11-755 Machine Learning for Signal Processing

Course Projects

Class 6. 9 Sep 2010

Administrivia

- Slides were not up last week
  - Should be up now
  - Problem generating handouts
- Homework questions?

Course Projects

- Covers 50% of your grade
- 10-12 weeks
- Required:
  - A seriously attempted project
  - Demo if possible
  - Project report
  - Poster presented in poster session
- Project complexity
  - Depends on what you choose to do
  - Complexity of project will be considered in grading
  - Projects can range from researchy to implementation of existing techniques
    - In the latter case, the implementation

Course Projects

- Projects will be done by teams of students
  - Ideal team size: 4
  - 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 the problem
- Grading will be done by team
  - Team members will grade one another
    - Final grade is combination of two

Projects

- A list of possible projects will be presented to you in the rest of this lecture
- This is just a sampling
- You may work on one of the proposed projects, or one that you come up with yourselves
- Teams must inform us of their choice of project by 21nd September 2010
  - The later you start, the less time you will have to work on the project

Projects from last year

- *Statistical Klatt Parametric Synthesis*
- Seam Carving
- Content-aware resizing for video applications
- Voice transformation with Canonical Correlation Analysis
- Talking Karaoke
- *Sound source separation and missing feature enhancement*
- Voice transformation
- Image segmentation
- ***Non-intrusive load monitoring***
- Counting blood cells in Cerebrospinal Fluid
- Determining Music Tablature
- Image Deblurring
- Face detection

A Theme this year

- Analyzing a movie
  - Who mining:
    - Form characters
      - What they look like, what they sound like
      - What kind of things do they say
  - Activity detection:
    - Identify different actions in the video
  - Story summarization

Potential Projects

- http://ayesha.lti.cs.cmu.edu/twiki/bin/view/Main/MLSP2010Projects
- Scene segmentation using video
- Scene segmentation/classification using audio
- Automatically clustering faces and voices
- Object detection and clustering
- Detecting/classifying actions
- Emotion detection from audio/images

Scene segmentation with video

- Automatically detect discontinuity in the narrative, from the video alone
  - Automatic shot change detection
    - Shot: sequence of images from a single camera operation

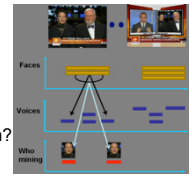

  - Scene change detection: A scene may have many shots

Scene segmentation with audio

- Identify change of scene from the audio alone
  - A set of characters speaks in a scene
    - Set of speakers is scene specific, rather than shot specific
    - The background conditions change
    - Detect when the change is significant and typical of scene change

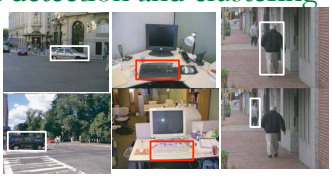
Automatically clustering faces and voices

- Individual shots have multiple faces
- Typically only one voice
  - Who does the voice belong to?
  - Can we cluster the faces?
    - Using voice as additional cue?
    - Not knowing face-voice association?



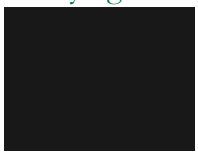
- A joint association-determination and clustering problem
  - Needs face detection, change point detection in voice and segmentation

Object detection and clustering



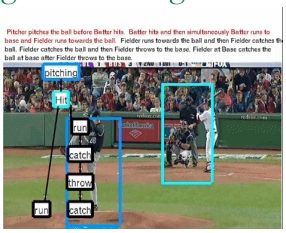
- Detect objects of various types in image
  - Supervised: Know what kind of objects to look for
  - Unsupervised: Detect objects based on motion
    - Cluster
  - Question: Perspective / view point ?

Detecting/classifying actions



- Detect and classify actions in video

Assigning semantic tags to video

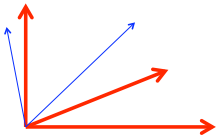


- http://www.cs.cmu.edu/~abhinavg/Home.html

Emotion detection from audio/images

- Detecting and recognizing the emotion in faces
- Emotion recognition in voices

Compressive Sensing

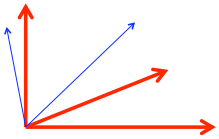


- A new and fast growing field
- If only a small number of components in a data instance are non-zero, the data are sparse
  - E.g., a 1-sparse data in 3-D space will lie only along the axes
    - All vectors will be of the form (x,0,0), (0,y,0), (0,0,z)
- When data are sparse a reduced number of measurements are sufficient
  - E.g., here knowing the projection of the data on two vectors is enough
    - Only 2 measurements

Compressive sensing

- Very important
- Data like MRIs are very sparse
- Must take many many many measurements for a single picture
  - Each measurement is expensive
  - Reduce the number of measurements taken
    - Use sparsity – Compressive sensing
- Goal: Adapt the measurements based on measurements taken so far
  - Will require even fewer measurements

Adaptive Compressive Sensing



- Identify the axes (blue lines) dynamically, based on
- Validating theory
- Data-driven measurement

CS projects

- Validating theory:
  - Have developed adaptive CS technique
  - Have developed mathematical models that predict its probability of making error
  - Must validate on real data
- Data-driven CS
  - Analyze lots of training examples
  - Use these to obtain adaptive measurement methods that require fewer measurements than current techniques

More Project Ideas

- Sound
  - Separation
  - Music
  - Classification
  - Synthesis
- Images
  - Processing
  - Editing
  - Classification
- Video
  - …
  - …

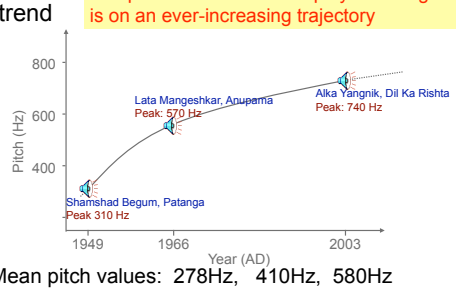
Ideas from Alan

- Synthesis/recognition of languages with no orthography
- Live voice transformation/mimicking. Convert a live voice with now training data to another voice as they are speaker.
- De-identification of speech
- Eigen voices for different speaker characters in a Virtual World (Alice) so people can choose child, adult, old, male, female ...
- In Let's Go data predict: if a call will be successful or not from the first utterance (based on acoustics, ASR output, signal to noise ratio etc
- Using Articulatory Features in parametric speech synthesis

A Strange Observation

- A trend

The pitch of female Indian playback singers is on an ever-increasing trajectory



- Mean pitch values: 278Hz, 410Hz, 580Hz

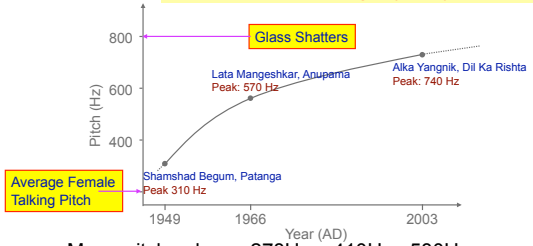
I'm not the only one to find the high-pitched stuff annoying

- Sarah McDonald (Holy Cow): ".. shrieking…"
- Khazana.com: ".. female Indian movie playback singers who can produce ultra high frequncies which only dogs can hear clearly.."
- www.roadjunky.com: ".. High pitched female singers doing their best to sound like they were seven years old .."

A Disturbing Observation

- A trend

The pitch of female Indian playback singers is on an ever-increasing trajectory



- Mean pitch values: 278Hz, 410Hz, 580Hz

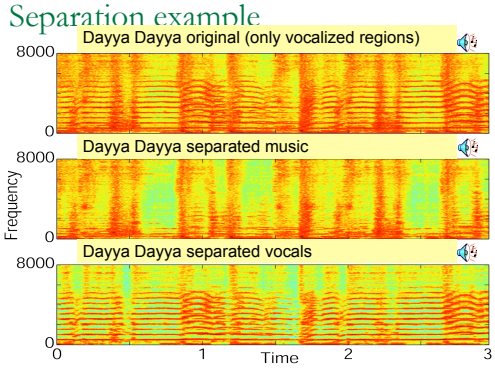
Subjectivity of Taste

- High pitched female voices can often sound unpleasant
- Yet these songs are very popular in India
  - Subjectivity of taste
- The melodies are often very good, in spite of the high singing pitch

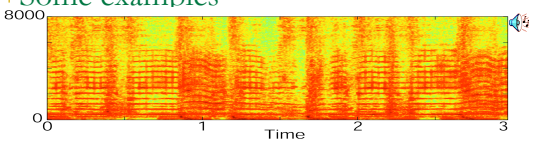
"Personalizing" the Song

- Retain the melody, but modify the pitch
  - To something that one finds pleasant
  - The choice of "pleasant" pitch is personal, hence "personalization"
- Must be able to separate the vocals from the background music
  - Music and vocals are mixed in most recordings
  - Must modify the pitch without messing the music
- Separation need not be perfect
  - Must only be sufficient to enable pitch modification of vocals
  - Pitch modification is tolerant of low-level artifacts
    - For octave level pitch modification artifacts can be undetectable.

Separation example

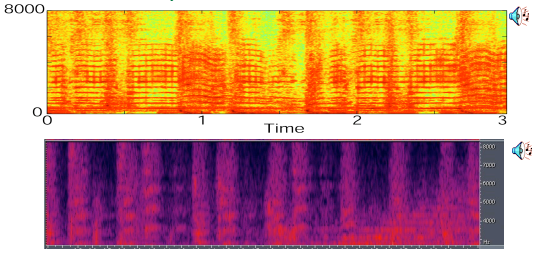


Some examples



- Example 1: Vocals shifted down by 4 semitones

Some examples



- Example 1: Vocals shifted down by 4 semitones
- Example 2: Gender of singer partially modified

Projects..

- Several component techniques
- Illustrate various ML *and* signal processing concepts
- Signal separation
  - Latent variable models
  - Non-negative factorization
- Signal modification
  - Pitch and spectral modification
  - Phase and phase estimation

Song "Personalizer"

- Modify vocals as desired
  - Mono or Stereo
  - "Knob" control to modify pitch of vocals
- Given a song
  - Separate music and song
  - Modify pitch as required
  - Adjust parameters for minimal artifacts
  - Add..
- Issues:
  - Separation
  - Modification
  - Use of appropriate statisical model and signal processing

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

The Doppler Ultrasound Sensor

- Using the Doppler Effect

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
  - Discovery attributed to Christian Doppler (1803-1853)

Person being approached by a police car hears a higher frequency than a person from whom the car is moving away

Observed frequency

- The relationship of actual to percieved frequencies is known
- Case 1: The source is moving with velocity *v*, but the listener is static
  - Observed frequency is:

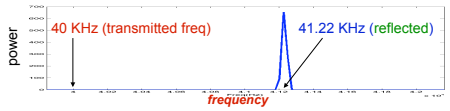
$$f' = \frac{c_{sound} f}{c_{sound} - v}$$

- Case 2: The observer is emitting the signal which is reflected off the moving object
  - Observed frequency is:

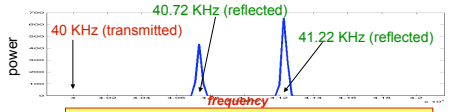
$$f' = \frac{(c_{sound} + v) f}{c_{sound} - v}$$

Doppler Spectra

- 40 Khz tone reflected by an object approaching at approximately 5m/



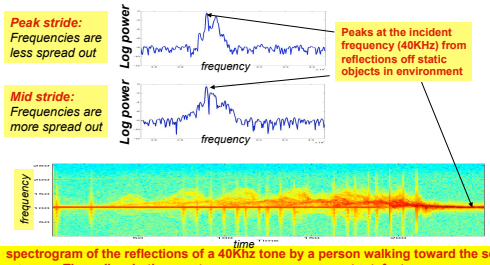
- 40 Khz tone reflected by two objects, one approaching at approximately 5m/s and another at 3m/s



Multiple velocities result in multiple reflected frequencies

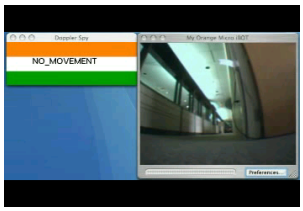
Doppler from Walking Person

- Human beings are articulated objects
- When a person walks, different parts of his body move with different velocities. The combination of velocities is characteristic of the person
  - These can be measured as the spectrum of a reflected Doppler signal



spectrogram of the reflections of a 40Khz tone by a person walking toward the sensor
The spikes in the spectrogram are measurement artefacts

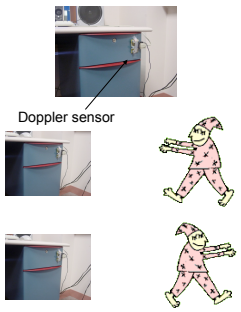
Identifying moving objects



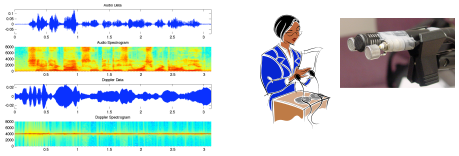
- Doppler spectra are signatures of the motion
  - The pattern of velocities associated with the movement of an object are unique

Gait Recognition

- Beam Ultrasound at a walking subject
- Capture reflections
- Determine identity of subject from analysis of reflections
- Issues:
  - Type of Signal Processing
  - Type of classifier
  - Hardware..

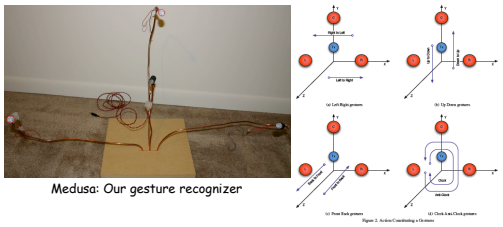


Identifying talking faces..



- Beam ultrasound on talker's face
- Capture and analyze reflections
- Identify subject

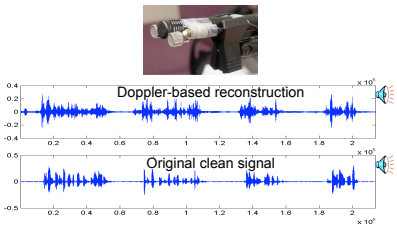
The Gesture Recognizer



Medusa: Our gesture recognizer

- Gesture recognizer
  - and examples of actions constituting a gesture

Synthesizing speech from ultrasound observations of a talking face



- Subject *mimes* speech, but does not produce any sound
- Can we synthesize understandable speech?

Sound Classification: Identifying Cars / Automobiles from their sound

- Sounds are often signatures
- Simple problem: Can we build a system that can identify the make (and possibly model) of a car by listening to it?
  - Can you make out the difference between a V6 and a V8?
    - What do you know of the underlying design that can help?
- Issues:
  - Gathering Training Data
  - Signal Represenation
  - Modeling

IMAGES

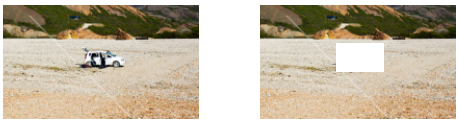
Face Recognition

- Similar to the face detector, but now we want to *recognize* the faces too
  - Who was it who walked by my camera?
- Can use a variety of techniques
  - Boosting, SVMs..
  - Can also combine evidence from an ultrasound sensor
  - Can be combined with face detection..

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

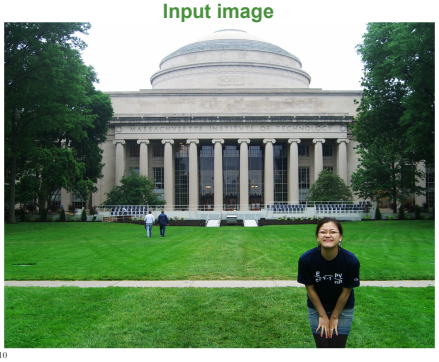
- Some objects are often occluded by other objects in an image
- Goal: Search a database of images to find the one that best fills in the occluded region

Image Manipulation: Modifying images

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

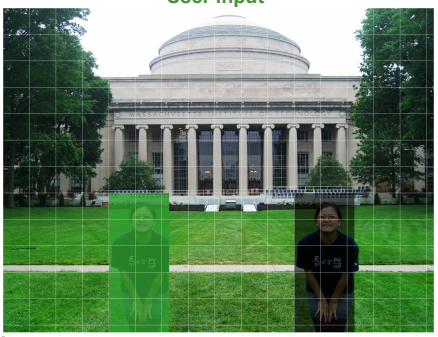
Applications – Subject reorganization♪

Input image



Applications – Subject reorganization

User input



Applications – Subject reorganization

Output with corresponding seams

Applications – Subject reorganization

Output image after Poisson blending

Image Composition

- Structure from Motion:
  - Given several images of the same person under different pose changes build a 3D face model.

Image Composition

- Solving for correspondence across view-point:
  - Given several faces images of the same person across different pose, expression and illumination conditions solve for the correspondence across facial features.
  - The frontal image will be labeled with 66 landmarks.
- Similar to patch models
  - Finding correspondences that match