An Immersive System with Multi-modal Human-computer Interaction

Rui Zhao¹, Kang Wang¹, Rahul Divekar¹, Robert Rouhani¹, Hui Su¹,², Qiang Ji¹
¹ RPI, ² IBM T. J. Watson Center

Overview
- An immersive system prototype that integrates face, gesture and speech recognition techniques to support multi-modal human-computer interaction capability.
- Understand users’ intention by resolving query like ‘who is doing what at where’ in real-time.

System Architecture

3-screen Display Set-up
Panoramic Display Set-up

Sensing Devices
- Kinect Camera, Pan-Tilt-Zoom Camera, Webcam, Lapel Microphone

Face Detection

Gesture Recognition

Head Pose Estimation

Use Case: Immersive Classroom
- Immersive Language Learning Experience through enabling natural multi-modal interaction
- User study shows benefit of the system

Use Case: User Registration
- System initiates registration request by voice prompt
- User responds by speech and gesture

Use Case: Meeting Assistance
- Assist information analysis and decision making through verbal and gesture interactions.
- Monitoring the progress through fusing multiple input modalities like speech, gesture and facial expression.
- Detect user attention through consensus of head pose.

Acknowledgement
This work is supported by Cognitive Immersive Systems Lab (CISL), a collaboration between IBM and RPI, and also a center in IBM’s Artificial Intelligence Network.

Contact author: zhaor@rpi.edu