Vision and graphics applications on Kinect

Yichen Wei
Visual Computing Group
Microsoft Research Asia
Kinect: a revolution

- Reliable 2.5D depth → new applications
More than body tracking

- Human action recognition
- Face recognition, pose estimation, tracking, expression recognition, modeling, animation...
- Hand/finger tracking, gesture recognition...
- 3D body scanning/modeling, accurate motion control, virtual character...
- 3D object modeling, recognition...
- 3D indoor modeling, virtual environment...
- Segmentation, denoising, super-resolution...
- ....
Working on Kinect and Xbox

- **Kinect**: moderate image quality
  - Small resolution: 640x480 for RGB, 320x240 for depth
  - Moderate quality/noises in RGB and depth

- **Xbox 360**: moderate performance
  - Consumer level hardware (released in 2005)
  - IBM PowerPC CPU: 3 cores, 3.2 GHz each
  - 1 MB L2 cache
  - 512 MB memory
  - GPU: moderate and mostly for 3D rendering only
Challenges in real world

- As robust, fast, smaller memory as possible

1. Platform level 0 function
   - Always running in system thread
   - ⫸⫸⫸ 30 FPS, favorably hundreds of FPS

2. Platform level 1 function
   - Called by games in need
   - ⫸ 30 FPS

3. Games
   - Running in an exclusive user thread
   - ⫸ 30 FPS
Works at MSRA

• Human action recognition
• Face recognition, pose estimation, tracking, expression recognition, modeling, animation...
• Hand/finger tracking, gesture recognition...
• 3D body scanning/modeling, accurate motion control, virtual character...
• 3D object modeling, recognition...
• 3D indoor modeling, virtual environment...
• Segmentation, denoising, super-resolution...
• .......
Kinect Identity

- A skeleton ↔ a game character / a player profile
- Seamless user experience
A demo
Technique: face recognition and fusion of multiple features

Kinect Identity: Technology and Experience, Tommer Leyvand, Casey Meekhof, Yichen Wei, Jian Sun and Baining Guo
Head pose estimation

- Body language
- Game control
A demo
Extensive training data capturing
Extensive training data capturing

neutral

mouth open

smile

dark & far  dark & near  bright & far  bright & near
Techniques and results

- Simple features
  - LBP + LDA
- Fast regression
  - kNN
- Per-frame estimation
- Promising accuracy
- Super fast: 2-3 ms

error in degrees: yaw 5, pitch 7.5, roll 5

Real Time Head Pose Estimation, Yichen Wei, Fang Wen, Jian Sun, Tommer Leyvand, Jinyu Li, Casey Meekhof, Tim Keosababian, pending patent
Avatar Kinect

- Chatting room
  - Hang out with friends on Xbox
- Live meeting
- Facebook, Youtube
  - Funny greetings, avatar comments
A demo
Improved AAM face tracking

- Temporal matching constraint
- better initialization for fast motion

AAM based Face Tracking with Temporal Matching and Face Segmentation, Mingcai Zhou, Lin Liang, Jian Sun and Yangsheng Wang, CVPR 2010
Improved AAM face tracking

• Temporal matching constraint
  • AAM model fitting constrained by feature matching

Frame t-1  Basic AAM  Result with temporal matching constraint
Improved AAM face tracking

- Depth Map Constraint
  - A soft constraint using depth based segmentation

Without depth constraint
Remove background
Our method
Skeleton Correction and Tagging

Depth Image → Background Removal → Skeleton Extraction → Skeleton Correction
A demo

Ground-truth

Kinect estimation

After correction
- Sometimes, only a gesture status value is needed
Regression from initial skeletons

- Random forest + Cascaded pose regression + temporal optimization

Exemplar-Based Human Action Pose Correction and Tagging, Wei Shen, Ke Deng, Xiang Bai, Tommer Leyvand, Baining Guo, Zhuowen Tu, CVPR 2012
Why is this possible and useful?

- Difficult in general
- Systematic errors under similar poses
- Perform correction case by case
- Used in Xbox gesture builder
Kinect Object Digitization

Transfer to Xbox, Minmin Gong, Xin Sun, Xin Tong and Baining Guo, MSRA
Techniques

- Data-Parallel Octrees for Surface Reconstruction, Kun Zhou, Minmin Gong, Xin Huang, Baining Guo, TVCG 2010
  - GPU based construction of octrees
  - Poisson surface reconstruction

- Highly optimized on Xbox
  - two scans of the object: front and back
  - 2 seconds for model creation
More in the future

• More accurate, robust, faster...

• Revolutionary user interaction experience
  • body, face, hand, eye, ...

• Virtual reality: games, social activities,...

• Beyond Xbox
  • PC, notebook, pad, and new wearable devices,...
Thank you!