A Hidden Markov Model Based Approach to Identify Emotion from Facial Expression Using a Combination of Emotion Probability Classifier and Facial Action Unit Intensity Classifier

Suvashis Das  
Department of Management and Information Systems Science  
Nagaoka University of Technology  
1603-1, Kamitomiokamachi, Nagaoka  
Niigata, Japan  
Email: suvashis.das.iuj2009@gmail.com

Koichi Yamada  
Department of Management and Information Systems Science  
Nagaoka University of Technology  
1603-1, Kamitomiokamachi, Nagaoka  
Niigata, Japan  
Email: yamada@kjs.nagaokaut.ac.jp

Abstract—The field of computerized emotion detection has received lot of attention from many researchers. It is well accepted that the human face holds a lot of information about the emotional state of the person concerned. The problem can be thought of in two stages, first being the measurement of facial actions and second, classifying the measured actions into emotion categories. In this work we deal with the classification part of the problem. This work focuses on developing a model to classify emotions from observation sequences in terms of Action Unit (AU) intensities from a facial image based on Hidden Markov Model (HMM) using our proposed classifiers: Emotion Probability Classifier and Facial Action Unit Intensity Classifier, to find a less computation expensive and accurate model for emotion interpretation. The novelty of this work is: 1) We propose a new and more accurate model for facial expression analysis by assigning a set of N HMMs to every AU where N is the number of Emotions that we are interested to study. Assigning N HMMs to each AU eliminates the chance of undermining the insignificant or non present AUs by calculating separately the probability contributions towards each emotion by every single AU in the entire AU set which is used later to calculate the joint probability for each emotion considering all AUs together. 2) Instead of considering the huge number of combinations that are possible for Action Units we coin a way to deal with the combination of different Emotions which will render much more information about the state of the subject rather than just the basic emotions. 3) This work also focuses on optimizing at runtime, the computation cost using a threshold variable \( \theta \).

Index Terms—FACS, Emotion, HMM, Action Units

I. INTRODUCTION

There are many ways that the human emotion is observed. It is sometimes done by bio-medical means, sometimes by speech or sometimes by facial expression among many other ways. Out of all the ways human emotion observed by the facial expression method is the most appropriate due to the fact that voluntarily or involuntarily emotions are very well depicted by the human face. It also brings with it an array of challenges and complexities. An accurate face tracking algorithm may get really complex due to the very high number of ways a facial expression can change over time. Even more difficult can be predicting the exact emotion expressed by a face at a given point of time. A lot of painstaking research has gone towards developing a systematic way to decipher emotion information from a facial image. The most remarkable among them is the Facial Action Coding System developed by P. Ekman and W. V. Friesen in 1978 [1,2]. The actions that render facial expressions were organized by Ekman and Friesen[1,2] into Action Units representing different muscles of the face and their movements. In year 2000 I. Cohen, A. Garg, & T. S. Huang[7] addressed Emotion Recognition problem using multilevel HMMs and using one HMM for each of the six basic emotions (Surprise, Sadness, Fear, Happiness, Anger and Disgust) having a common observation feed consisting of multiple AU sequences in a single stream. The research considered neither the importance of the non present AUs nor they attempted to look into the combinational effect of emotions thereby leaving a gap towards effective emotional interpretation. They did talk about computational efficiency but it was not about dynamic adjustment in run time which leaves the area of synchronous trade off optimization between the accuracy and computational efficiency. In the same year Maja Pantic and Leon J.M. Rothkrantz[4] discussed on various techniques of decoding emotion information from AUs and concluded that classification of emotions are done only for the six basic emotions and some of the reported results by researches are of little practical value. Azcarate, Hageloh, Sande and Valentii[8] in 2005 also researched on extracting emotion information from facial actions. Research was scarce towards combination of emotions or towards precise accuracy and robustness of the emotional interpretation but facial action unit measurements from image data received lot of attention. There are 64 main AUs depicted by Ekman and Friesen [1,2] and many others. These 64 AUs atomically or in combination

can represent a very wide variety of facial expressions. Around 7000 combinations are possible of action units. To identify these combinations M. Khademi, M.T. Manzuri-Shalmani, M.H. Kiapour and A.A. Kiae in 2010 applied a combination of HMM and Neural Networks [3]. Some researchers have tried to attack the problem of classification and interpretation by assigning one Hidden Markov Model (HMM) per AU and then passing it through a Support Vector Machine or a Neural Network to get to the final result in terms of the six basic emotions (Surprise, Sadness, Fear, Happiness, Anger and Disgust). In these approaches the effect of insignificant AUs and the not present AUs seem to be undermined due to the fact that only the AUs that are present or have significance are used to maximize HMM output probabilities. Also, the studies that have been done so far have sometimes tried to address the problem of the emotions due to multiple AU combinations but they have been done at the AU level which is painstaking and complex. To overcome the problems of considering non present and insignificant AUs, taking care of combination emotions and dynamically reducing computation cost a new model needs to be proposed.

II. THE PROPOSED MODEL

In this work we propose a three staged model to achieve a more accurate and non biased partial probability calculation, produce emotion interpretation output which is more tolerant to intensity variation of AUs, determine combinational emotions or atomic emotions depending on threshold and dynamically at run time reduce computation cost.

A. Stage 1

This stage consists of $M \times N$ HMMs where each set of $N$ HMMs are assigned to each of the $M$ AUs we intend to study. This stage finds the partial probability contributions of each emotion from the complete $M$ AUs representing one frame of input during testing and re-estimate the model parameters during training. For each of the $M$ AUs we implement $N$ HMM models $\lambda_{ij}$ where $0 \leq i \leq M$ and $1 \leq j \leq N$, each referring to one emotion ($\lambda_{i1} \rightarrow$ Anger, $\lambda_{i2} \rightarrow$ Contempt $\lambda_{i3} \rightarrow$ Disgust, $\lambda_{i4} \rightarrow$ Fear ... $\lambda_{i11} \rightarrow$ Surprise). Also, the observation symbol set $O_{ir} = \{v_{1r}, v_{2r}, ..., v_{7r}\}$ are the intensities of AU$_i$ graded from level 1 to level 7 where $1 \leq i \leq M, 0 \leq r \leq 7$ and $R$ is the total number of observable symbols per state in $\lambda_{i1}$. Actually, in the new FACS manual by Ekman and Friesen[2] the intensities of AU evidence are graded from A to E but for simplicity and to include the no presence situation of an AU we use the grading from 1 to 7 (e.g. AU1(2) is the weakest trace of AU1 and AU1(7) is the maximum intensity possible for the individual person and AU1(1) depicts that AU1 is not at all evident).

Let $V_{it} = (v_{i1t}, v_{i2t}, v_{i3t}, ..., v_{i7t})$ be the observation sequence of intensities for AU$_i$ over time 0 to $T$. Then the parameters of the proposed HMMs will be as follows,

- $S_{ij(k)}$ represents the hidden states for HMM $\lambda_{ij}$ where $1 \leq k \leq X, 0 \leq i \leq M, 1 \leq j \leq N$ and $X$ is the number of hidden states.
- $A_{ij(f,g)}$ represents the state transition matrix for HMM $\lambda_{ij}$ where $1 \leq f,g \leq X, 0 \leq i \leq M$ and $1 \leq j \leq N$.

$A_{ij(f,g)}$ is the probability of transition from previous state $S_{ij(t)}$ to the next state $S_{ij(t)}$. Thus, $A_{ij(f,g)} = P(q_t = S_{ij(g)} | q_{t-1} = S_{ij(f)}) = \text{the probability of } q_t = S_{ij(g)}$ given at time $t-1$, $q_{t-1} = S_{ij(f)}$ where $q_t$ is the state at time $t, 1 \leq t \leq T$

such that, $A_{ij(f,g)} \geq 0$, and $\sum_{g=1}^{X} A_{ij(f,g)} = 1$

- $B_{ij(d,e)}$ represents the observation symbol probability distribution. $B_{ij(d,e)} = P[V_{d} | O_{ie}]$ at time $t$ such that $q_t = S_{ij(d,e)}$ is the probability of observation symbol $O_{ie}$ for current state $q_t = S_{ij(d,e)}$ where $1 \leq d \leq X, 0 \leq e \leq R$

- $\pi_{ij(a)}$ represents the initial state distribution where $1 \leq a \leq X$ such that, $\pi_{ij(a)} = 1/X$ for all values of $a$. Due to the discrete nature of the observations it can be safely assumed that an AU observation is equally likely to start at any given state of the model and no bias needs to be considered. So using equal probabilities would be a safer approach than to use a randomized initial state distribution[10].

B. Stage 2

This stage is only active during the testing phase. Stage 2 of our model is bifurcated into two parts. The first being the Emotion Probability Classifier (EPC) block and the other block being a combination of a The Facial Action Unit Intensity Classifier (FAUIC) routine and a EPC routine. EPC routines in both blocks are functionally similar. For both of the routines two threshold parameters are predefined $\mu$ for EPC and $\theta$ for FAUIC. $\mu$ and $\theta$ are nothing but integer values depicting the number of maximum values that we will consider for generating the input for the next stage. Stage 2 calculates the final probabilities for both atomic and combination emotions in two methods and relays the output to the next and final stage.

The Emotion Probability Classifier (EPC) first finds the joint probabilities for $N$ emotions as a product of each $M$ probabilities found for each emotion$^2$. Then it sorts the set of $N$ probabilities representing each emotion. After the sorting process is over depending on threshold it selects $\mu$ maximum probabilities for the set and finds all possible combinations of the selected probabilities using multiplicative property of probabilities. It then passes the set of combined probabilities and the emotions they were derived from. For example, if the threshold is set to 2 and the top probabilities are of Happiness and Surprise then the output of EPC will be probabilities $P_1$, $P_2$ and $P_3$ representing Happiness, Surprise and Surprisingly Happy respectively. These probabilities along with the emotion(s) associated are passed on to the final stage.

The Facial Action Unit Intensity Classifier (FAUIC) first sorts the AU intensity inputs at time $t$ and selects the HMM groups to be used for the process depending on $\theta$. The value of $\theta$ is re-estimated after each run of the model in the test phase comparing the accuracy of the second block output with the first one. The FAUIC will then pass the selected HMM group outputs to its own EPC routine which in turn will calculate joint probabilities of the N emotions and then using $\mu$ find all possible emotion combinations and pass it to the final stage.

$^2$ As we consider each AU data to be streamed separately to the model, we can assume the output probabilities are statistically independent.
Stage 3

The third and final stage consists of a Maximum Likelihood Classifier block and an Emotion Index. This stage is responsible for providing the final output in terms of emotions or group of emotions.

The Maximum Likelihood Classifier (MLC) routine takes two sets of inputs from the two blocks in the previous stage and selects the maximum for each block data and passes two outputs to an emotion index and waits for acknowledgement signal from the Emotion Index routine.

The Emotion Index is under development. It will consist of a list of all valid atomic as well as combination emotions according to psychology researchers and practitioners that may arise in reality on a human face. Once the emotion index is build then it will search the list for a possible match for both the inputs from the MLC. If a match is found it will return positive signal to MLC or else a negative response will be issued and another set of data from MLC will be awaited.

Figure 1 and 2 shows the block diagram of the proposed model. Figure 1 shows stages 1 and 2. From the left the initial AU intensity input is fed into Stage 1 where the M*N HMMs decode the information in the form of partial probability contributions. Then the next stage the top EPC finds the joint probability of the respective emotions. Here the value of μ is taken to be 1 which means only atomic emotions will be considered. In the bottom block of Stage 2 the FAUC routine takes only one input as θ is 1. The max intensity is of input $V_M$ in this case and the outputs from the M*N set of HMMs are only passed to the bottom EPC block. Here again the EPC block takes on the atomic emotions as μ was specified as 1. Then both the outputs are compared and θ is re-estimated for the next run. The outputs are passed to Stage 3 shown in Figure 2. In Figure 2 the maximum probability emotions are identified and passed onto the Emotion Index when an acknowledgement is awaited for further action. Here the acknowledgement is positive so the model goes for the next set of AU intensity inputs at Stage 1.

III. IMPLEMENTATION

A. Datasets

Although there are many facial expression datasets available to researchers the Extended Cohn-Kanade Database (CK+) [6] is best suited for this research due to the emotion labels with respective AU codes. There were 327 emotion labeled observations representing the peak intensity of that subject for that emotion. Intermediate frames are in increasing order of intensity. By assuming that during the minor changes in intensity of detected AUs do not alter the final emotion depicted we got 2749 frames which were partitioned for training and testing phases as shown in Table I.

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATA DISTRIBUTION FOR TRAINING AND TESTING.</strong></td>
</tr>
<tr>
<td>Emotion</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Training</td>
</tr>
<tr>
<td>Testing</td>
</tr>
<tr>
<td>Sub-Total</td>
</tr>
</tbody>
</table>

B. Method of Training and Testing

After the data is segmented into training and test sets the training data is classified according to the emotions depicted and the corresponding AUs. Suppose in one frame AU1 and AU2 is dominant with intensities 4 and 5 and the labeled emotion is Y. We then train HMMs $\lambda_{1Y}$ and $\lambda_{2Y}$ with the AU intensity data and the other HMMs of the emotion Y are trained with zero intensity data for all other AUs thus taking care of parameter re-estimation for non present AUs. The Baum-Welch method of parameter re-estimation [9, 11] is used for training which readily provides us with the parameter updating equations based on observation sequences. After training the HMMs are ready for the emotion interpretation task. The M*N HMMs put into our model will render M probability outputs for each type of the N emotions considered [Figure 1]. The probability estimations for observation sequence $V_i$ will
be \(P[V_i | \lambda_j]\) where \(0 \leq i \leq M\) and \(1 \leq j \leq N\). This probability estimation is done using the Forward Backward procedure of probability estimation explained by Rabiner[9].

IV. RESULTS

After successful training we found that the results for single emotions (\(\theta = 1\)) were better than that obtained by Cohen, Garg[7]. The overall emotion recognition rate was slightly higher than Velusamy, Kannan[5]. The results are shown in Table 2. The overall average value of around 93 percent is quite high and reliable for this field of research which readily shows the efficiency of the proposed model.

Table 3 show the results for (\(\theta = 2\)). Partial results are shown due to page constraints. The table shows the top two combinations for each emotion category with respective values of percentage occurrence during the complete testing run. Although, we can infer interesting psychological phenomenon from these combination table, the ground truth data is available only for atomic emotion and hence the combination emotion labels cannot yet be verified.

Table 3 show the results for (\(\theta = 2\)). Partial results are shown due to page constraints. The table shows the top two combinations for each emotion category with respective values of percentage occurrence during the complete testing run. Although, we can infer interesting psychological phenomenon from these combination table, the ground truth data is available only for atomic emotion and hence the combination emotion labels cannot yet be verified.

Table 2

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Success Rate</th>
<th>Contempt</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happy</th>
<th>Sadness</th>
<th>Surprise</th>
<th>Overall Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Success</td>
<td>97.69</td>
<td>82.00</td>
<td>96.9</td>
<td>90.5</td>
<td>91.2</td>
<td>96.2</td>
<td>1</td>
<td>93.99</td>
</tr>
</tbody>
</table>

So combinational emotions is a must in understanding emotional state from a facial expression. Our model is well suited for that purpose. An interesting observation from Table 3 is that Sadness and Disgust are very closely associated. Psychological phenomenon can be inferred if ground truth data is available for combination emotions, the way it is available for atomic ones.

Table 3

<table>
<thead>
<tr>
<th>Ground Truth</th>
<th>Combination 1</th>
<th>% Occurrence</th>
<th>Combination 2</th>
<th>% Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>Anger, Disgust</td>
<td>67.6</td>
<td>Anger, Contempt</td>
<td>21.3</td>
</tr>
<tr>
<td>Contempt</td>
<td>Contempt, Disgust</td>
<td>58.2</td>
<td>Contempt, Fear</td>
<td>22.4</td>
</tr>
<tr>
<td>Disgust</td>
<td>Disgust, Contempt</td>
<td>44.8</td>
<td>Disgust, Anger</td>
<td>33.8</td>
</tr>
<tr>
<td>Fear</td>
<td>Fear, Anger</td>
<td>73.5</td>
<td>Fear, Surprise</td>
<td>14.2</td>
</tr>
<tr>
<td>Happy</td>
<td>Happy, Surprise</td>
<td>59</td>
<td>Happy, Contempt</td>
<td>34.1</td>
</tr>
<tr>
<td>Sadness</td>
<td>Sadness, Disgust</td>
<td>82.9</td>
<td>Sadness, Contempt</td>
<td>9.6</td>
</tr>
<tr>
<td>Surprise</td>
<td>Surprise, Fear</td>
<td>61.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As we earlier that we used a optimization variable \(\theta\) to optimize at runtime the computation cost as well as maximize accuracy. Figure 3 is a graphical representation of the variation of \(\theta\) and the corresponding computation reduction percentage at runtime and the percentage accuracy. We can observe from the graph that at values of \(\theta\) above 13 the accuracy with respect to the un-optimized results become steady at nearly 99 percent with a computation reduction of above 75 percent. This is quite good result for such a computation expensive task.

V. CONCLUSION

Using N HMMs per AU the model estimates of the emotion probabilities are more precise due to the contribution of all AUs irrespective of their significance level. In general, human emotion is rarely ‘pure’ e.g. 100 percent happiness[4] which means the probability of encountering a blend or combination is highly likely.

Figure 3. The effect of the value of \(\theta\) on the runtime and accuracy.

REFERENCES