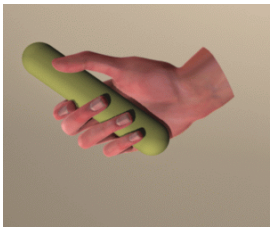
Hand Motion and Grasping: Capturing, Recognizing and Synthesizing

Daniel Thalmann
Institute for Media Innovation
Nanyang Technological University,
Singapore and EPFL, Switzerland

http://en.wikipedia.org/wiki/Daniel_Thalmann



**Photogrammetry Week 2015,
Stuttgart, Germany**

Hand Motion and object manipulation

- Virtual human using hands
- User grasping Virtual Objects
- Man-Machine Communication
- Challenges
  - Accuracy
  - Deformations
  - Coordination
  - Complexity
  - Multi-manipulation

Two movies

1987 Rendez-vous in Montreal

1992 IAD

https://en.wikipedia.org/wiki/Rendez-vous_in_Montreal

Skeleton-Based Deformations

Sub-Space Deformations

- Various names : Joint Local Deformations (JLD), then sub-space deformation (SSD), linear blend skinning, or smooth skinning.
- Method works first by assigning a set of joints j {1, ...,*n}* *with weights* to each vertex *v of the character.*
- *The deformation of a vertex v0 is then computed by weighted combination of transformation Tj of influencing joints j with the below* function:

$$S(v_0) = (\sum_{j=1}^{n} w_j T_j).v_0$$

N. Magnenat-Thalmann, R.Laperrière, D. Thalmann, **Joint-Dependent Local Deformations for Hand Animation and Object Grasping,** *Proc. Graphics Interface'88,* Edmonton, 1988, pp.26-33

Object Manipulation

- A very complex problem
- No general algorithm
- More complex:
  - Two people holding objects
  - Exchange of objects
  - Indirect manipulation
  - Not only grasping, e.g. door
  - deformations

J.P.Gourret, N. Magnenat-Thalmann, D. Thalmann, **Simulation of Object and Human Skin Deformations in a Grasping Task**, *Proc. SIGGRAPH '89*, Computer Graphics, Boston, Vol. 23, No 3, 1989, pp. 21-30

Automatic approach

- Heuristic grasping decision
- Inverse kinematics to find final arm posture
- Multi-sensors

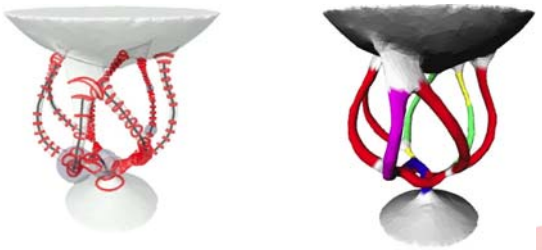
R. Mas, D. Thalmann, **A Hand Control and Automatic Grasping System for Synthetic Actors**, *Proc. Eurographics '94*, Oslo, September 1994

Problems

- Object can have various ways of grasping
- Dependent of situation: empty glass vs full glass
- Tubular structure to help grasping

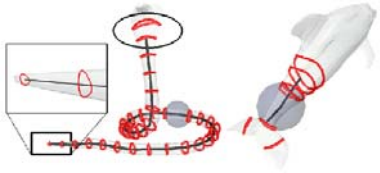
Another method of Grasping

- Method for grasping in an OM context
  - a tubular feature classification algorithm (Plumber)
  - a hand grasp posture generation algorithm
  - An object manipulation framework
- Works on objects with tubular or elongated parts
- Suitable for objects that are not designed manually (e.g. from laser scanning, mixed objects, etc.)



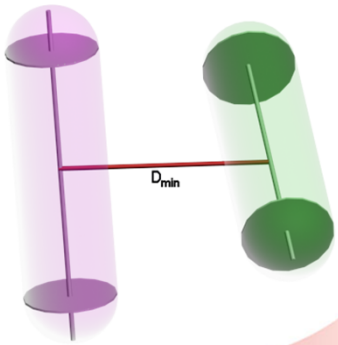
Plumber

- *(Mortara, 2004)*
- Analyses object shape
- Intersection of spheres centered at mesh vertices
- Identifies tube features of arbitrary section and axis shape



Touch Detection

- Approximate tube and finger segments with capsules
- Fast intersection testing: compute minimum distance (D_{min}) between two line segments
- Touch testing with tolerance (ε):
  - $D_{min} > R_{sum}$: Not touching
  - $R_{sum} \geq D_{min} > R_{sum} - \varepsilon$: Touching
  - $R_{sum} - \varepsilon >= D_{min}$: Inside



Real-time Manipulation





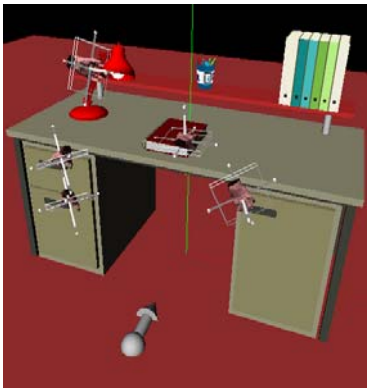
To grasp you need to reach

- Generate collision-free reaching motions
  - control of the entire body (e.g. leg flexion)
  - articulation limits
  - comfort criteria
  - Construct a *Probabilistic Reaching Roadmap*

M. Kallmann, A. Aubel, T. Abaci, and D. Thalmann, **Planning Collision-Free Reaching Motions for Interactive Object Manipulation and Grasping**, *Proc. Eurographics 2003*.

Semi-automatic approach: smart objects

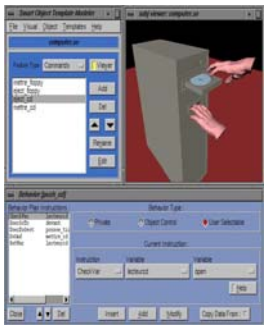
Geometric
Object Model
+
Modelling of
its Interaction
Features
=
Smart Object



M. Kallmann and D. Thalmann, **Modeling Behaviors of Interactive Objects for Virtual Reality Applications**, *Journal of Visual Languages and Computing,* Vol.13, 2002, pp.177-195.

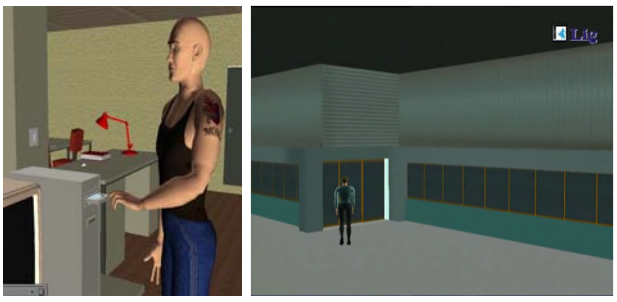
Actor-Object Interactions

Modeling Phase



Predefined Plans Referencing Geometric Parameters Relative to the Object.

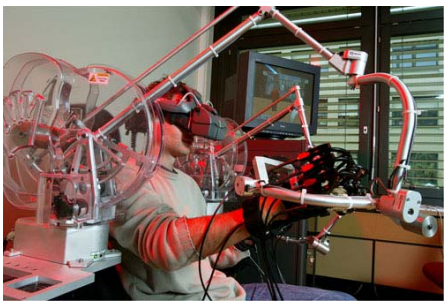
Simulation Phase



Distributed in the VE, Plans are followed to initialize motions when one is chosen.

Grasping in VR
hard without haptic feedback

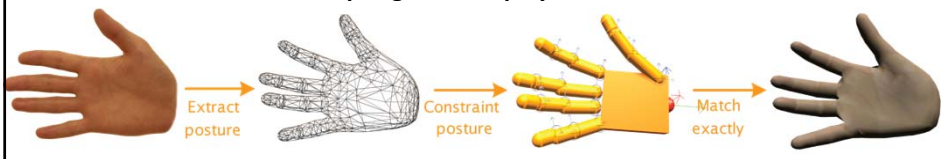
- The Immersion Haptic Workstation™



16

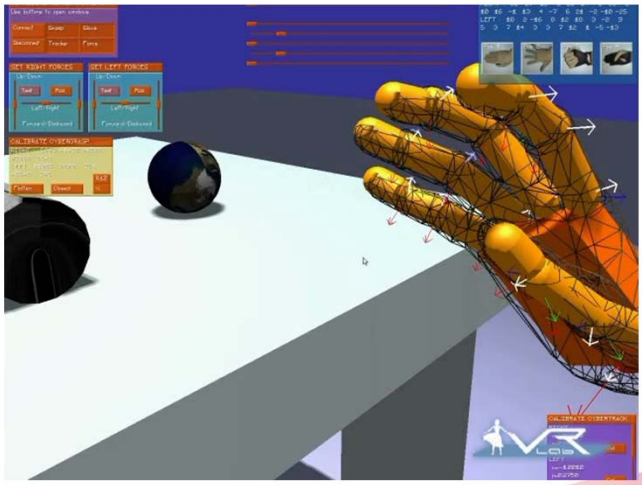
Two-handed force feedback computation

- Proxy based technique
  - **Real hand drives intermediate hand (wireframe)**
  - **Intermediate hand applies soft constraints on mass spring hand model.**
  - **Position of mass spring hand displayed to user**



Real Hand	Hand after Tracking and Calibration	Mass-Spring collision Hand	Visual Hand

17



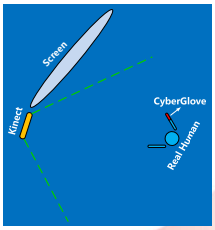
Body and hand gesture recognition

Hand feature extraction
(Cyber-glove)

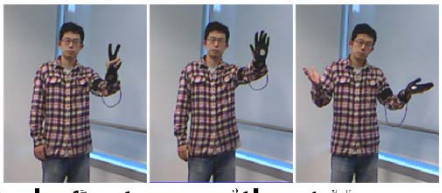
Upper body joints extraction
(Kinect)

Information fusion

Recognize upper body gesture



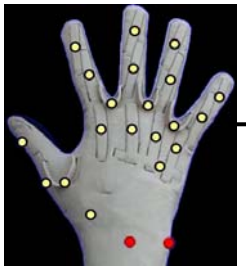
Yang Xiao; Junsong Yuan; Daniel Thalmann
Human-Virtual Human Interaction by Upper Body Gesture Understanding, Proc. ACM VRST 2013

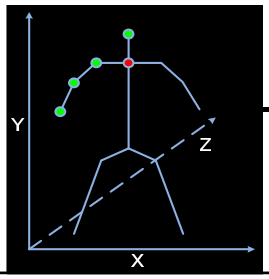


Upper Body Gestures without Human-object Interaction

Upper Body Gestures with Human-object Interaction

Upper Body Gesture Understanding





Feature Fusion

$$\widehat{F}_i = \frac{F_i - F_i^{min}}{F_i^{max} - F_i^{min}}$$

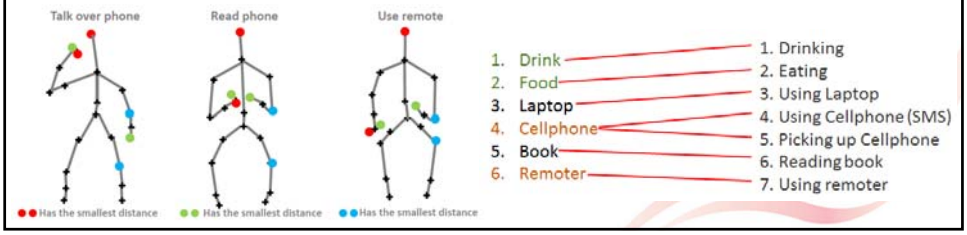
$$\widehat{F} = \left(\widehat{F}_{hand}, \widehat{F}_{body}\right)$$

Human Upper Body Gesture Dataset Samples

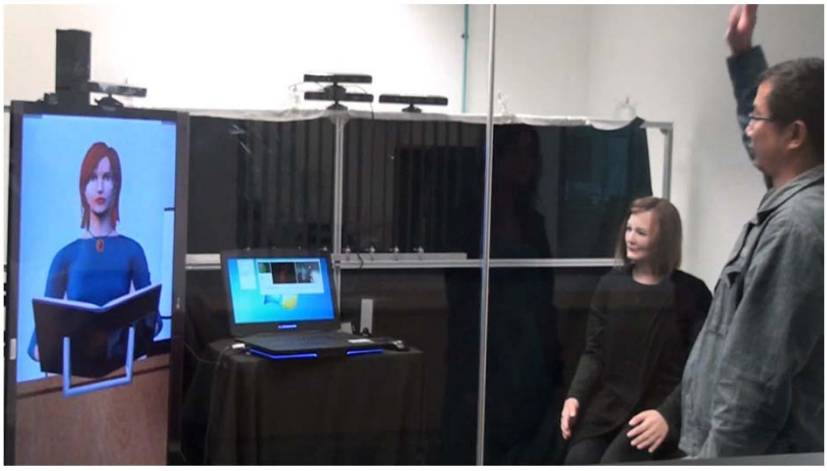
Human-Robot Interaction by Understanding Upper Body Gestures

Action localization based RGBD cameras

- Problem
  - continuous human-object interaction recognition based on RGBD (Kinect camera)
  - Determine start and end frame of one action
- Motivation
  - Combine skeleton and object information



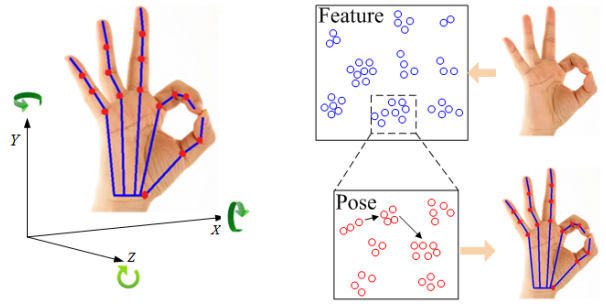
Demo: Body gesture recognition



Hand Gesture Recognition

Problem Description

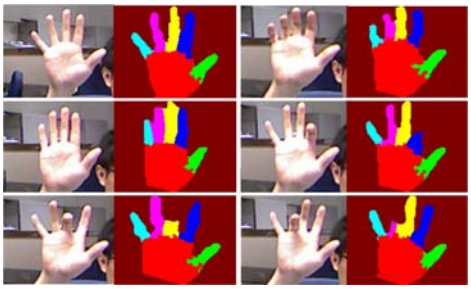
- Vision-based full DOF pose estimation for bare-hand input
- Hand pose parameterized as a 27D vector, including the global and local motion.
- Task: restore the pose vector for each input frame



(a) Parameterized pose space (b) Problem formulation

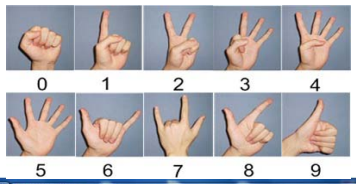
Hand Gesture Recognition

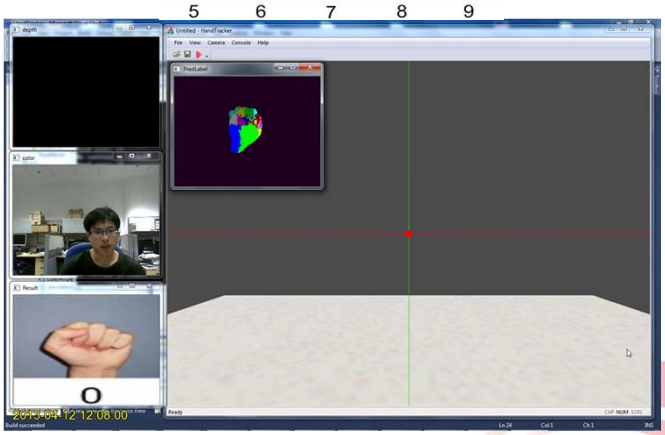
- Novelty: spatial and temporal constraints in a unified framework for hand parsing and fingertip detection.
- Result: more accurate compared to existing methods.



Hui Liang, et al. **3D Fingertip and Palm Tracking in Depth Image Sequences**, ACM International Conference on Multimedia 2012 (MM)
Hui Liang, Junsong Yuan and Daniel Thalmann, "**Model-based Hand Pose Estimation via Spatial-temporal Hand Parsing and 3D Fingertip Localization**", CGI 2013.

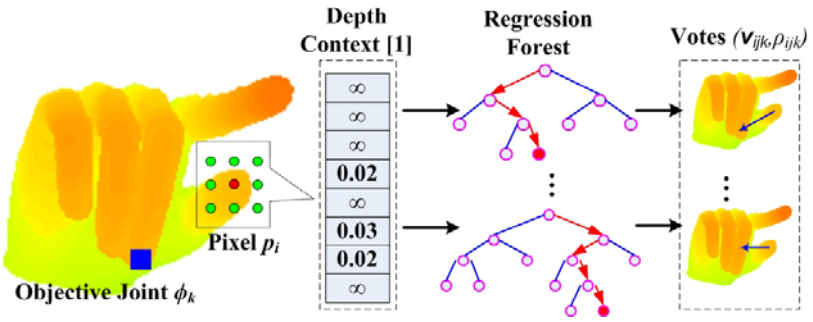
Demo: Hand Gesture Recognition





Regression forest = set of regression trees

- The Regression Forest is utilized to get the per-pixel prediction votes for each hand joint

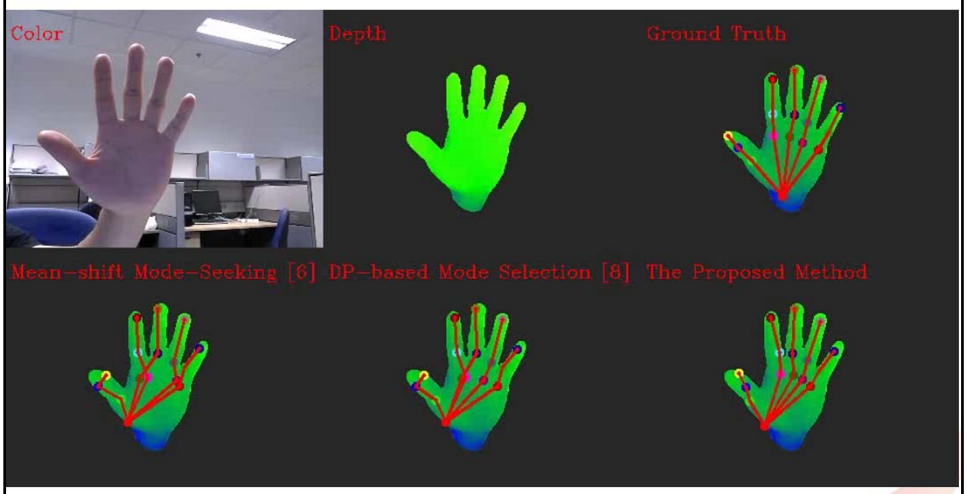


H. Liang, J. Yuan, D. Thalmann, **Parsing the Hand in Depth Images**, *IEEE Transactions on Multimedia*, Vol.16, No5, 2014, pp.1241-1253.

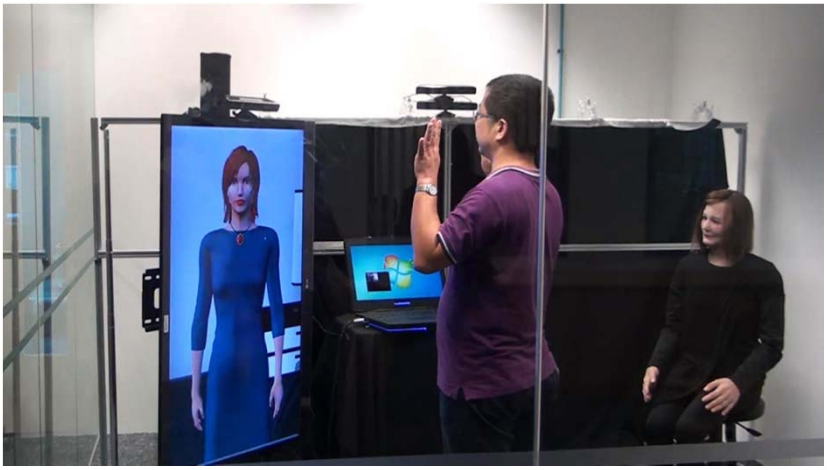
New Method: Multimodal Prediction Fusion

- Task: Predict the sixteen joint locations of the hand from single depth images
- Multimodal Prediction Fusion: Fuse the per-pixel predictions with the learned hand joint correlations using PCAs

H. Liang, J. Yuan, D. Thalmann, **Resolving Ambiguous Hand Pose Predictions by Exploiting Part Correlations**, *IEEE Transactions on Circuits and Systems for Video Technology*, 2015


Color
Depth
Ground Truth
Mean-shift Mode-Seeking [6]
DP-based Mode Selection [8]
The Proposed Method

Example



Thank you for your attention..