Team name	TUM-fortiss
Team leader name	Guang Chen
Team leader address, phone number and email	fortiss GmbH, Guerickestr. 25 , 80805 München, Germany +49.89.360.352239 chenhnu@gmail.com
Rest of team members	Yanping Zhang, Daniel Clarke, Alois Knoll
Team website URL (if any)	

Title of the contribution	Multi-Modality Gesture Recognition With Un-supervision and Randomization
General method description	1. We firstly segment the sample (video) into several candidate gesture clips using the skeleton data. To achieve this, we designed two features based on the skeleton data and trained a SVM model for each sample to classify the gesture frames and non-gesture frames. We then combine all SVM models together to classify the testing samples. Each frame of the testing sample will be assigned to label 1 (belong to gesture candidate clip) or 0 (belong to non-gesture clip) 2. We extract several spatio-temporal features using unsupervided learning algorithm. We apply the independent subspace analysis to depth video, grayscale video, gradient map video, normal(z) map video. Finally, we get four types of spatio-temporal features. 3. For each candidate gesture clip, we densly extract four types of features, encode the features by LLC algorithm. We split the clip into 8 intervals using maximum pooling. 4. We densely sample enough video cuboids, and use random forest and SVM to find the cuboids that contain discriminative information. Specifically, we randomly sample 200 cuboids for each node of the tree, then use SVM to select the best cuboids. We train 100 trees for each type of

Describe data preprocessing techniques applied (if any)	Segment video sample into candidate gesture clips using skeleton information.
Describe features used or data representation model (if any)	Unsupervised spatio-tempral features (ISAdepth, ISAgray, ISAnormal, ISAgradient)
Data modalities used, i.e. depth, rgb, skeleton… (if any)	Skeleton, Depth video, Grayscale video, Normal Map video, Gradient Map video
Fusion strategy applied (if any)	Late fusion
Dimensionality reduction technique applied (if any)	

Temporal clustering approach (if any)	No
Temporal segmentation approach (if any)	Based on skeleton data
Gesture representation approach (if any)	Discriminative Sampled Cuboids Minied by Random Forests and SVM
Classifier used (if any)	RF, SVM
Large scale strategy (if any)	No

Transfer learning strategy (if any)	No
Temporal coherence and/or tracking approach considered (if any)	No
Other technique/strategy used not included in previous items (if any)	No
Method complexity analysis	Training: about 48 hours on Linux Server, RAM (64G), 12 Cores for parallel computing Testing: about 36 hours on Linux Server, RAM (64G), 12 Cores for parallel computing

Qualitative advantages of the proposed solution	Novel, effective and efficient
Results of the comparison to other approaches (if any)	No
Novelty degree of the solution and if is has been previously published	It is the first time to apply the independent subspace analysis for multi modality data (e.g. Gradient Map data and Normal Map video data) to learn the spatio-temoral features. We avoid the laborious process of hand-crafting features. We design a new approach(not published) by combining un-supervision, randomization and discrimination for multi modality gesture recognition.

Language and implementation details (including platform, memory, parallelization requirements)	**Matlab, Python, C++** **Using Python to convert the video data to .mat format data** **Using Matlab mex C++ to train the RF model** **Parallelization: matlab with 12 cores.** **Memory: 64GB**
Human effort required for implementation, training and validation?	No
Training/testing expended time?	Training: 48h Testing: 36h
General comments and impressions of the challenge	Very Good