

# UW CENTER FOR PATTERN ANALYSIS AND MACHINE INTELLIGENCE

## GRADUATE SEMINAR SERIES

### Audio-Visual Feature-Decision Level Fusion for Spontaneous Emotion Estimation in Speech Conversations

**Speaker:** Rodrigo Araujo

**Date:** August 7, 2013

**Time:** 4:30pm – 5:00 pm

**Place:** E5 (5128) Refreshments will be served

#### **Abstract :**

Recognition of spontaneous human affect has gained a lot of interest recently after a shift in research focus from the traditional six prototypical emotions. In this paper, a combined bi-modal feature-decision fusion approach is proposed to enhance the performance of estimating emotions from spontaneous speech conversations. First, a feature vector consisting of audio information extracted from the whole speech sentence is combined with video features of the individual key frames representing that sentence. Then, the final estimate is calculated by a decision level fusion of predictions from all corresponding frames. The performance is compared with two fusion approaches, the decision level fusion using weighted linear combination, and a simple feature level fusion. The experimental results show improvement in correlation between the emotion estimates and the audio references. In addition, we evaluated the performance of supervised PCA (SPCA) for dimensionality reduction on this audio and video emotion estimation application, which led to better results compared to traditional principle component analysis (PCA).