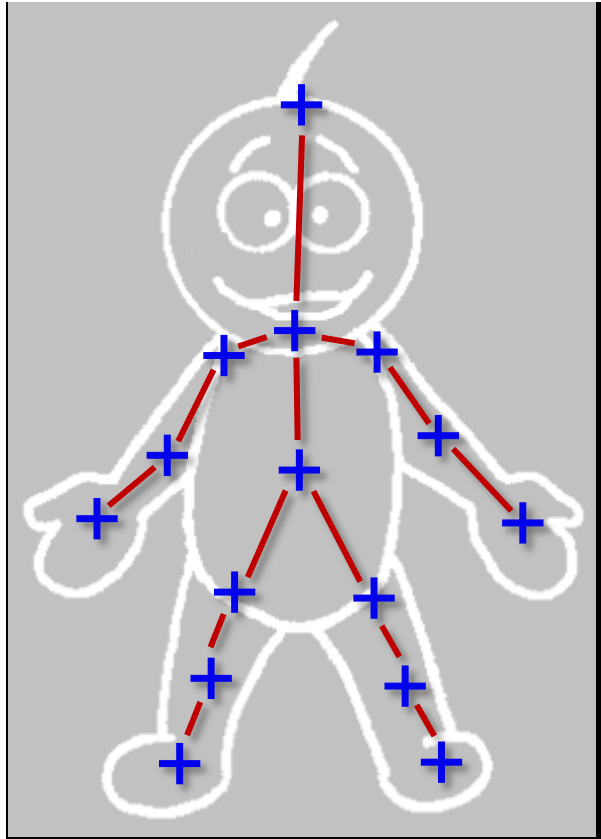
## DEPTH

- ✓ Works in low light
- ✓ Person 'pops' out from bg
- ✓ Scale known
- ✓ Uniform texture
  - ✓ easy to simulate
- ✗ Shadows, missing pixels

# Outline

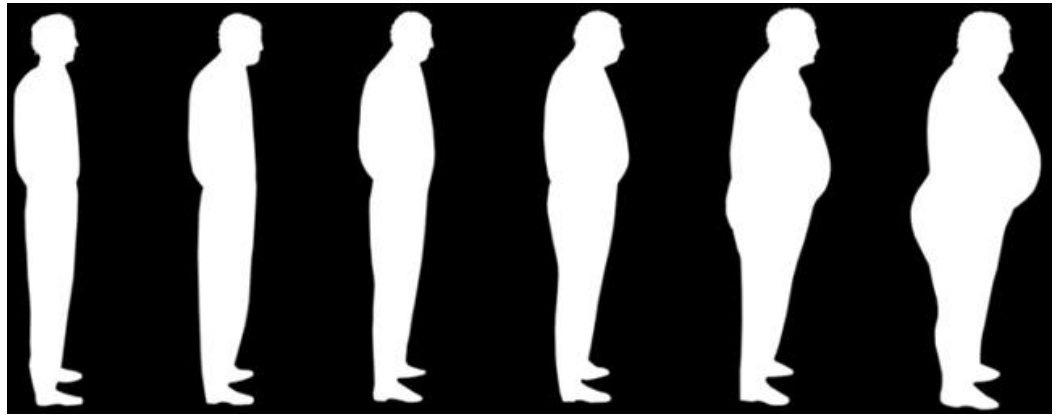
- Kinect in the PC
- Audio processing
- Depth Sensor
- **Skeletal Tracking**
- Head pose & facial expression tracking

# Human pose estimation



Kinect tracks 20 body joints in real time.

# Why is it hard?



# Skeletal Tracking (Jamie Shotton & MSRC)



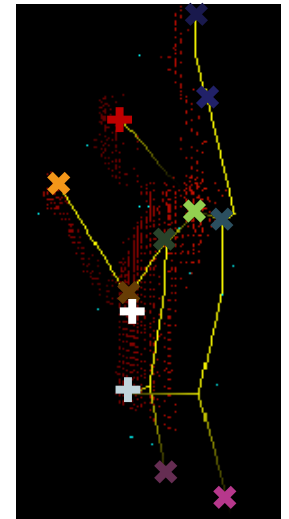
1. capture  
depth image



2. infer  
body parts



3. hypothesize  
body joints



4. track skeleton  
(3D side view)



# Body part recognition



**input depth image**



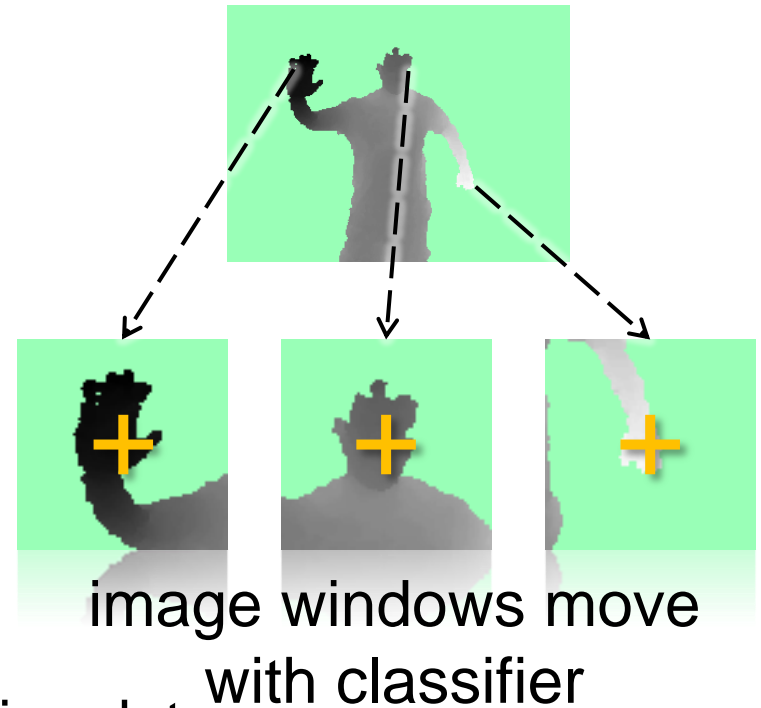
**inferred body parts**

# Classifying pixels

- Compute  $P(c_i | w_i)$ 
  - pixels  $i = (x, y)$
  - body part  $c_i$
  - image window  $w_i$

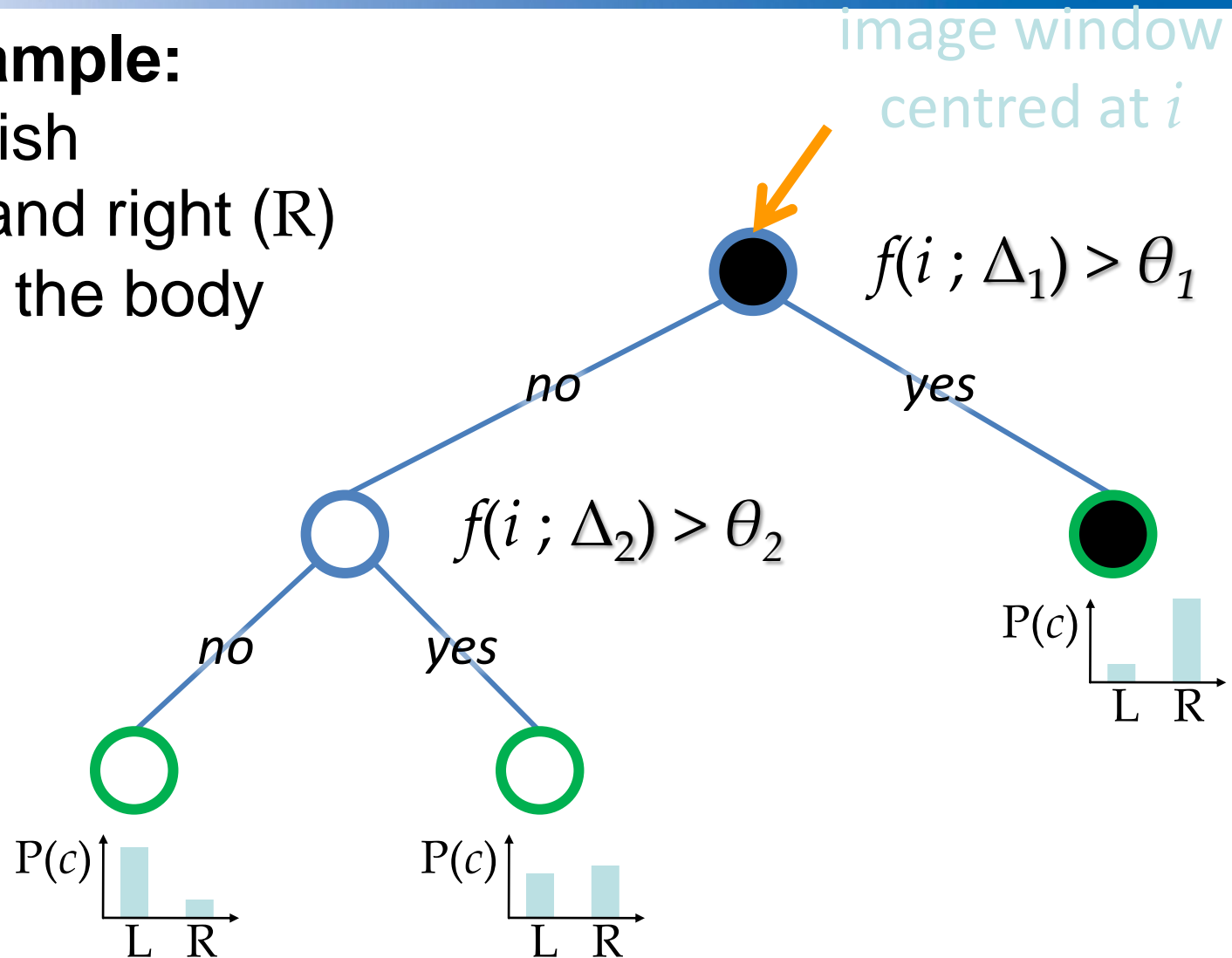
- Discriminative approach

- Learn classifier  $P(c_i | w_i)$  from training data



# Decision tree classification

**Toy example:**  
distinguish  
left (L) and right (R)  
sides of the body



# Depth of trees

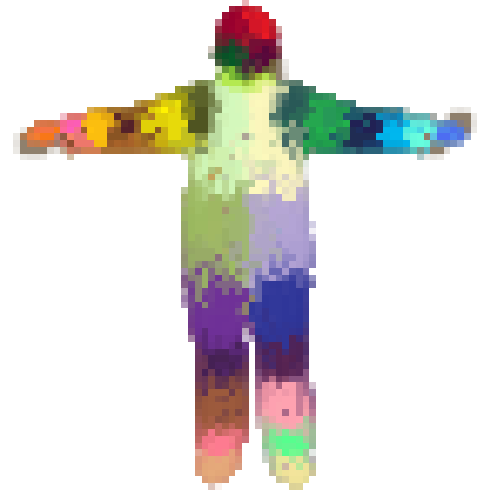
input depth



Correct parts  
(ground truth)



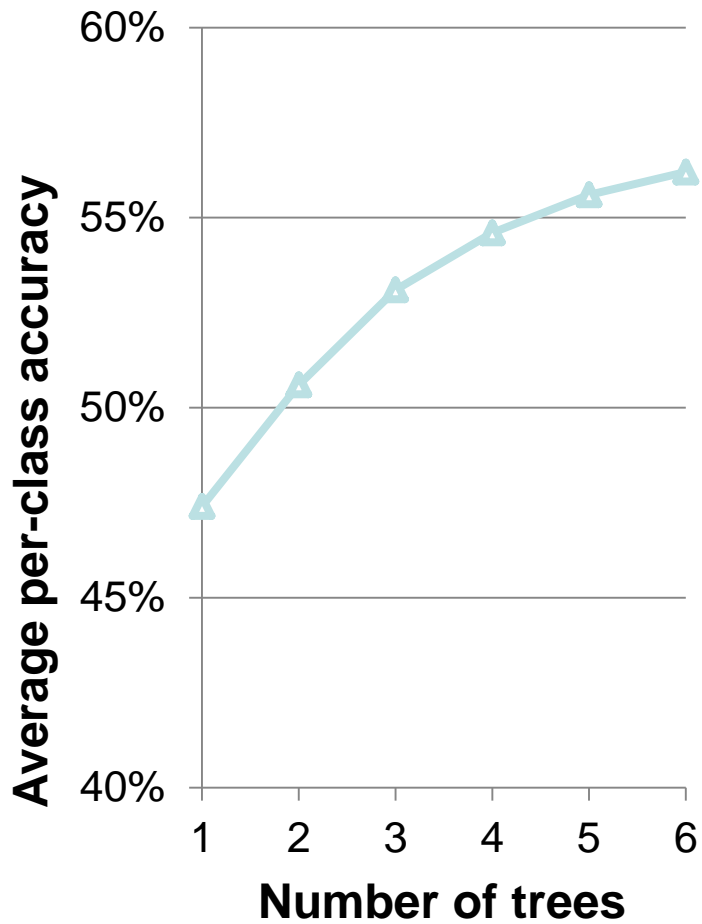
inferred parts (soft)



depth 18



# Number of trees



input



ground truth



inferred body parts (most likely)

1 tree



2 trees



3 trees



# Body parts to joint hypotheses

- Depth image & probability mass
- Localize body parts in 3D
  - global centroid of prob. mass
  - local modes of density (mean shift)
- Map body parts to skeletal joints
  - many parts map directly to joints



# 3D joint hypotheses

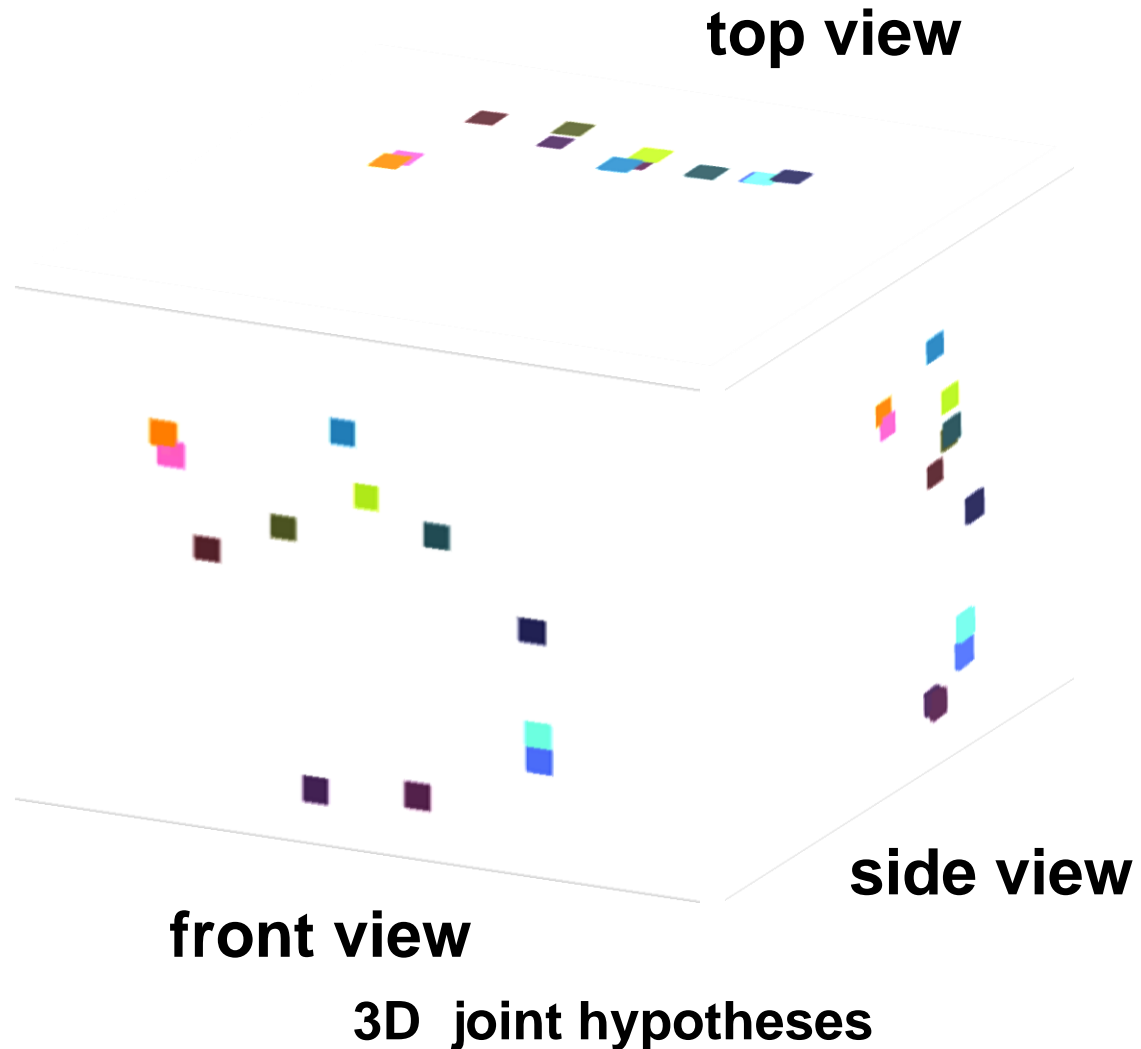
NB No tracking  
or smoothing!



input depth image



inferred body parts &  
overlaid joint hypotheses



# Outline

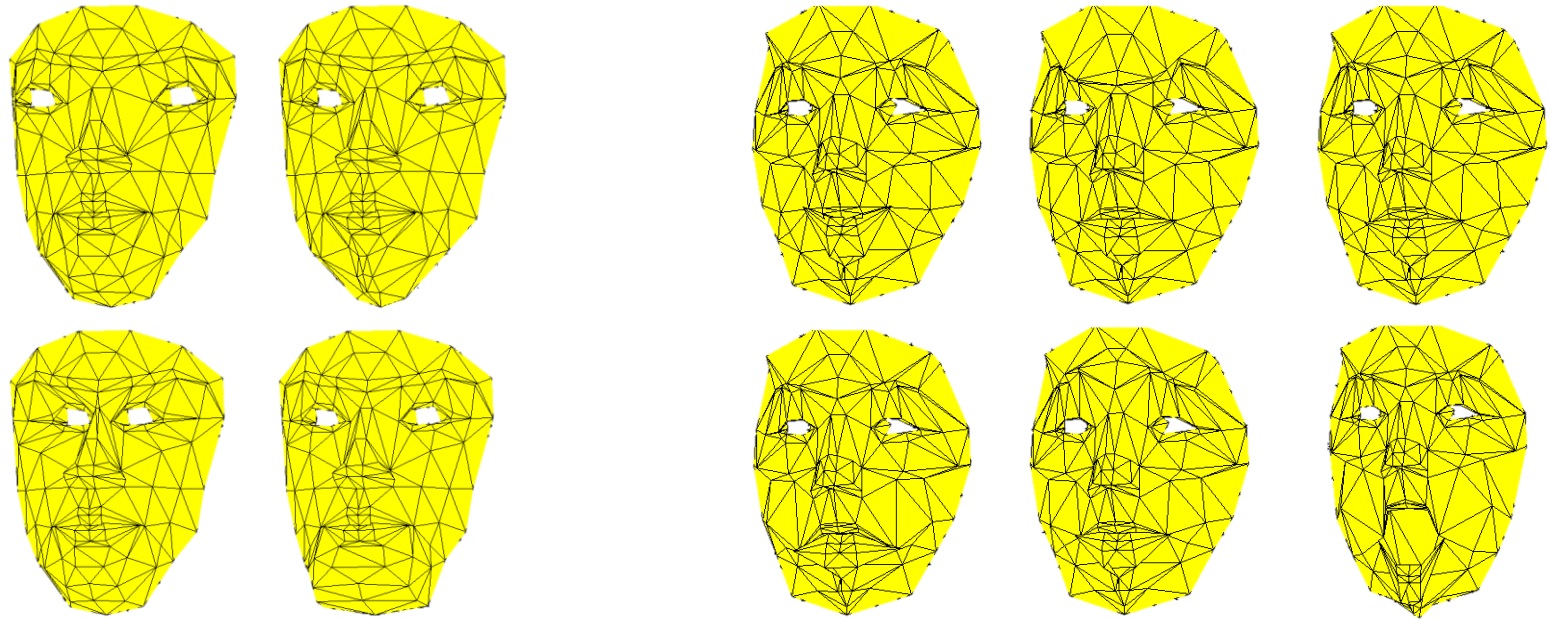
- Kinect in the PC
- Audio processing
- Depth Sensor
- Skeletal Tracking
- Head pose & facial expression tracking



# Avatar Kinect

- Avatar Kinect team
- MSR-R: Qin Cai, Cha Zhang, Zhengyou Zhang
- MSR-A

# Linear Deformable Model



Static deformations

Action deformations

(**Artist rendered** linear deformable model)

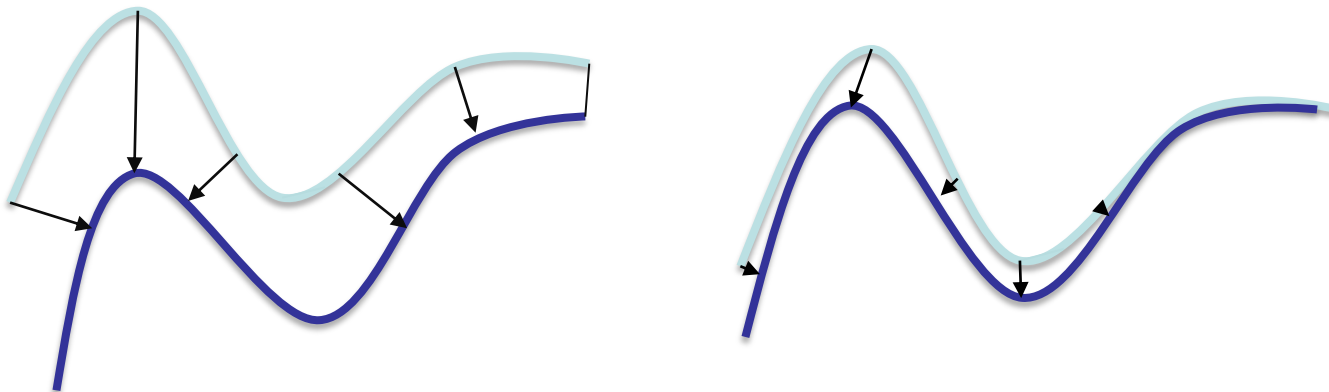
$$\begin{bmatrix} q_1 \\ \vdots \\ q_K \end{bmatrix} = \begin{bmatrix} p_1 \\ \vdots \\ p_K \end{bmatrix} + \mathbf{A} \begin{bmatrix} r_1 \\ \vdots \\ r_K \end{bmatrix} + \mathbf{B} \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}, \text{ where } \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_K \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}$$

# Maximum Likelihood DMF

- Formulation,  $(\mathbf{q}_k, \mathbf{g}_k)$  correspondence pair:

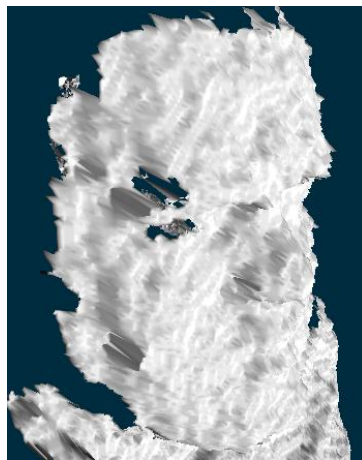
$$\mathbf{R}(\mathbf{p}_k + \mathbf{A}_k \mathbf{r} + \mathbf{B}_k \mathbf{s}) + \mathbf{t} = \mathbf{g}_k + \mathbf{x}_k$$
$$\mathbf{x}_k \sim N(\mathbf{0}, \Sigma_{\mathbf{x}_k})$$

- Iterative closest point
  - Assume closest points correspond
  - Compute transformation
  - Iterate until convergence

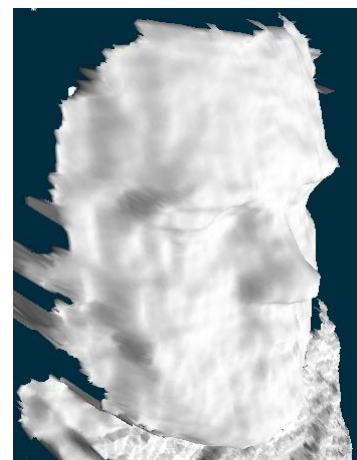


# Model Initialization

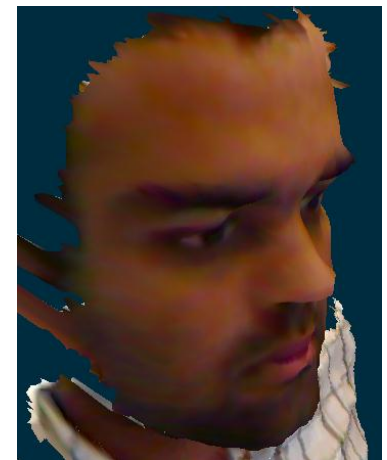
- Initialization
  - Assume multiple neutral face frames available
  - Action deformations set to zero
  - Jointly solve shape deformations and rotation/translation of each frame



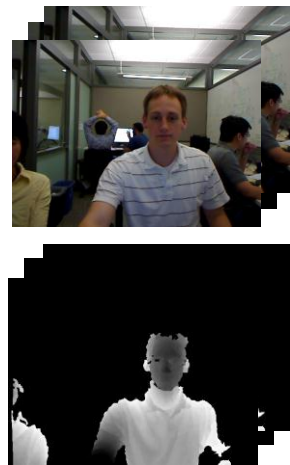
single frame



30 frames aligned and averaged



# Model Initialization



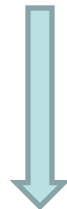
Input



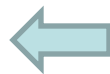
Face  
Detection



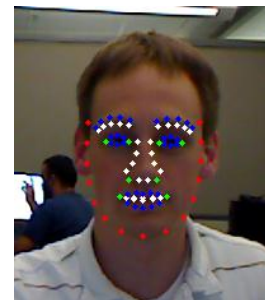
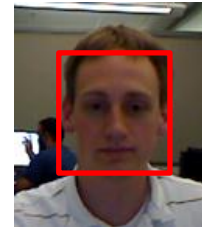
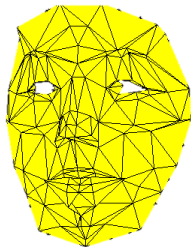
Face  
Alignment



Model  
Initialization



Output



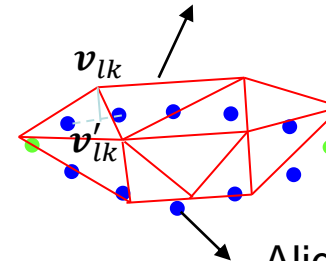
Green dots: point-to-point  
distance

Blue dots: point-to-plane  
3D distance

Red dots: point-to-plane  
2D distance

White dots: unused

Deformable model projected  
onto the texture image



Alignment points

# Summary

- Kinect has revolutionized gaming and beyond...
- Voice and Gestures are natural for users:
  - But very hard to get right
- This is V1: help us improve the technology

Thank you