Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network

Vineetha G R, Sreeji C
Department of Computer Science & Engineering
Sree Buddha College Of Engineering, Pattor ,
Alappuzha, India
chachu.gr@gmail.com,mail2sree1@gmail.com

Lentin Joseph
ASIMOV Robotics Pvt Ltd, XII/82A
Kochi, India
lentin007@gmail.com

Abstract**—** Face expression detection is very common in computer-human interface. So by using the sensor Microsoft Kinect we are detecting the face expression with the help of Artificial Neural Network. Here we are using Supervised Learning to train the network. By Microsoft Kinect sensor it will detect the corresponding expression. In this paper we present a method for face expression distinctly by using MS Kinect in 3D from the input image. Kinect features us by providing the depth information of the foreground objects. We present our experience of using a Kinect depth camera for detection of some common face expressions. Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs.Microsoft has developed Software Development Kit(SDK) free to use for non commercial purposes. The SDK includes not only drivers but also APIs, device interfaces, installer documents and resource materials. It's another exciting milestone for a technology that has captured the imagination of millions, and has become the fastest selling computer electronics device of all time.

Keywords—Face expression, MS Kinect, Supervised Learning, Depth values, human computer interface, Software Development Kit.

1. INTRODUCTION

Human face recognition have been conducted vigorously (Fasel &Luettin, 2003; Yang et al., 2002; Pantic & Rothkrantz, 2000a; Zhao et al., 2000; Hasegawaetal., 1997; Akamatsu, 1997). Such studies are aimed at the implementation of an intelligent man-machine interface. Especially, studies of facial expression recognition for human machine emotional communication are attracting attention (Fasel & Luettin, 2003; Pantic & Rothkrantz, 2000a; Tian et al., 2001; Pantic & Rothkrantz, 2000b; Lyons et al., 1999; Lyons etal., 1998; Zhang et al., 1998).The shape (static diversity) and motion (dynamic diversity) of facial components such as the eyebrows, eyes, nose, and mouth manifest expressions. Considering facial expressions from the perspective of static diversity because facial configurations different among people, it is presumed that a facial expression pattern appearing on a face when facial expression is manifested includes person-specific features. In addition, from the viewpoint of dynamic diversity, because the dynamic change of facial expression originates in a person-specific facial expression pattern, it is presumed that the displacement vector of facial components has person-specific features. The properties of the human face described above reveal the following tasks. The first task is to generalize a facial expression recognition model. Numerous conventional approaches have attempted generalization of a facial expression recognition model. Here we are using Supervised Learning Technique which comes under Artificial Neural Network. First train the Network with the help of Supervised Learning and with the help of MS Kinect the system will detect that expression

Human gestures can be identified by observing the different movements of eyes, mouth, nose and hands. In this paper we are focusing on the human face for tracking and detection. Face is tracked using MS Kinect which uses 1.5 SDK.This makes use of the depth map to create a 3D wire frame model of the face. From this, facial features can be extracted – eye position, mouth position etc

II. RELATED WORK

Face plays an important role in human communication. Facial expressions and gestures incorporate non verbal information which contributes to human communication on verbal behaviour such as body posture ,gaze facial expressions and gesture plays a critical role in human communication These behaviours can convey a multitude of information about the attention,emotions,attitudes and physiological states of conversation participants[1] [2] [3] . By recognizing the facial expressions from facial images, a number of applications in the field of human computer interaction can be facilitated. The work inspired by many researchers to analyse facial expressions in 2D by means of image and video processing. Where by tracking of facial features, they attempt to classify different facial expressions. Almost all of the methods developed use 2D distribution of facial features as inputs into a classification system, and the outcome is one of the facial expression classes. Charles Darwin was one of the first scientists to recognize that facial expression is one of the most powerful and immediate means for human beings to communicate their emotions, intentions, and opinions to each other. In addition to providing information about affective state, facial expressions also provide information about cognitive state, such as interest, boredom, confusion, and stress, and conversational signals with information about spee ch emphasis and syntax[13].A fairly recent survey of techniques can be found in [4].However this paper to attempts as to recognizes 3D face tracking and detecting the corresponding facial expressions.

The paper is organized as follows. Section III, explains Materials and Method description which consists of the overview of the Microsoft Kinect .Section IV consists of Proposed Method .Section V consists of Experimental results. Finally the conclusion in Section VI.

III. MATERIALS & METHODS

In this paper we present our experience of using a MS Kinect depth camera for detection of some common face expressions. Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs. Based around a webcam-style add-on peripheral for the Xbox 360 console, it enables users to control and interact with the Xbox 360 without the need to touch a game controller, through a natural user interface using gestures and spoken commands

The MS Kinect sensor which contains a depth sensor , a colour camera (RGB) and a four microphone array that provides full body 3D motion capture , facial recognition and voice recognition capabilities

Figure 1 shows the MS Kinect sensor

Kinect interprets a 3D scene information using a projected infrared structured light which this mainly employs Light Encoding System. This MS Kinect sensor is a horizontal bar connected to a small base with a motorized pivot and is designed to be positioned above or below the video display. This has an RGB camera which gives the colour image.

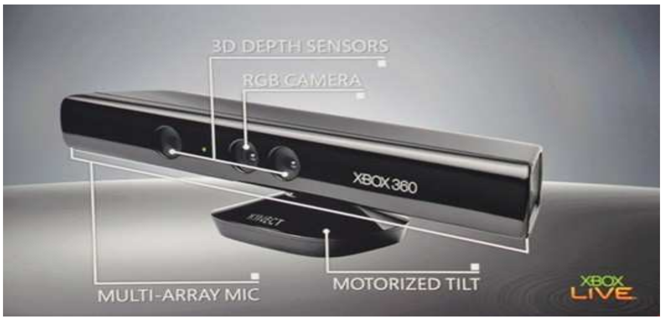

Figure 2 shows the arrangement of kinect[14]

This consists of IR Projector the colour Camera and IR Camera. The depth sensor consists of the IR Projector combined with the IR Camera which is a monochrome complementary semiconductor device (cmos).

The depth sensing technology is licensed from the Israeli company Primesense. The exact technology is not disclosed yet. It is based on the structured light principle. So it is called Light Coding technique. The IR Projector is an IR laser that moves through a diffraction grating and turns into a set of IR dots.

Microsoft has developed an official software Development kit (SDK) free to use for academics and now for commercial purposes. This is still in public beta form and is only available in the windows 7 platform.SDK of Microsoft includes APIs, device interfaces, installer documents and resource materials. This SDK becomes the fastest selling computer-electronic device of all time.SDK can be able to access to key pieces of MS Kinect system such as audio technology, skeletal tracking system applications program interface and direct control of the Kinect sensor .mainly used in vision for the natural user interface-interaction between people and computer. The SDK includes windows 7 compatible pc devices for Kinect. This provides Kinect capabilities to develop to build applications with c++, c# or visual basic by using Microsoft visual studio 2010

IV.PROPOSED METHOD

The proposed system is divided into five modules as shown in the following figure.

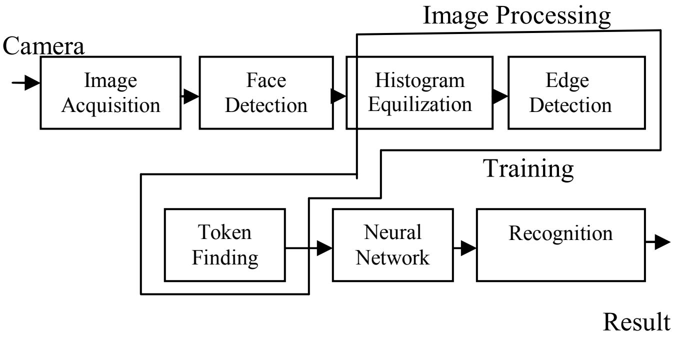

Figure 3 Simple architecture of Gesture Recognition

Each box is treated as one module in figure 3. The first module describes about capturing the image using the MS Kinect. Second module is for detection of face, which can detect the human face from the captured image. The pre-processing block is represented by a set of modules bounded by a boundary line. This mainly consists of histogram equalization, edge detection, thinning, and token generation modules. The next module describes the training module to store the token information that comes from the image Pre-processing module. This training is implemented with the help of back propagation neural network. The last module describes token matching and decision making called recognition module which gives the final result. The following flow chart shows how all the modules works.

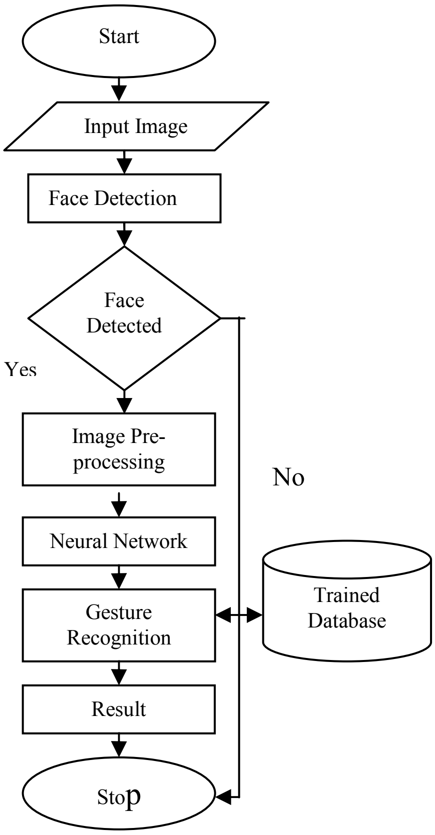

Figure 4 Flow chart of Gesture Recognition System

A Face Detection

The process that aims to locate a human face in an image is Face detection. The process can either be applied on stored images or images from the kinect. Human face is different from one person to another. This variation in faces mainly occurs due to race, gender, age, and other physical characteristics of an individual. Therefore face detection is a challenging task in computer vision. Because of the additional variations in scale, orientation, pose, facial expressions, and lighting conditions this becomes a more challenging task. There are so many methods that have been proposed to detect faces. Some common methods are neutral networks, skin locus, and color analysis. It is important to get rid of non-facial information in the image , because these detected faces become an input to the recognition of the gestures. In this paper, we are using the wire frame model to detect the face.

The following figure represents the detected faces with their corresponding input images.

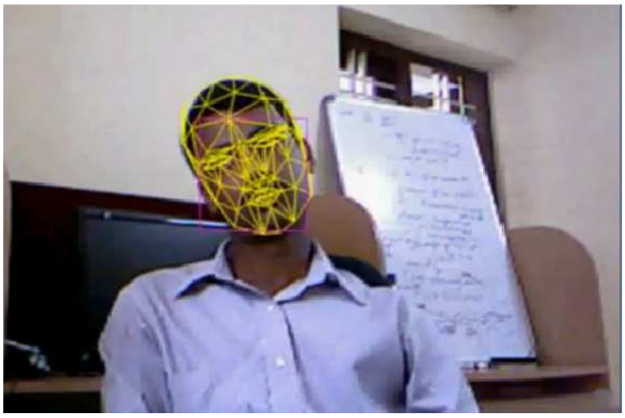

Figure 5.Wire Frame Model of the Face

A. Image Pre-processing

Four different modules are present in this block. Input is taken as a face image and tokens are produced as output. To enhance the image quality is the first step in this block. For performing this histogram equalization is done. This is then followed by the edge detection process. Tokens are generated when edge detection is over. Segmentation is the process of partitioning a digital image into its constituent parts or objects or regions [5].These regions share some common characteristics based on colour, intensity, texture etc. To produce something like a line drawing of an image is the goal of edge detection. The boundaries of objects tends to produce sudden changes in the image intensity. For example, different objects are usually different colors or hues and this causes the image intensity to change as we move from one object to another [6].

Different surfaces of an object receive different amounts of light, which again leads to intensity changes. So the intensity boundary information that we extract from an image will tend to indicate object boundaries, but not always. This method makes use of the following function to detect the edge and a linear filter [6] to remove the noise: function h=proposed edge(im,thr,T)Where im is an input image, this is a threshold between 0-1, T is the thickness of the line to indicate the edge and h is a uint8 black and white image with values of 0 and 255[7]. It is observed that when the threshold value T is less than 0.65 or greater than 0.70then the edges are not detected properly. When the threshold value is in between 0**.65<=T<=0.70** then the edges are determined properly and we found that at T=0.68 sharp and accurate edges are determined. A new edge detection technique is proposed which detects the accurate and sharp edges that are not possible with the existing techniques. This method with different Threshold values for given input image is shown that ranges between 0 and 1 and it are observed that when the threshold value is 0.68 the sharp edges are recognized properly

Figure 6: Edge detection

In image processing and computer vision edge detection is a terminology, particularly in the fields of feature detection and feature extraction, which mainly deals with identifying points in a digital image at which the image brightness changes sharply or more formally has discontinuities. Thinning has to be performed after edge detection Thinning is done to reduce the width of an edge to single line as shown in the following figure.

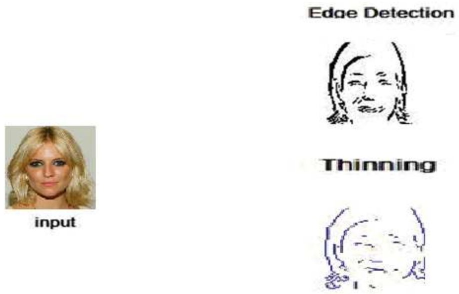

Figure 7 Images obtained before and after applying Thinning process

Tokens have been generated after the thinning process. For subsequent processing tokens mainly divides a data set in to the smallest unit of information used. Fig. 7 shows a part of the face which has been already processed and thinned. The line shows the shape of the eye image after performing successful edge detection & thinning. A point on the shape of the eye image is represented by a square box and the blue line joins the centre of two squares from which the cosine and sine angles are calculated. The representation of an eye token is a line connecting one box to another. A small right-handed triangle shown on this image and the summary of all triangles of a face image are the representation of the tokens.

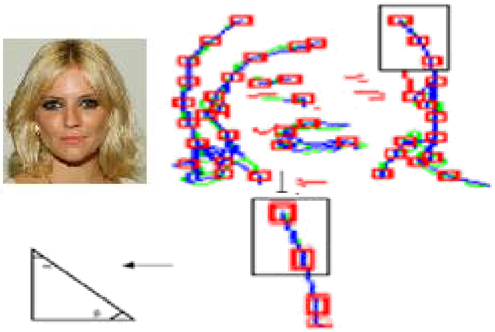

Figure 8: Token Generation and Presentation..

B. *Training And Back Propagation Network*

In the past, Image recognition has been done using image pixels to train a neural network via back-propagation. N inputs and one or more outputs are there for a typical ANN shown in fig. 8. The input layer is consists not of full neurons, but rather composed simply of the values in a data record, which constitutes inputs to the next layer of neurons. The next layer is called a hidden layer and there may be more than one hidden layers. The output layer is the final layer, where there is one node for each class. A single sweep forward through the network results in the assignment of a value to each output node, and the record is assigned to whichever class's node had the highest value. These actual pixels are given into the network as the inputs. With fixed orientation and scale, this approach works great when trying to recognize textures or objects. At different scale and orientation, it doesn't give encouraging results. Tokens(smallest unit of information)of an image are used for training the network.

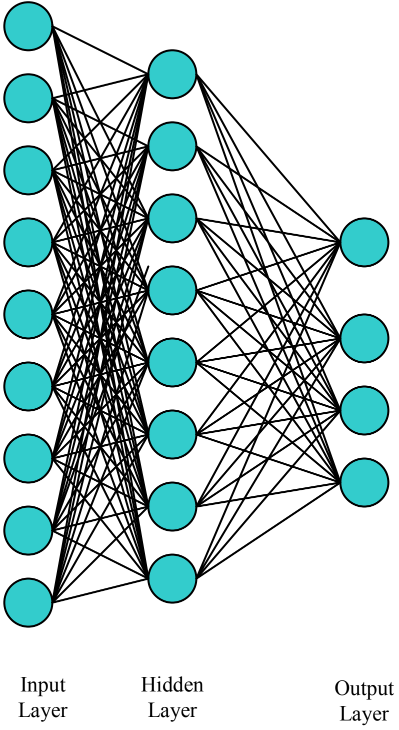

Figure 9 Example of simple Feed-Forward Neural Network

The network is trained to associate outputs with input patterns, during training. It identifies the input pattern and tries to output the associated output pattern, when the network is trained. To perform some task, we have to train a neural network by adjusting the weights of each unit in such a way that the error is reduced between the desired output and the actual output. This process needs that the neural network to compute the error derivative of the weights (EW). It must also calculate how the error changes as each weight is increased or decreased slightly. The most widely used method for determining the EW is the back propagation algorithm. During testing, the power of neural networks is realized when a pattern of tokens is given as an input and it identifies the matching pattern it has already learned during training

C. Recognition

After the training process, the network is ready to recognize gesture presented at its input. There are two options for recognizing the gesture of a face are provided. If user can recognize the gesture of existing image, then that image should be loaded from memory. As the user selects the image, the face recognition method works and returns the face part of the image. Another option is to capture the live image with the help of kinect sensor. When asked to recognize, this captures the image and finds the face part in it. Then the edge detection, thinning, and token generation are performed. After the above processs it classifies the given tokens into one of three gestures it learned during training. After training this gives percentage of recognition to each gesture with highest percentage closely matching and lowest to the farthest matching and the closest match is considered as the final result. The recognition process is as per the outline given in the flow chart in fig. 4.

V. EXPERIMENTAL RESULTS

To test the proposed solution, different set of gestures are captured with the help of kinect sensor. The images should be taken at different time with different gestures. Even certain gestures was closer to the existing or has different orientation. When a single face was exposed to the network, it was able to recognize the corresponding gesture with a high percentage of matching. There exist many gestures of the face. The network was trained mainly of three facial gestures-happy, sad and thinking faces. When this gestures are given to the network it will undergo training process and the gesture with the highest percentage will be matched. The gesture with the highest percentage is the corresponding result.

VI. CONCLUSION

Human gesture recognition in color image with different gestures was presented. Human face is detected first using the technique described by MS Kinect. Then edge detection, thinning, and token detection are performed. Then, recognition is performed. There are some positive and negative detections are found, the simplicity and robustness of the system is significant. The user is recognizing the gesture by giving the input threshold value for the detection of tokens. This is a difficult task to decide the best threshold value to generate the tokens. In this paper, a simple method for facial expression recognition using back propagation was proposed. The experimental results show that back propagation algorithm with the facial features extracted method can recognize well the appropriate facial expressions with the higher percentage than another facial expressions. The expression of sadness and disgust are more difficult than the others recognize. Generally speaking, online and spontaneous expression recognition is a difficult task. We focus to tackle the recognition of subtle spontaneous facial expressions.

.

REFERENCES

[1] M.Argyle and M.Cook,Gaze and Mutual Gaze. Cambridge:Cambridge Unviersity Press,1976

[2] P.Ekman and E.L. Rosenberg, What the face reveals: Basic and applied studies of spontaneous expressions using the Facial Action Coding System (FACS). New York: Oxford University Press ,1997.

[3] A. Kendon, "Language and Gesture: Unity or Duality," in Language and Gesture: Window into Thought and Action . Cambridge:Cambridge Unviersity Press,2000,pp 47-63

[4] P. Turaga, R.Chellappa, V.S. Subrahmanian and O.Udrea,"Machine Recognition of Human Activities: A Survey ," *IEEE Transactions on Circuits and Systems For Video Tecnology.* vol. 18, no 11,pp. 1473–1488, 2008

[5] A Novel Threshold Based Edge Detection Y. Ramadevi et al. / International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 6 June 2011

[6] A. Sloman, M. Croucher, "Why Robots will have Emotion", Proceeding of the 7th International Conference Artificial Intelligence, pp. 197-202, 1981.

[7] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple features", Conference on computer vision and pattern recognition, 2001.

[8] Chang Jyh-Yeong, Chen Jia Lin, Automated Facial Expression Recognition System Using Neural Net-works, Journal of the Chinese Institute of Engineers,Vol. 24, No.3, pp.345-356, 2001

[9] Xin Chen; Houjin Chen(2010). *A Novel Color Edge Detection Algorithm in RGBColor Space*, School of Electronic and Information Engineering(SEIE) 2010 IEEE.

[10] D. Beymer, P. McLauchlan, B. Coifman and J. Malik: "A Real-time Computer Vision System for Measuring Traffic Parameters," ComputerVision and Pattern Recognition pp. 495-501, 1997.

[11] Raheja, J. L., Das, K., Chaudhary, A., An Efficient Real Time Method of Fingertip Detection, Proceedings of 7th International Conference on Trends in Industrial Measurements and Automation (TIMA 2011), CSIR Complex, Chennai, India, 6-8 Jan, 2011, pp. 447-450.

[12] Md. Z. Uddin, N. D. Thang, T. S. Kim, "Human activity recognition via3 – D joint angle features and Hidden Markov Models", International Conference on Image Processing, pp. 713 – 716, 2010.

[13] Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. Marian Stewart Barlett,Gwen Littlewort, Ian Fasel, Javier R.Movellan

[14] Gesture Recognition using Microsoft Kinect. K.K Biswas,Saurav Kumar Basu.Procedings of the 5th International Conference on Automation,Robotics And Applications,Dec 6-8,2011,Wellington,NewZealand