# Machine Learning for Signal Processing

## Project Ideas

Class 4a. 9 Sep 2014

Instructor: Bhiksha Raj

# Administrivia

- Second TA:  Rahul Rajan
  - rahulraj@andrew.cmu.edu
  - SV campus
  - Office hours: TBD
- Homework questions?
  - If you have any questions, please feel free to approach TAs or me

# Administrivia

- On Thursday: Dr. Griffin Romigh of AFRL
  - Student of MLSP.. ☺
- Will talk about methods for estimating HRTFs
- Outstanding thesis on the use of data-driven methods to reduce measurements needed to compute HRTFs
  - By an order of magnitude!

# Course Projects

- Covers 50% of your grade
- 10-12 weeks of work
- Required:
  - Serious commitment to project
  - Extra points for working demonstration
  - Project Report
  - Poster presented in poster session
  - Graded by anonymous external reviewers in addition to the course instructors

# Course Projects

- Projects will be done by teams of students
  - Ideal team size: 3
  - Find yourself a team
  - If you wish to work alone, that is OK
    - But we will not require less of you for this
  - If you cannot find a team by yourselves, you will be assigned to a team
  - Teams will be listed on the website
  - All currently registered students will be put in a team eventually
- Will require background reading and literature survey
  - Learn about the problem

# Projects

- Teams must inform us of their choice of project by 25th September 2014
  - The later you start, the less time you will have to work on the project

# Quality of projects

- Project must include aspects of signal analysis and machine learning
  - Prediction, classification or compression of signals
  - Using machine learning techniques
- Several projects from previous years have led to publications
  - Conference and journal papers
  - Best paper awards
  - Doctoral and Masters' dissertations

# Projects from past years: 2013

- Automotive vision localization
- Lyric recognition
- Imaging without a camera
- Handwriting recognition with a Kinect
- Gender classification of frontal facial images
- Deep neural networks for speech recognition
- Predicting mortality in the ICU
- Human action tagging
- Art Genre classification
- Soccer tracking
- Image manipulation using patch transforms
- Audio classification
- Foreground detection using adaptive mixture models

# Projects from previous years: 2012

- Skin surface input interfaces
  - Chris Harrison
- Visual feedback for needle steering system
- Clothing recognition and search
- Time of flight countertop
  - Chris Harrison
- Non-intrusive load monitoring using an EMF sensor
  - Mario Berges
- Blind sidewalk detection
- Detecting abnormal ECG rhythms
- Shot boundary detection (in video)
- Stacked autoencoders for audio reconstruction
  - Rita Singh
- Change detection using SVD for ultrasonic pipe monitoring
- Detecting Bonobo vocalizations
  - Alan Black
- Kinect gesture recognition for musical control

# Projects from previous years: 2011

- Spoken word detection using seam carving on spectrograms
  - Rita Singh
- Bioinformatics pipeline for biomarker discovery from oxidative lipdomics of radiation damage
- Automatic annotation and evaluation of solfege
- Left ventricular segmentation in MR images using a conditional random field
- Non-intrusive load monitoring
  - Mario Berges
- Velocity detection of speeding automobiles from analysis of audio recordings
- Speech and music separation using probabilistic latent component analysis and constant-Q transforms

# Project Complexity

- Depends on what you want to do
- Complexity of the project will be considered in grading.
- Projects typically vary from cutting-edge research to reimplementation of existing techniques. Both are fine.

# Incomplete Projects

- Be realistic about your goals.
- Incomplete projects can still get a good grade if
  - You can demonstrate that you made progress
  - You can clearly show why the project is infeasible to complete in one semester
- Remember: You will be graded by peers

# Projects..

- Several project ideas routinely proposed by various faculty/industry partners
  - Sarnoff labs, NASA, Mitsubishi

# From Griffin Romigh..

- Projects on HRTFs
  - Head-tracking and prediction of anthropometric parameters
    - head size, pinna height, pinna angle, etc.
  - Improved prediction of efficient HRTF model from anthropometric parameters
  - HRTF measurement using a single speaker and a head tracker
  - HRTF-based sound source localization/segregation from a binaural recording
    - many recordings available

# Alan Black: Potential Projects

- Find F0 in story telling
  - F0 is easy to find in isolated sentences
  - What about full paragraphs
  - Storytellers use much wider range
- Find F0 shapes/accent types
  - Use HMM to recognize “types” of accents
  - (trajectory modeling)
  - Following “tilt” and Moeller model

# Alan Black: Parametric Synthesis

- Better parametric representation of speech
  - Particularly excitation parameterization
- Better Acoustic measures of quality
  - Use Blizzard answers to build/check objective measure
- Statistical Klatt Parametric synthesis
  - Using “knowledge-base” parameters
  - F0, aspiration, nasality, formants
  - Automatically derive Klatt parameters for db
  - Use them for statistical parametric synthesis

# Alan Black: TTS without Text

- Speech processing without written form
  - Derive symbolic form from speech (done-ish)
  - Discover “words”/”syllables”
  - Derive speech translation models
- Build a cross linguistic synthesizer
  - Hindi text in, but speaks in Konkani

# Alan Black: UPMC "APT" Projects

- Speech Translation for zero-resource languages
  - Collect cross linguistic speech prompts
  - Learn mapping at (near)sentence level
- Working with refugee populations at UPMC

# Gary's Work

Digit Classification on the Street View House Numbers (SVHN)

Dataset. http://ufldl.stanford.edu/housenumbers/

- Students could explore features, classification methods, deep learning, normalizations etc.

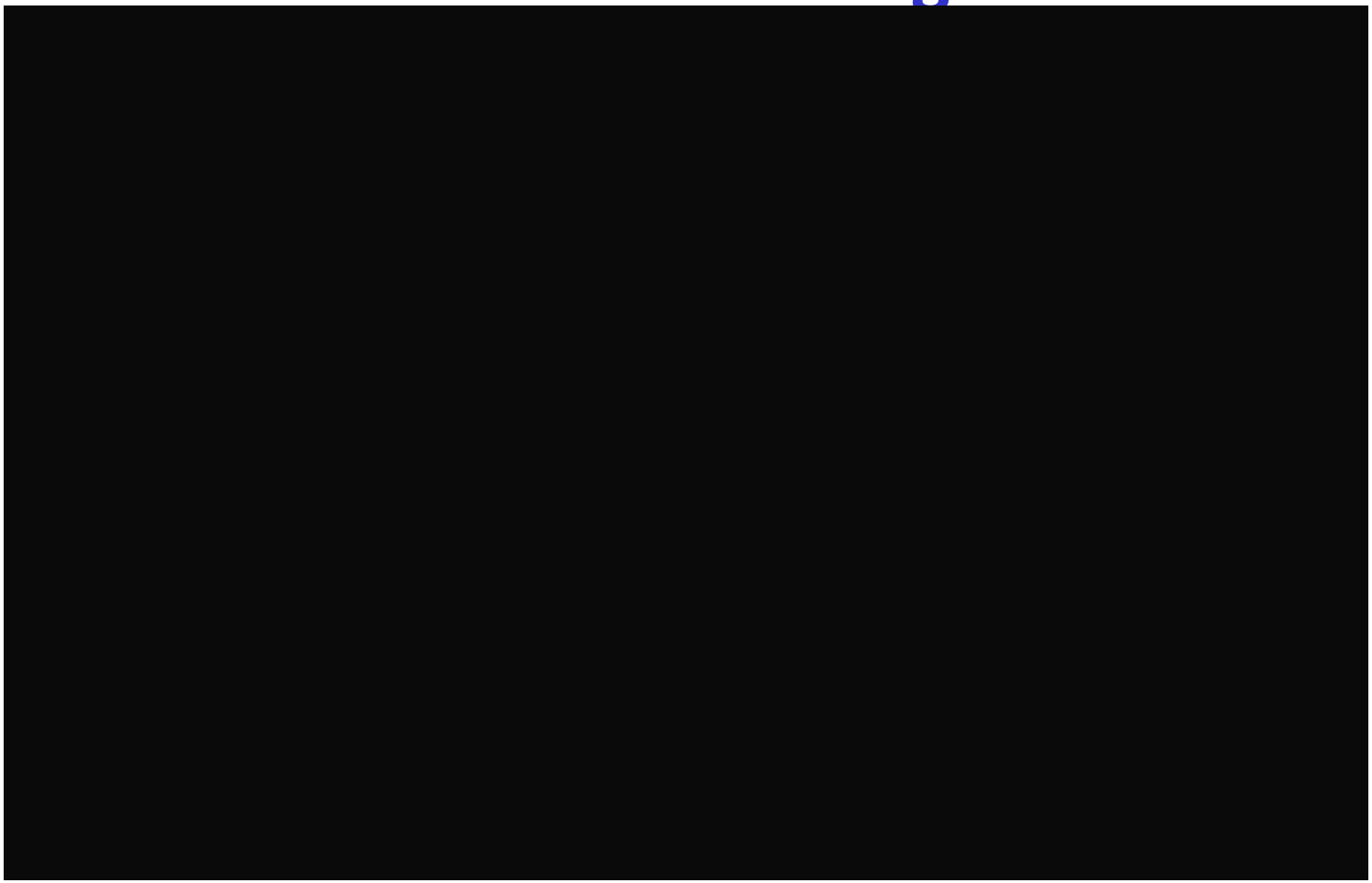
# Suggested theme : health

- http://physionet.org/
- Data of various kinds
  - Static snapshots
  - Time-series data
- For various health markers
  - Timing measurements, e.g. Gait
  - Electrical measurements, e.g. ECG, EKG
  - Images: Magnetic Resonance

# Problems

- Signal enhancement
  - Measurement is noisy, can you clean it
- Classification
  - Does this person have Parkinsons
  - Does this person have a cardiac problem
- Prediction
  - Rehospitalization: What fraction of these patients will go back to hospital in the next N days

# User Guided Sound Processing: A fun demo from Paris Smaragdis



# Talk-Along Karaoke

- Pick a song that features a prominent vocal lead
  - Preferably with only *one* lead vocal
- Build a system such that:
  - User talks the song out with reasonable rhythm
  - The system produces a version of the song with the user *singing* the song instead of the lead vocalist
    - i.e. The user's singing voice now replaces the vocalist in the song
- No. of issues:
  - Separation
  - Pitch estimation
  - Alignment
  - Pitch shifting

# Plagiarism Detection

- Youtube videos..

- e.g. Are the first bars in these two identical to merely close or copied?

  http://www.youtube.com/watch?v=iPqsix_wm6Y

vs.

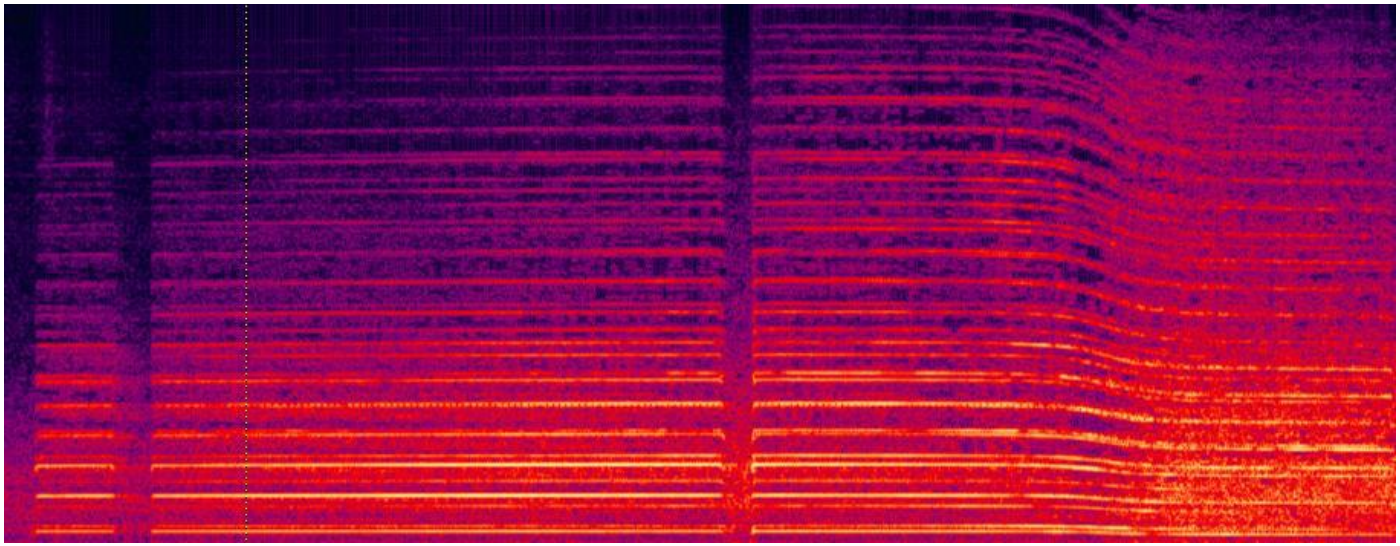
  http://www.youtube.com/watch?v=RhJaVvyanZk

- Cover song detection

# The Doppler Effect

- The observed frequency of a moving sound source differs from the emitted frequency when the source and observer are moving relative to each other


# The Doppler Effect

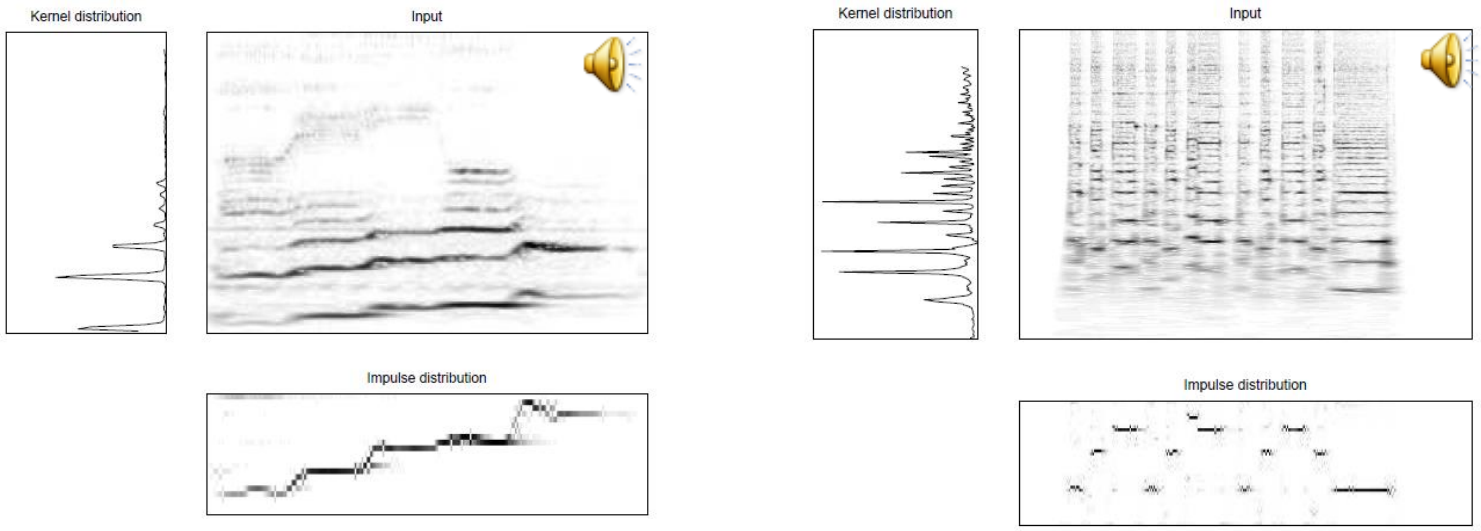


- Spectrogram of horn from speeding car
  - Tells you the velocity
  - Tells you the distance of the car from the mic

# Problem

- Analyze audio from speeding automobiles to detect velocity using the Doppler shift
- Find the frequency shift and track velocity/position
- Supervisor: Dr. Rita Singh

# Pitch Tracking




- Frequency-shift-invariant latent variable analysis
- Combined with Kalman filtering
- Estimate the velocity of *multiple* cars at the same time

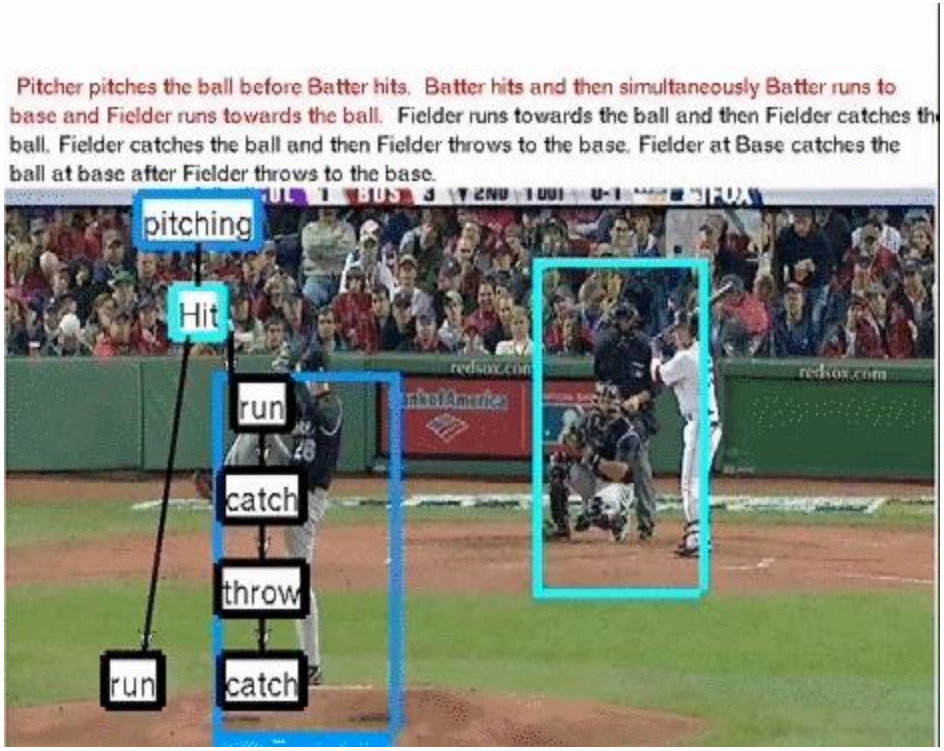
# New Doppler Problem

- Can we learn to derive articulator information from speech by considering its relationship to Doppler signal
- Can this be used to improve automatic speech recognition performance

- Procedure
  - Learn a deep neural network to learn the mapping
  - Use the network as a feature computation module for speech recognition
    - Augments conventional features

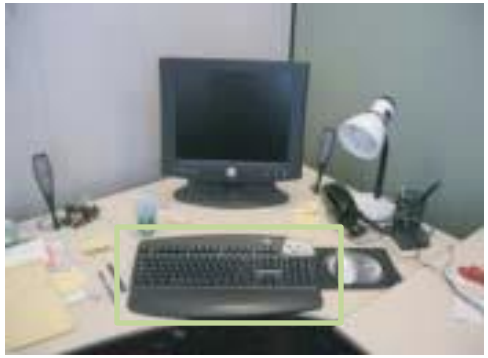
- Supervisor: Bhiksha Raj

# Assigning Semantic tags to multimedia data




- http://www.cs.cmu.edu/~abhinavg/Home.html
- Dan Ellis' website..

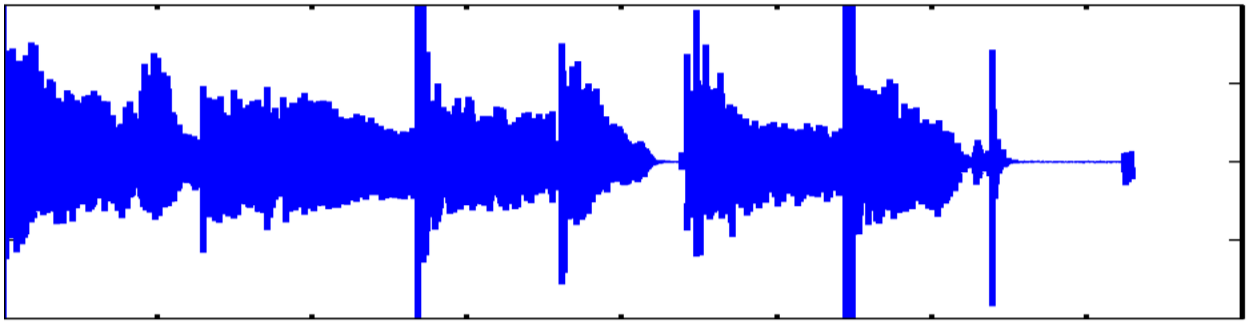
# Object detection and Clustering


- Detect various types of objects in images
  - Supervised: You know what objects to detect
  - Unsupervised: Detect objects based on motion

- **Required for content-based description**
- Semi-knowledge-based clustering, supervised/semi-supervised learning

# *Audio object* detection and Clustering



- Learn to detect various sound phenomena in multimedia recordings from “the wild”
  - YouTube style data
- Even when they occur concurrently with other sounds
- Including sound phenomena for which we may have no training instances!

# Geolocation

- Different places *look* different
- And *sound* different

- Problem: Given an image, video or audio recording, can we localize it geographically
  - E.g. identify the town / country / continent
  - Localize it qualitatively
    - E.g. Its in a high-traffic area / Near the sea / at A windy place / “Sounds like Chicago..”

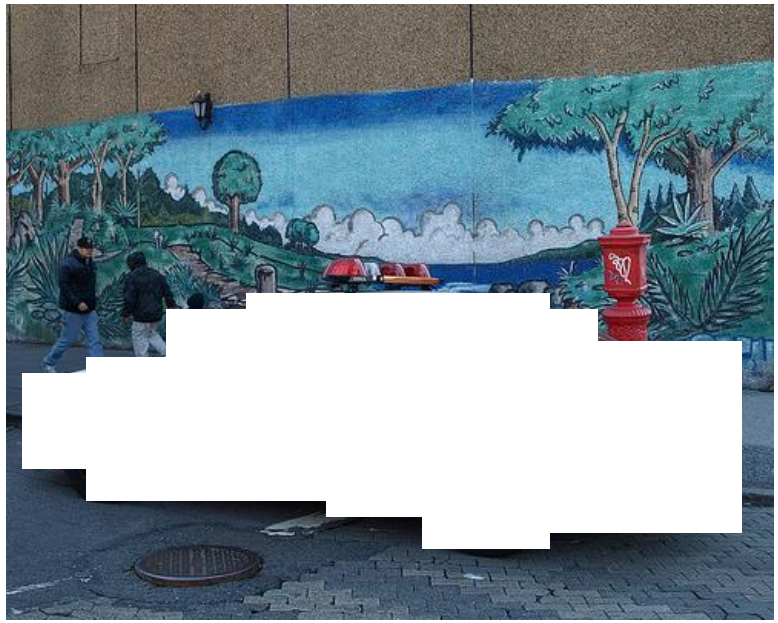
# Recognizing Gender of a Face

- A tough problem
- Similar to face recognition
- How can we detect the gender of a face from the picture?
  - Even humans are bad at this

# Image Manipulation: Filling in

- Some objects are often occluded by other objects in an image
- Goal: Search a database of images to find the one that best fills in the occluded region

# Image Manipulation: Filling in

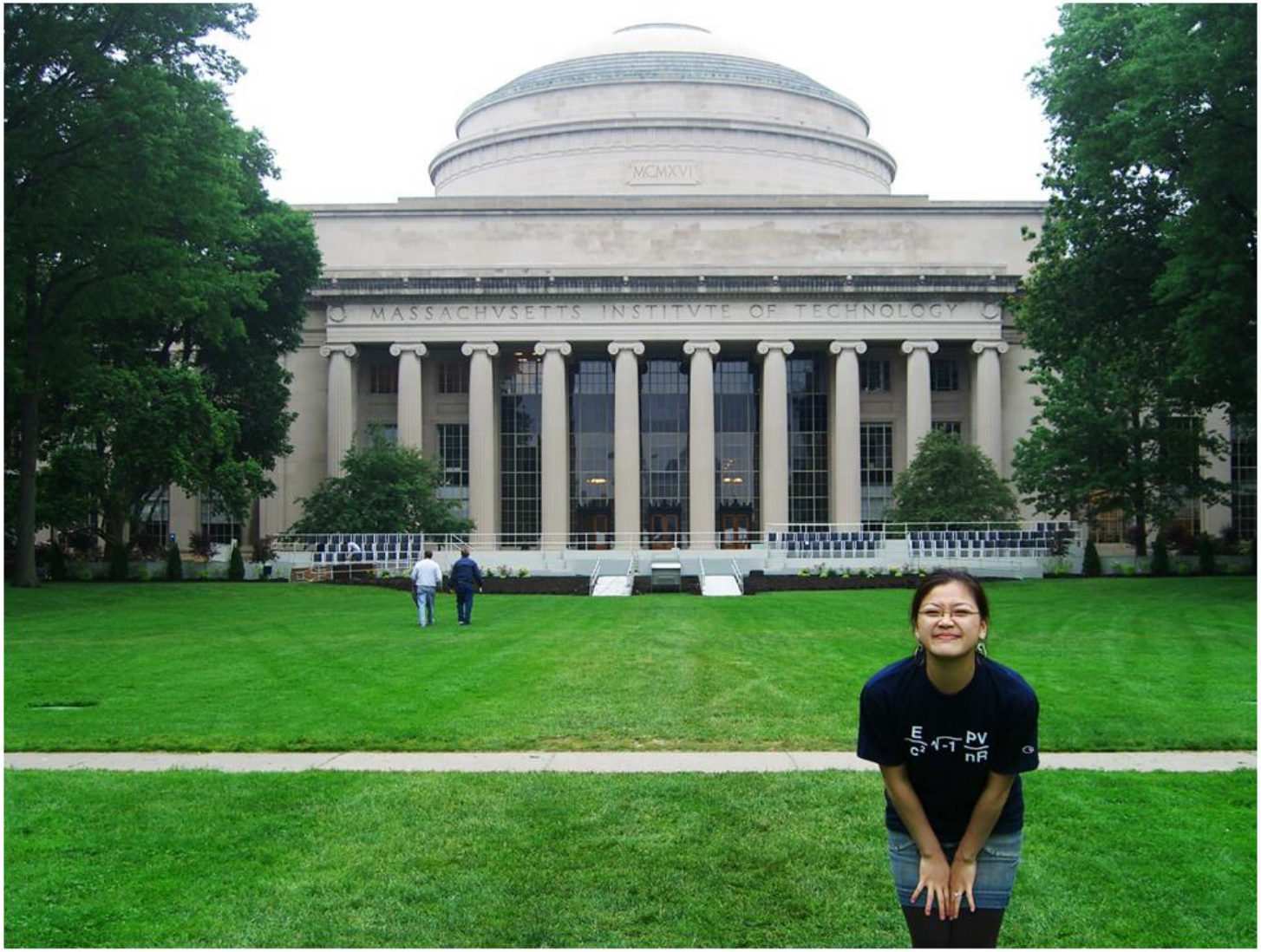
- Some objects are often occluded by other objects in an image
- Goal: Search a database of images to find the one that best fills in the occluded region

# Image Manipulation: Modifying images

- Moving objects around
  - "Patch transforms", Cho, Butman, Avidan and Freeman
  - Markov Random Fields with complicated a priori probability models

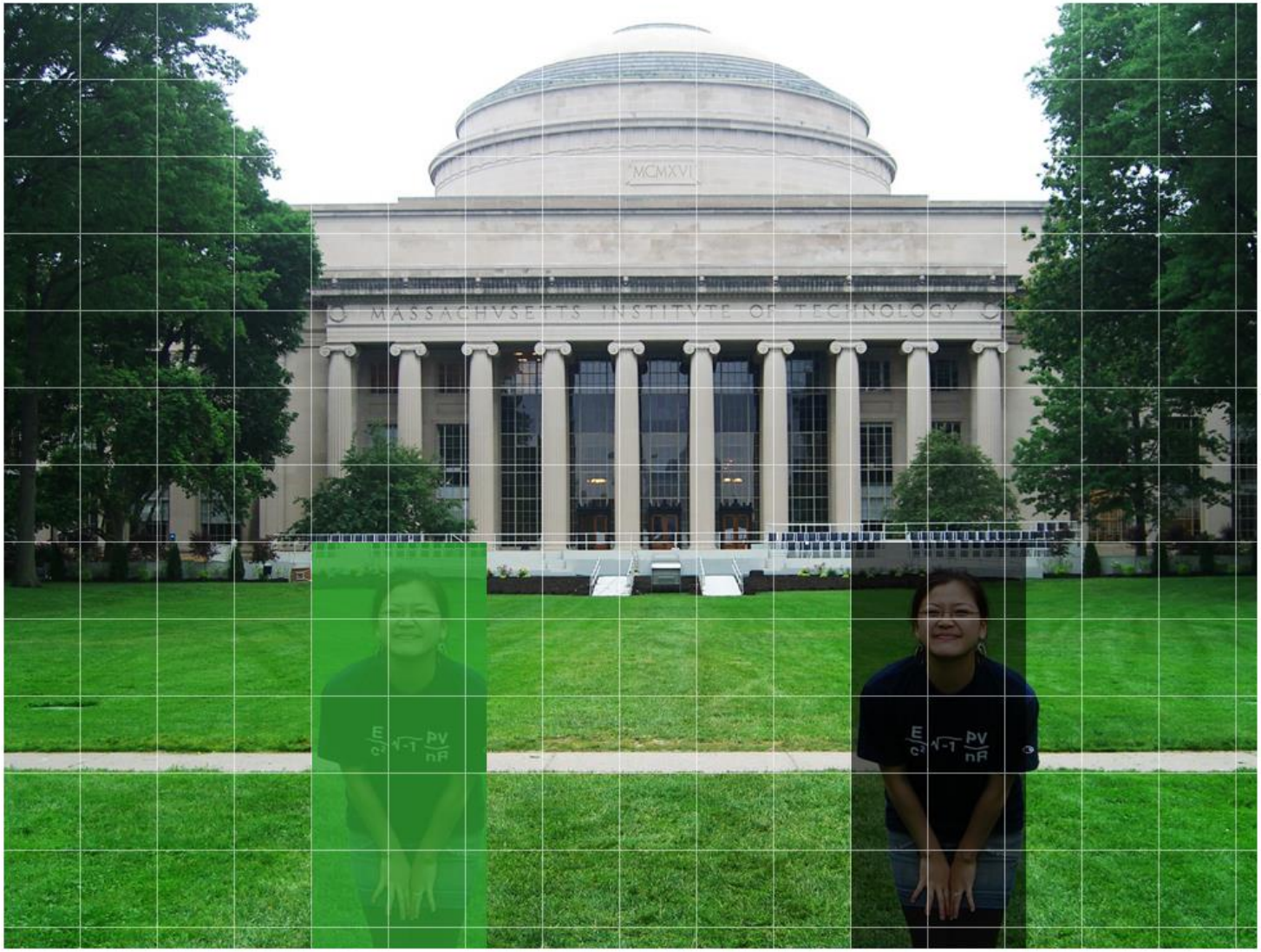
# Applications – Subject

## Input image



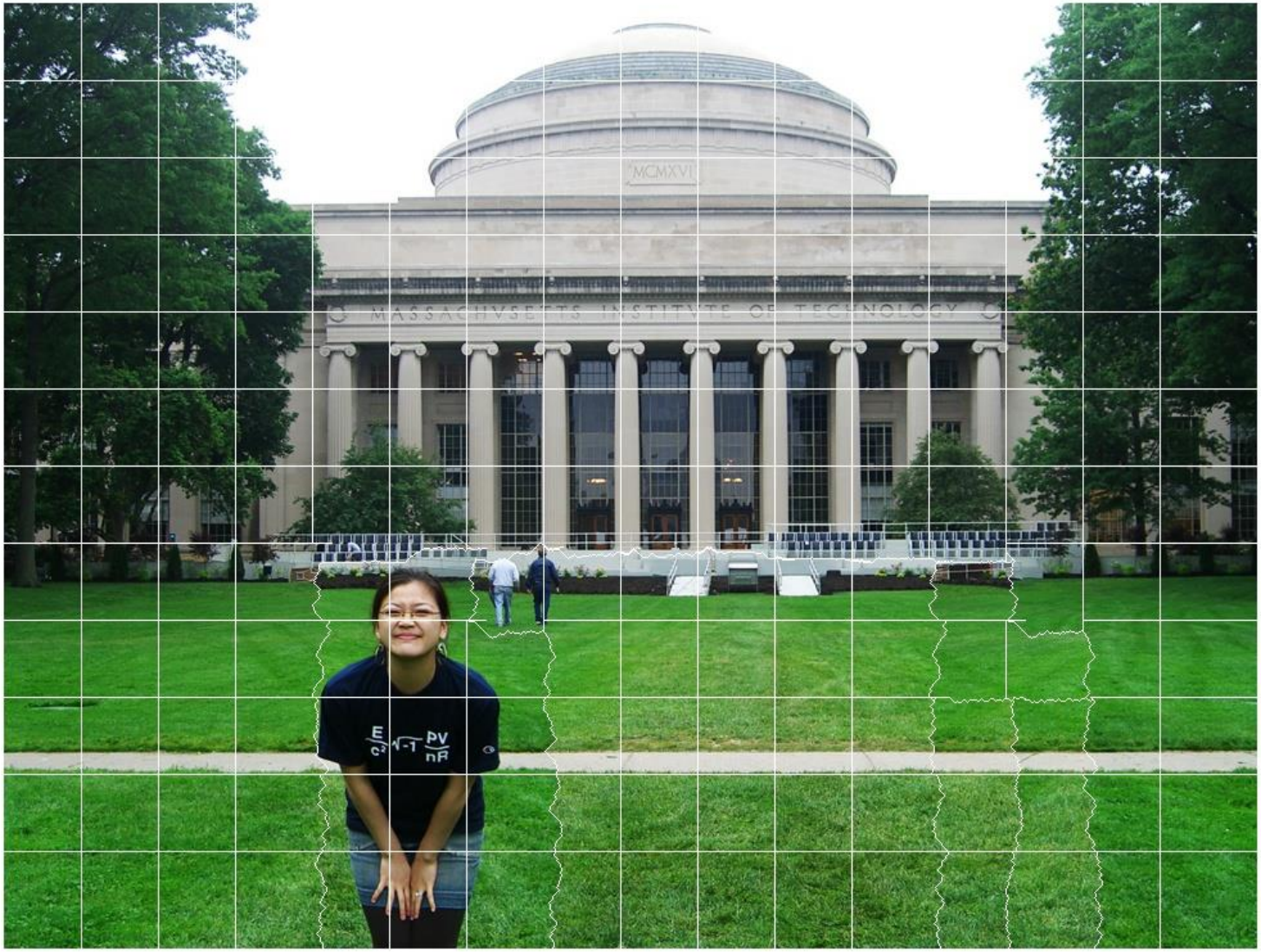
# Applications – Subject reorganization

## User input



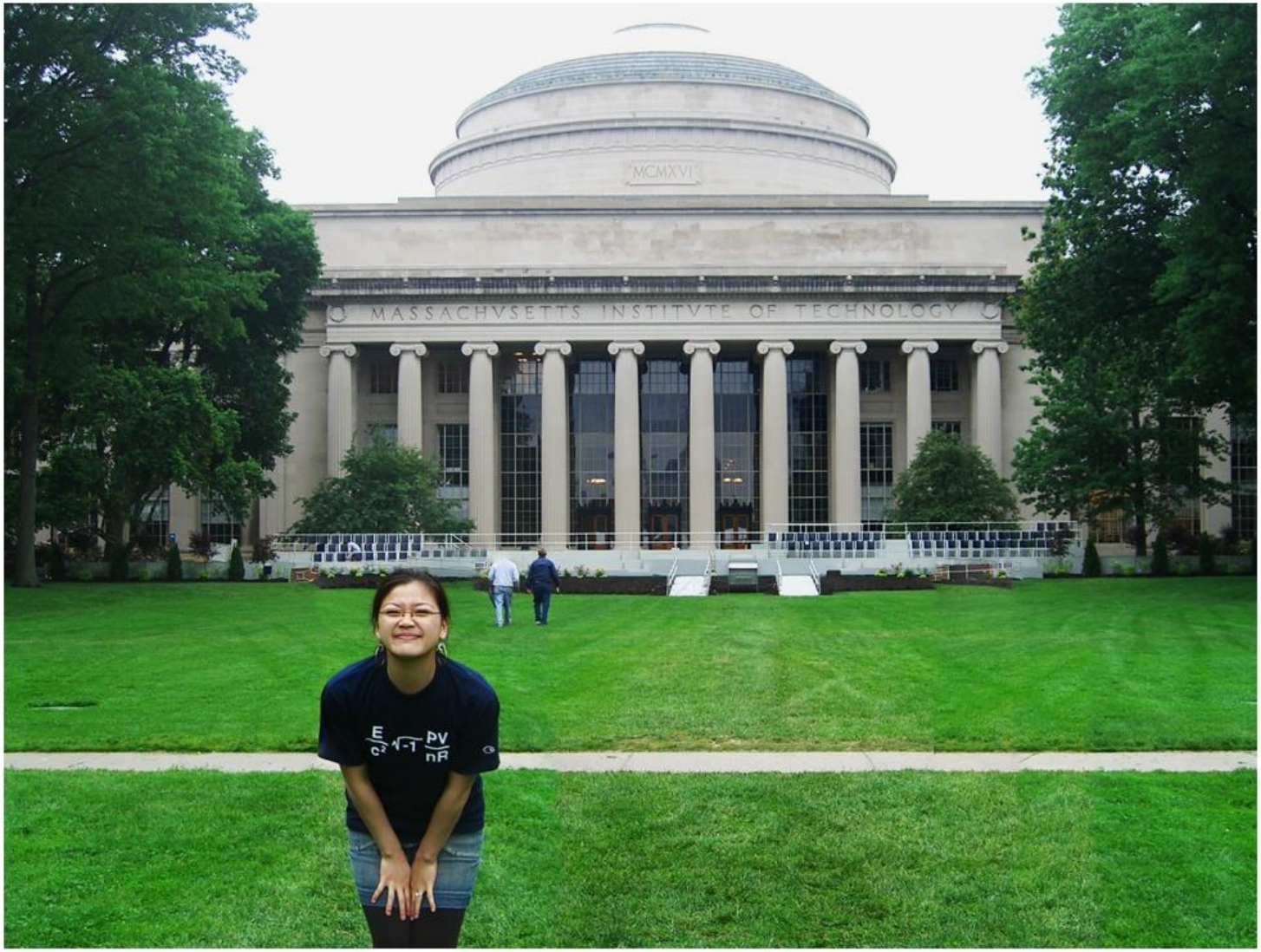
# Applications – Subject reorganization

## Output with corresponding seams



# Applications – Subject reorganization

## Output image after Poisson blending



# You get the idea

- You may pick any of these problems or come up with a fun one of your own
- They *must* exercise your MLSP skills

- Please form teams and inform me and TAs of teams asap
  - Or we will assign you to a team
- Please send us project proposals before 25th
  - Try to break down the steps in solving your problem in your proposal
  - Needed to evaluate feasibility