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

Final Poster Presentations

Instructor: Bhiksha Raj

TAs: Zhiding Yu

Fall 2014

1. IMPROVING SPATIALIZATION ON HEADPHONES FOR STEREO MUSIC

Muhammad Haris Usmani, Ramon Cepeda

Music is mixed and mastered for playback on near and far-field speakers. While this is a standard practice, today there's a huge and growing proportion of listeners that use headphones. Playing legacy stereo mixes on headphones places the stereo image inside the listener's head as three isolated lobes (left, right and phantom center) and makes the image appear ultra-wide. This is favorable to separate the center and stereo content, but otherwise is extremely deteriorating to the sound in terms of the spatialization of the music content. It results in making headphone listening unnatural and uncomfortable, although we- as listeners- have learned to accept it. In this project, we will work towards providing a better sound image to headphone listeners and will assess, using subjective evaluation, if headphone externalization is a desired feature.

There are two approaches to the problem: The *Post-Master* approach where stereo mixes are processed to provide better spatialization over headphones, and the *Recording Approach* where the artist specifically makes a mix for headphone listening. This project takes the post-master approach and proposes a system that enhances stereo mixes to re-create the studio's soundstage over headphones without making any perceivable spectral changes.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

2. PREDICTING THE OUTCOME OF ROULETTE

Matthew Bauch

Roulette wheels are widely regarded as near-perfect random number generators. In fact, roulette is one of the most popular betting games in the world and casinos stake their livelihood on the unpredictability of the game. Despite this, several groups have claimed success in exploiting the inherently deterministic, albeit complex, nature of roulette to gain an advantage over the house using portable electronic devices. In this project, we attempt to reproduce the reported results of others. To that end, a selection of techniques suggested in the literature as well as some novel approaches are evaluated for effectiveness in predicting the outcome of a game of roulette from limited data.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

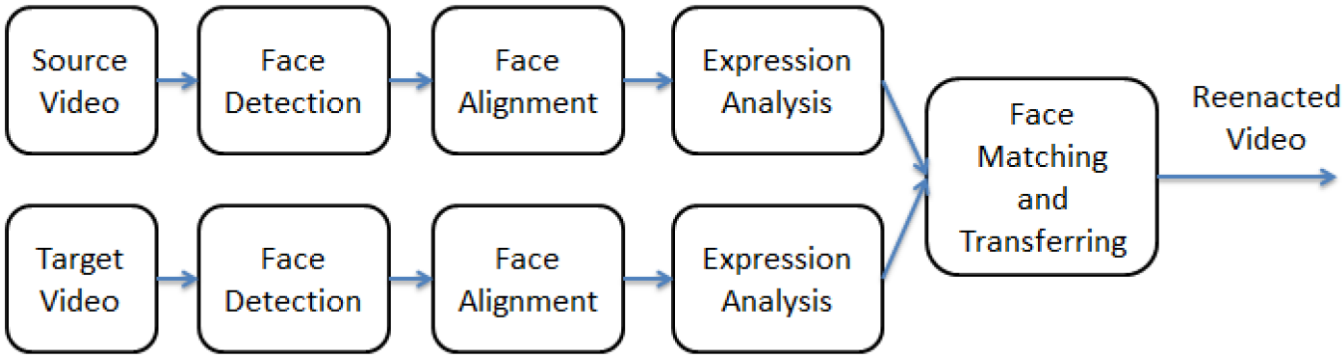
Overall: 1-5 (5 is highest):

3. FACIAL REPLACEMENT IN VIDEOS

Cathey Wang, Chengxiong Ruan, Matthew Tay

Facial replacement is a popular procedure used heavily in today's computer-generated imagery works. However, the procedure is usually not automated and requires large amounts of human labor. In this project, we build an automatic facial replacement system that replaces the face of an individual in an existing target video with the face of another individual from a source video, while preserving the original target's facial expressions. Our system is able to produce a convincing video with replaced faces from a short source video where the user performs multiple arbitrary facial expressions.

The pipeline for our system is shown as follows:



We first locate the faces in the video frames using the Viola-Jones' cascaded framework. Then, we locate the facial landmarks with boosting regression and shape-indexed features. Following that, facial expressions are analyzed with local binary pattern descriptors, and based on temporal clustering, matching faces are selected. Finally, faces from source frames are transferred to target frames based on piecewise warping, and image blending is employed to make transferred faces look more natural.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

4. ISOLATED SIGN WORD RECOGNITION SYSTEM

Jian Gong, Abelino Jimenez

According to some researches, an estimated 360 million people worldwide suffer from hearing loss. Since most of hearing individuals do not understand sign language, it is pretty necessary to have a low cost system for translating sign language into spoken language.

The goal of this project is the recognition of isolated words of the American Sign Language using a standard webcam that observes the signer from a frontal view. We focus on the most common 50 words.

Recognition is based on Non Negative Factorization and Dynamic Time Warping. Several feature vectors taking into account information about 2D position of the hand with respect to the face and the expression of the face are examined. To improve the recognition rate we implemented a method for tracking the hands, using the trajectory, non-negative factorized finger spelling and some additional features.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

5. ACCENTED ENGLISH DIALECT CLASSIFICATION

Joyce Chung

This project is an exploration of several well-known and well-established dialect classification techniques. The two back-end frameworks compared were GMM-UBM (Gaussian Mixture Models-Universal Background Model) and an i-vector-based (Eigenvoice) framework. As both frameworks proved to be successful in NIST LRE tasks, a secondary goal of this project was to first reproduce comparable results with NIST 2007 LRE data, and then to test both models on new sets of foreign accented English data and verify that similar results could be reproduced with other dialects.

In addition to two back-end frameworks, three front-end frameworks were also compared: MFCC (Mel-Frequency Cepstral Coefficients), PLP (Perceptual Linear Prediction), and RASTA-PLP (Relative Spectral Transform – Perceptual Linear Prediction). Preprocessing of all training and test data was also done, and this involved SAD (speech activity detection), dropping of frames in order to maintain a high SNR (signal-to-noise-ratio), and CMVN (cepstral mean and variance normalization).

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

6. BRAIN IMAGE CLASSIFIER

Senbo Fu, Chenyang Wu, Qinghao Hou

Magnetic Resonance Imaging (MRI) is a medical imaging technique to detect potential disease of the body. In this project, we obtained 90 T2-weighted MR brain images of 18 classes and classified them using various machine learning algorithms.

We first extracted features by Discrete Wavelet Transform (DWT), and reduced the features by Principal Component Analysis (PCA). After preprocessing, the original 256 * 256 images were reduced to 30 * 1 features, which greatly simplifies the computation and improves the efficiency. Then we applied a five-fold cross validation to multiply classifiers to classify (and also clustering) the images. The classifiers include: Density Based Spatial Clustering of Application with Noise (DBSCAN), Naive Bayesian, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN) and Random Forest. The efficiency and accuracy of different classifiers are compared and analyzed.

The experimental results show that all these classifiers have high accuracy (about 90%) in classifying these brain images. Also by using PCA, most classifiers enjoy higher classification performance than those without using PCA analysis, which demonstrate that PCA can not only reduce the data dimensions to improve the computational efficiency, but also remove some noise to improve the classification accuracy.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

7. FACIAL EXPRESSION RECOGNITION

Chieh-Han Wu, Hao-Ping Ho, Ting-Yu Chen, Tse-Han Huang

The project is one of the challenges of the ICML 2013 Workshop "Challenges in Representation Learning": the facial expression recognition challenge. This contest introduced the Facial Expression Recognition 2013 (FER-2013) dataset, which consists of seven categories (angry, disgust, fear, happy, sad, surprise, and neutral).

The goal of this project is to design a system for recognizing which emotion is being expressed in a photo of a human face. We extracted Eigenfaces by applying Principal Component Analysis (PCA), and represented each image as a linear combination of the Eigenfaces to categorize each face into one of seven categories. We compared the classification performances of various classifiers such as k-nearest neighbor (KNN), and Support Vector Machine (SVM).

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

8. SPEECH RECOGNITION AND TRANSLATION USING UTTERANCE CLASSIFICATION FOR LOW-RESOURCE LANGUAGES

Sai Sumanth Miryala

Speech translation is a challenging problem for low-resource languages. Since the written notation is usually not available, the conventional paradigm of recognition and machine translation cannot be used. In this project we are attempting a novel scheme of speech recognition in a limited domain. By working with a set of sentences appropriately chosen to fit a scenario, we use utterance classification as a speech recognition algorithm. Utterance classification is achieved using cross-lingual, language independent phonetic labeling. Also, since we are only working with a limited set of selected phrases, the translation part is trivial. The set of phrases in our dataset are relevant to taking a doctor's appointment at a hospital. We chose Telugu as a test language, as it has characteristics similar to many low-resource languages.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

9. MOOD BASED CLASSIFICATION OF SONGS TO IDENTIFY ACOUSTIC FEATURES THAT ALLEVIATE DEPRESSION

Adhithi Aji

Depression is increasingly affecting the US population with nearly 14.8 million American adults suffering from depression in a given year. Research shows that music therapy has been very effective in curing depression by elevating the mood. The objective of this paper is to understand the acoustic features that characterize the songs which alleviate depression.

In order to achieve this, a two-step approach was taken. The first step was to extract certain acoustic features of songs and classify them based on the mood. For this purpose, the MIREX 2010 mood based labeled dataset was used for training and classification into two classes – Happy and Sad. The second step was to establish a correlation between the mood and the songs that are known to help depression. These songs were collected from the internet to represent the test data.

A number of classification techniques such as SVM and Neural Networks were experimented for the mood based classification. The impact of a number of acoustic features were analyzed and narrowed down to eight main features namely zero crossing rate, RMS, roll off, entropy, energy, MFCC, flux and tempo. A fivefold cross validation was implemented on the test data to render a classifier that would identify new songs which would help depression.

The results provided interesting insights. Counter intuitively, it was seen that the songs that belong to the sad category had a better correlation to songs that alleviate depression. This aspect is also supported by a number of research articles on depression confirming that acoustic features of sad songs would help in reducing depression.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

10. PERSON IDENTIFICATION THROUGH FOOTSTEP-INDUCED FLOOR VIBRATION

Shijia Pan, Carlos Ruiz

Personal identity information is crucial for several applications in smart buildings, including customer behavior analysis, senior housing monitoring, etc. Here, we introduce our indoor person identification system, which senses footstep-induced floor vibration. Since floor vibration can be measured without interrupting human activities, our system is suitable for many smart building applications.

Briefly, this is an explanation of how our system works. First, it measures the vibration of the floor at different points and detects footstep events through signal processing techniques. Then, the system extracts features from the vibration signals during the event, which represent characteristics of each person's gait pattern. Once the features have been extracted, the system conducts support vector machine (SVM) based hierarchical classification at an individual step level, as well as at a trace (i.e., set of steps from the same person) level. On average, our system achieves over 83% identification accuracy. Furthermore, if the system discards traces whose confidence level is below a threshold, the identification accuracy increases (e.g., 96.5% accuracy when 50% of the traces are discarded).

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

11. DETECT HUMAN HEAD-ORIENTATION BASED ON CONVOLUTIONAL NEURAL NETWORK AND DEPTH CAMERA

Xin Wang, Ningning Wang

In this paper, we present a method to detect in-car driver's head orientation relative to the view of camera. The method is based on Convolutional Neural Network (CNN) and depth camera. Automatically detecting human head orientation is essential in many applications such as attention analysis of human, interfaces for human-computer interaction and surveillance. In smart vehicle, the camera based in-car driver monitoring system requires a high-performance attention analysis engine in order to enable adventurous safety features. If the driver's head orientation changes frequently or the driver's head faces rear/right/left for an unacceptable long time, the driver's safety may be in danger. According to statistics, reckless driving accounts for 33% of all deaths involving major car accidents in 2013. By detecting driver's head orientation, the monitoring system will be able to warn a driver who is driving unsafely. To build a fully automated system that detects head orientation, a key step is to develop efficient algorithms to identify head orientations. Compared to traditional methods of human head orientation using shape information and SIFT, Convolutional Neural Network based method can learn the geometrical and texture features automatically, even without much domain knowledge background and provide better robustness based on its deep architecture. In this paper, we propose a CNN model for in-car driver head orientation detection. The model consists of input layer of R, G, B, depth and direction label data, convolutional network layer, hidden layer and logistic regression layer(output layer). The training and testing images data are collected from depth camera with both RGB and depth information which can help eliminating background interference. Experiment shows our approach can provide consistent performance of 92.12% accuracy on distinguishing driver head-orientation in 3 directions (i.e driver looking at left, right and forward), which provides motivation for future research on more sophisticated driver posture detection based on deep neural network.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

12. NEURAL NETWORK BASED SLUDGE VOLUME INDEX PREDICTION

Si Wang, Yingrui Zhang

Wastewater disposal is one important part of environment control in the cities, in which activated sludge treatment is often used as an efficient method. Activated sludge treatment reduces the organic pollutants of wastewater when they are absorbed and dissolved by the microorganisms within the sludge. But the problem of sludge bulking not only cause major sludge loss, but will also lead to the collapse of the entire wastewater treatment system and cause serious damage in severe cases.

Sludge volume index (SVI) is an important indicator of the activated sludge consistency and settling performance. The value of SVI reflects the degree of sludge bulking, but the process is highly non-linear, time-consuming and uncertain, which makes SVI hard to be detected and modeled. This project uses real data from sewage plant and builds a neural network, achieving real-time measurement, which is absent from current measurement. The main content of this project is as follows:

(1) A SVI soft sensor model is designed basing on neural network. By analyzing the characteristics of sludge sedimentation, easily measured variables related to SVI are identified (temperature, PH value, COD and BOD). The soft sensor model between these variables and SVI is established.

(2) A SVI soft sensor method is proposed. It is aimed at solving the problem that a normal neural network finds it hard to express the highly nonlinear, uncertain, timeconsuming sludge settling process. Through the designing of the method, a neural network that could dynamically adjust itself is established. It greatly improved the self-adaptive capacity and accuracy of SVI soft sensor model. Simulation experiments were conducted with actual sewage plant data. The results show that the SVI soft sensor model based on this neural network could achieve SVI prediction.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

13. 8-BIT MUSIC NOTE IDENTIFICATION - TURNING MARIO INTO METAL

Haley Dalzell, Dale McConachie

Non-negative matrix factorization is a relatively new technique with many different application areas. By utilizing the idea of decomposing an image or piece of audio into additive "building blocks", we can extract meaningful semantics from the data. We utilize this technique to analyze 8-bit music from video games, deriving both the original synthesis technique used as well as the musical score.

To develop a set of building blocks we took advantage of fact that each musical piece was generated using MIDI synthesis techniques. This allowed us to develop our dictionary of building blocks by synthesizing every MIDI note using a variety of synthesis techniques. Each note was then converted to its spectral equivalent with each note having it's own unique spectral signature. This dictionary of building blocks is then used to transcribe sample audio into individual notes in MIDI format. To analyze the success of our transcription we synthesize the resulting MIDI track back into audio using an electric guitar as our instrument of choice.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

14. A CONTEXT RELEVANT MODEL OF SENTIMENT ANALYSIS: AN EMPIRICAL STUDY ON SOCIAL MEDIA

Chia-Chuan Wu, Hao Ni, Ben Bamberger

Sentiment analysis is becoming more and more popular in today's world, a place where social media is growing explosively and becoming more intensively used by users to express their opinions. For this project, we built a powerful sentiment classifier with Bag-of-word model (BOW) and Conditional Random Field (CRF). A sentiment classifier is capable of detecting the polarity of positive and negative emotions from a document. This classifier assigns a document into either the positive or negative category according to its content. However, in the traditional method, BOW is applied to convert a document into a word vector and each word in this vector is independent of each other. BOW model ignores the sequential relationship of words in the document; it thus degrades the classification accuracy. In this paper, we proposed a method based on graph model that is called CRF. CRF can sequentially tag the sentiment of words in a document and we built a classifier based on this method. In addition, we also designed another classifier based on BOW and combined it with CRF, and we found these two classifiers can compensate each other and form an even better classifier.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

15. RECOGNITION OF HOUSING NUMBERS IN REAL-WORLD IMAGES USING CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURES

Yuanchi Ning, Shiny Singh, Evgeny Toropov

We consider the problem of number recognition in real world images. This problem has numerous valuable applications. One such application is distinguishing house numbers in real-world images for the purpose of geo-tagging. Given an image containing a few house numbers (if any) the problem includes 1) detecting numbers in the image and 2) recognizing those numbers. In particular, these methods came useful for Google Maps for tagging house numbers in Google Street View for improving the accuracy of house numbering in the service.

We are focusing on the second part of the problem, namely recognizing the number given a region of interest in an image by utilizing convolutional neural network (CNN). CNN is sensitive to extract good features from small sub-regions of visual field, so it's great architecture to use when training images. We are fixing the length of the street number to 3 digits. The number is considered correctly recognized if and only if all digits are correctly recognized.

We use the publicly available SVHN dataset to train our network. The dataset contains about 50000 street numbers, along with bounding boxes for individual digits, out of which about 10000 street numbers are of length 3. We preprocess the data to get a 54x54 bounding box enclosing the street number (all digits).

Different networks are trained for different digits in the house number, though the training data is the same for every network (image region in the bounding box). The collective output of all the networks is used to determine the overall accuracy. Each network architecture consists of 3 convolutional hidden layers, 2 Rectified Linear Units (ReLU) layers, 2 fully connected layers and one softmax layer. Each convolutional layer includes max pooling.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

16. STREET VIEW HOUSE NUMBER RECOGNITION BASED ON CONVOLUTIONAL NEURAL NETWORKS

Hang Yuan, Chuhan Yan, Chi Zhang

Convolution Neural Networks (CNN) is biologically-inspired variants of artificial neural network. It simulates animal cortex which contains cells of different functions acting as filters to extract features from and recognize images. CNN has been proved to be efficient in image classification such as digit recognition. In this paper we provide an implementation of CNN classifier training to recognize street view house numbers. We use LeNet as the basic structure of CNN which contains two levels of convolution filters and pooling. To improve classification performance, on more layer of convolution was added and the images are also properly pre-processed. Approximately 90% precision can be achieved in our street house number test dataset.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

17. TRAIN-BASED INFRASTRUCTURE MONITORING

Peng Gong, Anuva Kulkarni, George Lederman

Globally, rail infrastructure is a vital asset for economic prosperity, but condition assessments tend to be subjective and infrequent. The lack of objective information leads to suboptimal capital replacement projects, and the information lag prevents timely repair. One solution is to instrument the tracks and track structures, but given the expanse of our rail network, the installation and maintenance costs of such a sensor network would be prohibitively high. A second solution is to use a custom instrumentation vehicle capable of monitoring the infrastructure as it moves. However, such dedicated vehicles tend to be expensive, particularly in rail where monitoring is more specialized, so to keep costs down, infrastructure owners use these vehicles infrequently.

To address this, we propose to monitor the track from the vibrations of a passing train. The concept of using ordinary vehicles to probe multiple types of infrastructure assets has been proposed previously, but most work has been limited to simulations, laboratory tests, and short-term experimentation. The main barrier has been the complexity and the noisiness of the data. We have applied Machine Learning for Signal Processing methods to better process the data and learn from the signals. We have been able to successfully localize changes in the track, which means that more targeted maintenance could be done.

Currently, we have tried to localize a process known as tamping (Fig. 1), which is a maintenance activity to correct track geometry. Just as we find where the track changes from a state of bad repair to a state of good repair, we could equally identify the reverse, an area where the track changed from a state of good repair to a state of bad repair. This type of work can save both money and lives, as timely maintenance can prevent trains from derailing.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

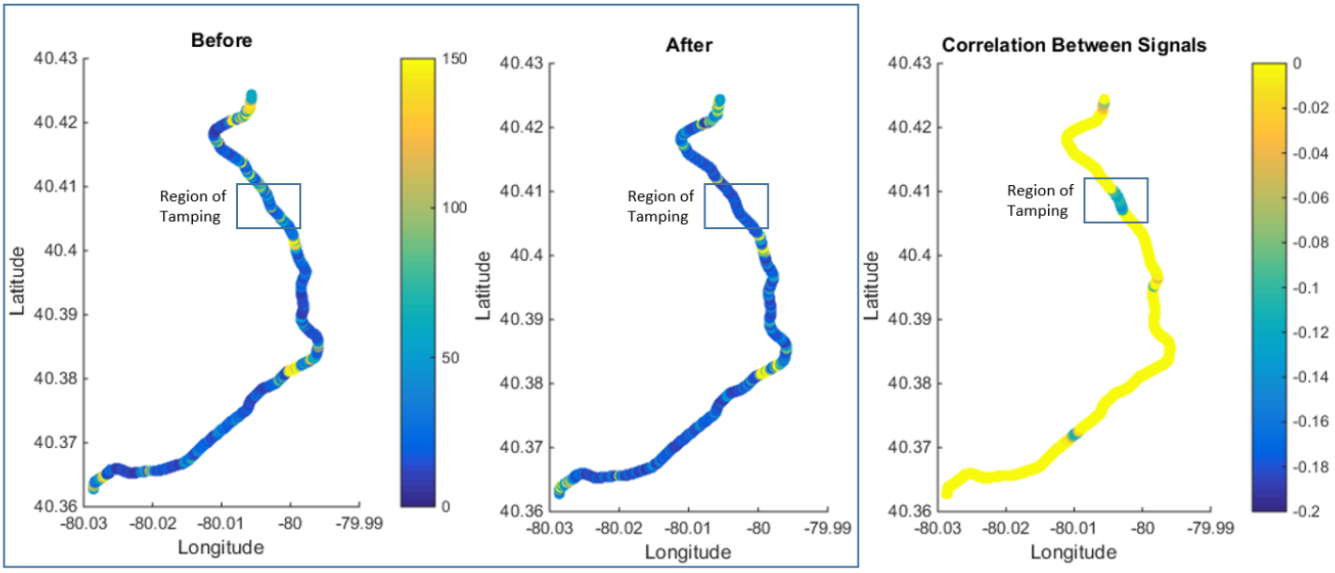


Figure 1 : Using correlations to localize the site of tamping

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

18. MANIFOLD INTERPOLATION OF X-RAY RADIOGRAPHS

Arjun Kumar

The use of X-ray computed tomography (CT) imaging in engineering has become more prevalent to study the nanoscale properties of materials. The principle behind X-ray CT is the reconstruction of a 3D image from 2D X-ray radiographs at various angles. Some of the challenges of this are the long scan times required and sample motion. This poster presents a technique to interpolate between projections to account for these issues through the use manifold learning.

Manifold learning is a nonlinear dimensionality reduction, which unwraps the manifold that a set of high dimensional data lies in. In this work, we use it to find the trend between the acquired 2D radiographs. We can then interpolate along the manifold to artificially create the required number of 2D radiographs.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

19. A SMARTPHONE BASED INDOOR POSITIONING SYSTEM AUGMENTED WITH INFRARED SENSING

Chaitanya Poolla, Sumeet Kumar

Although there have been significant advances in indoor positioning technologies, most new applications are confined to a research environment mainly due to manual fingerprinting or sophisticated additional hardware. In addition, multi path propagation of radio signals, interference, and presence of obstacles and movement of people pose additional challenges.

This work aims to address some of the above issues at a low cost by leveraging smartphone enabled in-built sensing technologies integrated with an ultra-low cost Infrared sensing via a dongle. In particular, the Infrared (IR) sensor is a novel addition for positioning purposes since it does not penetrate walls (unlike WiFi) and hence reduces interference and mis-classification. The location extraction problem with various sensor combinations is formulated in both classification as well as regression frameworks. Signal Processing methods are employed to encode time-series sensor measurements into algorithmic inputs. Supervised learners are then trained and tested over the collected dataset. Experimental results indicate almost accurate (~100%) room level classification and meter level location accuracy. Further, tradeoffs in accuracy between various sensor combinations are provided for comparison and recommendation purposes.

Our main contributions include:

a. A working Android application (Available @ http://tinyurl.com/k6dvdvk)
b. Data collection (via App) and Signal processing of IR, WiFi and Magnetometer readings.
c. Analysis using various Machine Learning algorithms.
d. Comparison of various sensor combination accuracies.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

20. MOTION TRACKING AND PROXEMICS EXTRACTION FROM THERMAL-SENSOR ARRAY

Chandrayee Basu

Indoor person tracking has all-pervasive applications beyond mere surveillance, for example in education, health monitoring, marketing, energy management and so on. Image and video based tracking systems are intrusive. Thermal array sensors, on the other hand, can provide coarse-grained tracking while preserving privacy of the subjects. GridEYE is an 8 x 8 array of PIR sensors. It measures the relative temperature differences between the objects and the background within its field of view. The goal of the project is to facilitate proxemics and behavioral studies by tracking motion and distance between subjects using a GridEYE sensor array.

The ground coverage of a GridEYE array, mounted 3 meters above the ground, is 2.5 meters x 2.5 meters. Our experiments suggest that human head is approximately 12°F - 20°F warmer than the rest of the scene in an office environment. The number of pixels occupied by one subject and distinguishable from the background varies with the height of the person, temperature of the body parts and background temperature. The measured temperature drops with increasing distance from the sensor. Detection and tracking with GridEYE, thus, poses challenges due to low-resolution and noisy temperature data.

Prior research has only focused on estimating the number of people in static scenes using thermal array. In this work, we can extract distances between the subjects within the field of view of GridEYE. We also leverage the spatial information provided by the array sensors to detect motion between the scenes captured. Motion is then used as a context to further support the above human detection task.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

21. RECURRENT NEURAL NETWORK FOR SEQUENCE LABELING

Bing Liu, Xin Wan

Sequence labeling refers to the task of mapping an observation sequence x1, x2 … xn to a sequence of discrete labels y1, y2 … yn. Well-known applications include speech recognition, part-of-speech tagging, and semantic role labeling. Recurrent neural network (RNN) is a class of artificial neural network that contains at least one feedback connection, enabling the network to learn sequence structure.

In this project, we implemented RNN for slot filing, one of the key problems in spoken language understanding. Standard ATIS dataset was used as training and evaluation sets, in which the word sequence was taken as input, and slot labels were predicted on the output side. Precision, recall and F1-score were measured and compared with a conditional random fields (CRF) baseline. Several variations of RNN with different word context and word embedding were also evaluated. Best F1-score achieved using RNN with word embedding is 94.36, advancing the reported CRF based state-of-the-art performance (91.09) substantially. Using word embedding trained on part of Google News dataset (about 100 billion words) without fine tuning gives best F1-score of 92.56, lower than the best score of 94.36 achieved with fine tuned word embedding from random initialization. Analysis shows that the task-specific word embedding learned by RNN contributes to the advanced performance. In future work, we plan to further explore more comprehensive sequence feature representations incorporating knowledge of part-of-speech tags and named entities.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

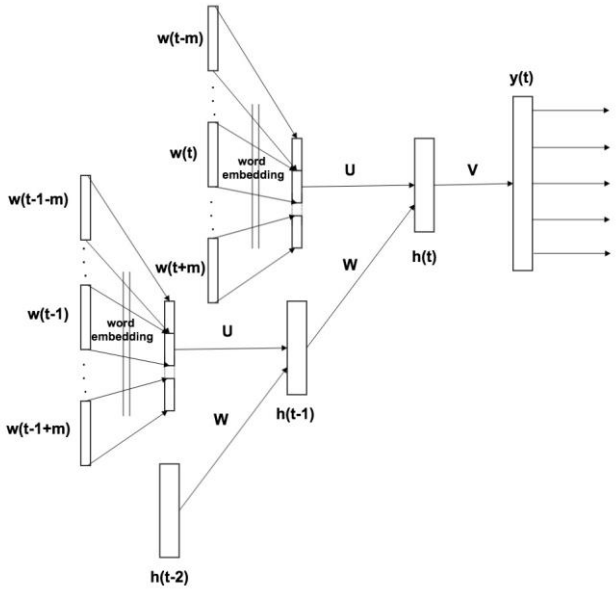


Fig 1. Recurrent Neural Network

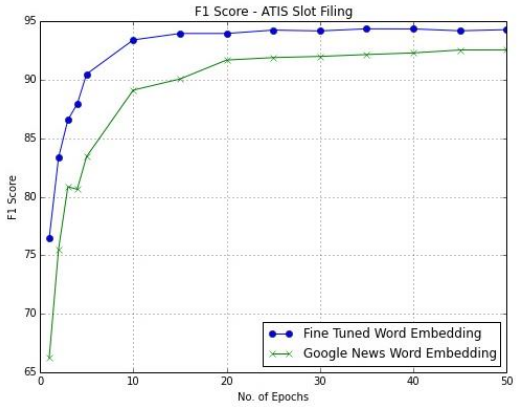


Fig 2. F1-Scores on Slot Filing

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

22. ROCK, PAPER, SCISSORS -- HAND GESTURE RECOGNITION

Qin Wei, Yu Wu, Jialiang Lin

Hand gesture recognition is a widely used application of machine learning in daily lives. With the help of camera, computers can understand human body language and interact naturally without any mechanical devices. In this project, we aim to solve a classic gesture classification problem to classify three signs in the famous game: Rock, Paper and Scissors. Dataset is selected from Cambridge Hand Gesture Data set with 3 types of hand signs. We study the issue of feature sets for gesture recognition, showing that Histogram of Oriented Gradient (HOG) descriptors provide relatively good performance. Linear multiclass SVM is used as a baseline classifier for simplicity and speed. Finally, limits and possible improvements on this project are discussed.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

23. LANGUAGE MODELS WITH SEMANTIC CONSTRAINTS

Jin Sun, Chenchen Zhu

We want to find a matrix that transforms words to embeddings with semantic meanings. The semantic embeddings of words with closer meanings should be close in distance. Previous language models proposed by NLP researchers have some interesting clustering behavior on semantic vectors, even without any semantics constrains applied to training process. We want to train a language model and explicitly address the semantic constrains, and observe if these constrains can make a difference.

We train a linear language model base on 5-gram architecture. The inputs are four embedded words in a low dimensional space. The output of the model predicts the one hot representation of the 5th word given previous four words. We impose the constraint on word semantic similarities and word relationship similarities. The training method is stochastic gradient descent. The perplexity of our model is compared to standard 5-gram model, and the word similarities are compared to Mikolov's word2vec project.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

24. DATA-DRIVEN BOUNDARY DETECTION

Aayush Bansal, Xinlei Chen

In recent past, the computer vision community saw a huge performance in improvement of various tasks with the use of more data. The performance of current state-of-the-art object detection algorithm improved from 40.7% to 58.5% mAP on PASCAL VOC-2007 when using more 1.5 M data from ImageNet to train neural network. Learning from this achievement, we use extra data for the problem of edge detection. The major challenge which restricted the use of extra data is the unavailability of human-labeled ground truth boundaries. In this work, we introduce an amalgam of supervised and unsupervised approaches which makes it possible to exploit extra data to do edge detection. Finally, we evaluate the performance of our approach on Berkeley Segmentation Dataset and compare it with existing baselines.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

25. LEARNING TO PREDICT WHERE A DRIVER LOOKS

Chihiro Suga

Driver's inattention of an obstacle on road is major cause of traffic accident (NHTSA, 2008) yet there's no system to help it. The difficulty of developing application for obstacle warning system is to understand the driver's attention. In this project we challenged to build a model to predict a driver's visual saliency. Gaze data is collected from a subject doing emulated driving task and used as the ground truth. The features of the driving scene and the subject head orientation were used as input features. Logistic Regression and Neural Network were applied to predict the attained location. The best error rate at this moment is 60% with Logistic Regression.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):

26. REAL TIME MONITORING OF STUDENT'S LEARNING PERFORMANCE

Parag Agrawal, Haohan Wang

In this project, we built a system that monitors a student's learning performance with real time captured visual data from the student's vision. With visual data captured by a portable camera set near a student's eyes, the system could understand student's learning performance like whether the students is concentrating on school work or casually browsing entertainment reading materials. To achieve this, our model separates into two parts, signal processing part to represent visual data into strings and machine learning part to decide whether the students are working on school work and entertainment readings.

In data representation part, the images captured from camera could be very noised, we make the assumption that students are reading normal sized books and detect the edges of book. Then, we focus on the text part of the book, because we believe the difficulty of our goal is to distinguish textbook and novels (rather than comic books). We applied line detection to detect text area and rotate the text lines to be horizontal of the image. Our last step is to apply OCR to get words and sentences.

For classification (machine learning part), we use a pruned decision tree, trained on unigram and trigram model. The tree is allowed to grow recursively until the stopping criteria is met. We use a hinge loss stopping criteria on mutual information gain for pruning the tree. With this approach, using a 10fold cross validation we get a value of 3.58164027 for Average Log Likelihood and Perplexity is 11.97235946. The same approach (pruned decision tree with hinge loss on information gain) but with LOOCV (Leave One Out Cross Validation), gives us Average Log Likelihood equal to 3.58491009 and Perplexity equal to 11.99980006. In both the cases we are able to get an accuracy of 71.8732% on synthetic test dataset.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):