11-755/18-797 Machine Learning for Signal Processing

Final Poster Presentations

Instructor: Bhiksha Raj

TAs: James Ding and Varun Gupta

Fall 2013

1. USING SPEECH SYNTHESIS TECHNIQUES TO MODEL MUSICAL ARTICULATION

Andrew Russell, Tina Liu, Vinay Vemuri

Sound synthesis is very common in today's popular music. Popular music will often use synthesized instruments playing in various styles. However, the synthesis techniques used do not accurately capture the intricacies of the desired style. Our system better understands a specific style of music, and more accurately captures the style's unique elements while synthesizing.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

2. AUTOMOBILE VISION LOCALIZATION

Ying Sheng, ChiYu (Donny) Dong, Xiaobo Hu

The GPS signal could be blocked in the downtown area. Thus, we need a new method to take the place of GPS when it cannot give good instruction. Alternatively, the vehicle could locate itself by using the nearby environment. Intuitively, the vehicle takes picture when it runs on the road. Then a program will match this input picture to a database, in which a sequence of pictures is taken with their coordinates. In the ideal case, the program could perfectly find a picture from the database such that it matches the input picture very well, and return the coordinates with the picture. Finally, this coordinate could be used to obtain the vehicle position.

Instead of making a 3D map or combining range information, we only rely on single camera images which were captured along a highway. SURF was used to extract feature points. In the case of suffering from high dimensions, we applied a bag-of-words algorithm (a.k.a. "Texton") to dramatically decrease dimensions. (*i.e.* 2000x64 decrease to 1x100). Thus, one frame could be described by a single 100-element vector. Therefore, we could not only build a map with a little size, but also easily compare input images to the map.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

3. LYRIC RECOGNITION

Serim Park and Suyoun Kim

The goal of this project is to build a lyric recognition system, which takes a segment of vocal song as an input and generates the lyric of that line as an output.

This system is analogous to the speech recognition system in that it tries to find the correct words/sentences of the given acoustic signal. The natural question to ask first would be whether current speech recognition systems can understand the lyrics of songs, and the answer is no. The way English language is treated in songs and speech are totally different, as are the inherent acoustic characteristics of songs and speech. For example, grammatical orders break down in songs even more than the spoken language, word pronunciation in songs depends on musical context rather than its original pronunciation, and frequency of used words is completely different for that of song compared to speech. Moreover, background music in songs interferes with the recognition procedure, and extracting only the vocal line from the background music isn't as easy as removing noise from a speech signal. Background music, which has its own distinctive spectrogram pattern compared to background noise, repeatedly appears in a song and confuses the recognition system, which cannot distinguish if the current segment of signal is a word or a drum sound. Due to such reasons, lyric recognition remained as an unsolved problem. However, if we can wisely take advantage of its similarity with speech recognition system and adapt the system according to the underlying properties of songs rather than the speech, lyrics recognition wouldn't be infeasible, as it seems like. We note that the lyric recognition system in our project is much more simplified version than the normal speech recognition system, modified especially for lyric recognition.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

4. IMAGING WITHOUT A CAMERA

Nikhil Balchandani

By illuminating a scene with a projector, and generating measurements with a photodetector, it is possible to build an image of the scene. Using compressed sensing we can reduce the number of measurements needed, and speed up the process. With multiple photodetectors we can generate a 3 dimensional depth map of the scene. 3D sensing without a camera enables the sensor to operate in any light spectrum. Photodetectors come in a variety of operating wavelengths and the LED in the projector can be replaced with ones in different spectrums.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

5. VIDEO CONTENT EXTRACTION REVIEW

Aadil Khan

Currently, the world is finding itself producing large amounts of digital information taking the form of film, YouTube videos, home videos taken by personal camcorders, and many more multimedia representations. There is high motivation to extract valuable and unique information from video data with ease of access. Speech and textual data may be developed enough to procure information, but video data owns extra complexity in deciphering features that best assist in content extraction. Here, we review techniques that explore segmenting video scenes into units that help combine into a story. Stories are linked together through relative correlation of semantics. Additionally, indicating objects and actions as points of interest in the video helps complete our understanding of what is going on its content.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

6. HANDWRITING RECOGNITION WITH KINECT

Tsu-Hao Kuo, Wei-Ting Yeh, Yin-Chen Chang, Dwight Dyi Tse Liu

Handwriting recognition is a widely-used application of machine learning in daily living. With the help of Microsoft Kinect, we can enhance the user experience of interaction with machine. In this project, we propose a system to perform handwriting recognition with the help of Kinect, giving users a way to interact with computer by writing in the air. This provides the users with a type of new handwriting input interactive experience.

We aim to solve a classic handwriting classification problem through the natural use of our figures to write in the air. By using Kinect we attempt to track the finger movements of people in real time. We could then identify all 26 alphabets through the trajectory with machine learning methods

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

7. GENDER CLASSIFICATION FROM FRONTAL FACE IMAGES

Chandrasekhar Bhagavatula and Jonathan Carreon

Various facial recognition techniques can perform unique identification well under constrained circumstances nowadays. While the desired subject may appear near the top of a rank list percentage wise, other attributes remain unavailable to those wishing to use this technology. Attributes such as gender and ethnicity recognition can be used to reduce the list of possible matches by excluding those who don't match the proper soft biometrics. We show that by normalizing input images by the eye coordinates and using simple classifiers, we are able to achieve high gender classification accuracies in the 90% range even when testing across different databases and demographics under widely varying constraints.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

8. DNN ACOUSTIC MODEL COMBINATION FOR EFFICIENT SPEECH RECOGNITION

Wonkyum Lee, Jungsuk Kim, Srikanth Kalluri, Farak Yavuz

The goal of this project is to find out the best way to build acoustic models that are computationally efficient and accurate. The focus was on acoustic model combination in order to improve the speech recognition performance. Vietnamese telephony speech data was used for evaluation of the models. Three Gaussian Mixture Model (GMM)/Hidden Markov Model (HMM) Acoustic Models (AMs) were built. Expectation Maximization (EM) algorithm was used to estimate parameters for the GMM from training data. Also, three Deep Neural Network (DNN)/HMM AMs were built using non-linear sigmoid and softmax functions. In order to build three distinct AMs, we extracted three different features, which were fed into each GMM/HMM AMs and DNN/HMM AMs. Using this approach better modeling capacity was achieved. In order to build the best single AM, these three features were combined. Because of the large number of features, this AM was computationally inefficient. By combining 3 small DNN models, a better performance (WER 64% vs WER 64.4%) with less computational complexity was achieved. This was the main motivation for extending the work on combining AMs.

Various conventional combining schemes for multi-stream combination methods were researched and implemented. The experiments suggest that the arithmetic mean averaging method outperforms geometric and harmonic mean averaging models. Moreover, it is observed that diversity in model structures and features (DNN, GMM) help improve the performance notably. A novel AM combination strategy was proposed and implemented using joint DNN AM training. At the fine tuning stage, back propagation was replaced with pre-training of each neural network separately. After that, all the individual networks were jointly trained and the parameters for joint combination of networks were learnt. The proposed approach achieves better Word Error Rate (WER) than the conventional Multi-Stream AM combination (WER 55.3% vs WER 56.7%). The results provide motivation for further research on AM combination.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

9. PHYSIONET CHALLENGE 2012: PREDICTING MORTALITY IN ICU

Wei Zhang, Haibo Wang, Vishrut Gupta, Chen Chen

The project comes from the PhysioNet/Computing in Cardiology Challenge 2012, which topic is Predicting Mortality of ICU Patients. Given a set of discrete and continuous variables observed individually for each patient in hospital, and the observations of their mortality during and after hospitalization, a model should be learned, and mortality prediction should be made for unseen patient given the observations during their stay in ICU.

The project is challenging in three aspects:1. incomplete data. Not all the variables are observed during time. 2. effective feature extraction from a mixture of discrete and continuous variables. 3. Based on the result reported in the Challenge, the best binary classification measurement result is around 50% (minimum of precision and recall). In this project, we plan to revisit this problem, and try to improve the prediction accuracy by exploring the feature effectiveness and time series characteristics of the data.

We will first observe the dataset, do some statistics and find correlations between variables and the labels, and then we evaluate one or several models one the data. In this project, we want to examine the effectiveness of several models based on different considerations of the data, and experimental results are shown.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

10. EFFICIENT AUTOMATIC HUMAN ACTION TAGGING

Xiaoyi Fu, Xin Yu

Most of the current high performing methods for human action classification and recognition are based on SVM classification using low-level features based on color and texture. For real world applications, the human body appears at various scales and perspectives in images, which takes effort to capture and recognize.

To resolve this problem, our work presents a framework which is applicable to arbitrary input images containing different scales of human body: Firstly, a semantic understanding of human appearance in arbitrary input image in terms of seven human limbs (body, left upper arm, right upper arm, left lower arm, right lower arm) is learned in a supervised manner using an optimized graph-cut based algorithm and semi-automatically generated ground truth. Second, scale invariant features of each limb are generated based on the masked limb segments and the concatenated bag of visual words of all body limbs are used as feature vector for each observation (training image).

Penalized Linear Discriminant Analysis (PLDA) is then used for the action classification of high feature dimension setting to overcome the problem that traditional Linear Discriminant Analysis is not appropriate. The non-convex PLDA problem is efficiently optimized using a minorization-maximization approach when convex penalties are applied to the discriminant vectors. Our algorithm is trained on twelve classes of 1,200 images from PASCAL Action dataset. Ten-fold cross- validation is conducted and comparably good result is obtained from the initial experiments.

The method we proposed does not rely on the accurate bounding box of human body provided as input for the algorithm. Moreover, it provides a fine-grained understanding of human body by representing it using seven different limbs. Promising results are shown in empirical experiments on a standard dataset.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

11. MACHINE LEARNING BASED ART GENRE CLASSIFICATION

Srivignesh Rajendran, Pradeep Prabakar Ravindran, Lavanya Viswanathan, Xiaoyun Yang

In this work, we investigate the task of automated classification of art forms according to their genres, based on image characteristics. Our contribution to this problem domain is two-fold. Firstly, we extract diverse features that we hypothesize as being representative of different aspects of an art form. Secondly, we study the effectiveness of various flavours of machine learning algorithms in classifying the given art forms using the extracted features.

To perform this study, we considered a set of six art genres. Our corpus consisted of a training and test set of images, crawled from the internet. Some features that we examined were those based on edge, texture, color histogram, SIFT, GIST, HOG and Hough transform. We experimented with different classification algorithms like ensemble classifiers, SVM and regression and the bag-of-words approach. We also performed feature selection using PCA, ICA and sparse coding. We then analyzed as to which features were most informative for each art form and which classifier performed most accurately on which classes.

Preliminary results suggest that the performance of the classifiers was encouraging, given that we were dealing with a challenging six-class classification problem.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

12. SOCCER TRACKER

Phil Bailey, Eli Murphy-Trostsky, Christopher Leaf

Soccer is a complex sport with many people and a ball all moving simultaneously. A game is full of movement and thus has rapidly changing variables, making it hard to follow even for the inexperienced human eye. However, a coach or analyst needs a way to sort through this vast amount of rapidly changing information to improve their strategy and understanding of a team's formations. Armed with statistical information about a team's formations, tendencies while moving the ball, and general player and ball location data, they could make insightful discoveries that could influence future games.

The Soccer Tracker project takes a large sample of freely available soccer footage and uses machine learning techniques such as weak learners and *k*-means in order to track the location of the ball on the field as well as the location of the players on the field from each frame of the video. From these locations statistical models can be created for likely movement patterns in which teams will deploy their players. Along with this, simple statistics for analysis can be generated, such as ball possession, shots on goal, successful passes, and interceptions.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

13. IMAGE MANIPULATION:MODIFYING IMAGE BY APPLYING PATCH TRANSFORM

Quio Shi

This report gives a method of modifying images by applying patch transform. This method breaks the original image into hundreds of small patches, tries placing the patches on different positions and uses the locally optimal patch placement to reconstruct the modified image. The locally optimal placement can be computed from Markov Random Fields. The report also shows manipulation results on different images and gives analysis why some images show good results and why others turn out to be a failure.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

14. AUDIO EVENT CLASSIFICATION

Khaled El Glaind, Robert Kotcher

The goal of this project was to investigate recent developments in audio event classifications. We chose two techniques for comparisons. The first is Deep Belief Network which is composed of multiple Boltzmann machines stacked on top of each other. A Boltzmann machine is a Markov Random Field associated with a bipartite undirected graph that can be trained to model the joint distribution of the input data. DBN is initially pre-trained to initialize the weight of the network in the neighborhood of the global minimum. Then, back propagation is used to fine tune the weights.

In the second technique, Bag of Audio Words, we use the visual bag of words model to classify audio events through their spectrograms. Our process involved performing a Scale-Invariant Feature Transform (SIFT) on the spectrograms of each file being used in the process. Next, a codebook was formed over millions of uniformly sampled feature vectors. Vector quantization over the SIFT features was used to build a histogram over the codewords. These models can then be used to build a discriminative model for each of the audio classes.

In the first phase of this project, we were able to successfully implement and test both of these techniques on small data sets, in the range of a few hundred audio files. In the second phase, we were aiming for a larger data set, in the range of thousands. We chose one technique: bag of audio words, and developed it in C. For the sake of comparison, we used results achieved using MFCC features as reference.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

15. FOREGROUND DETECTION USING ADAPTIVE MIXTURE MODELS

Rentaro Matsukata

Real time motion tracking in video data has many useful applications. It is common practice to first preprocess each frame before applying image recognition and tracking algorithms due to the large amount of data needed to be processed in real time. One method used to simplify the subsequent recognition and tracking processes is to first reduce the problem size by subtracting out background pixels, pixels that are not a part of a moving object, and processing only the remaining foreground pixels of an image. What is presented here are the results of an implementation of Stauffer and Grimson's method of adaptive background mixture models to classify background pixels.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):