



Microsoft® Research

# FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS



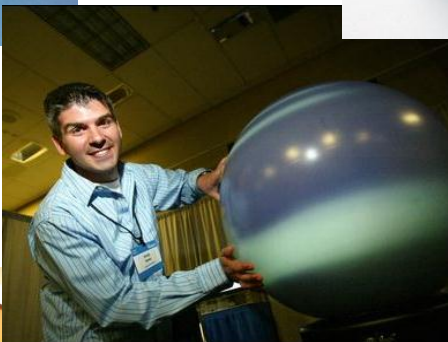
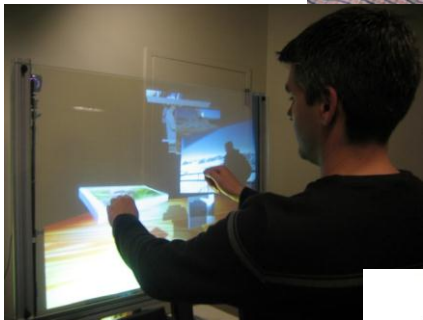
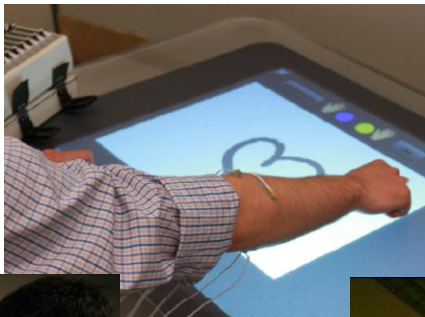
Microsoft® Research

# FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

## High-Fidelity Augmented Reality Interactions

Hrvoje Benko  
Researcher, MSR Redmond



# New generation of interfaces

Instead of interacting through **indirect** input devices (mice and keyboard), the user is interacting **directly** with the content.

Direct un-instrumented interaction

Content is the interface

# Surface computing



# Kinect



# New generation of interfaces

Direct un-instrumented interaction.

Content is the interface.

# New generation of interfaces

Bridge the gap between  
"real" and "virtual" worlds...



A photograph of a computer monitor with a black bezel and a silver stand. The monitor is centered in the frame and displays a white screen with yellow text. The text is centered and reads "... but still confined to the rectangular screen!". The monitor is placed on a light-colored wooden desk. Two black speakers are visible on either side of the monitor, and a small green light is visible on the bottom bezel of the monitor.

... but still confined to  
the rectangular screen!

# An opportunity...



Depth camera

Projector

Enable interactivity on **any available surface** and **between surfaces**.

# MicroMotoCross



Wilson, 2007

# Augmented reality



Spatial

"Deviceless"

High-fidelity

# Depth Sensing Cameras

# Depth sensing cameras

Color + depth per pixel: RGBZ

Can compute world coordinates of every point in the image directly.



# Three basic types

- Stereo
- Time of flight
- Structured light

# Correlation-based stereo cameras

Binocular disparity

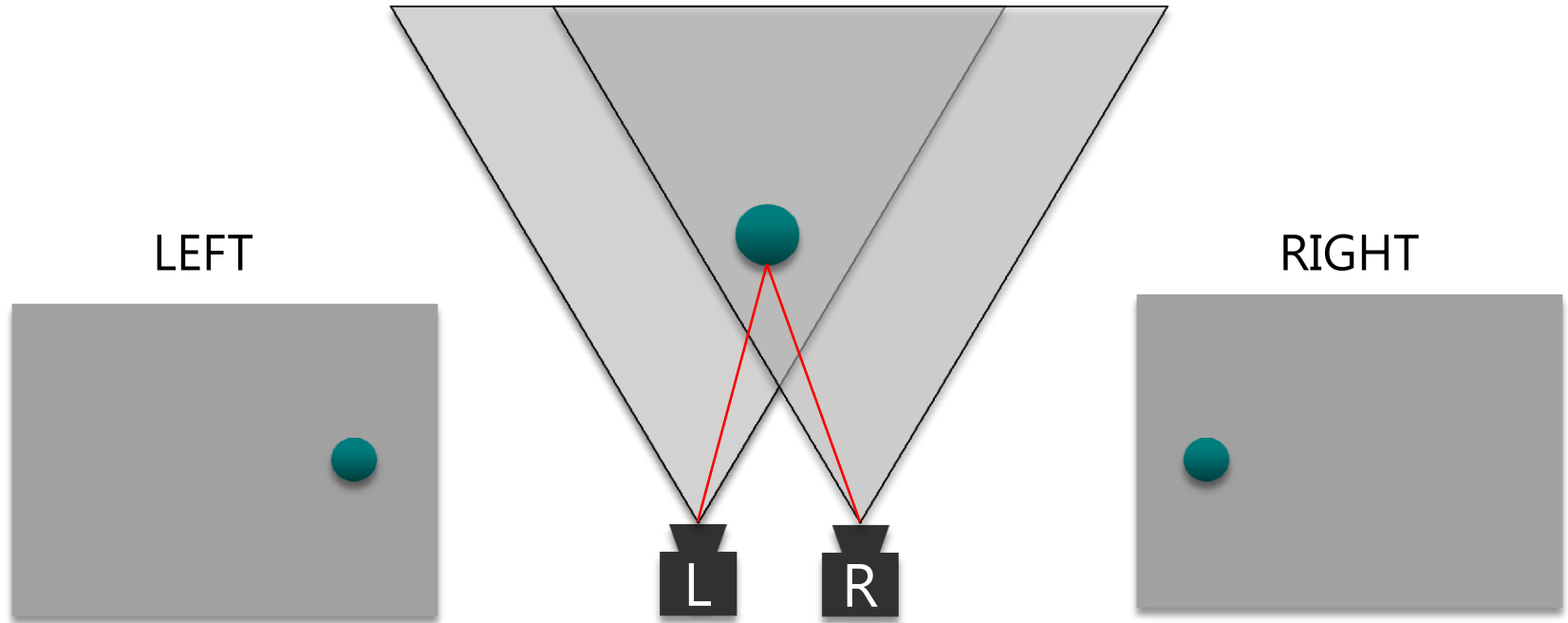


TZYX <http://www.ty zx.com/>

Point Grey Research <http://www.ptgrey.com>



# Correlation-based stereo



# Stereo drawbacks

- Requires good texture to perform matching
- Computationally intensive
- Fine calibration required
- Occlusion boundaries
- Naïve algorithm very noisy

# Time of flight cameras

## 3DV ZSense

Infrared camera +  
GaAs solid state shutter

RGB camera



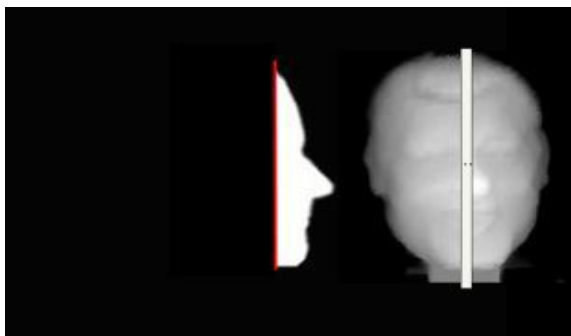
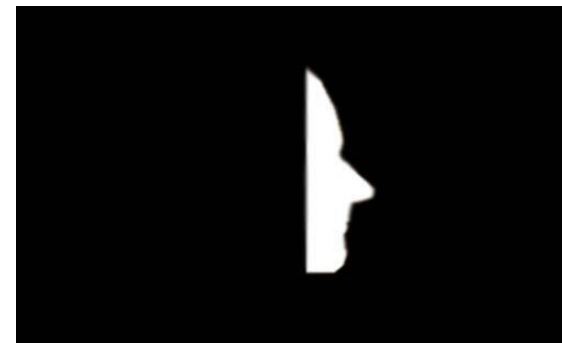
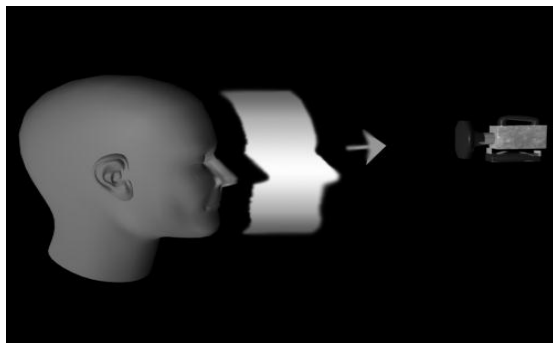
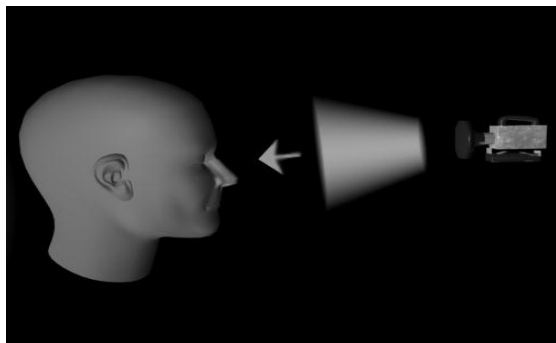
Pulsed infrared lasers

3DV, Canesta (no-longer public)

PMD Technologies <http://www.PMDTec.com>

Mesa Technologies <http://www.mesa-imaging.ch>

# Time of flight measurement



# Structured light depth cameras



**KINECT™**  
for  XBOX 360.

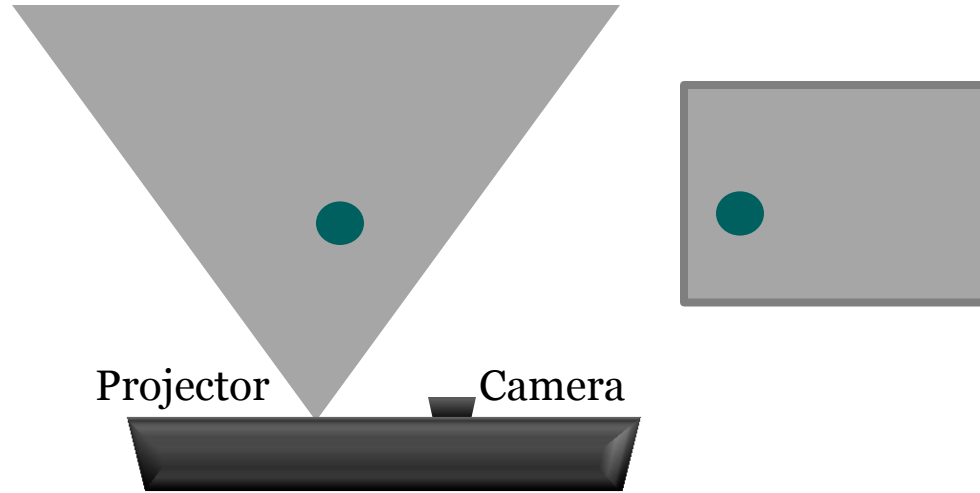
PrimeSense <http://www.primesense.com>

Microsoft Kinect <http://www.microsoft.com/kinect>

# Structured light (infrared)



# Depth by binocular disparity



- Expect a certain pattern at a given point
- Find how far this pattern has shifted
- Relate this shift to depth (triangulate)

# Kinect depth camera

- Per-pixel depth (mm)
- PrimeSense reference design
- Field of View 58° H, 45° V, 70° D
- Depth image size VGA (640x480)
- Spatial x/y resolution (@ 2m distance from sensor) 3mm
- Depth z resolution (@ 2m distance from sensor) 1cm
- Operation range 0.8m - 3.5m
  
- Best part – It is affordable - \$150



**KINECT**  
for XBOX 360.



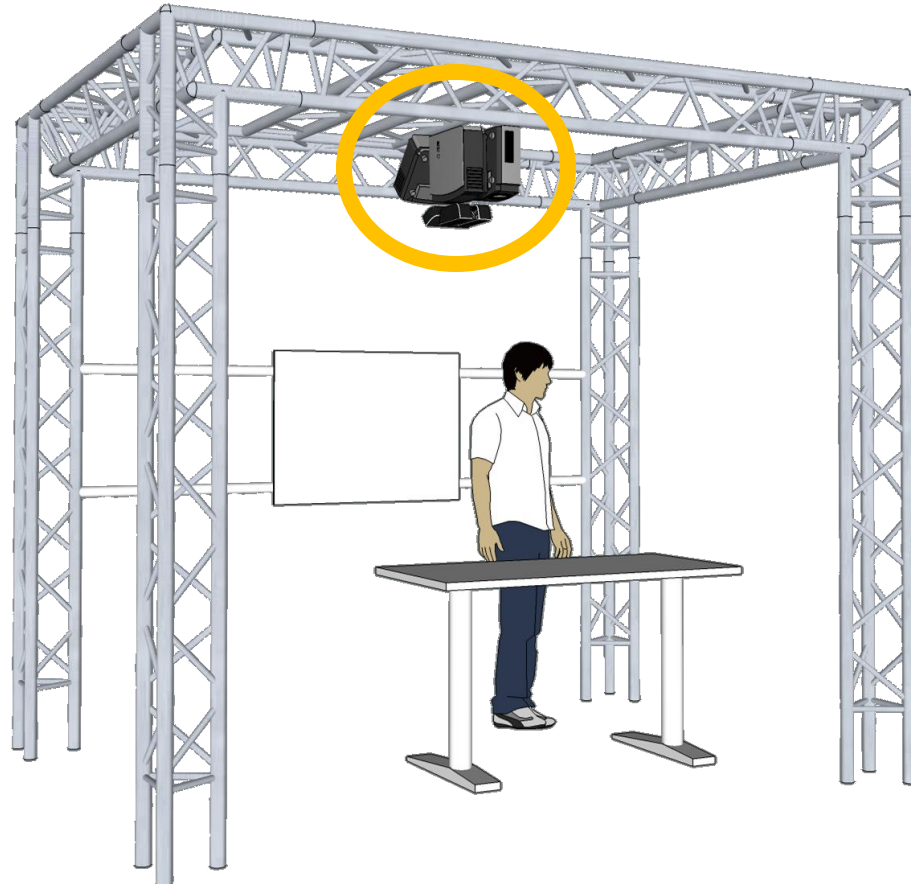
# Why sense with depth cameras?

Requires **no instrumentation** of the surface/environment.

Easier understanding of physical objects in space.

**Enabling interactivity  
everywhere**

# LightSpace



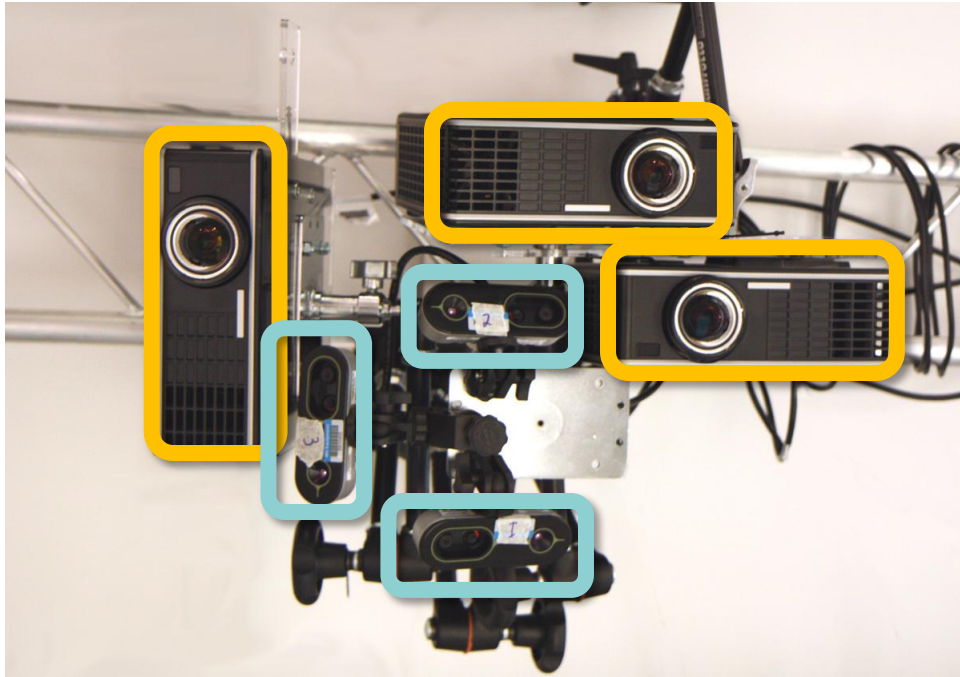
# LightSpace

## LightSpace

Combining Multiple Depth  
Cameras and Projectors for  
Interactions On, Above, and  
Between Surfaces

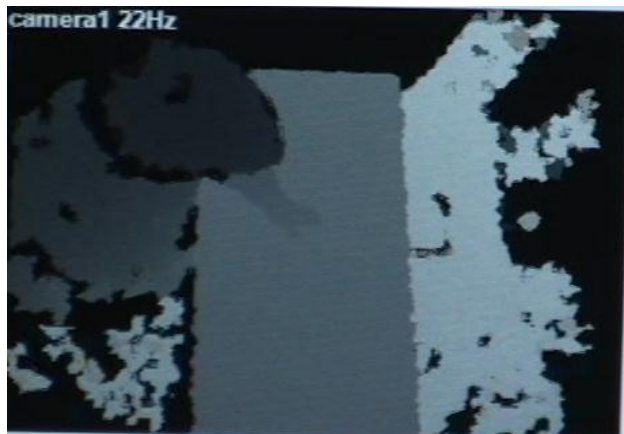
# LightSpace Implementation

**Projectors**



**PrimeSense  
Depth  
Cameras**

# PrimeSense depth cameras



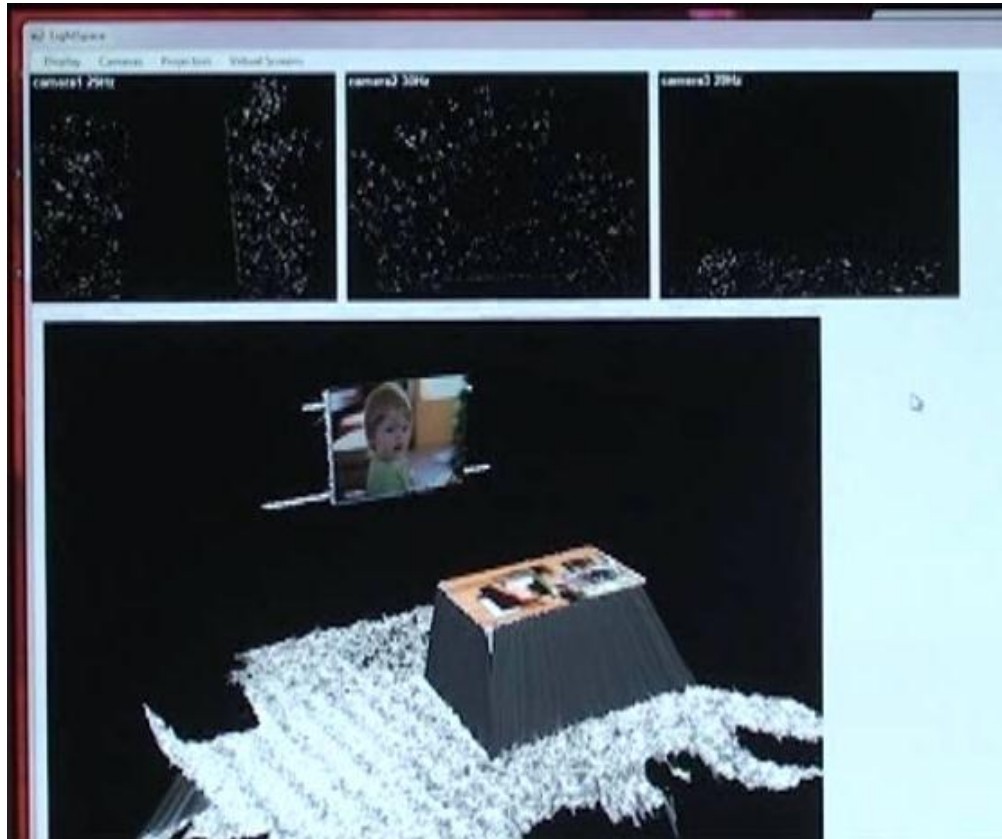
320x240 @ 30Hz

Depth from projected structured light

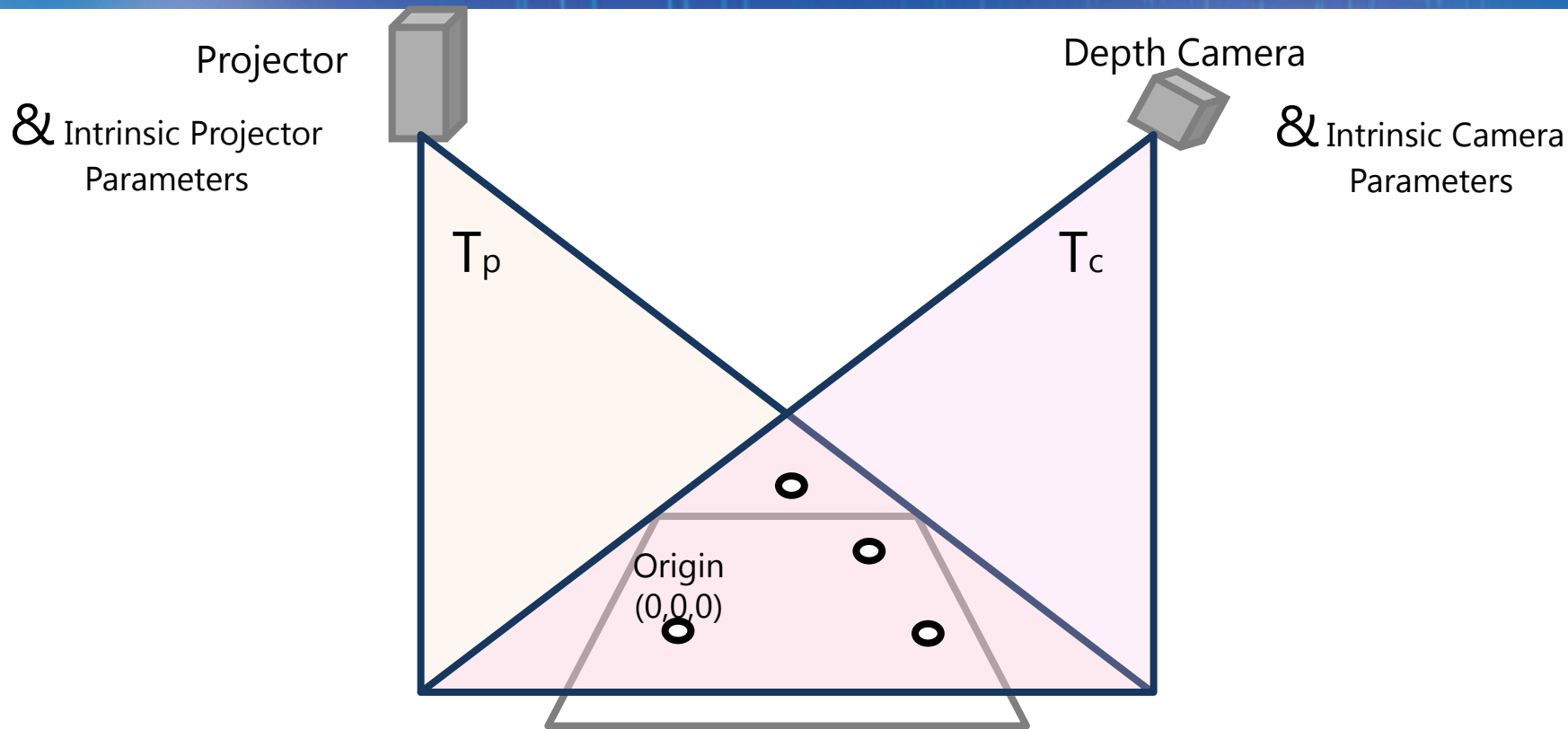
Small overlapping areas

Extended space coverage

# Unified 3D Space

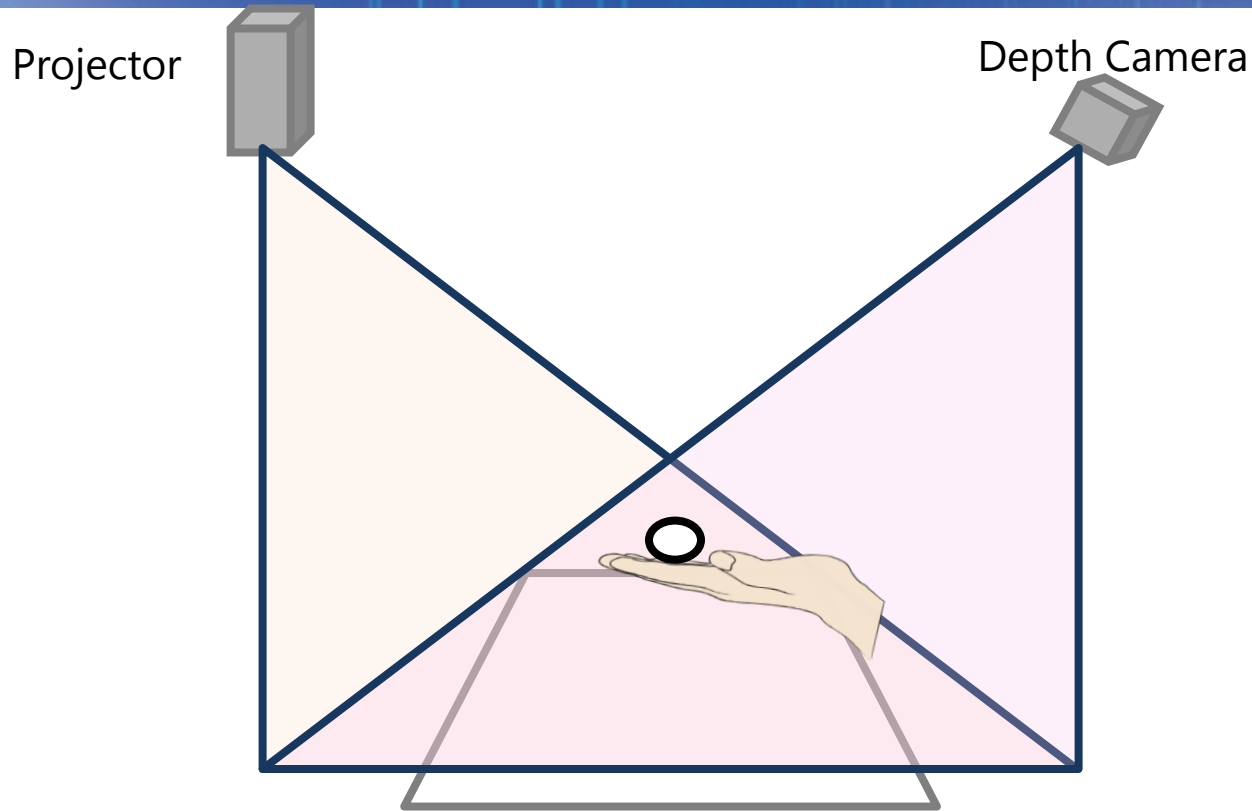


# Camera & projector calibration





# Camera & projector calibration



# LightSpace authoring

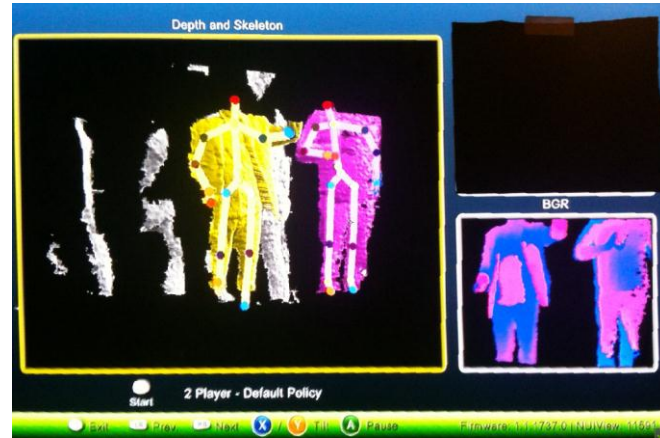
All in real world coordinates.

Irrespective of “which” depth camera.

Irrespective of “which” projector.

# Supporting rich analog interactions

# Skeleton tracking (Kinect)



# Our approach

Use the full 3D mesh.

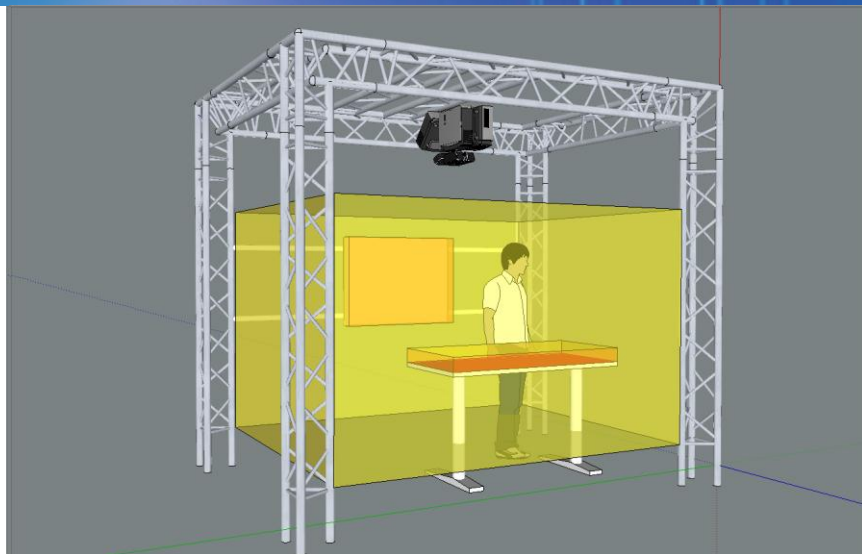
Preserve the analog feel through physics-like behaviors.

Reduce the 3D reasoning to 2D projections.

# Pseudo-physics behavior



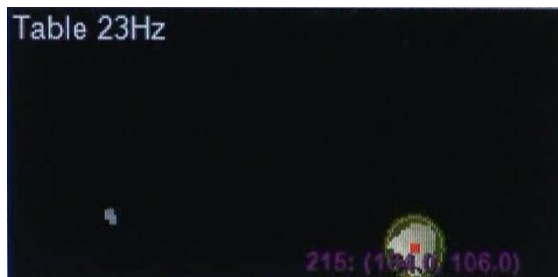
# Virtual depth cameras



Space 29Hz



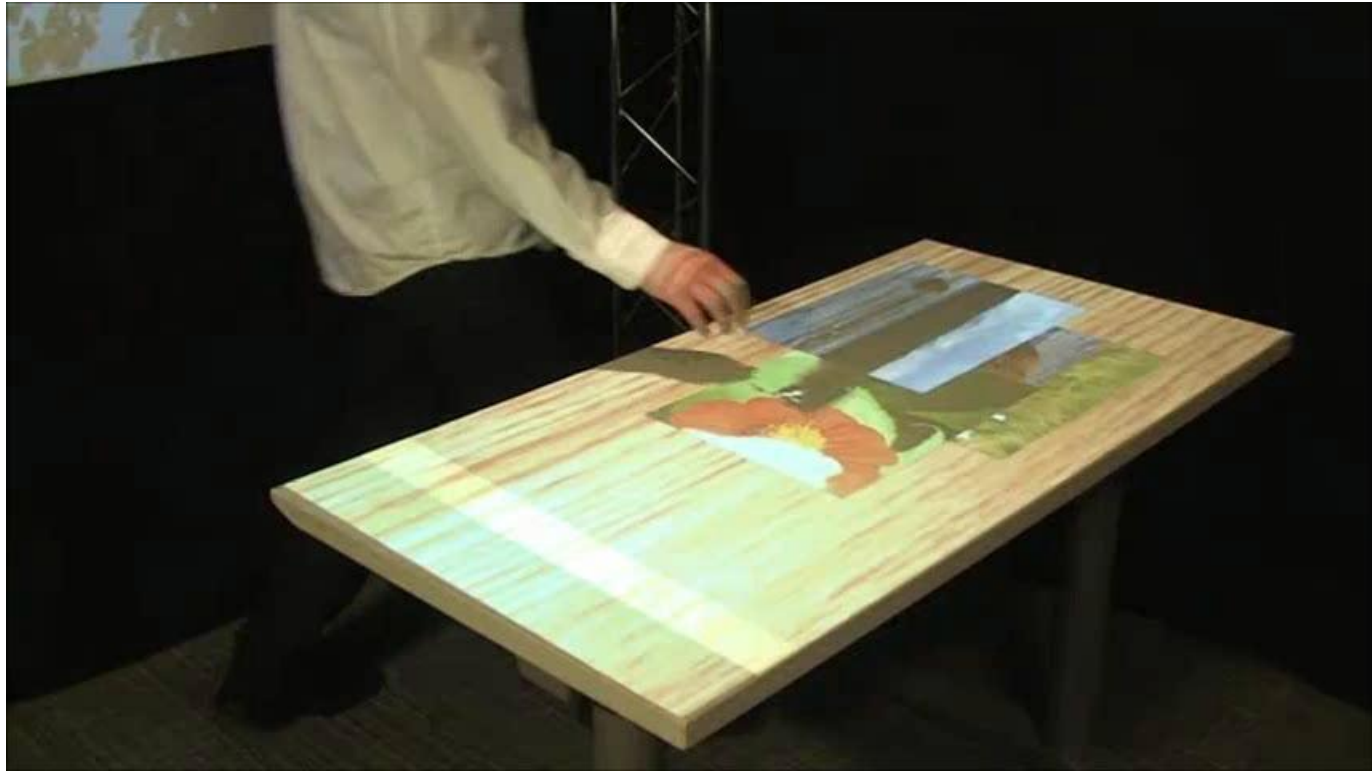
Table 23Hz



Wall 23Hz



# Simulating virtual surfaces





# Through-body connections

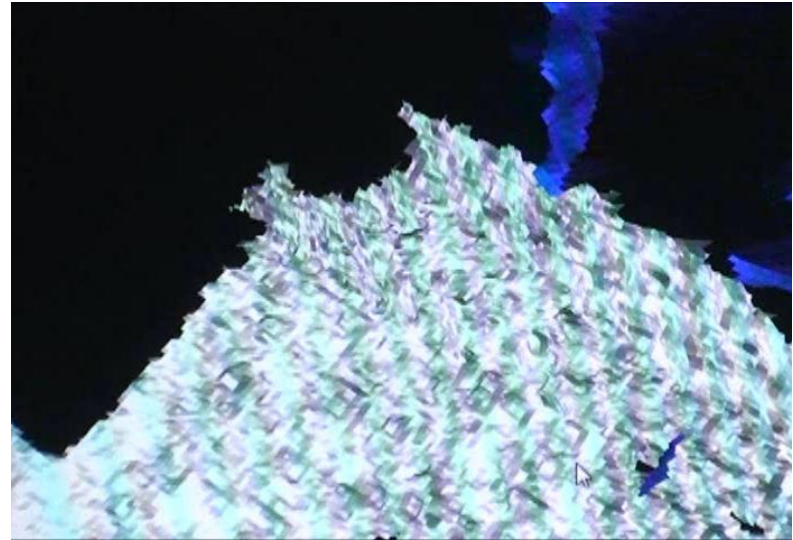


# Physical connectivity



# Spatial widgets

User-aware, on-demand spatial menu



# What is missing?

## LightSpace

- “Touches” are hand blobs
- All objects are 2D
- Very coarse manipulations

## Ideally

- Multi-touch
- 3D virtual objects
- Full hand manipulations

**Touch on every surface**

# Problem of two thresholds



# How to get surface distance?

## Analytically

- Problems:
  - Slight variation in surface flatness
  - Slight uncorrected lens distortion effect in depth image
  - Noise in depth image



# How to get surface distance?

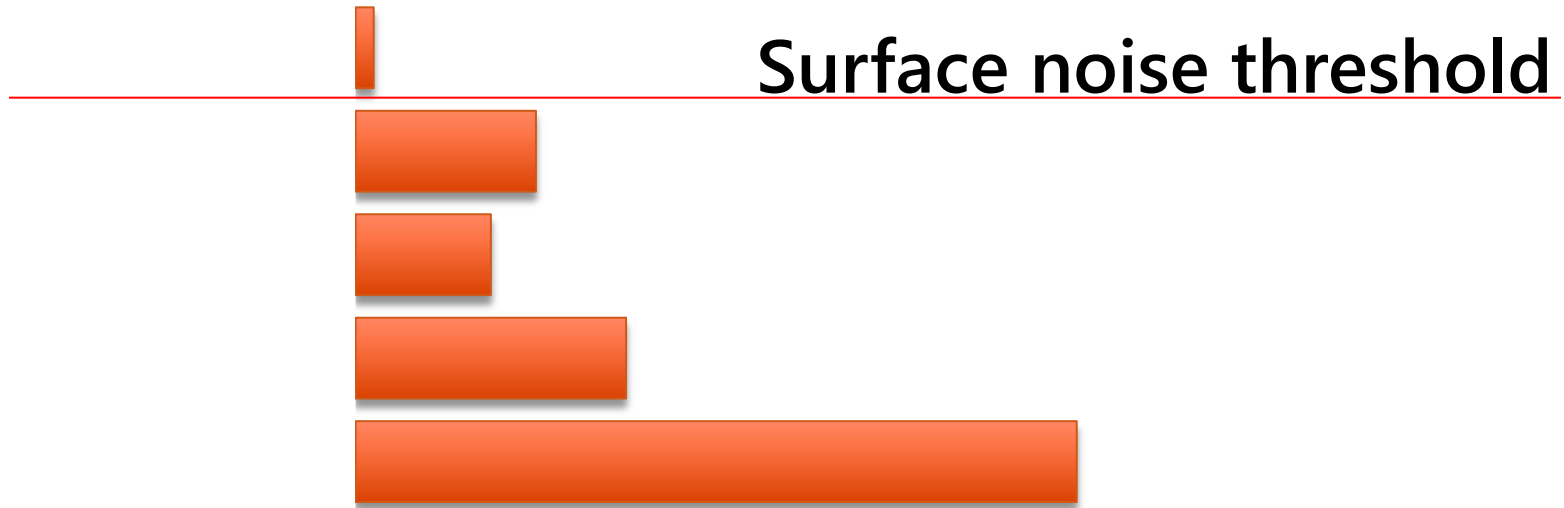
## Empirically

- Take per-pixel statistics of the empty surface
  - Can accommodate different kinds of noise
  - Can model non-flat surfaces
- Observations:
  - Noise is not normal, nor the same at every pixel location
  - Depth resolution drops with distance



# Modeling the surface

Build a surface histogram at every pixel.

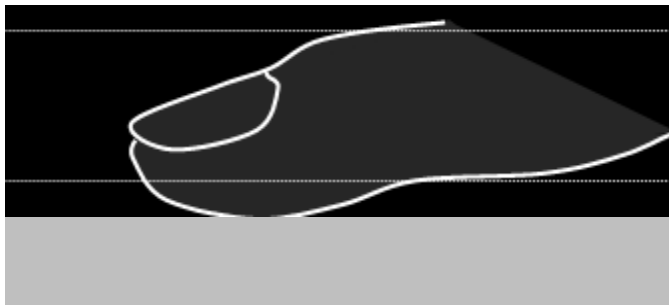


SURFACE

# Setting reasonable finger thickness

Must make some assumption about anthropometry, posture, and noise.

# How good can you get?



Camera above surface	0.75m	1.5m
Finger threshold	14mm	30mm
Surface noise	3mm	6mm

# KinectTouch

Camera at 1.5m above table

# But these are all static surfaces

How to allow touch on **any (dynamic)** surface?

- Dynamic surface calibration
- Tracking high-level constructs such as finger posture, 3D shape
  - Take only the ends of objects with physical extent (“fingertips”)
  - Refinement of position

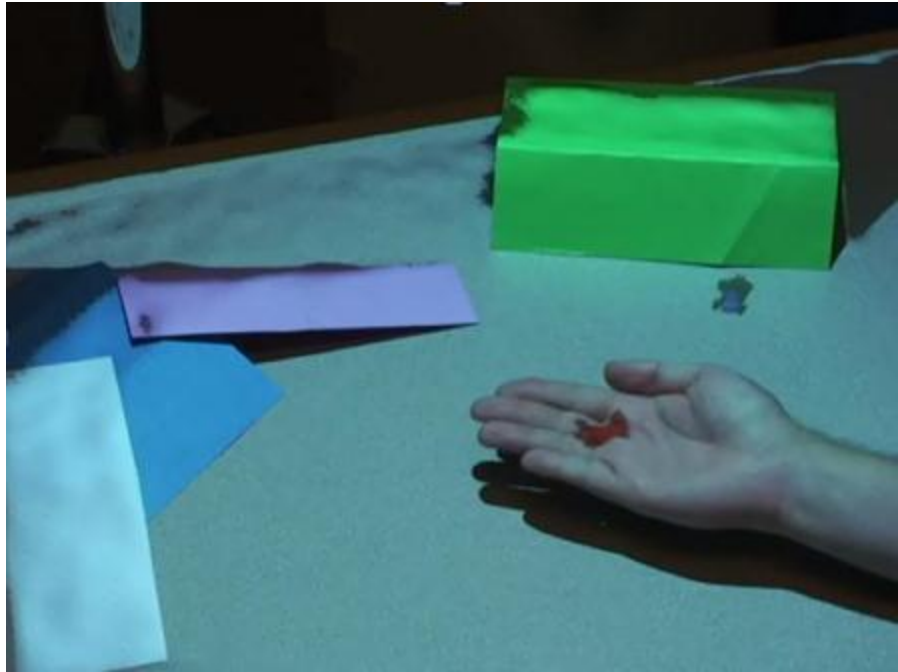
Depth camera touch sensing is almost as good as conventional touch screen technology!

Works on any surface!

(curved, flexible, deformable, flat...)

# Interacting with 3D objects

# Previous approaches were 2D



Micromotocross



LightSpace



Can one hold a virtual 3D object in their hand?

And manipulate it using the **full dexterity** of your hand?

If you know the geometry of the world, you **should be** able to **simulate** physical behaviors.

PhysX<sup>®</sup>  
by NVIDIA

newton™  
GAME DYNAMICS

# Problems with physics and depth cameras

Dynamic meshes are difficult

- Rarely supported in physics packages

No lateral forces!

- Can't place torque on an object

Penetration is handled badly

- Can't grasp an object with two fingers



# Particle proxy representations



Wilson 2007

But can you see 3D in your hand?

# 3D perception

## Many cues:

- Size
- Occlusions
- Shadows
- Motion parallax
- Stereo
- Eye focus and convergence



Can correctly simulate **if** you know:

- The geometry of the scene
- User's view point and gaze

# Depth camera is ideal for this!

Can easily capture scene geometry

Can easily track user's head



# MirageBlocks

3D Projector  
(Acer H5360)

Depth Camera  
(Kinect)

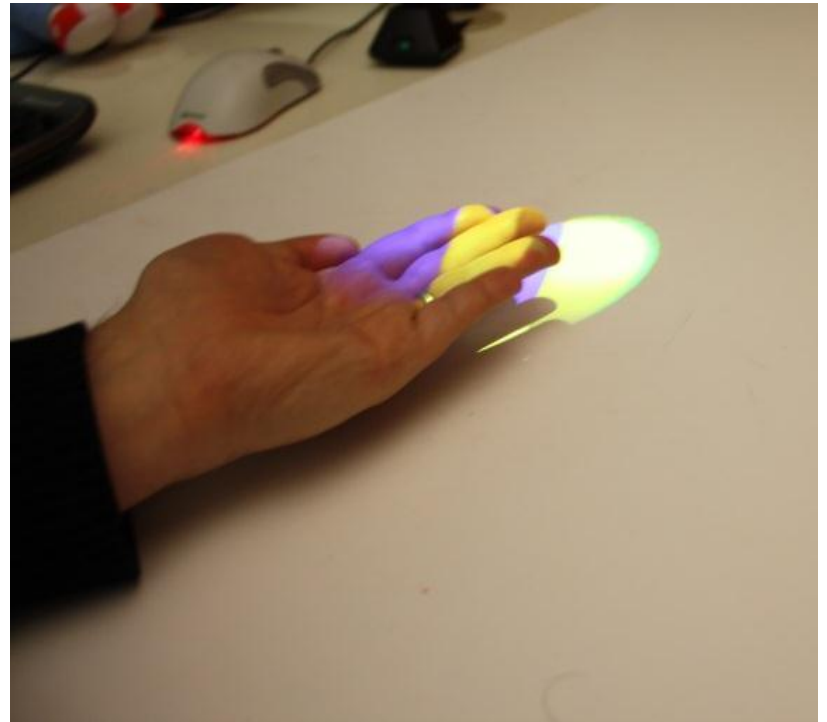
Shutter Glasses  
(Nvidia 3D Vision)



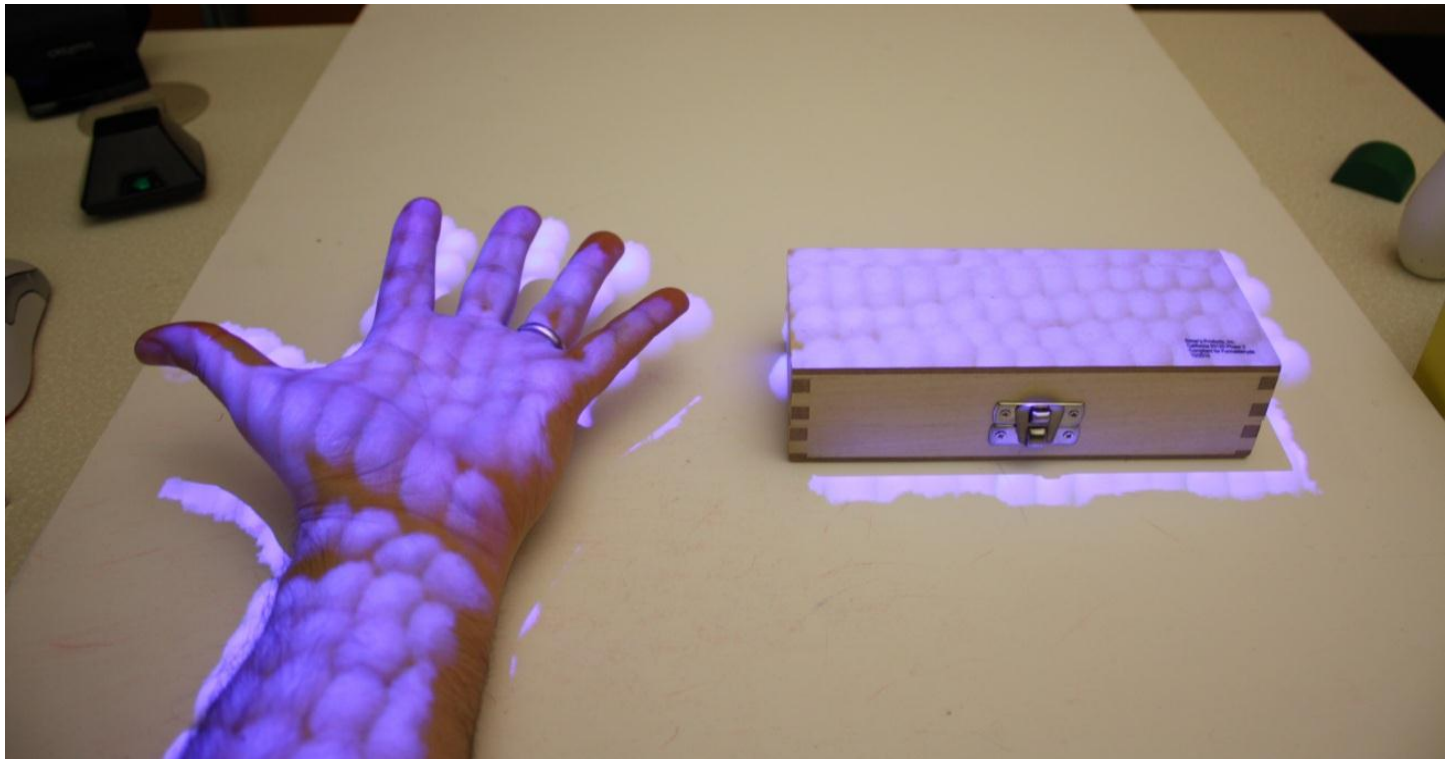
Benko, Costa, and Wilson, 2011



# A single user experience!



# Particle proxies



# MirageBlocks

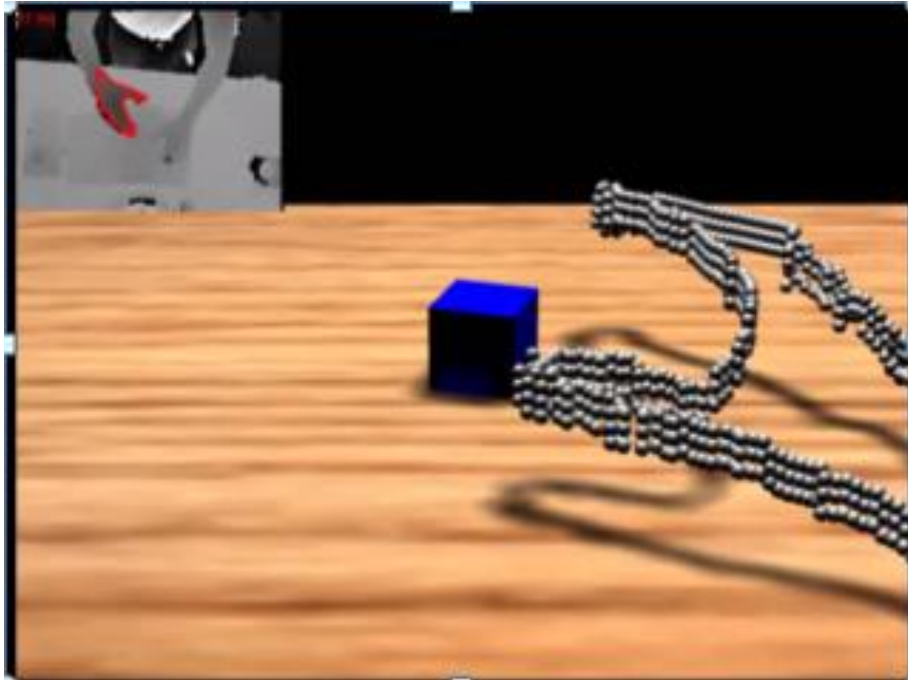
# MirageBlocks

Hrvoje Benko, Ricardo Costa, Andy Wilson

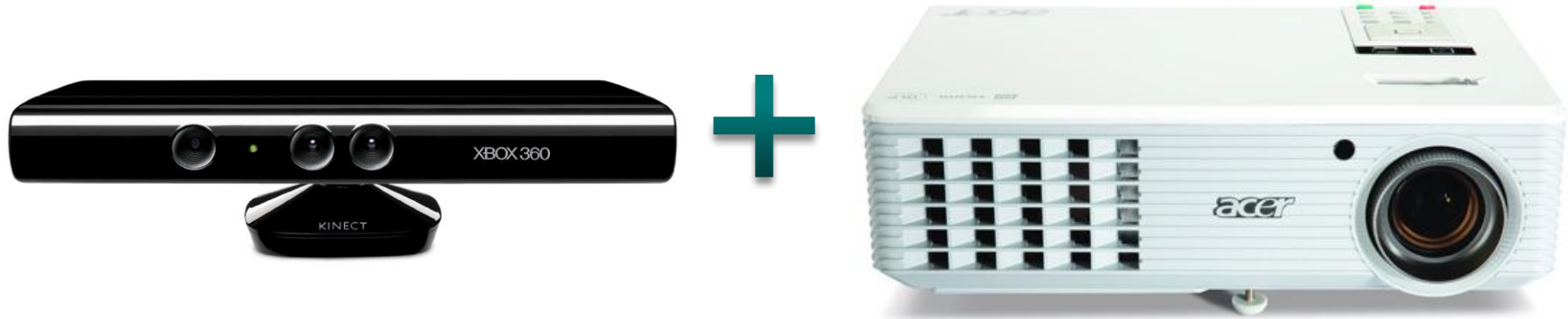
Microsoft Research 2011

# Next: Grabbing

Very hard problem – Working on it!



# Summary



1. Interactivity everywhere
2. Room and body as display surfaces
3. Touch and 3D interactions
4. Preserve the analog feel of interactions

# Come to try it yourself!



MirageBlocks demo

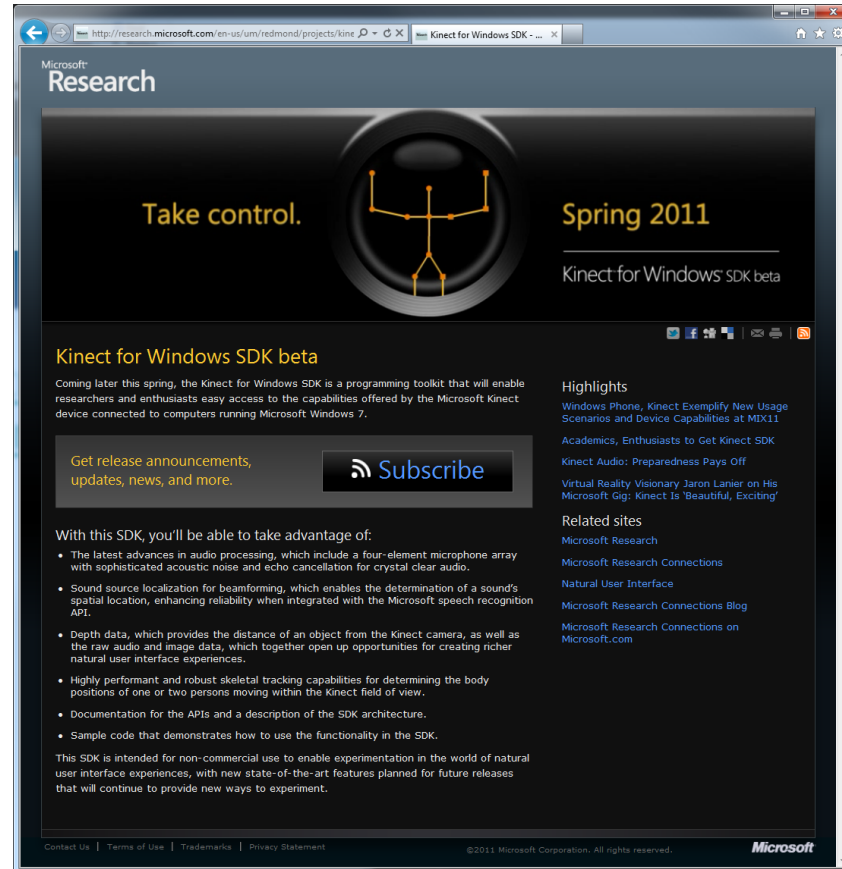
Friday 10am – 1pm

# Resources to consider

# Resources

## Kinect for Windows SDK

- <http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk>



Microsoft Research

Take control.

Spring 2011

Kinect for Windows SDK beta

Kinect for Windows SDK beta

Coming later this spring, the Kinect for Windows SDK is a programming toolkit that will enable researchers and enthusiasts easy access to the capabilities offered by the Microsoft Kinect device connected to computers running Microsoft Windows 7.

Get release announcements, updates, news, and more.

Subscribe

With this SDK, you'll be able to take advantage of:

- The latest advances in audio processing, which include a four-element microphone array with sophisticated acoustic noise and echo cancellation for crystal clear audio.
- Sound source localization for beamforming, which enables the determination of a sound's spatial location, enhancing reliability when integrated with the Microsoft speech recognition API.
- Depth data, which provides the distance of an object from the Kinect camera, as well as the raw audio and image data, which together open up opportunities for creating richer natural user interface experiences.
- Highly performant and robust skeletal tracking capabilities for determining the body positions of one or two persons moving within the Kinect field of view.
- Documentation for the APIs and a description of the SDK architecture.
- Sample code that demonstrates how to use the functionality in the SDK.

This SDK is intended for non-commercial use to enable experimentation in the world of natural user interface experiences, with new state-of-the-art features planned for future releases that will continue to provide new ways to experiment.

Highlights

- Windows Phone, Kinect Exemplify New Usage Scenarios and Device Capabilities at MIX11
- Academics, Enthusiasts to Get Kinect SDK
- Kinect Audio: Preparedness Pays Off
- Virtual Reality Visionary Jaron Lanier on His Microsoft Gig: Kinect Is 'Beautiful, Exciting'

Related sites

- Microsoft Research
- Microsoft Research Connections
- Natural User Interface
- Microsoft Research Connections Blog
- Microsoft Research Connections on Microsoft.com

Contact Us | Terms of Use | Trademarks | Privacy Statement

©2011 Microsoft Corporation. All rights reserved.

Microsoft



# Resources

## NVIDIA PhysX SDK

- <http://developer.nvidia.com/physx-downloads>
- <http://physxdotnet.codeplex.com/> (.NET wrappers)



## Newton Physics Game Engine

- <http://newtondynamics.com/forum/newton.php>



# Resources

## NVIDIA 3D Vision

- <http://www.nvidia.com/object/3d-vision-main.html>



## DLP Link

- <http://www.dlp.com/projector/dlp-innovations/dlp-link.aspx>
- <http://www.xpand.me/> (3D glasses)



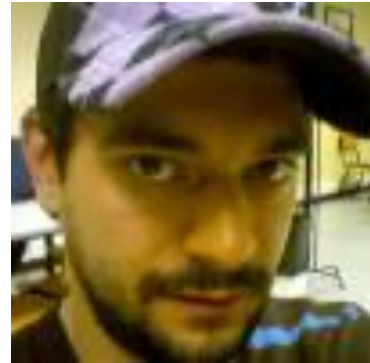
# My collaborators



Andy Wilson



Chris Harrison



Ricardo Costa Jota



Hrvoje Benko

[benko@microsoft.com](mailto:benko@microsoft.com)

<http://research.microsoft.com/~benko>



***Microsoft***<sup>®</sup>

© 2010 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.  
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.  
MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.