Facial Emotion Recognition from Kinect Data – An Appraisal of *Kinect Face Tracking Library*

Tanwi Mallick, Palash Goyal, Partha Pratim Das and Arun Kumar Majumdar
Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur, 721302, India
{*tanwimallick, palash1992*}*@gmail.com*, {*ppd, akmj*}*@cse.iitkgp.ernet.in*

Keywords: Facial Expression Recognition, Emotion Recognition, Kinect Face Tracking Library (KFTL), Facial Action Coding System (FACS), Action Units (AU), Artificial Neural Network (ANN).

Abstract: Facial expression classification and emotion recognition from gray-scale or colour images or videos have been extensively explored over the last two decades. In this paper we address the emotion recognition problem using Kinect 1.0 data and the Kinect Face Tracking Library (KFTL). A generative approach based on facial muscle movements is used to classify emotions. We detect various *Action Unit*s (AUs) of the face from the feature points extracted by KFTL and then recognize emotions by *Artificial Neural Networks* (ANNs) based on the detected AUs. We use six emotions, namely, *Happiness*, *Sadness*, *Fear*, *Anger*, *Surprise* and *Neutral* for our work and appraise the strengths and weaknesses of KFTL in terms of feature extraction, AU computations, and emotion detection. We compare our work with earlier studies on emotion recognition from Kinect 1.0 data.

1 INTRODUCTION

Facial emotions represent an important aspect of human communication, particularly in conveying the mental state of an individual. The ability to automatically recognize facial emotion is useful for Human-Computer Interaction (HCI). Potential applications include (Kołakowska et al., 2013) video games, educational software, auto-mobile safety, mental health monitoring and many others.

In order to solve this problem, researchers use different techniques to detect face and then extract features of various expressions that build emotions. These techniques (Section 2, Table 1) are mainly based on the color, shape and motion of the face and facial points like eyes, eye-brows, nose, cheek, and lips. It has been customary to use gray-scale and colour intensity information in images and videos to recognize facial emotions. Though depth data have been available for nearly a decade, there is no reported work on this till 2013.

In this paper we use Kinect[1] 1.0 for recognizing emotions. Besides capturing depth, it also provides the *Kinect Face Tracking Library* (Microsoft, 2014) (KFTL) with capabilities for basic feature extraction and tracking of faces. In 2013, some work ((Youssef et al., 2013), (Wyrembelski, 2013), (Nelson, 2013)) have been reported on emotion recognition from Kinect data. We build up on these with the specific target of appraising the performance of KFTL for effective use in emotion recognition.

We capture facial emotion images by Kinect 1.0 and use KFTL to detect and track the face, and to extract its basic features (like face points). We then compute various Action Units (AUs) of the face. Finally we use the well-accepted Candide-3 FACS (Linköping, 2012) model to recognize emotions by *Artificial Neural Networks* (ANNs). We use six emotions – *Happiness*, *Sadness*, *Fear*, *Anger*, *Surprise* and *Neutral* for our work and appraise the strengths and weaknesses of the KFTL in terms of feature extraction, AU computations, and emotion detection.

The remainder of the paper is organized as follows. Section 2 gives an overview of the state of the art techniques for recognizing emotions. In Section 3 we describe the Candide-3 FACS (Linköping, 2012) emotion model. A brief description of Kinect Face Tracking Library is given in Section 4. Section 5 describes the architecture of the emotion recognition system. Then in Sections 6 and 7 we discuss its stages in depth. We present the results in Section 8. Finally we conclude in Section 9.

[1] *Kinect for XBox One* has been released a while after this work was completed. This is called Kinect 2.0 now.

Mallick, T., Goyal, P., Das, P. and Majumdar, A.
Facial Emotion Recognition from Kinect Data âĂŞ An Appraisal of *Kinect Face Tracking Library*.
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP*, pages 525-532
ISBN: 978-989-758-175-5

Table 1: Chronological Survey of Work in Emotion Recognition.

Ref. (Year)	Input	Output	Remarks
(Essa and Pentland, 1997)	Gray-Scale	**(UE)** w/o Fear w/ Raise Brow	Features are extracted by motion-based dynamic model using optical flow. Classification is done using 2D motion energy model.
(Yoneyama et al., 1997)	Gray-Scale	**(UE)** w/o Fear & Disgust	Optical flow are measured in partitions of face and then a discrete Hopfield NN is used for classification.
(Black and Yacoob, 1995)	Gray-Scale	**(UE)**	Local parametric motion feature are used for classification.
(Cohn et al., 1998)	Gray-Scale	**(UE)**	Optical flow for selected facial points are used as features.
(Yang et al., 1999)	RGB	**(UE)**	Fuzzy rule-based classifier is used to recognize emotions from deformation of 18 feature points wrt neutral face.
(Tsapatsoulis and Piat, 2000)	MPEG-4 Video	**(UE)** w/o Fear w/ Raise Brow	Scale invariant distances of 15 salient points are detected on face for each emotion and a fuzzy classifier is used for classification.
(Tian et al., 2001)	RGB	Action Units	Tracks feature points on face based color, shape and motion to compute AUs. These can be used for emotion analysis.
(Cohen et al., 2003)	Video	**(UE)** w/ Neutral	Uses wire-frame model and facial motion measurements for feature extraction. HMM and Bayesian classifiers are used for classification.
(Kim and Bien, 2003)	RGB	**(UE)**	Skin-colour segmentation and T-based template matching are used for detecting face and extracting features. A fuzzy NN is used for classification.
(Kim et al., 2005)	RGB	**(UE)** w/o Fear	Features are extracted by using fuzzy color filter, virtual face model and histogram analysis. Then fuzzy classifier is used for classification.
(Pantic and Patras, 2006)	Video	27 Action Units	Shape and location-based feature points are tracked using particle filters. It can handle temporal dynamics of AUs.
(Youssef et al., 2013)	Kinect Data	**(UE)** w/ Neutral	Based on 121 3D points and their deformation, emotions are classified by SVM and k-NN classifiers.
(Wyrembelski, 2013)	Kinect Data	**(UE)** w/o Fear & Disgust	Based on AUs from KFTP (Microsoft, 2014) the emotions are classified by k-NN classifiers.
(Nelson, 2013)	Kinect Data	Assorted Emotions	Based on AUs from KFTP (Microsoft, 2014) six emotions – Surprise, Sad, Kissing, Smiling, and Anger w/ Mouth open & closed – are classified by decision tree.

Set of **Universal Emotions (UE)** *(Hung et al., 1996) include – Fear, Surprise, Anger, Disgust, Happiness, and Sadness. In addition, Smile and Raise Brow are used by some researchers.*

2 SURVEY OF FACIAL EMOTION RECOGNITION

Though there have been sporadic interests in modelling facial emotions (Ekman and Friesen, 1978) and their constituent expressions from the 1970's, serious research in computer analysis and synthesis of facial emotions started in the mid-1990's. Initially it used gray-scale images and later grew with colour images and video sequences. Recently some researchers ((Youssef et al., 2013), (Wyrembelski, 2013), (Nelson, 2013)), like ours, have attempted to use depth and RGB data from Kinect data for this purpose.

Based on the set of features the work in emotion recognition has been broadly divided into three categories (Wu et al., 2012). We briefly review these below and in Table 1 present a chronological survey of some representative work in this area since 1995.

- *Deformation Features*: In these methods ((Yang et al., 1999), (Youssef et al., 2013), (Wyrembelski, 2013), (Nelson, 2013)) some facial deformation information like *Geometric deformation* or *Texture changes* caused by the changing expressions are extracted. The techniques based on Candide-3 FACS Model (Linköping, 2012) fall in this category as Action Units are estimated based on deformations from the neutral face under an emotion.
- *Motion Features*: These methods ((Essa and Pentland, 1997), (Yoneyama et al., 1997), (Black and Yacoob, 1995), (Cohn et al., 1998), (Tian et al., 2001), (Cohen et al., 2003)) use sequential expression images and extract some feature points or motion information from the regions of the features. Common methods include: *Feature point tracking* and *Optical flow*.
- *Statistical Features*: In these methods ((Tsapatsoulis and Piat, 2000), (Pantic and Patras, 2006), (Kim and Bien, 2003), (Kim et al., 2005)) the characteristics of emotion are described by typical statistics – *Histogram* or *Moment Invariant*.

The classification algorithms are chosen based on the extracted features. These include *Artificial Neural Network (ANN)*, *Support Vector Machine (SVM)*, *k-Nearest-Neighbor (kNN)*, *Bayes' Classifiers*, *Fuzzy Rule-Based Classifier*, *Fuzzy Neural Network (FNN)*, *Hidden Markov Model (HMM)*, *Spatial and Temporal Motion Energy Templates Methods*. However, in recent years, HMM, ANN, Bayesian classification, and SVM have become the mainstream methods for facial emotion recognition.

In the next section we review the work with Kinect in depth.

2.1 Facial Emotion Recognition from Kinect Data

Recently some studies / systems have been reported on facial expression recognition that use Kinect depth (as well as RGB) data.

Youssef et. al. (Youssef et al., 2013) use Kinect depth video with SVM & *k*NN for detecting Autism Spectrum Disorders (ASDs) in children. They consider six universal emotions and report the best recognition rate of 39% with SVM. In (Wyrembelski, 2013), Wyrembelski report an Emotion Recognition System using Kinect data with *k*NN classifier. The AUs from KFTL are used here as features. Nelson, in her thesis (Nelson, 2013), presents an emotion recognition system for six emotions. Unlike the usual practice of using the universal emotions, Nelson uses a different set – *Surprise*, *Sadness*, *Kissing*, *Smiling*, *Anger with mouth open*, and *Anger with mouth closed*. Again the AUs from KFTL are used as features and the classification is done by a decision tree. However, no data on test or accuracy is reported in (Wyrembelski, 2013) and (Nelson, 2013).

We use AUs in this work. So our approach belongs to *Deformation Features* category. We use KFTL for early processing, and ANNs to recognize AUs and finally the emotions. Before discussing our approach, we briefly present the FACS model and KFTL in the next two sections.

3 FACIAL EMOTION MODELLING

The formations and transitions of facial expressions and the ensuing emotions were first encoded in (Ekman and Friesen, 1978) by Ekman and Friesen in 1978. Realizing that facial expressions are resultant of combined contractions and relaxations of various facial muscles, they worked on a system to systematically encode the same and relate muscles to movements (Table 2). Since the muscle movements behind every expression gets too detailed, they defined *Action Units* (**AU**s) as combinations of groups of muscles (Figure 1) that cause constituent movement behaviour for various expressions. They called it the *Facial Action Coding System* (FACS) (Ekman and Friesen, 1978). It has now become the de-facto standard (Mellon University, 2015) in describing facial behaviours.

Table 2: Action units in terms of Facial muscles.

AU	Description	Facial muscle	Example image
1	Inner Brow Raiser	Frontalis, pars medialis	
2	Outer Brow Raiser	Outer Brow Raiser	
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii	
20	Lip stretcher	Risorius with platysma	

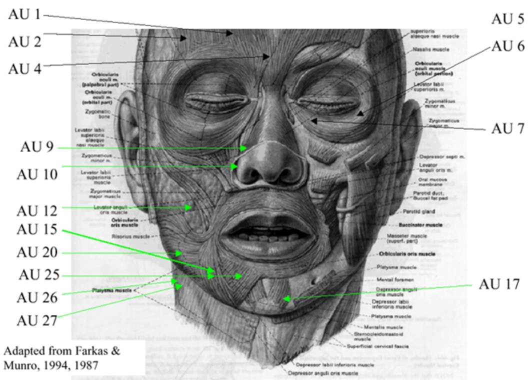


Figure 1: Selected FACS Action Units.

FACS encodes an emotion as a combination of AUs. For example, *happiness* is described by AU6 (*Cheek Raiser*) and AU12 (*Lip Corner Puller*). Duration, intensity, and asymmetry also add to the formation of emotions. In Table 3 we list the AU compositions for emotions that we consider in this paper.

FACS finds extensive use in synthesizing emotions while various computer animation techniques for ascribing emotions to avatars are worked out. The most accepted form of FACS is implemented as Candide-3 model (Linköping, 2012) where there are 44 AUs. Like many RGB-based techniques mentioned above, we also use the AUs defined in Candide-3 to model the face and to recognize emotions in terms of AUs. Unless otherwise mentioned the AU numbers in this paper refer to the common FACS scheme (Mellon University, 2015).

Table 3: Composition of Emotions in terms of Action Units.

Emotion	Component AUs
Happiness	AU6, AU12
Surprise	AU1, AU2, AU5, AU10
Sadness	AU1, AU4, AU15
Fear	AU1, AU2, AU4, AU5, AU10, AU20
Anger	AU4, AU5, AU23
Neutral	AU1, AU2, AU4, AU5, AU6, AU10, AU12, AU23, AU26

This table is derived from the emotion coding in FACS (Mellon University, 2015) and uses the same AU numbers.

4 KINECT FACE TRACKING LIBRARY (KFTL)

We intend to use KFTL[2] (Microsoft, 2014) to detect and track a face, and extract its features. KFTL takes depth and RGB frames as input and tracks the human face to compute the following:

1. *Tracking Status*: It outputs the status to indicate if face tracking is successful or if it has failed and its reason.
2. *2D Tracked Points*: It tracks the 100 2D points on the face including a bounding rectangle around the head. These points are returned in an array and are defined in the coordinate space of the RGB image (in 640 x 480 resolution) returned from the Kinect sensor. These are used as *Feature Points* (Figure 2).
3. *3D Head Pose*: It captures the pose of the head by three angles – *Pitch*, *Roll*, and *Yaw*. These represent the 3D orientation of the head. The tracking works when the pitch, roll, and yaw angles are less than 20^o, 90^o, and 45^o respectively.
4. *Action Units (AUs)*: It outputs the weights of 6 AU[3]s from Candide-3 model (Linköping, 2012).
5. *Shape Units (SUs)*: It return 11 SU[4]s that are related to the Candide-3 model (Linköping, 2012).

[2]Kinect Windows SDK version 1.7

[3]In SDK these are referred to as *Animation Units*. These are deltas from the neutral shape that can be used to morph targets on animated avatar models so that the avatar acts as the tracked user does.

[4]The SUs estimate the particular shape of the user's head: the neutral position of their mouth, brows, eyes, and so on.

5 EMOTION RECOGNITION SYSTEM

In this paper, we describe a multi-stage emotion recognition system for near-frontal faces. The architecture of the system is shown in Figure 2. An automated emotion recognition system first needs to extract facial features and then recognise emotions in terms of these features. Hence we architect our system in three stages:

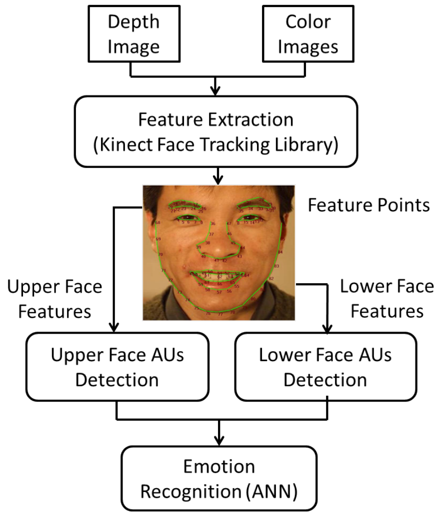


Figure 2: Architecture of emotion recognition system.

1. *Feature Extraction using KFTL*: KFTL takes depth and RGB images as input and tracks 100 2D points on the human face. Various features like eye-width, eye-height, distance between lips etc. are extracted from these tracked points and / or taken from SUs (Section 4). These (local) features are used in the next stage of the system for detection of AUs.
2. *Detection of Action Units*: AUs are defined in terms of local features of the face. Hence we use the extracted features to detect various AUs. This is done by Decision Algorithms and ANN classifiers as discussed in Section 6.
3. *Emotion Recognition*: Finally, we use Candide-3 emotion model to recognise target emotions from AUs detected above. This is done by ANN classifiers and is detailed in Section 7.

6 DETECTION OF ACTION UNITS

Initially we attempt to use the Kinect AUs (Section 4) as detected by KFTL for our system. Hence we perform experiments to evaluate them. The AUs as detected by KFTL for our data set, are tabulated in Table 4 as fraction of total samples. We note that for most AUs the detection rate is rather poor. No AU, with the exception of Neutral, is detected with even 40% accuracy and therefore cannot be used to reliably recognize emotions. Further, KFTL detects only 6 AUs. Hence we fail to build our system using Kinect AUs.

Table 4: Confusion Matrix for Detection of Kinect AUs using SDK.

	Detected AUs					
Actual AUs	**Neutral**	**AU0**	**AU1**	**AU3**	**AU4**	**AU5**
Neutral	**0.67**	0.10	0.23	0.00	0.00	0.00
AU0 (AU10)	0.14	**0.36**	0.11	0.07	0.20	0.12
AU1 (AU26/27)	0.17	0.05	**0.21**	0.12	0.18	<u>0.27</u>
AU3 (AU4)	0.02	0.08	<u>0.30</u>	**0.13**	0.09	0.20
AU4 (AU13/15)	0.28	0.09	0.14	0.08	**0.34**	0.07
AU5 (AU2)	0.14	0.13	0.18	0.11	0.18	**0.26**

AU numbers shown here are from KFTL. The corresponding AU numbers from Candide-3 model are shown within parentheses. In KFTL, Neutral face is detected when all AUs are 0. AU2 (AU20 - Lip Stretcher) has not been considered in the experiment.

Next we focus to design our own algorithms to detect AUs from KFTL feature points. Fortunately, the 2D tracked or feature points from KFTL are found to be moderately accurate. Various features like eye-width, eye-height, distance between lips etc. can be reliably extracted from these tracked points. As some of the features like eye-height and lip-width differ across people, we normalize these features by the features of the neutral face. These features are then used along with the annotated frames in the detection process.

We also notice that the AUs in the upper and lower face are relatively independent and they behave in distinct ways. Hence, we use different algorithms for the detection of upper and lower face AUs. These algorithms are discussed in the next two sections.

6.0.1 Detection of Upper Face AUs

Each upper face AU (like AU1, AU2, AU4, AU5, and AU6) is a monotonic function of a single feature. For example, *upper lid raiser* is generally accompanied by *brow raiser* which leads to giving *brow* to *lower eye lid* distance a reasonably high weight. Hence, we use a function X_{MAX} (x value for which y is maximum) as a monotonic functions for recognising the upper face AUs. The features, as used, are shown in Table 5.

Table 5: Features for Upper Face AUs.

Action Unit	Feature	Lower limit	Upper limit
AU1 (Inner brow raiser)	Inner brow to eye	1.000	1.270
AU2 (Outer brow raiser)	Outer brow to eye	1.000	1.220
AU4 (Brow lowerer)	(Inner brow to eye)$^{-1}$	1.000	1.200
AU5 (Upper lid raiser)	Eye height	1.000	1.235
AU6 (Cheek raiser)	(Eye height)$^{-1}$	1.000	1.950

This table uses AU numbers from FACS (Mellon University, 2015)

6.0.2 Detection of Lower Face AUs

Lower face AUs (like AU12, AU15, AU20, and AU23) are non-monotonic in terms of the feature points. Hence we use an ANN to recognize 4 lower face AUs. The input and output layers for the ANN are set as:

- *Input layer*:
  - Lip width / Neutral lip width,
  - Lip height / Neutral lip height, and
  - Lip angle (Angle between lines joining end points of lip to mid-point of lower lip) / Neutral lip angle.
- *Output layer*:
  - AU12 (Lip Corner Puller),
  - AU15 (Lip Corner Depressor),
  - AU20 (Lip Stretcher), and
  - AU23 (Lip Tightener)
- *Hidden layer*: 1 hidden layer with 6 neurons
- *Steepness of activation function*: 0.01

Results for the detection of AUs are given in Tables 6 and 7 in Section 8.2.

7 EMOTION RECOGNITION

After extracting the AUs, we use ANNs to learn Emotions from AUs in the input layer. We do this in two ways:

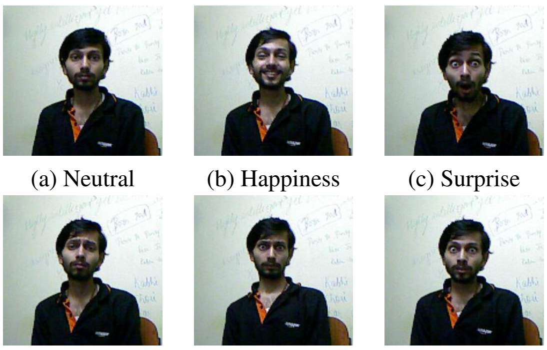
1. *Multiple Neural Networks:* We build 6 ANNs, one for each emotion having suitable AUs in the input layer. For example for *Happiness* the ANN is:
   - *Input layer*:
     - AU6 (Cheek raiser)
     - AU12 (Lip Corner Puller)
   - *Output layer*:
     - Happiness
   - *Hidden layer*: 1 hidden layer with 6 neurons
   - *Steepness of activation function*: 0.01

This approach has the advantage that unnecessary weights to some of the AUs is avoided.

2. *Single Neural Networks:* We build a single ANN with the input layer consisting of all the AUs and the output layer containing all the emotions. we use 1 hidden layer with 6 neurons and steepness of 0.01 for activation function. This approach has the advantage that it removes the subjectivity. It may be difficult to predict the dependence between individual emotions and the AUs making this approach more robust.

8 EXPERIMENTS AND RESULTS

We investigate the effectiveness of our framework using six emotions: *Happiness*, *Sadness*, *Surprise*, *Fear*, *Anger*, and *Neutral*. Figure 3 shows sample images as captured by Kinect.



(a) Neutral (b) Happiness (c) Surprise

(d) Sadness (e) Fear (f) Anger

Figure 3: Six Emotions used in our work.

8.1 Data Set

There is no benchmark available for facial expression classification experiments based on depth data of faces. Hence, we first generated a data set that can be used for training as well as testing. 25 volunteers participated in data generation. About 4 minutes' video was recorded for each volunteer. This comprised about 6000 to 7000 image frames. 10 out of the 25 volunteers are drama actors and each of them enacted the target emotions twice. In addition, all volunteers were asked to perform some of the Action Units. Nearly equal number of frames of all the emotions were obtained to avoid any bias.

8.2 Results

We use 80% of the data to train the system and 20% to test for detection of AUs as well as emotion recognition. An open-source library FANN (Fast Artificial Neural Network) (Nissen, 2014) is used for learning the ANNs. First, we detect the upper and lower face AUs from the facial features. Tables 6 and 7 show the results of the AUs detection. After extracting the AUs, we use the data to recognize emotion using multiple and single ANNs. The results are shown in Tables 8 and 9. Bold entries along the diagonal of the table show correct recognition rate while underlined entries off-diagonal show misclassification rate.

Table 6: Confusion Matrix for Detection of Upper-Face AUs.

	AU0	AU1	AU2	AU4	AU5	AU6
AU0	**0.67**	0.10	0.23	0.00	0.00	0.00
AU1	0.14	**0.86**	0.00	0.00	0.00	0.00
AU2	0.17	0.00	**0.83**	0.00	0.00	0.00
AU4	0.26	0.00	0.00	**0.74**	0.00	0.00
AU5	0.14	0.00	0.00	0.00	**0.86**	0.00
AU6	0.63	0.00	0.00	0.00	0.00	**0.37**

This table uses AU numbers from FACS (Mellon University, 2015). AU0 denotes Neutral Face. Note that AU6 (Cheek raiser) is grossly misclassified as Neutral face.

Table 7: Confusion Matrix for Detection of Lower-Face AUs.

	AU0	AU12	AU15	AU20	AU23
AU0	**0.49**	0.04	0.44	0.03	0.00
AU12	0.02	**0.76**	0.01	0.21	0.00
AU15	0.50	0.02	**0.46**	0.02	0.00
AU20	0.01	0.16	0.03	**0.80**	0.00
AU23	0.00	0.00	0.00	0.00	**1.00**

This table uses AU numbers from FACS (Mellon University, 2015). AU0 denotes Neutral Face. Note that AU15 (Lip Corner Depressor) is often confused with Neutral face.

Table 8: Confusion Matrix for Emotion Recognition using Multiple ANNs.

	Neutral	Happiness	Surprise	Sadness	Fear	Anger
Neutral	**0.76**	0.00	0.00	0.15	0.07	0.02
Happiness	0.30	**0.20**	0.01	0.31	0.08	0.10
Surprise	0.01	0.00	**0.88**	0.01	0.08	0.02
Sadness	0.10	0.04	0.00	**0.84**	0.01	0.01
Fear	0.10	0.00	0.06	0.54	**0.20**	0.10
Anger	0.03	0.02	0.10	0.02	0.12	**0.71**

Note that Fear is often misclassified as Sadness and Happiness is confused with Sadness or Neutral emotion. These are due to the weak discrimination of AU6 (Table 6) and AU15 (Table 7).

Table 9: Confusion Matrix for Emotion Recognition using a Single ANN.

	Neu-tral	Happi-ness	Sur-prise	Sad-ness	Fear	Anger
Neutral	**0.64**	0.12	0.00	0.15	0.07	0.02
Happiness	0.25	**0.40**	0.01	0.21	0.08	0.05
Surprise	0.03	0.04	**0.81**	0.01	0.08	0.02
Sadness	0.10	0.06	0.00	**0.82**	0.01	0.01
Fear	0.14	0.00	0.01	0.59	**0.14**	0.12
Anger	0.02	0.01	0.10	0.02	0.12	**0.74**

The recognition of Happiness improves over the multiple ANN model (Table 8), but Fear still behaves poorly.

8.3 Discussion

Using *Multiple ANN* approach, *Fear* is frequently misclassified as *Sadness*, *Happiness* as *Neutral* or *Sadness* (Tables 8). The use of *Single ANN* does not significantly increase the overall accuracy of the system, although the happiness categorization improves somewhat.

The recognition accuracies of *Happiness* and *Fear* are not satisfactory. Figures 3(b) and 3(d) are both being recognized as *Sad*. The reason for inaccurate recognition of *Happiness* is the inaccuracy in Cheek Raiser detection. As for detection of *Fear*, the lip movements involved in fear are not the typical *lip stretch* (AU20). Furthermore, Kinect 1.0 is unable to track the lip points corresponding to *Fear*. It loses track of end points which leads to erroneous output. The reader might think that variation in eye features should be enough to detect *Fear* but the variations are subtle which Kinect 1.0 is unable to capture.

We analyze the 2D points as extracted by KFTL (and the corresponding RGB image, for visual understanding) to get insight of why *Fear* and *Happiness* are not properly recognized. A close look at the data reveals that the lip points are often falsely tracked, that is, Kinect is unable to track the lip points when the emotional state of the person is *Fear*. Also we see that the eyes are not tracked properly. This is due to the significant distortion of lips and eyes. With this the *Active Appearance Model* (AAM) that Kinect uses, is unable to form a suitable mesh. Improving the resolution or using another algorithm for tracking points may solve the problem.

9 CONCLUSION

We present a facial emotion recognition system using Kinect 1.0 data and the KFTL. We use the Candide-3 FACS model (Linköping, 2012) for this work and achieve the tasks in three stages: Feature extraction, AUs detection and Emotion recognition. We use KFTL for feature extraction. Next we detect the AUs. For this we first tried to use KFTL but failed. So we define AUs in terms of local features of the face and develop algorithms to detect AUs separately in upper and lower face regions. We attempt to detect 10 AUs of which 8 are detected accurately. But the detections of AU6 (Cheek Raiser) & AU15 (Lip Corner Depressor) have been poor. Finally, we use the detected AUs to recognise six emotions. This is done by multiple as well as single ANN classifiers. Single ANN behaves better, though the recognition rates of *Fear* and *Happiness* are unsatisfactory.

In the course of this work we observe the following characteristics of the KFTL:

- As such KFTL is inadequate for emotion recognition as it detects only 6 AUs. Also, some of the AUs as detected by KFTL are unstable.
- 2D Points as detected are accurate and stable.
- Some other extracted parameters are unstable.
- Some typical key information (like iris) are not available.

The performance of our system compares favourably with earlier work with Kinect data. Compared to the 39% accuracy reported by Youssef et. al. (Youssef et al., 2013), our system achieves an accuracy of 40% or more for five out of six emotions (excluding *Fear*), and over 64% for four of them (further excluding *Happiness*). While we also use KFTL as in (Wyrembelski, 2013) and (Nelson, 2013), our own detectors for AUs and ANN-based classifiers perform better than other methods. Interestingly, in (Nelson, 2013), Nelson suggests[5] that more specific bounds for facial expressions and neural networks be used for improvement of KFTL. Our work with detection of AUs (specifically in lower face) already implements this and experimentally supports this observation.

We thus conclude that better accuracy of emotion recognition cannot be achieved with the current Kinect library. Specifically, the tracking of lip points and the lowering of eyes – critical for the discrimination of *Fear* from *Sadness* and for characterization of *Happiness* – need separate processing. Thus, in future, we intend to develop separate iris and lip corner detectors to improve the emotion recognition performance.

[5]http://themusegarden.wordpress.com/2013/05/16/kinect-face-tracking-results-thesis-update-4/

ACKNOWLEDGEMENT

The authors acknowledge the TCS Research Scholar Program for financial support.

REFERENCES

Black, M. J. and Yacoob, Y. (1995). Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Conf. on Computer Vision*, pages 374–381.

Cohen, I., Sebe, N., Garg, A., Lew, M. S., and Huang, T. S. (2003). Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding*, 91.

Cohn, J. F., Zlochower, A. J., Lien, J. J., and Kanade, T. (1998). Feature-point tracking by optical flow discriminates subtle differences in facial expression. *International Conference on Automatic Face and Gesture Recognition*, pages 396–401.

Ekman, P. and Friesen, W. (1978). Facial action coding system: A technique for measurement of facial movement. *Palo Alto, CA.: Consulting Psychologists Press*.

Essa, I. A. and Pentland, A. P. (1997). Coding, analysis, interpretation, recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:757–763.

Hung, D., Kim, H., and Huang, S. (1996). Modeling six Universal Emotions. in *Human Facial Modeling Project* at Cornell University. http://www.nbb.cornell.edu/neurobio/land/oldstudent projects/cs490-95to96/hjkim/emotions.html. Last accessed on 27-Sep-2015.

Kim, D. and Bien, Z. (2003). Fuzzy neural networks (fnn)-based approach for personlized facial expression recognition novel feature selection method. *IEEE International Conference on Fuzzy Systems*, 2:908–913.

Kim, M. H., Joo, Y. H., and Park, J. B. (2005). Emotion detection algorithm using frontal face image. In *Computer Applications in Shipbuilding (ICCAS 2005), 12th International Conference on*, pages 2373–78.

Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., and Wróbel, M. R. (2013). Emotion recognition and its application in software engineering. In *Human System Interaction, 6th International Conference on*.

Linköping, U. (2012). Candide3 Model. http://www.icg.isy.liu.se/candide/main.html. Last updated on 24-May-2012. Last accessed on 27-Sep-2015.

Microsoft (2014). Face Tracking SDK. http://msdn. microsoft.com/en-us/library/jj130970.aspx. Last accessed on 27-Sep-2015.

Nelson, A. (2013). Facial expression analysis with Kinect. http://themusegarden.wordpress.com/2013/02/02/ facial-expression-analysis-with-kinect-thesis-update-1/ and the linked updates. Last accessed on 27-Sep-2015.

Nissen, S. (2014). Fast Artificial Neural Network. http://leenissen.dk/fann/wp/. Last accessed on 27-Sep-2015.

Pantic, M. and Patras, I. (2006). Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Systems, Man, and Cybernetics - Part B: Cybernetics*, 36:433–449.

Tian, Y., Kanade, T., and Cohn, J. F. (2001). Recognizing action units for facial expression analysis. *IEEE transactions on pattern analysis and machine intelligence*, 23(2).

Tsapatsoulis, N. and Piat, F. (2000). Exploring the time course of facial expressions with a fuzzy system. *National Technical University of Athens*.

Mellon University, C. (2015). Facial Action Coding System. http://www.cs.cmu.edu/ face/facs.htm. Last updated on 24-Apr-2014. Last accessed on 27-Sep-2015.

Wu, T., Fu, S., and Yang, G. (2012). Survey of the facial expression recognition research. *Advances in Brain Inspired Cognitive Systems, Lecture Notes in Computer Science*, 7366.

Wyrembelski, A. (2013). Detection of the selected, basic emotions based on face expression using Kinect. Unpublished Report. http://stc.fs.cvut.cz/pdf13/2659.pdf. Last accessed on 27-Sep-2015.

Yang, D., Kunihiro, T., Shimoda, H., and Yoshikawa., H. (1999). A study of realtime image processing method for treating human emotion by facial expression. *International Conference on System, Man and Cybernetics (SMC99)*.

Yoneyama, M., Iwano, Y., Ohtake, A., and Shirai, K. (1997). Facial expressions recognition using discrete Hopfield neural networks. *International Conference on Information Processing*, 3:117–120.

Youssef, A. E., Aly, S. F., Ibrahim, A. S., and Abbott, A. L. (2013). Auto-optimized multimodal expression recognition framework using 3d kinect data for asd therapeutic aid. *International Journal of Modeling and Optimization*, 3.