

# IMI RESEARCH SEMINAR

**DATE:** 10 December 2014, Wednesday

**TIME:** 11:00 am – 12:30 pm

**VENUE:** IMI Seminar Room, Research Techno Plaza, XFrontiers, Level 03-01  
50 Nanyang Drive, Singapore 637553

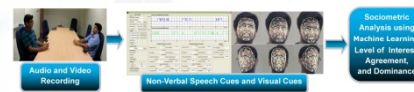
\*Lunch will be served



**Dr CHEN Renjie**  
Visiting Researcher,  
IMI

## *Wide Field of View Compressive Light Field Display Using a Multilayer Architecture and Tracked Viewers*

We will discuss an intuitive extension to compressive multilayer light field displays that greatly extends their field of view and depth of field. Rather than optimizing these displays to create a moderately narrow field of view at the center of the display, we constrain optimization to create narrow view cones that are directed to the few viewers' eyes. These narrow view cones follow the viewers, creating a wide apparent field of view. The view cones can be scaled to match the positional error and latency of the tracking system. Using more efficient optimization and commodity tracking hardware and software, we demonstrate a real-time, glasses-free 3D display that offers a 110x45 degree field of view.



**Yasir TAHIR**  
PhD Student,  
EEE / IMI

## *Real-Time Sociofeedback Using Non-Verbal Audio-Visual Cues*

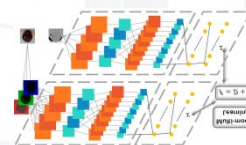
Apart from spoken words, human communication is greatly facilitated through non-verbal behavioural cues which contain a wealth of information regarding their current social state. The Socio-feedback system captures audio from lapel microphones and RGB and depth data from Microsoft Kinect devices to extract non-verbal speech and visual cues non-invasively. The system leverages on these non-verbal cues to quantitatively assess speaking mannerisms of each participant. Furthermore, the speech and visual cues are incorporated as features in machine learning algorithms to quantify various aspects of social behaviour including the level of interest, dominance and agreement. Numerical tests through leave-one-person-out cross-validation indicate that the accuracy of the algorithms for inferring the sociometrics is around 85%. The objective is to provide real-time sociofeedback where speakers are provided feedback about their social behaviour in an ongoing discussion to improve communication, thus enhancing the overall interaction experience.



**LI Bingbing**  
PhD Student,  
MAE / IMI

## *Human Motion Imitation Using Third Perspective View Optimization*

Motions such as postures and gestures are used daily communication of human being to enrich the meaning of message other than the words itself. To keep the maximum meaning of the message delivered by a human operator in the scene of using tele-presence robot, it is important to retain message of the body language. In this work, the human being's position, who is communicating with the robot, is determined. His perspective, which is third perspective in terms of the tele-presence robot, is used to determine the position of the robot to compensate the difference due to different joint sizes of human operator. A gradient optimization method is used to determine the orientation of robotic hand orientation to reduce the difference in visual effect due to the mismatch of degree of freedoms in robot and human body.



**WANG Anran**  
PhD Student,  
SCE / IMI

## *Multi-modal Feature Learning for RGB-D Object Recognition*

Most of the existing feature learning based methods for RGB-D object recognition either simply combine RGB-D data together as the learning input or learn features from color and depth separately, which do not well consider different characteristics of the two modalities as well as the relationship between them. In this paper, we propose a general CNN based multi-modal learning framework for RGB-D object recognition. We first construct deep CNN layers for color and depth separately, and then connect them with a carefully designed multi-modal layer, where we not only search for most discriminative features for each modality, but also take into account the relationship between two modalities to obtain more complementary information. The results of the multi-modal layer are back-propagated to update the parameters of CNNs. The multi-modal feature learning and the back-propagation are iteratively performed till convergence. Experimental results on two widely used RGB-D object datasets show that our method for general multi-modal learning achieves comparable performance to state-of-the-art methods that are specifically designed for RGB-D data.