Microsoft Research

Microsoft Research Asia
Faculty Summit 2012

Sign Language Recognition and Translation Based on Kinect

Xilin Chen

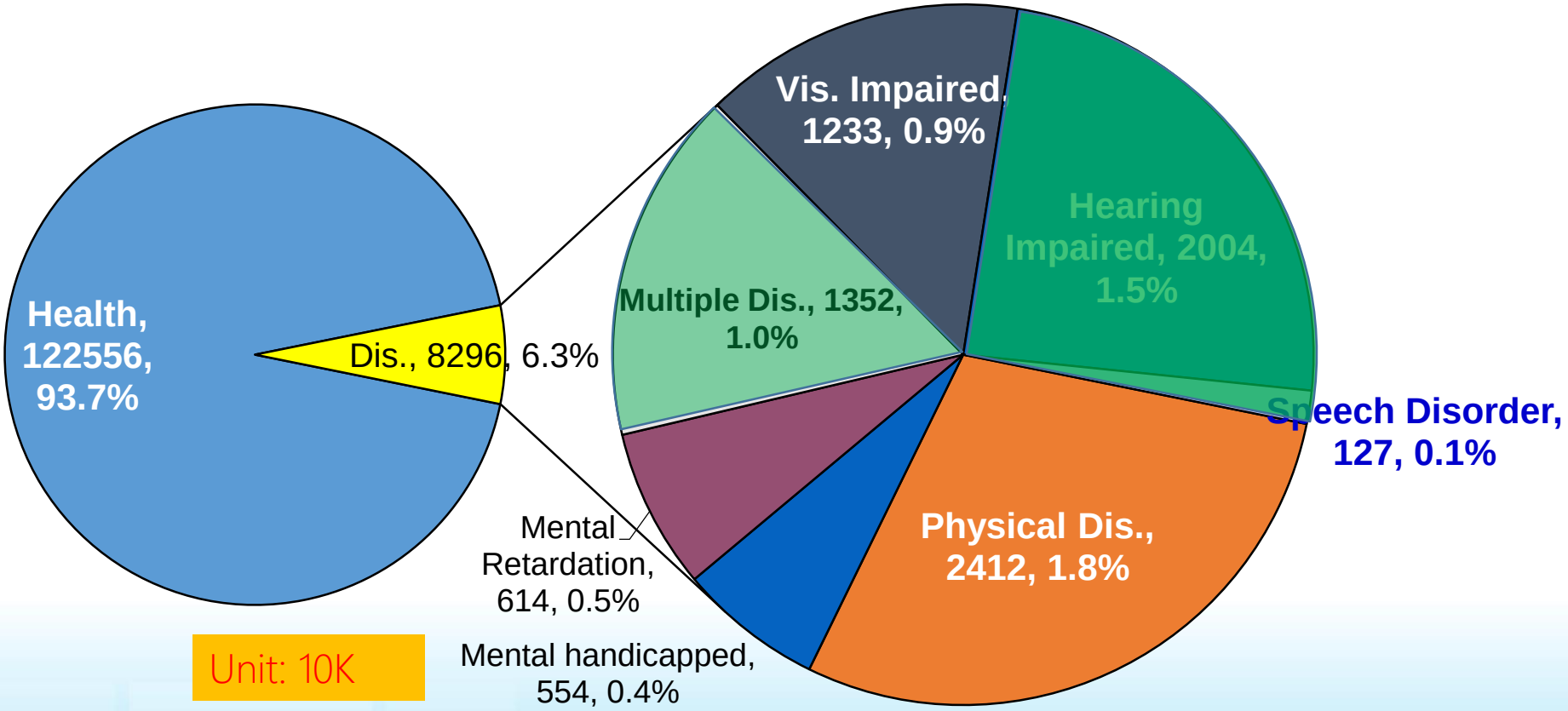
Institute of Computing Technology,
Chinese Academy of Sciences

Microsoft Research Asia
Faculty Summit 2012

Acknowledgement

- This is a joint work with
  - Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Jialu Zhu, Xiujuan Chai from ICT, CAS
  - Hanjing Li from Beijing Union University
  - Xin Tong, Zhuowen Tu, Jian Sun, Ning Xu, Guobin Wu, Ming Zhou from MSRA
  - Zhengyou Zhang from MSR
- Thanks for those students who make big contribution on data collection from BUU, especially thanks for Hui Liu , and Dandan Yin

Disabled People in China

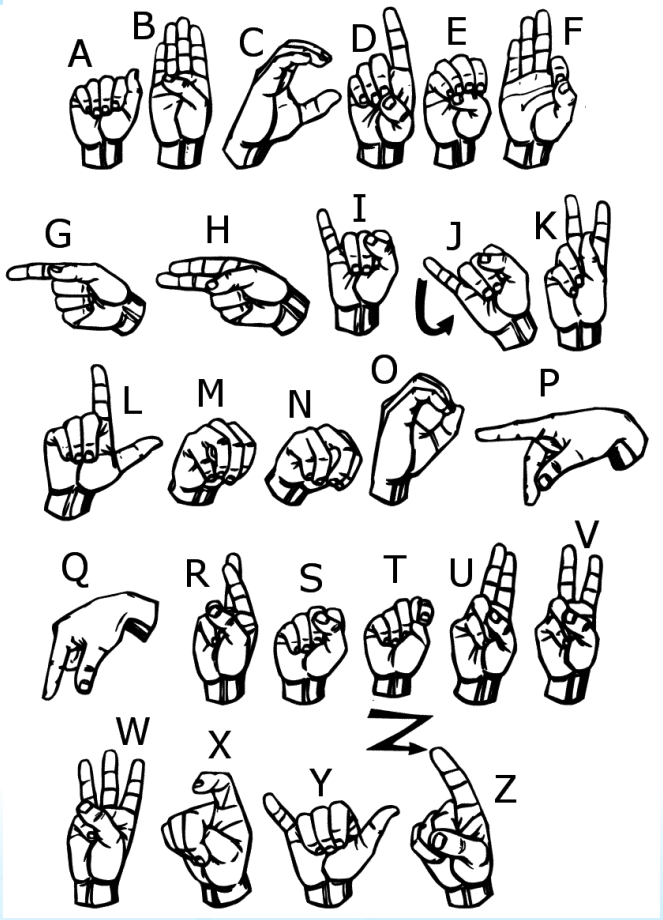


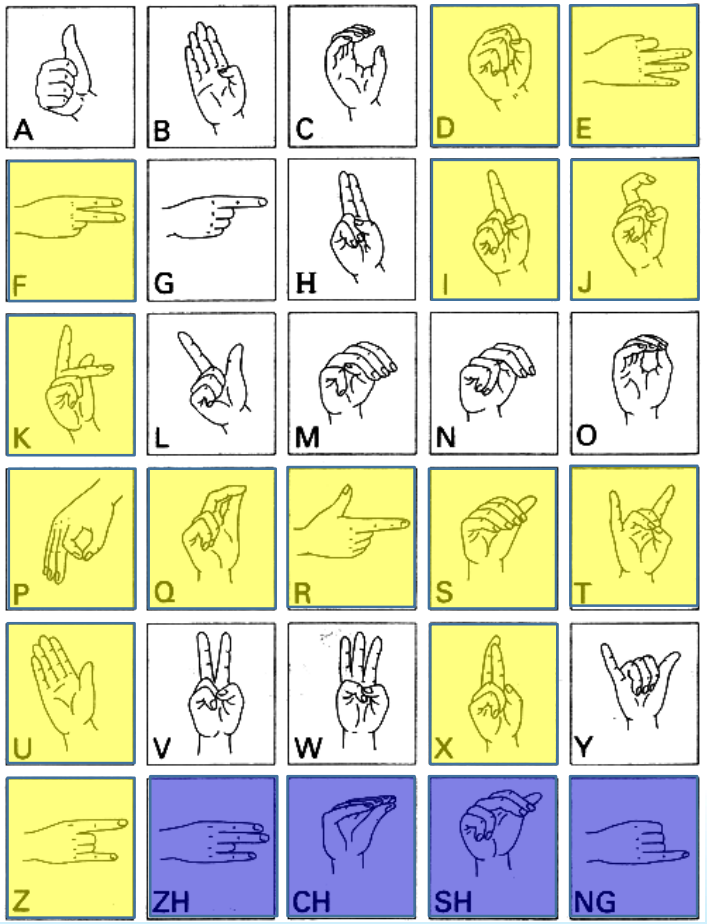
Source: 2nd census of disabled people in China, 2006

Sign Language

- 100 million people use sign language in China and 200 million people in the world
- Sign language is recognized as a natural language in many countries
- Language barrier between deaf-mute and health people
  - Human sign language translator is a hot job
- Automatic sign language translator
  - Automatic sign language recognition and generation

Alphabets in American / Chinese SL

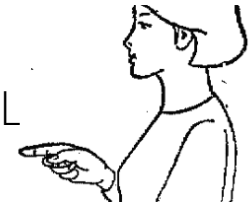




Some words in ASL / CSL

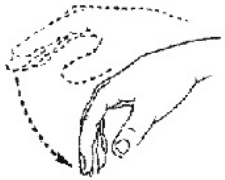
ASL

CSL



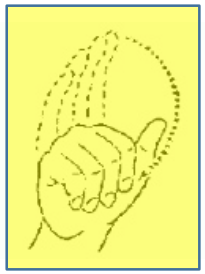
你(You)

好 (Good)



来(Come)

我(Me)



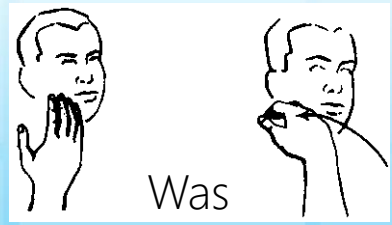
能(Can)

请(Please)

不(No)

Are

Be



Was

是 (be/is/are/was/were)

Challenges in SL Translation

- A large vocabulary set for recognition
  - 5000+ words in Chinese Sign Language

Challenges in SL Translation

- A large vocabulary set for recognition
- Motion and posture in different scale
  - Some words with only one posture
  - Some words only with fingers motion, e.g. 谢谢 (thanks)
  - Some words with significant hand / arm motion, e.g. **大家**(everyone)

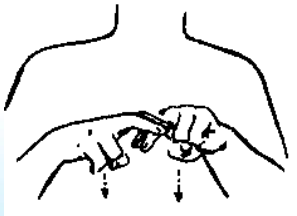
五(Five)

谢谢(thanks)

大家(everyone)

Challenges in SL Translation

- A large vocabulary set for recognition
- Motion and posture in different scale
- Vocabulary set is relatively smaller than spoken language
  - Thousands words vs. 100+ thousands ones
  - Many to one mapping
    - Sit / Chair → same gesture



Challenges in SL Translation

- A large vocabulary set for recognition
- Motion and posture in different scale
- Vocabulary set is relatively smaller than spoken language
- Grammar is different
  - English: I like to fly small planes.
  - Sign: SMALL PLANES — FLY — LIKE ME

Lessons from Previous Works

- SL recognition with video camera
  - Only works on a small vocabulary set
  - Segmentation is a big challenge
  - Sensitive to lighting change

Lessons from Previous Works

- SL recognition with video camera
- Data-glove based sign language recognition
  - Input: Data-glove + Location Sensor
  - Recognition Model: HMM
  - Merits
    - Stable Input
    - Supportable to large vocabulary set (5000+ words)

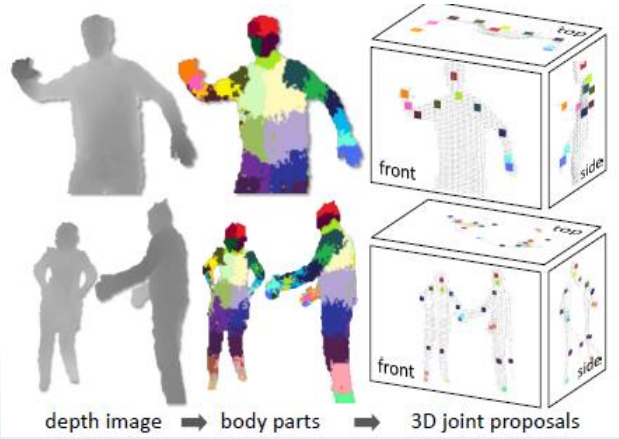
CSL Recognition with Data-glove

Lessons from Previous Works

- SL recognition with video camera
- Data-glove based sign language recognition
  - Input: Data-glove + Location Sensor
  - Recognition Model: HMM
  - Merits
    - Stable Input
    - Supportable to large vocabulary set (5000+ words)
  - Demerits
    - Too expensive
    - Extra accessories
    - Easy damaged

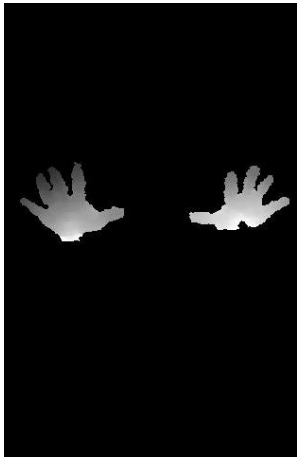
Kinect – an opportunity for SL Recognition

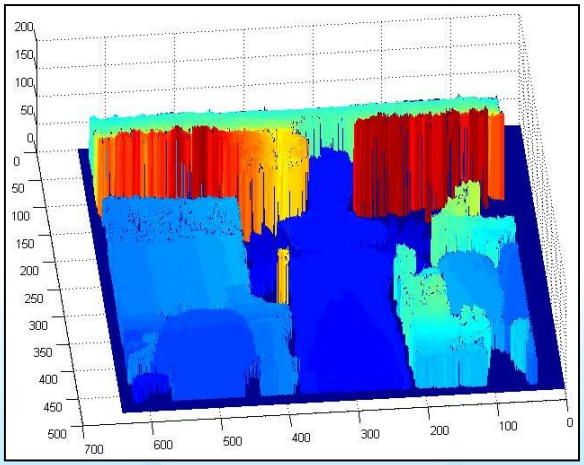
- Depth provides additional robust information
  - Body segmentation / tracking
- Balance between data-glove and pure visual camera
  - Cost
  - Robustness
  - Understandable to raw data

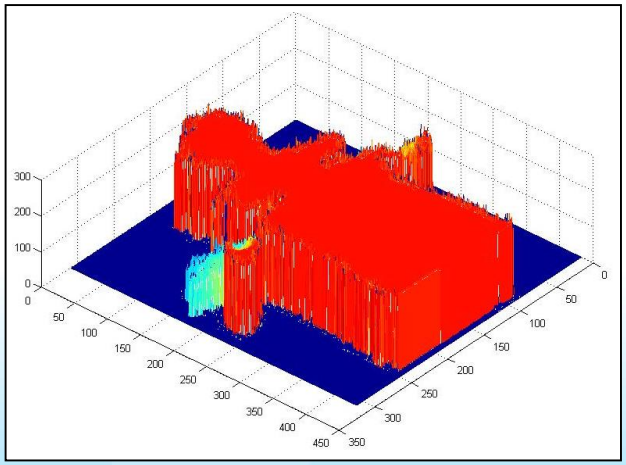


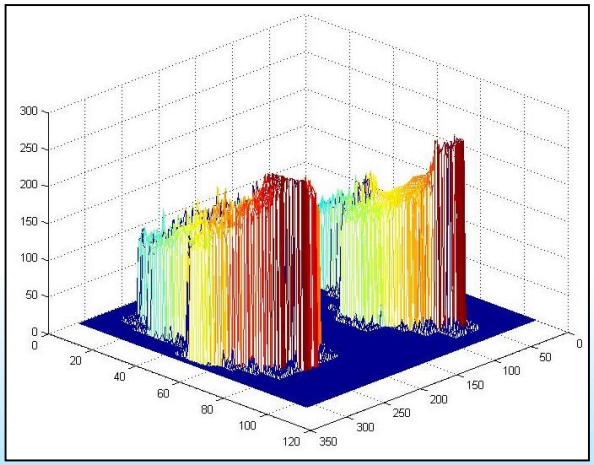
Shotton et al. CVPR11

An Example from Kinect









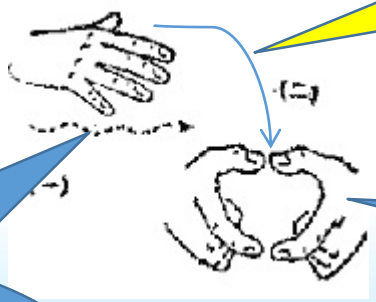
Basic Idea

- SL = Hand Motion + (Face expression)
- Hand Motion = Trajectory + Key postures
- Basic idea from SL dictionary
  - Postures + a few trajectories

Basic Idea

- SL = Hand Motion + (Face expression)
- Hand Motion = Trajectory + Key postures
- Basic idea from SL dictionary
  - Postures + a few trajectories

Even some clips aren't essential elements in SL, they still encode important context



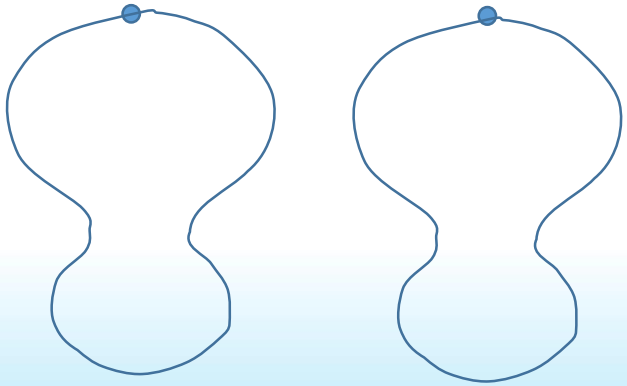
水果 (Fruit)

Some clips of the trajectory are essential elements in SL

Postures are basic elements in SL

Recognizing SL from trajectory

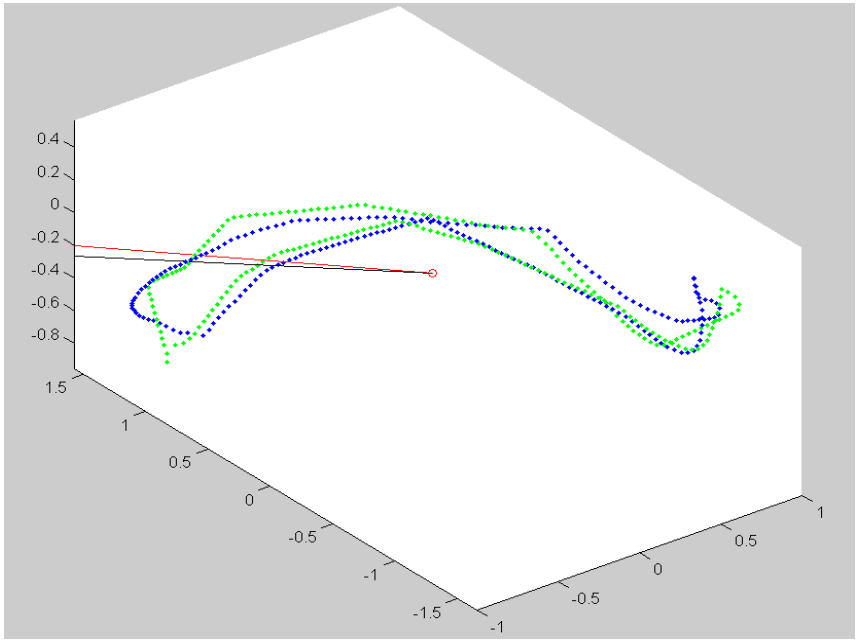
- Basic task
  - $D = f\ (c_1, c_2)$, where c_1 and c_2 are two curves in 3D space
  - Manifolds matching and distance measuring
- People play SL in different cases
  - Speed (duration) to play a sign
  - Height of the signer
  - Slightly different in pose



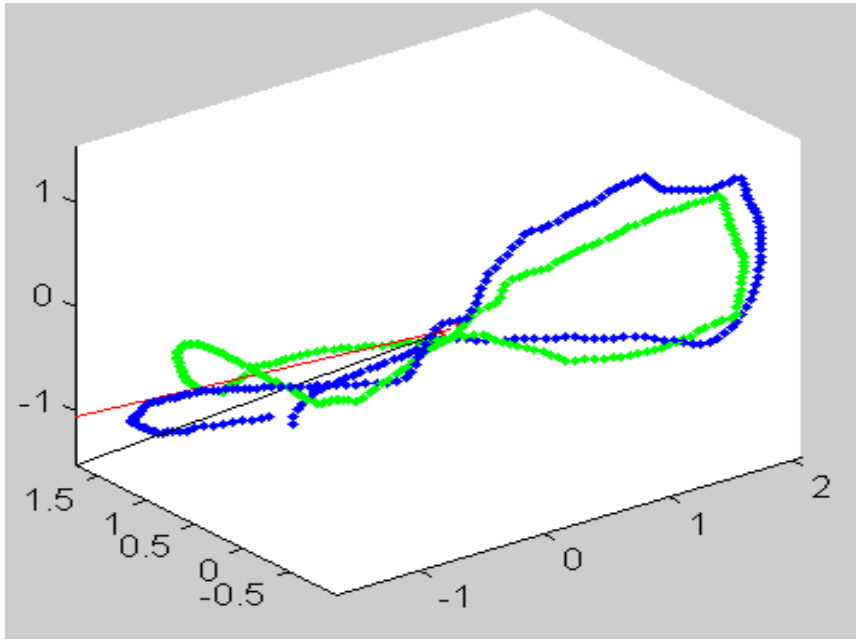
Alignment of Trajectories

- A essential step to deal with various distortions
  - Speed (duration) to play a sign
  - Height of the signer
  - Slightly different in pose
- Noise remove to improve robustness
- Trajectory interpolation
  - Improve the performance on different speed
- Trajectory length normalization
  - Improve the performance between different signers (height)
- Calculation principle direction
  - Independent with pose

Examples of Aligned Trajectories



Everyone(**大家**)

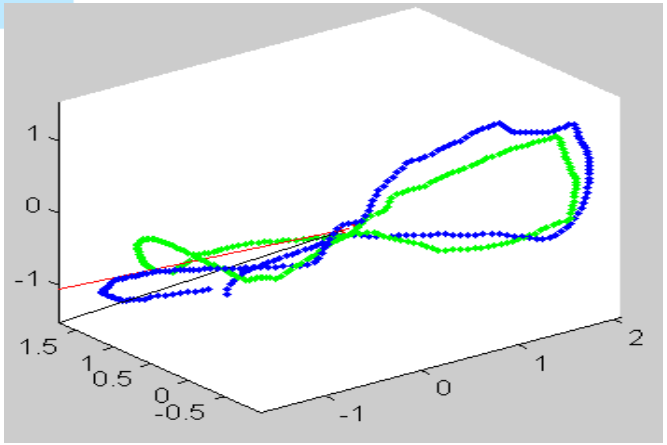


On purpose (**故意**)

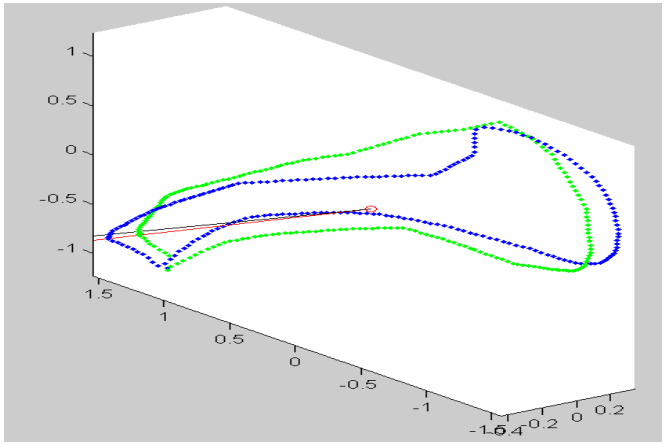
Black line: principle direction of blue curve
Red line: principle direction of green curve

*All trajectories above from right hand

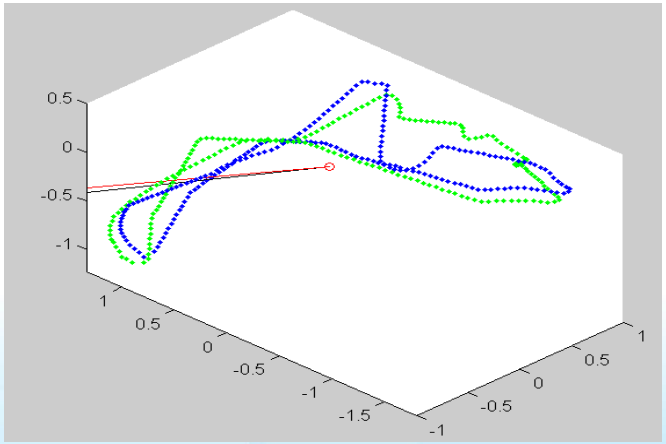
Matching Same Word Trajectories



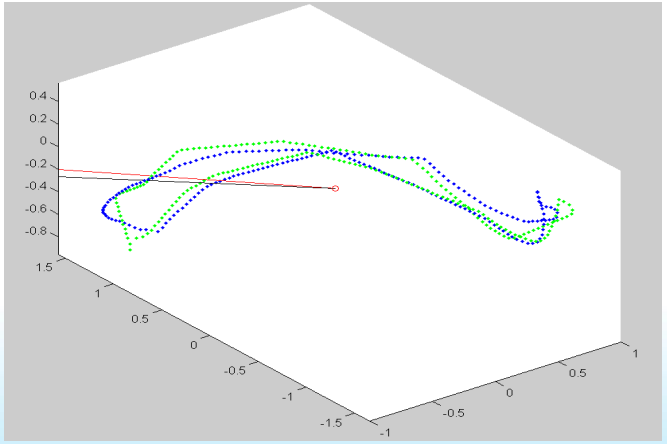
Everyone (**大家**) (d = 561)



Reach(**到**) (d= 162)



Reserve(**保留**) (d=400)



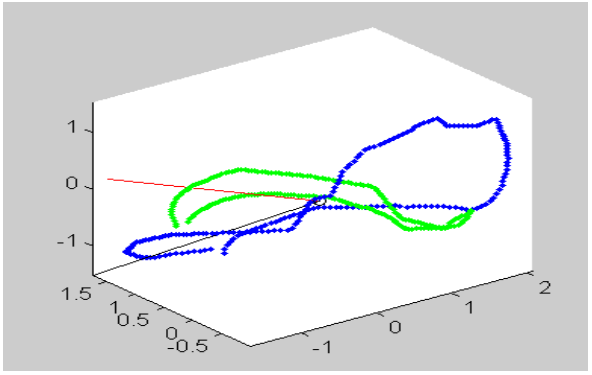
On purpose(**故意**) (d=212)

Matching Different Word Trajectories

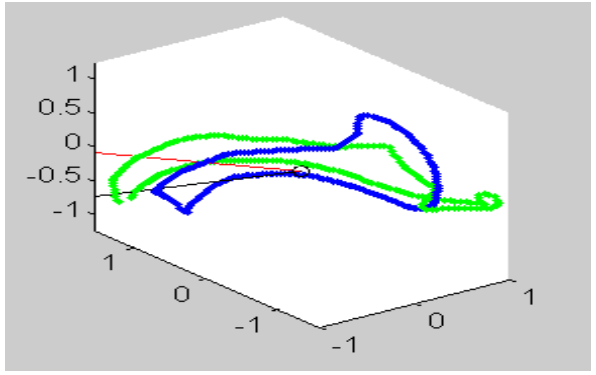
Everyone(blue)

Reach (Blue)

On Purpose (Green)

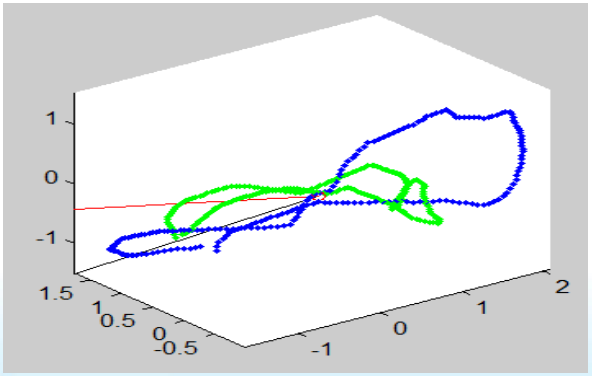


d=1,079

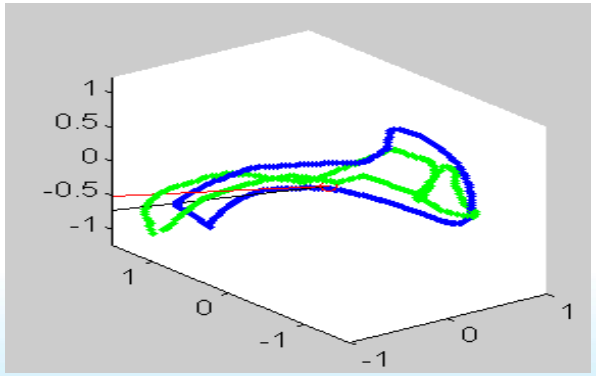


d=380

Reserve (Green)



d= 41,149



d=40,508

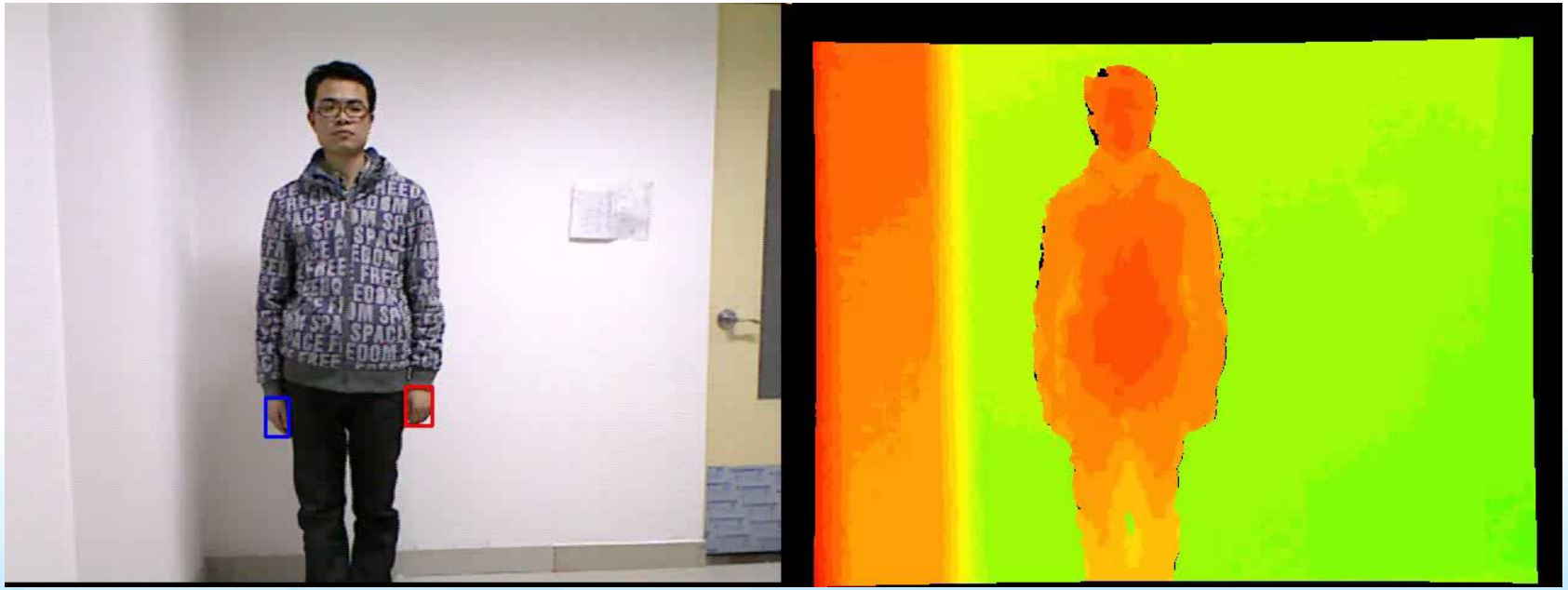
Trajectory-based Recognition Result

rank	count	rate
1	180	75.3%
5	225	94.1%
10	232	97.1%
20	235	98.3%
50	237	99.2%

- Vocabulary set size: 239

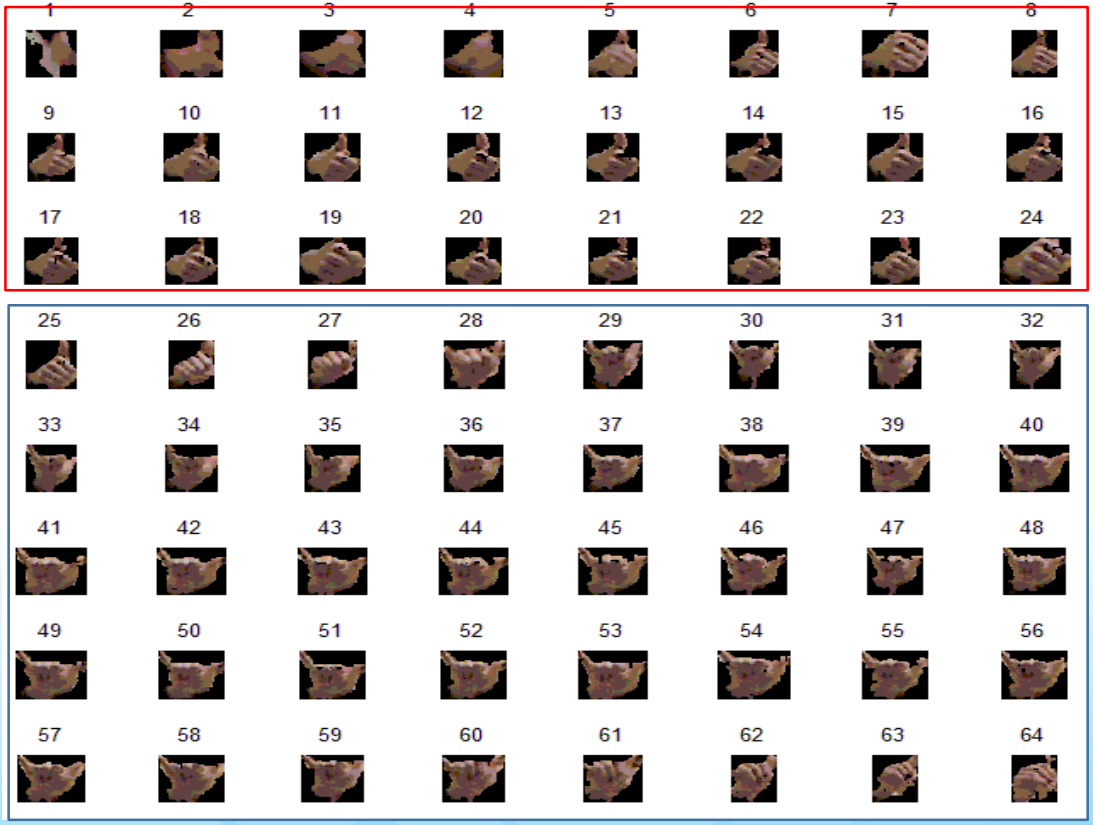
Posture Recognition

- Key posture detection
- Key posture recognition



Posture Recognition

- Key posture detection
  - Intersection-union ratio



Posture Recognition

- Key posture detection
  - Intersection-union ratio
- Key posture recognition
  - PCA used for orientation normalization
  - Normalize hand size to 64*64
  - HOG feature
    - block size(8*8)
    - cell size(8*8)
    - 9 bins
  - LDA use for recognition

Towards Better Communication with Kinect

Institute of Computing Technology, CAS
Beijing Union University
Microsoft Research Asia

Oct. 2012

Microsoft Research

Thank you!

Microsoft Research Asia
Faculty Summit 2012

Microsoft Research

Microsoft Research Asia
Faculty Summit 2012