Intelligent Transportation Systems

In-Vehicle Hand Gesture Recognition using Hidden Markov Models

李万里
李仁发教授

2017.06.16

论文相关信息

- Title: In-Vehicle Hand Gesture Recognition using Hidden Markov Models
- Authors: Nachiket Deo, Akshay Rangesh and Mohan Trivedi

  @University of California San Diego

- Published in: Intelligent Transportation Systems (ITSC),

  2016 IEEE 19th International Conference on
- Date of Conference: 1-4 Nov. 2016

问题　In-Vehicle Hand Gesture Recognition

Hand Gesture Recognition

- temporal and postural variability
- different users

In-Vehicle

not a controlled indoor environment

- the driver or the passenger
- rapid illumination changes and shadow artifacts

相关研究

- SVM based gesture classifier (2014)
- using a 3D Convolutional Neural Network (2015)

模型　Hidden Markov Models

- 对时间序列建模

  generative models inherently capable of modeling time series

- Hidden Markov Models (HMMs)

- For Automatic Speech Recognition
  - the spectral and temporal variability of speech signals

- For hand gesture recognition
  - variations in hand posture, the temporal variability

方法

- 1) optimal hyperparameters -- topology and training
- 2) shape descriptors -- fetures
- 3) reducing overfitting
  - Dimensionality reduction
  - Data Augmentation

实验

- 从dataset说起

  The VIVA hand gesture dataset

- grayscale and depth videos of dynamic hand gestures
- using a Microsoft Kinect device (115 * 250 pixels)

- 19 different gestures, 8 different subjects
- 885 gesture videos

A. HMM parameter sweep

- 1) Number of States
- 2) Number of mixture components

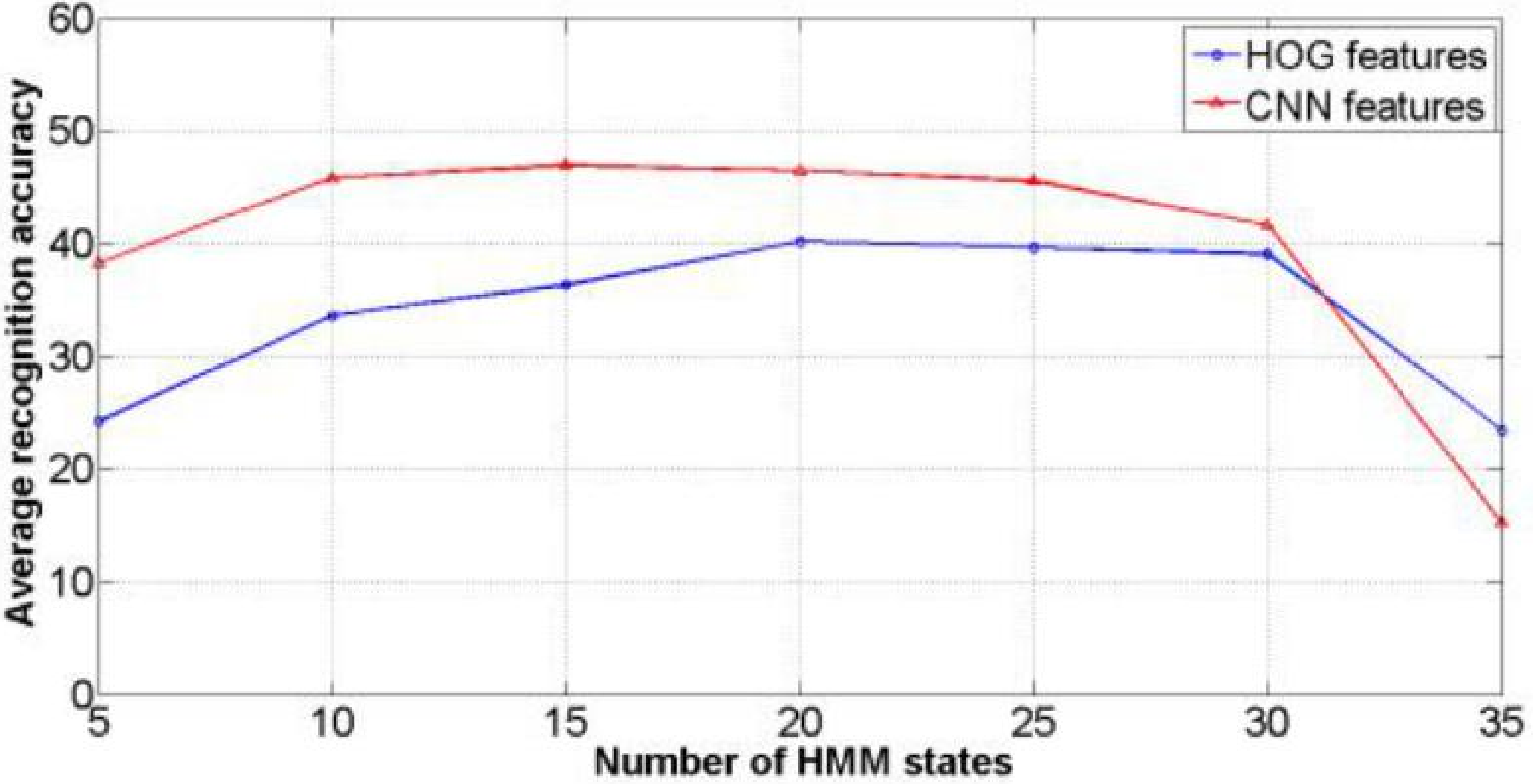

Fig. 1. Effect of varying number of HMM states

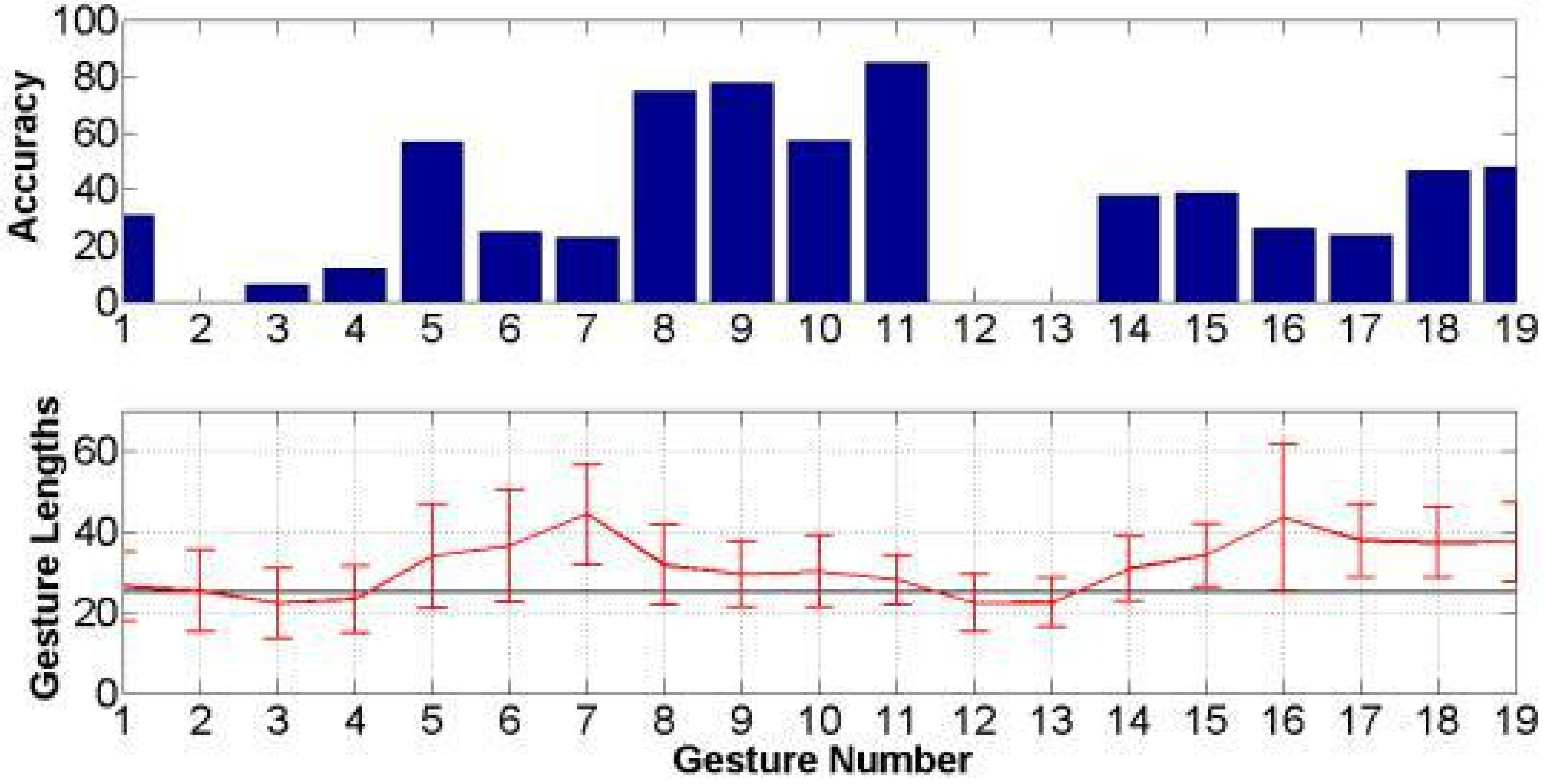

Fig. 2. Comparison of gesture wise recognition accuracies and average gesture lengths for 25 HMM states

B. Comparison of features and modalities

- 1) HOG features VS CNN features
- 2) depth, grayscale, or both

Average recognition accuracies & their standars deviations

Features	Modality		
	Depth	**Grayscale**	**Both**
HOG	**41.51 ± 11.56**	18.02 ± 7.22	38.27 ± 11.83
CNN	46.48 ± 10.31	39.14 ± 10.09	**54.76 ± 12.7**

C. Reducing overfitting in the HMM

- 1) Dimensionality reduction using PCA

- 2) Data Augmentation

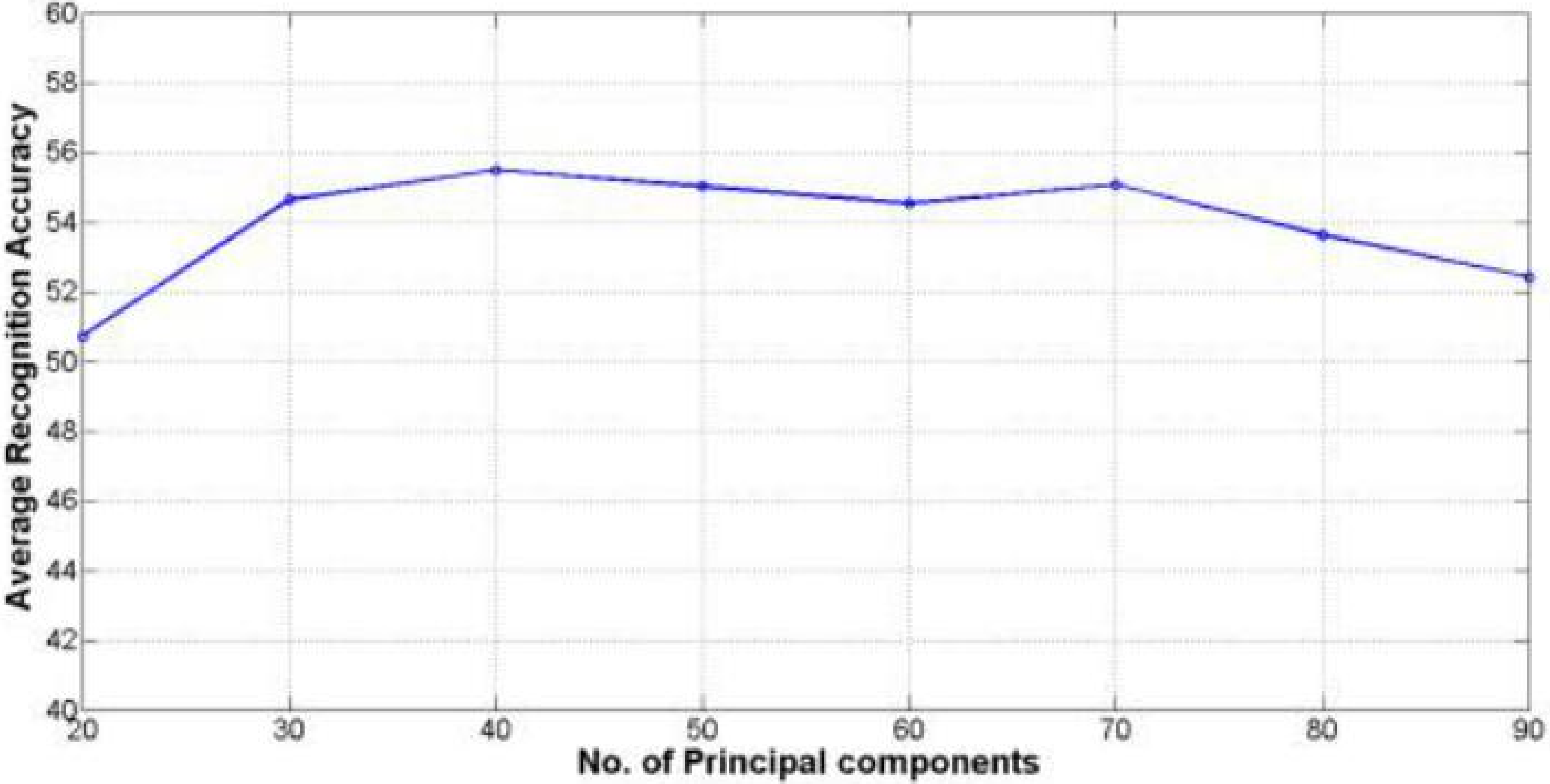

Fig. 6. Effect of varying number of retained principal components

Effect of Data Augmentation

Modality	**Without data augmentation**	**Ordering / Orientation transfromation**	**Affine transformation**
Depth	53.55 ± 12.23	49.03 ± 14.53	55.32 ± 14.13
Grayscale	42.28 ± 10.97	44.37 ± 7.73	46.75 ± 9.44
Both	55.49 ± 12.65	53.52 ± 10.11	$\mathbf{55.71 \pm 10.40}$

CNN-HMM hybrid

- a trained CNN replaces Gaussian mixture models
  as the state emission probability estimator of the HMM

- the ImageNet trained VGG-16 network

CNN as feature extractor	**CNN-HMM Hybrid**
55.71 ± 10.40 %	**57.50 ± 13.05 %**

Discussion

- 1) 他山之石 —— 语音识别
- 2) Hybrid framework 方法的融合
- 3) not the best result, 如何自圆其说？
  - 讲清工作；指明意义；大牛影响…
- 4) 实验。扎实。

Thank you