ISSN No: - 2456 - 2165

Human Behaviour and Facial Expression Detection using Xbox Kinect Sensor

Aditya Ravi,
VIT University, Vellore 632-014

Mohammad Saad Rashid,
VIT University, Vellore 632-014

S. Sankar Ganesh
VIT University, Vellore 632-014

Abstract – The project aims at developing a module which consists of Microsoft Kinect Version 2 to identify the human face and various human facial expressions. Using the commonly available gaming console sensor, Kinect for Xbox One, we plan to create a simple windows application which utilizes various data streams from the sensor. The major sensor data steams we are using for facial detection and emotion detection are the colour stream, IR stream, depth stream and the body joints stream. Microsoft Visual Studio is used along with Microsoft Kinect SDK Version 2 to make the application.

Keywords *– Face detection, Facial expression recognition, Face Detection, Xbox Kinect.*

I. INTRODUCTION

In a discussion between two individuals, body language and facial expression play a major role. Although words contain the essence of the conversation a lot more is understood about the motions and physical state of a person through their facial expressions. A camera is most common optical instrument for capturing images. The digital camera we use on a day to day basis including the average cell phone camera works on the principle of capturing the RGB colors, i.e. Red-Green-Blue. Several algorithms have been used in the past with cameras for facial recognition where as with the Kinect, it a step ahead with its IR camera which helps in capturing depth, which is a new dimension added to these images. The four components together i.e. RGB and IR work in together and hence provide more useful information and hence refinement of the present algorithms is absolutely vital for better and more accurate results in facial recognition. Facial expression can send across information to the receiver which may or may not be aware of. But getting a better understanding of these facial expressions play a vital role in communication process.

Facial expression detection has become increasing popular in recent years and several developments have been made in research in this field. The advancements have happened in the field of pattern recognition, computer vision and machine learning . This makes detection of facial expressions even more crucial for the next generation computing tools including automation in areas where humans work, tutoring services, automated counselling, patient wellness monitoring,, etc. Every human is different from one another hence making identification of human beings and standardization procedures more difficult. Human faces vary due to a variety of reasons ranging from genetic, age, racial group and other physical characteristics of the individuals. All this makes face detection a very big challenge.

Shown below is the usual representation of steps in face detection:

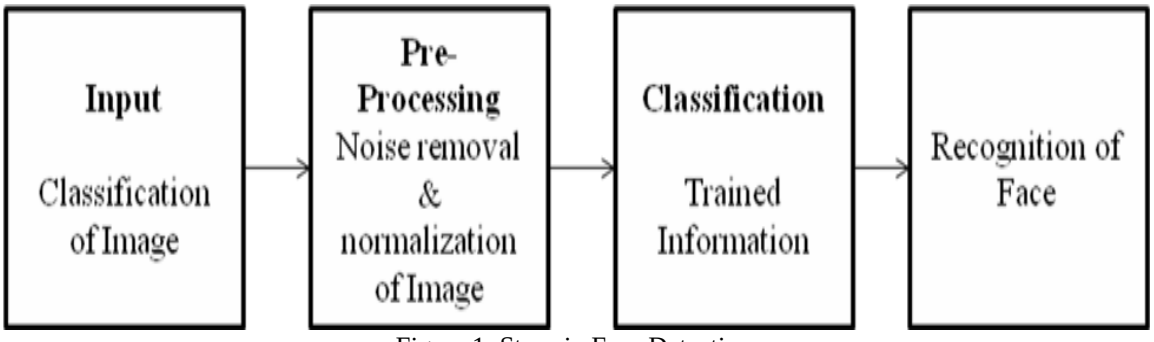

Figure 1: Steps in Face Detection

In the face detection process, the input block contains the image which has the human face which needs to be found from the image. The pre-processing block process the input image to remove any disturbances and imperfections in the image such as noise and also normalizes the image. The output of this block is then fed to a trainer which trains the image to decide whether the image belongs to the face group or not and will provide information about the recognized face.

ISSN No: - 2456 - 2165

II. FACIAL DETECTION– PROCEDURE OVERVIEW

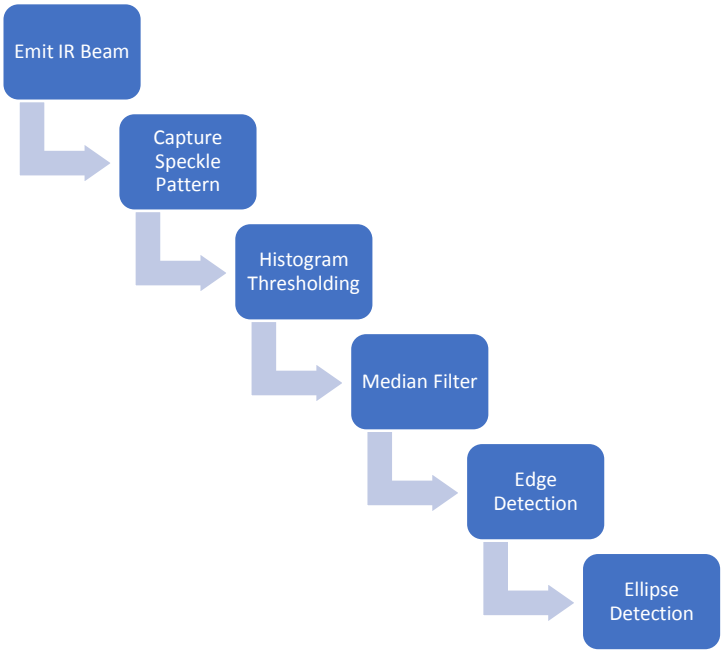

Figure 2: Face Detection Algorithm

For the face detection algorithm we use utilize the IR depth frame along with RGB frame to enhance the performance of the face detection process. The addition of the IR frame mainly benefits this process as it removes the dependency of the surrounding lighting conditions since IR capturing works flawlessly even in very low light conditions. The IR depth map analysis method involves an IR light beam which is emitted from the Kinect camera to capture a speckle pattern as a grayscale image. There are two main speckle patterns which can be detected by the Kinect camera namely primary and secondary speckle pattern. We require only the primary speckle pattern for our facial detection process. The speckle pattern gives us a better understanding of the textures and contours of the image captured. This speckle patternhelps us comprehend and hence identify human faces much more easier in the image and 3D space.

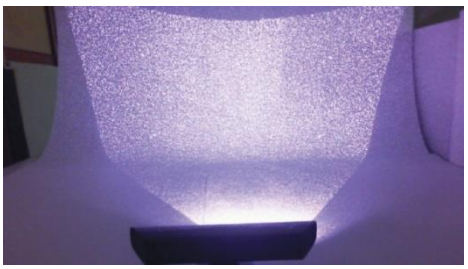

Figure 3: Speckle Pattern from Kinect Sensor

This is followed by Histogram Thresholding and Median Filtering to remove any sort of noise that arises in the grayscale image data feed. This is followed by Edge Detection and Ellipse Detection which is useful for detecting the objective block which includes the face and other body parts. Ellipse detection is required to identify the face structure pattern of a normal human face of a person.

III. FACIAL EXPRESSION RECOGNITION – PROCEDURE OVERVIEW

As mentioned earlier in a conversation between two or more people, the body language and emotions of the speaker play a vital role in conveying the message as much as the words do. Cameras are very common gadgets in todays word wherein almost every cellphone has a RGB camera capable of taking photographs and videos. Also almost every adult has a cellphone and hence indirectly a camera. The digital cameras we use today work on the principle of capturing colour in the form of RGB from the bayer matrix present on the pixels of the camera. There are several algorithms available which utilize this RGB data feed to go about facial recognition but the Xbox Kinect sensor benefits from having both a traditional RGB camera and an IR camera or imaging source which can also be used in measuring the new dimension of depth in images and videos. Hence the RGB data works along with the IR data available from the IR imaging source.When there is more data available there is

ISSN No: - 2456 - 2165

also an additional need for refined algorithms which can perform better than previously available algorithms for facial recognition. Extracting facial expression data accurately is a vital task and needs to be done as accurately as possible else wrong results may not reflect the true message. Several developments have been made in the field of facial expression detection and the research work available is also increasing day by day since this is the next big domain of computer and machine vision in the upcoming years.

The first step to implementing a facial expression detection system is face detection and alignment, which was discussed in the above section. This is followed by image normalization, feature extraction and then classification. There are several existing methods which are available for recognizing facial expressions. In an efficient algorithm from a research done in the past which was based on recognizing human emotion state detection which was done with facial expression data. This research utilized the basic optical flow method which extracts all the information required from motion vectors. This was done with monitoring the image changes with movement or motion in a fixed period of time. This algorithm works on segmented image frames and gives the result from the motion vectors. The strongest similarity with respect to the action unit standard database gives us the motion at that particular instant of time. This paper utilizes IR depth map method for facial recognition.

IV. METHDOLOGY

In this paper the methodology used for facial expression and behaviour detection takes place in four steps. These steps are as follows - face recognition, pre-processing, principle component analysis and classification.

The first step is facial detection which is performed by IR depth map analysis. This integrates IR frame with colour frame to create a coordinate mapper which converts a 2D colour space into a 3D space. The algorithm used was explained in detail above. This method utilizes a database of similar images expressing various emotions and facing different directions and different people which is used for training and testing purpose. The various emotions for which the library was built were happy, sad, neutral, looking away and yawning. The proposed technique depends on coding and decoding of these images which will then compare it to the trained database. The primary thing done is extracting all the information from the image and then encoding them followed by matching them to the database of the model. This is followed by pre-processing block in which the image runs some operations in order to normalize it and also remove any disturbances to it such as noise. Using the eigenface library for reference, the database is divided into two sets one which will be used exclusively for training and the other used for testing purposes. The images which are used for training are used to create a low dimensional face space. This is done in the next block which in done by Principal Component analysis of images in the training set and taking Eigen vectors with large eigen values. This is done so that the projected projected versions of all training images are created. Like the training images, the test images are also projected onto face space and a Euclidian distance is calculated so that the least possible value is chosen and this is inorder to track out which train image has highest similarity to the testing image.

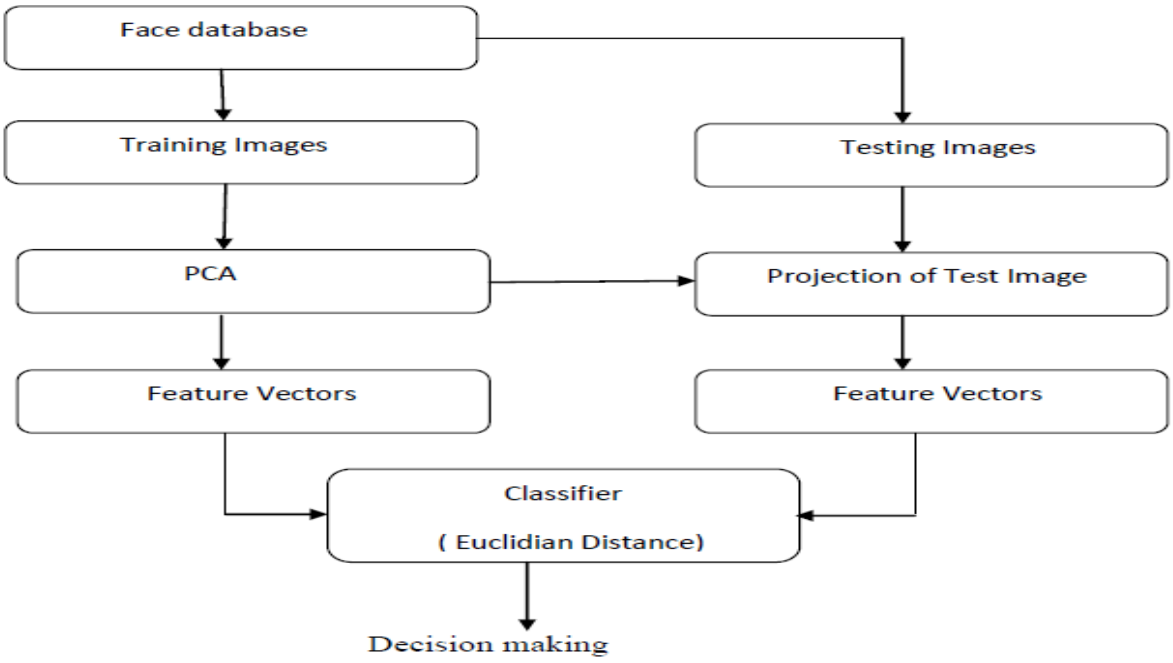

Figure 4: Decision Making Process

V. RESULT

A. Detection Accuracy

Emotion	Accuracy Percentage
Happy	78.85%
Sad	78.85%
Neutral	87%
Looking Away	82%
Yawning	91%

VI. CONCLUSION

Face detection in this project was done using IR depth map analysis face detection algorithm. The expression detection was performed by training and testing of images of various expressions. The training was performed with taking 10 individuals performing various expressions which was used. For testing we used 5 new individuals expressing various expressions and the accuracy of the expressions detection was tabulated. The results were more than satisfactory. This PCA method results in very high accuracy without comprising the consistency. The recognition of gestures was fast, accurate and works well in various types of environments.

References

[1]A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. PRL, 28(14):1885–1906, 2007.

[2]K. W. Bowyer, K. Chang, and P. Flynn. A survey of ap- proaches to three-dimensional face recognition. In ICPR, volume 1, pages 358–361, 2004.

[3]E. Corvee and F. Bremond. Body parts detection for peo- ple tracking using trees of histogram of oriented gradient de- scriptors. In AVSS, pages 469–475, 2010.

[4]T. Huynh, R. Min, and J. L. Dugelay. An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In ACCV, 2012.

[5]B. Y. L. Li, A. S. Mian, W. Liu, and A. Krishna. Using kinect for face recognition under varying poses, expressions, illumination and disguise. In WACV, pages 186–192, 2013.

[6]Y. Park, V. Lepetit, and W. Woo. Texture-less object tracking with online training using an RGB-D camera. In ISMAR, pages 121–126, 2011.

[7] A. R. Nagesh-Nilchi and M. Roshanzamir "An Efficient Algorithm for Motion Detection Based Facial Expression Recognition using Optical Flow" International Journal of Engineering and Applied Science 2006.

[8] P. Saudagare, D. Chaudhari, " Facial Expression Recognition using Neural Network – An Overview", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volue-2, Issue-1, March 2012.