

ECE 417: Multimedia Signal Processing

Fall 2018

Machine Problem #4

Due: Thursday, December 6, 2018

1 Overview

In this machine problem, you will get familiar with face modeling and animation techniques, learn to use ANNs (artificial neural networks) to map features of speech signals to facial animation parameters, and produce facial animation sequences from the audio tracks.

2 The Audio-Visual Database

The pre-processed database will be provided in the Matlab MAT file format, namely, **ECE417_MP4_AV_DATA.mat**. This file contains the following four Matlab variables:

1. **av_train**

av_train is a structure variable storing the audio-visual data for training the ANN. It has the following elements:

av_train.audio is a matrix of the audio features. The k th column of the audio feature matrix can be accessed as `av_train.audio(:, k)`. It represents the audio feature vector of the frame k of the audio.

av_train.visual is a matrix of the visual features. The k th column of the visual feature matrix can be accessed as `av_train.visual(:,k)`. It represents the visual feature vector of the frame k of the video. A visual feature vector contains three numbers. The first one is the Δ width of lips ($\Delta w = w - w_0$). The last two numbers are the Δ height of the upper lip ($\Delta h_1 = h_1 - h_{10}$), and the Δ height of the lower lip ($\Delta h_2 = h_2 - h_{20}$) where w_0 , h_{10} , h_{20} are the width and heights of the neutral lips (see Figure 1).

2. **av_validate**

av_validate is a structure variable storing the audio-visual data for validation of ANN. It has elements:

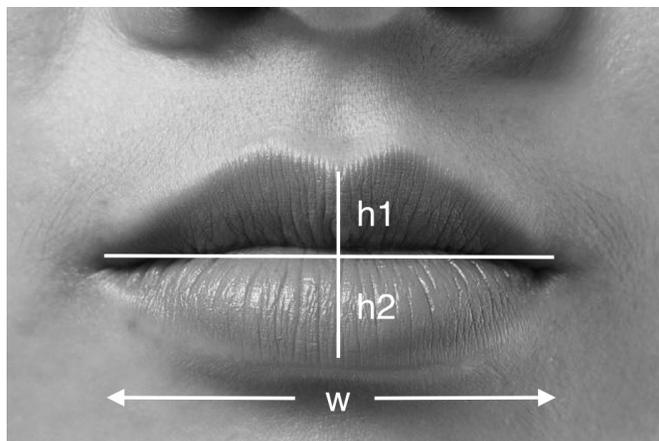


Figure 1: The visual features

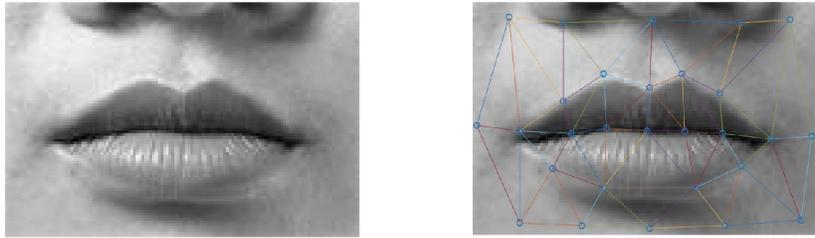


Figure 2: Mouth image (left); Mouth image with triangular mesh (right)

av_validate.audio is the audio feature matrix. Each column `av_validate.audio(:,k)` represents the audio feature vector of frame `k`.

av_validate.visual is the visual feature matrix. Each column `av_validate.visual(:,k)` represents the visual feature vector of frame `k`.

3. testAudio

`testAudio` is the test audio data matrix. Each column is an audio feature vector.

4. silenceModel

`silenceModel` will be used to decide if an audio frame corresponds to silence.

3 Data for Producing Animation

In this machine problem, facial animation is achieved by image warping. Two files are provided for image warping. In addition, a waveform file is provided as the sound track corresponding to `testAudio` for making the final movie file.

1. mouth.jpg

A neutral mouth image will be provided (see Figure 2). You will use it to generate a mouth animation image sequence.

2. mesh.txt

A triangular mesh that triangulates the mouth area in `mouth.jpg` (see Figure 2). You will use this mesh and the mouth image to generate new mouth images through image warping. The format of the file `mesh.txt` is:

Number of vertices

x coordinate of vertex 1, y coordinate of vertex 1

x coordinate of vertex 2, y coordinate of vertex 2

...

Number of triangles

vertex 1, vertex 2, vertex 3 (of the 1st triangle)

...

3. test.wav

The waveform file corresponding to the audio feature matrix `testAudio`.

4 Tasks

1. Write your image warping code in MATLAB. The code takes the visual features as input and synthesizes new mouth images. (We recommend you to do this part first.)
2. Load pre-processed training data from ECE417_MP4_AV_DATA.mat.
3. Use the training data set to train a set of ANNs as the mapping from audio features to visual features. The MATLAB code ECE417_MP4_AV_train is provided.
4. Apply the mapping to the test audio features and obtain synthetic visual features. The MATLAB code ECE417_MP4_AV_test is provided.
5. Produce image sequence for the synthesized visual features.

5 Detailed Descriptions

1. **Image warping:** First, you need to deform the mesh according to the visual features. The deformation of the mesh can be decided by interpolation from the visual features. A MATLAB function `interpVert` using linear interpolation will be provided. Then write a warping function to generate the deformed mouth images using the given mouth image (See Figure 2). For the pixels outside the mesh and pixels in the holes inside lips, leave them black.
2. Load pre-processed data.
3. ANNs training and testing:
Matlab function `ECE417_MP4_train` will be provided. One parameter (number of hidden units) can be adjusted to get good mapping results.
Matlab function `ECE417_MP4_test` will be provided.
4. Use the estimated visual feature from test data, the triangular mesh, and the mouth image to generate mouth image sequences.
5. Produce an animation movie file.
 - (a) Firstly generate face images from visual features.
 - (b) Save the images in JPEG format and name them as `test_####.jpg`. `####` is the frame number of the image, starting from 0. For example, for the 15th frame, the file name is `test_0014.jpg`.
 - (c) Use the provided executable `DxBMP.exe` to convert the image sequence into a movie file. If `DxBMP.exe` is in the same directory as the images, the command line is `'DxBMP -framerate 30 test_*.jpg test.avi.'` The output movie file is `test.avi`. More information about `DxBMP.exe` can be found in the provided `DxBMP.htm`. OR use matlab's `VideoWriter` object, and `writeVideo` command, to create a movie inside matlab.
 - (d) Open Windows Movie Maker, or VirtualDub, or imovie, or any other movie program of your choice. Import the video `test.avi`, and the audio `test.wav`. Save the resulting file, with both audio and video.

6 Extra Credit [Up to 20%]

Here are some suggestions that you could do to improve the photo animation. You can choose up to two of them, for earning up to 10% extra credit each:

1. Replace `interpVert.m` with a function that computes the lip positions in some other way. Try to come up with a function that makes more realistic mouth shapes than `interpVert.m`.
2. Implement the forward neural network in python, and show that you can use the same network weights to generate the same outputs from your python code.

3. Animate a new face image: either your own face, or any face that you download from the internet.
4. The open mouth is unrealistic, in part because there is no tongue or teeth. Find a somewhat realistic way to show the tongue and teeth in your animated image.
5. The neural net sometimes fails to correctly guess lip positions because sometimes the audio signal doesn't contain enough information. Can you think of some auxiliary signal you could give, that would tell the lips when to open? Can you find a way to incorporate that into your animation?

7 Note

You may need to download the 'nntool' Toolbox in Matlab for training and testing ANN mapping, and 'Image Processing' Toolbox for some image manipulation commands.

8 Submission

You will submit (1) the movie file mp4.wmv you generated (2) a report in PDF format, and (3) a zip file containing your code along with a Readme file to Compass. The report should describe how you do this machine problem, and discussion about your results or any problems you have met. If you choose to do the extra credits, please document that at the end of your report. Clearly indicate which extra credit you choose, describe how you did the task and show your results as well as the analysis. You must name your report as <Lastname>.<Firstname>.report.pdf and your zip file as <Lastname>.<Firstname>.code.zip. If you are working as a team, only one person should upload the report. Please make sure that the title page includes the names of all the team members.