

# Αναγνώριση Συναισθήματος σε Οπτικά δεδομένα

Ειδικό Θέμα 2012-13

Βάσωφ Αθανάσιος

AEM 666

## 1.Εισαγωγή

Κατά την διάρκεια της ομιλίας, το ανθρώπινο πρόσωπο περιέχει, πλούσια σε συναίσθημα πληροφορία. Σκοπός της εργασίας είναι η εισαγωγή στα βασικά στοιχεία της αναγνώρισης προτύπου, και η καταγραφή των αποτελεσμάτων ενός απλού πειράματος μηχανικής μάθησης με χρήση Gaussian Mixture Model πάνω σε οπτικά δεδομένα ανθρώπινου συναισθήματος.

## 2.Βάση Δεδομένων

Σαν πρώτο βήμα της εργασίας έγινε έρευνα για την κατάλληλη βάση δεδομένων. Τελική επιλογή ήταν η βάση δεδομένων Interactive Emotional Dyadic Motion Capture (IEMOCAP), η οποία περιέχει μεγάλο όγκο οπτικοακουστικής πληροφορίας(12 ώρες από 10 διαφορετικούς ηθοποιούς), μεταξύ της οποίας video και motion capture. Η βάση αυτή περιεχέει εκτός των άλλων για κάθε frame την 3D (x,y,z) θέση από 46 σημεία κανονικοποιημένα ως προς το κεντρικό σημείο του προσώπου.

Λόγω του μεγάλου όγκου πληροφορίας έγινε κατάλληλη αναζήτηση και συρραφή πληροφορίας σε αρχεία κειμένου ώστε να είναι εύκολη η περαιτέρω επεξεργασία των δεδομένων. Για τον σκοπό αυτό χρησιμοποιήθηκε unix shell script. Παρακάτω παραθέτω ένα παράδειγμα:

```
mkdir new
for myfile in `find Ses*`; do
myfile = `find Ses*|head -1`
echo $myfile
grep Ses0 $myfile>temp
cut -f 2,3 temp>./new/$myfile
rm $myfile
done
mv new session5
```

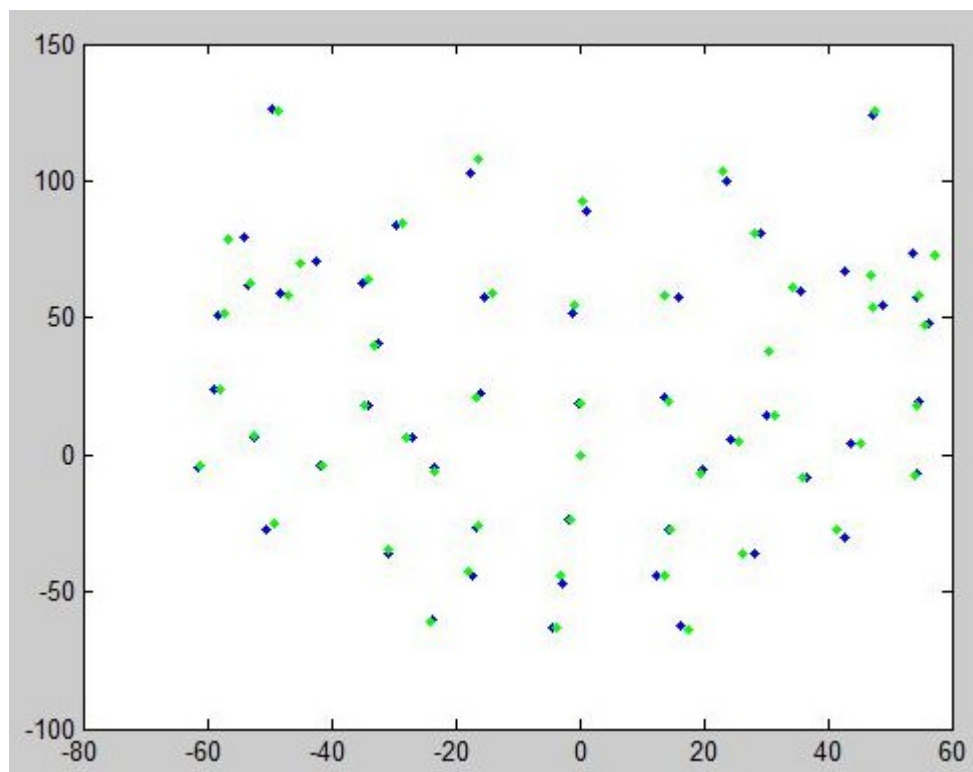
Τελικά η βάση οργανώθηκε σε φακέλους ανά συναίσθημα που ο κάθε ένας περιεχέει σε αρχεία κειμένου την οπτική πληροφορία μόνο των ομιλητών σε συγκεκριμένη συναισθηματική κατάσταση.

## 3.Εξαγωγή χαρακτηριστικών

Για την εξαγωγή χαρακτηριστικών, χρησιμοποιήθηκε ο μετασχηματισμός Principal Component Analysis(PCA) για την μείωση της διαστατικότητας των αρχικών δεδομένων αλλά και για την παραγωγή γραμμικά ανεξάρτητων διανυσμάτων για το training των GMM με χρήση διαγώνιου πίνακα.

Σε πρώτο στάδιο γίνεται η συρραφή όλων των δειγμάτων του training set,μετά γίνεται ένα average ανά 3 frames (πολύ υψηλό αρχικό frame rate) για μείωση του όγκου δεδομένων. Τέλος γίνεται το PCA και η αποθήκευση των 30 πρώτων

συντελεστών του(>95% variance) σε ένα πινάκα από struct που περιέχει τα δεδομένα και το ground truth για το κάθε συναίσθημα. Παρακάτω δίνετε ένα τυχαίο frame (σε 2 διαστάσεις) και το reconstructed frame από τους 30 πρώτους συντελεστές μετά τον PCA :



Για τον pca χρησιμοποιήθηκε το statistics toolbox του matlab, παρακάτω δίνονται κομμάτια από τον κώδικα της συνάρτησης που κάνει το feature selection:

```
%perform pca
[coef,score,latent] = princomp(lower_framerate_data);
%use 30 coefficients
coef=coef(:,1:30);

%save scores to feature table
training(k).features=score2;
```

#### 4.Training

Για το training χρησιμοποιήθηκε το μοντέλο των Gaussian Mixture Models με 8 mixtures για κάθε συναίσθημα. Τα συναισθήματα που χρησιμοποιήθηκαν τελικά ήταν 4 (angry,neutral,sad,happy). Παρακάτω είναι η συνάρτηση του Statistics Toolbox:

```
%training
ang=gmdistribution.fit(training(1).features,8,'CovType','diagonal');
hap=gmdistribution.fit(training(2).features,8,'CovType','diagonal');
neu=gmdistribution.fit(training(3).features,8,'CovType','diagonal');
sad=gmdistribution.fit(training(4).features,8,'CovType','diagonal');
```

## 5. Classification

Για το classification το πείραμα που δοκιμάστηκε ήταν 10-fold leave-one-speaker-out cross validation, δηλαδή το test set αποτελούνταν κάθε φορά από τις φράσεις του ομιλητή που δεν χρησιμοποιήθηκε στο training. Από την βάση δεδομένων κρατήθηκαν τα 100 πιο μικρά αρχεία για κάθε συναισθήμα, και η συνάρτηση που υλοποιεί το PCA αλλάχτηκε κατάλληλα ώστε να επιστρέφει για κάθε ομιλητή ένα πίνακα με τα χαρακτηριστικά για κάθε φράση ανά συναισθήμα. Παρακάτω είναι ο κώδικας που υλοποιεί το classification ελέγχοντας το pdf κάθε frame με το mixture model:

```
%classification
for n=1:length(test)
    angry=0;
    happy=0;
    neutral=0;
    sadd=0;
    data=test(n).features;
    [M N]=size(data);

    for i=1:M
        dat = data(i,:);
        a = pdf(ang,dat);
        b = pdf(hap,dat);
        c = pdf(neu,dat);
        d = pdf(sad,dat);

        [~,maxp]=max([a,b,c,d]);

        if (maxp==1)
            angry=angry+1;
        end
        if (maxp==2)
            happy=happy+1;
        end
        if (maxp==3)
            neutral=neutral+1;
        end
        if (maxp==4)
            sadd=sadd +1;
        end
    end

    [~,maxp]=max([angry,happy,neutral,sadd]);

    if strcmp(test(n).truth,'ang')

        if (maxp==1)
            ang_angry=ang_angry+1;
        end
        if (maxp==2)
            ang_happy=ang_happy+1;
        end
        if (maxp==3)
            ang_neutral=ang_neutral+1;
        end
        if (maxp==4)
            ang_sad=ang_sad +1;
        end

        ang_count=ang_count+1;
    end

    if strcmp(test(n).truth,'hap')
        if (maxp==1)
```

```

        hap_angry=hap_angry+1;
    end
    if (maxp==2)
        hap_happy=hap_happy+1;
    end
    if (maxp==3)
        hap_neutral=hap_neutral+1;
    end
    if (maxp==4)
        hap_sad=hap_sad +1;
    end

    hap_count=hap_count+1;

end

if strcmp(test(n).truth,'neu')
    if (maxp==1)
        neu_angry=neu_angry+1;
    end
    if (maxp==2)
        neu_happy=neu_happy+1;
    end
    if (maxp==3)
        neu_neutral=neu_neutral+1;
    end
    if (maxp==4)
        neu_sad=neu_sad +1;
    end

    neu_count=neu_count+1;

end

if strcmp(test(n).truth,'sad')
    if (maxp==1)
        sad_angry=sad_angry+1;
    end
    if (maxp==2)
        sad_happy=sad_happy+1;
    end
    if (maxp==3)
        sad_neutral=sad_neutral+1;
    end
    if (maxp==4)
        sad_sad=sad_sad +1;
    end

    sad_count=sad_count+1;

end

```

## 6.Αποτελέσματα

Παρακάτω είναι τα αποτελέσματα του πειράματος που εκτελέστηκε στα πλαίσια του ειδικού θέματος:

ANGRY 41.17%  
 HAPPY 34.44%  
 NEUTRAL 7.9%  
 SAD 19.5%

## 7.Βιβλιογραφία

**Audio-Visual Emotion Recognition using Gaussian Mixture Models  
for Face and Voice**

Angeliki Metallinou, Sungbok Lee and Shrikanth Narayanan

**VISUAL EMOTION RECOGNITION USING COMPACT FACIAL REPRESENTATIONS AND  
VISEME INFORMATION**

*Angeliki Metallinou, Carlos Busso, Sungbok Lee and Shrikanth Narayanan*