

The paper introduces an approach to identify three social interactions, dialogue, discussion and monologue, by using face detection in first person video. The idea is that the face indicates important information during social interaction. Attention pattern of people's faces can be associated with certain events when people pay attention to someone or something. Many social interactions don't evolve activity interaction. Also detecting face is much easier than detecting the activity or the whole object.

Face orientation and location are easily computed. People's attention is computed by MRF. The unary potential is the likelihood of looking at a grid cell. It consists of three terms. The first term is Gaussian function which indicates that a person has larger likelihood to look around himself/herself in the face orientation. The second function is sigmoid function which avoids a face to look at itself. The third function means that a face has increasing chance to look at another face. Then binary potential defines that an object is more likely to be looked at if there are more people to look at it. A clustering algorithm assigns each face's attention to a particular location. Each face has four features based on the attention information. Then each social interaction is associated with one attention pattern by computing the k-mean over the training data. HCRF is used to train the model with dense optical flow.