International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT19527

Analysis of Customer Emotion from Video based Feedback of a Product

Bharathi E[1], Bagyalakshmi M[1], Ambika G[1], Priyanka G[2]

[1] Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimabatore, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Sri Krishna College of Technology, Coimabatore, Tamil Nadu, India

ABSTRACT

Due to the high levels of competition in a global market, companies have put more effort on building strong customer relationships and increasing customer satisfaction levels. Now-a-days due to technological improvements in information and communication technologies gives a highly anticipated key contributor to improve the customer experience and satisfaction in service episodes is through the application of video analytics, such as to evaluate the customer's emotions over the complete service cycle. Currently, emotion recognition from video could be a difficult analysis space. One of the foremost effective solutions to deal with this challenge is to utilize each audio and visual part as two sources contained within the video knowledge to form an overall assessment of the emotion. The combined use of audio and visual knowledge sources presents further challenges, such as determining the optimal data fusion technique prior to classification. In this paper, we propose an audio–visual emotion recognition system to detect the universal six emotions (happy, angry, sad, disgust, surprise, and fear) from video data. The detected customer emotions are then mapped and translated to provide client satisfaction scores. The projected client satisfaction video analytics system will operate over video conferencing or video chat. The effectiveness of our proposal is verified through numerical results.

Keywords : Viola Jones, Local Binary Pattern , Energy Entropy, K-NN classification.

I. INTRODUCTION

Recently, video-based calls (e.g., Skype Video) have become more popular as a mode of communication among people. As the technology matures and becomes more advanced, the interactive communication capabilities of contact centers are anticipated to extend from text and audio or speech communications to video-based communications. A highly anticipated key contributor for improving the customer experience for service episodes in contact centers is through the application of video analytics to evaluate the customer's emotions over the full service cycle. The processing of video-based data for analytics requires new technical challenges to be overcome. Video-based data contains two data components or sources: audio-based data and visual-based data. The use of two forms of data sources presents additional challenges such as determining the optimal data fusion technique prior to classification. According to the applications such as video surveillance, object detection and tracking is the first step. The main aim of this work is Face detection and tracking system with a static camera has been developed to estimate velocity, distance parameters. We used image difference algorithm for object detection and tracking based on vision system. Then the speech of the person is recognized to get

the feedback from the corresponding person. This process focuses on detection of human in a scene and then speech signal processing was done.

The overall goal of the process is to create a system which pre-screens video surveillance feed and assists the user in identifying feedback activity. Towards that goal, we designed an automated scheme capable of performing three large-scale tasks: identifying contacts, tracking contacts, and characterizing contact behavior. In order to get an enhanced image or to extract some useful information from the image we want to do image processing on that image. It is one of the types of signal processing in which image is an input and image or characteristics/features associated with that image are an output.

Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps:

- Importing the image via image acquisition tools
- Analyzing and manipulating the image
- Output is based on image analysis

Over all aim of this system is to find the emotion of the customers from their feedback. First the video is separated into audio and visual data. Then the video is converted into frames and preprocessing is done so that image is resized to get the exact face region. By using face detection algorithm, the face was detected. Using feature extraction algorithm the features are extracted and classified into six types of emotions such as happy, sad, angry, fear, disgust and surprise. The audio data for that particular video is taken and given into the audio feature extraction algorithm. From this the audio features are extracted. Now both the visual and audio extracted features are given into classification algorithm to predict the emotions.

II. RELATED WORK

The work by Richards and Jones [1] proposed a definition of CRM as "a set of business activities supported by both technology and processes that is directed by strategy and is designed to improve business performance in an area of customer management." The CRM often involves using technology to organize, automate, and synchronize a company's sales, marketing, customer service, technical support operations, etc.

The work by Michael D et al. [2] proposed to explore how these unreliable information sources can be used for robust multi-person tracking. This algorithm detects a large number of dynamically moving people in complex scenes with occlusions, does not rely on background modeling, requires no camera or ground plane calibration, and only makes use of information from the past. But each original video samples was used only once.

Another work by Vidrascu and Devillers [3] used a 10-h dialogue corpus in French recorded in a French Medical emergency call center and obtained a correct detection rate of about 82% between negative and positive emotions using paralinguistic cues and compared several classification methods including support vector machines (SVM) and decision trees.

More recently, Pappas et al. [4] proposed a method to classify speech windows into either containing the anger emotion or not. Their technique does not require speech recognition, or utterance segmentation, and can thus be applied toward real-world recordings from call centers.

III. PROPOSED SYSTEM

In this paper, we propose a face tracking algorithm with temporal-spatial information and trajectory of

confidence. The whole process is divided into Video and Speech association. Trajectories with high confidence are associated with the detection result of the current frame during local association, whereas trajectories with low confidence are associated with the detection results of the current frame are not matched during global association. We determine the association results using a combined model of digital image processing and digital signal processing. The major steps carried out are, Detection and recognition.

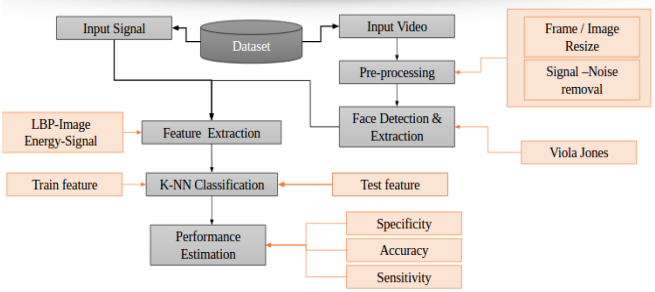

Fig 1. Block Diagram

Dataset Collection

The dataset collected consist of videos that we recorded from consumers giving good and bad reviews about the product. The participants were mainly college students giving feedback about the product. In this work the product considered is various brands of mobile phone.

Extraction of audio and video

This is the first step in this system and here the video is given as input. The input video is the feedback of any product used by the consumer. The input video information from consumer varies from seconds to minutes depending upon their product review. The customer feedback includes both positive and negative comments. That input video is separated into two models one is visual and another one is audio. It is done by using any online audio video extractor.

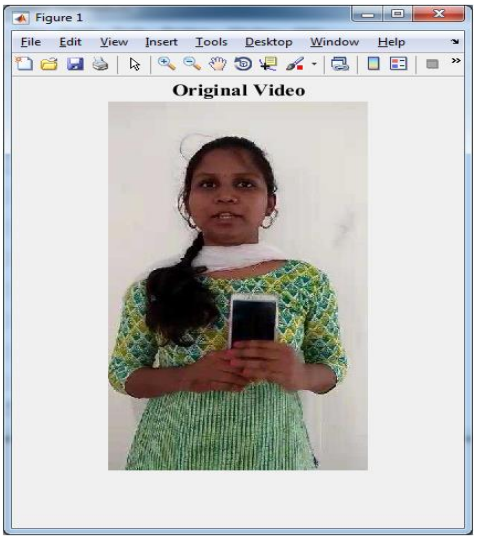

Fig 2. Original Video

IV. VIDEO PROCESSING

In Video processing ,the original video is preprocessed and it is given as input to the face detection algorithm. Then feature extraction is used to extract the features from the data.

Preprocessing

Preprocessing is done to remove the noise from the data. Preprocessing steps include:

- Frame Resize
- Noise Filtering

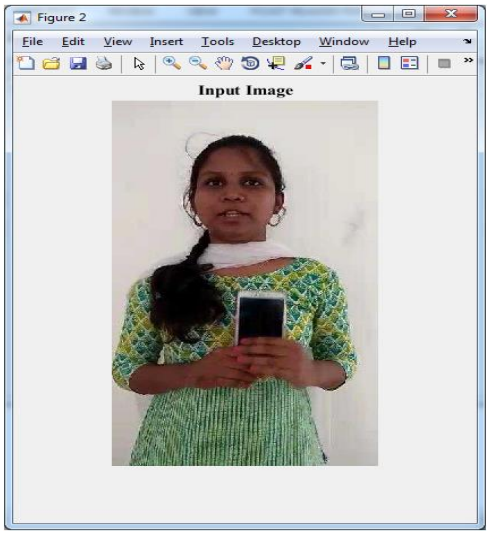

Fig 3. Input Image

Frame Resize

In computer graphics and digital imaging, image scaling refers to the resizing of a digital image. In video technology, the magnification of digital material is known as up-scaling or resolution enhancement. When scaling a vector graphic image, the graphic primitives that make up the image can be scaled using geometric transformations, with no loss of image quality. When scaling a raster graphics image, a new image with a higher or lower number of pixels must be generated. In the case of decreasing the pixel number (scaling down) this usually results in a visible quality loss. From the standpoint of digital signal processing, the scaling of raster graphics is a two-dimensional example of sample-rate conversion, the conversion of a discrete signal from a sampling rate (in this case the local sampling rate) to another.

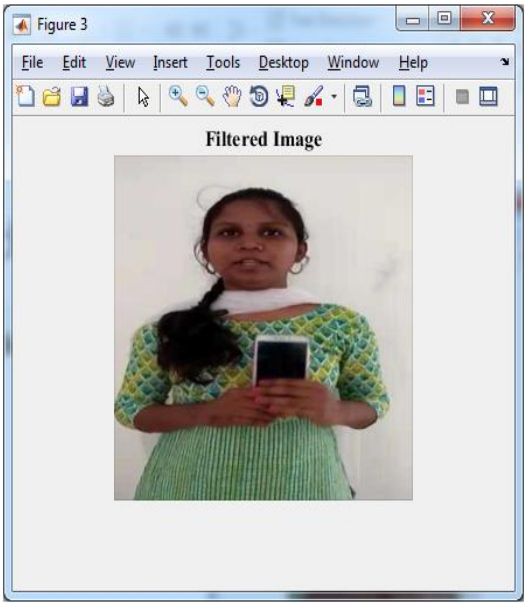

Fig 4. Filtered Image

Noise Filtering

Image processing is basically the use of computer algorithms to perform image processing on digital images. Digital image processing is a part of digital signal processing. Digital image processing has many significant advantages over analog image processing. Image processing allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing of images. Wavelet transforms have become a very powerful tool for de-noising an image. One of the most popular methods is wiener filter. In this work four types of noise (Gaussian noise, Salt & Pepper noise, Speckle noise and Poisson noise) is used and image de-noising performed for different noise by Mean filter, Median filter and Wiener filter. The above figure shows the result of preprocessing.

Face Detection

The Preprocessed visual data is given as input to the face detection algorithm to detect the exact face region of the customer. Here ,Viola Jones face detection algorithm is used to detect the face from the video frame.

Viola Jones Algorithm

The viola Jones algorithm is a majorly used mechanism for object detection. The major property of this algorithm is that training is slow, but detection is fast. Viola Jones uses Haar basis feature filters, so it does not use multiplications. The efficiency of this algorithm can be significantly increased by first generating the integral image.

$$II(y,x) = \sum_{p=0}^{y} \square \sum_{q=0}^{x} Y(p,q)$$

In this algorithm, the integral image allows integrals for the Haar extractors to be calculated by adding only four numbers. For example, the image integral of area ABCD is calculated as $II(y_A,y_A) - II(y_B,y_B) - II(y_C,x_C) + II(y_D,y_D)$.

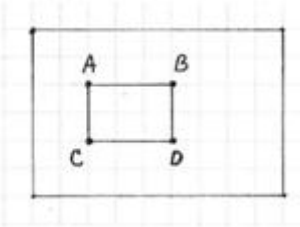

Fig 5. Computation of integral image

In this algorithm, each face recognition filter (from the set of N filters) contains a set of cascade-connected classifiers. Each classifiers looks like a rectangular subset of the detection window and determines if it looks like a face. If it is done, the next classifier is applied. If all classifiers give a positive answer the face is recognized. Otherwise the next filter in the set of N filters is run.

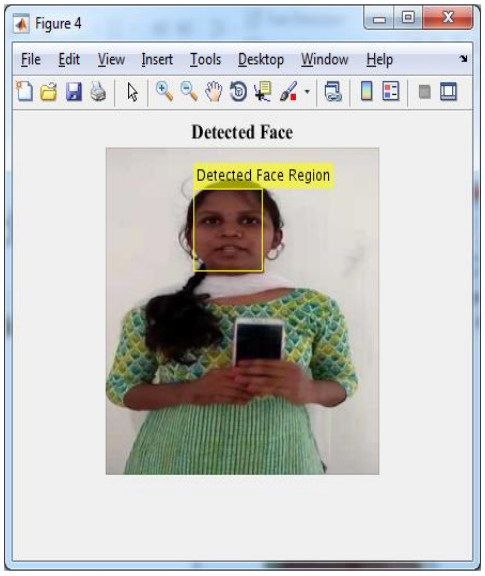

Fig 6. Detected Face

Fig7. Exact Detected Face

Feature Extraction

Feature Extraction is used to reduce the amount of resources required to describe the large set of data. Here the preprocessed visual data is given as the input to the feature extraction algorithm. Local Binary Pattern algorithm is used for feature extraction.

Local Binary Pattern

LBP is a type of visual descriptor. It is used for classification in computer vision. It is the particular case of the Texture Spectrum model. The LBP feature vector, in its simplest form, is created in the following manner:

Divide the examined window into cells (e.g. 16x16 pixels for each cell).For each and every pixel in a cell; compare the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e. clockwise or counter-clockwise. Where the center pixel's value is greater than the neighbor's value, write "0". Otherwise, write "1". This gives an 8-digit binary number (which is converted to decimal).To compute the histogram, over the cell, of the frequency of each "number" occurring (each combination of the pixels which are smaller and greater than center). This histogram can be seen as a 256-dimensional feature vector. Optionally normalize the histogram and then concatenate the normalized histograms of all cells. It gives a feature vector for the entire window. The feature vector can now be processed using the Support vector machine, extreme learning machines, or some other machine-learning algorithm to classify images. These classifiers can be used for face recognition or texture analysis.

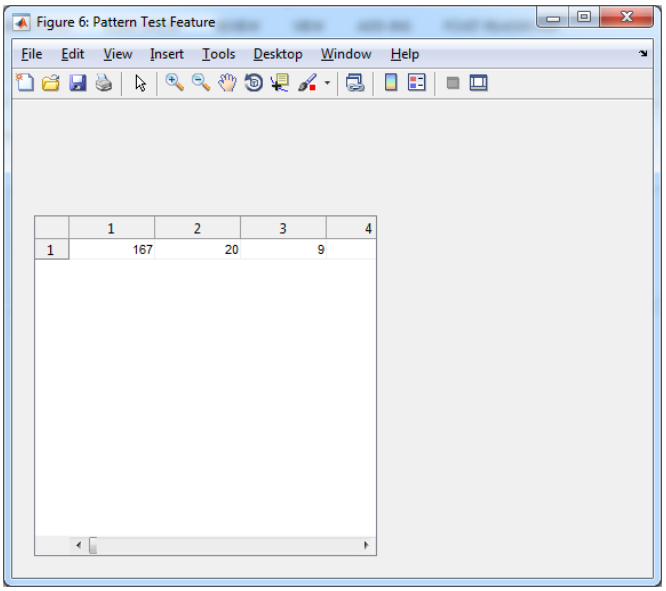

Fig.8 Pattern Test Feature

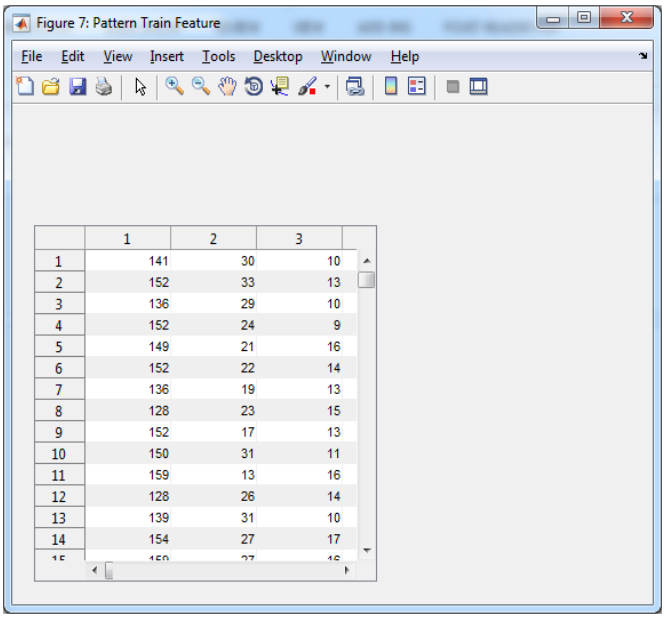

Fig.9 Pattern Training Feature

V. AUDIO PROCESSING

In audio processing the extracted audio data is given as the input to the audio feature extraction algorithm to get the extracted audio data.

Feature Extraction

The feature extraction of the audio data is done by using the Energy Entropy algorithm.

Energy Entropy

Energy entropy algorithm can be used to describe the uncertainty distribution and the complexity characteristics of the signal, which can quantitatively describe the internal information characteristics that contain in the signal. Therefore, entropy characteristic can be used to extract the internal features of the signals. Feature extraction algorithm based on entropy theory can distinguish different communication signals through describing the distribution state characteristics of the signals. The details of those signal characteristics are unnecessary, and the calculation is relatively simple. Feature extraction algorithm based on entropy is suitable for the signal recognition under higher SNR environment. when the signal is completely submerged in the noise, the recognition method based on the complexity of the signals by the characteristics of entropy is difficult when accurately identifying the signal due to the big overlap interval between different signal entropy characteristics. In information theory, “entropy” can measure the uncertainty of the things.

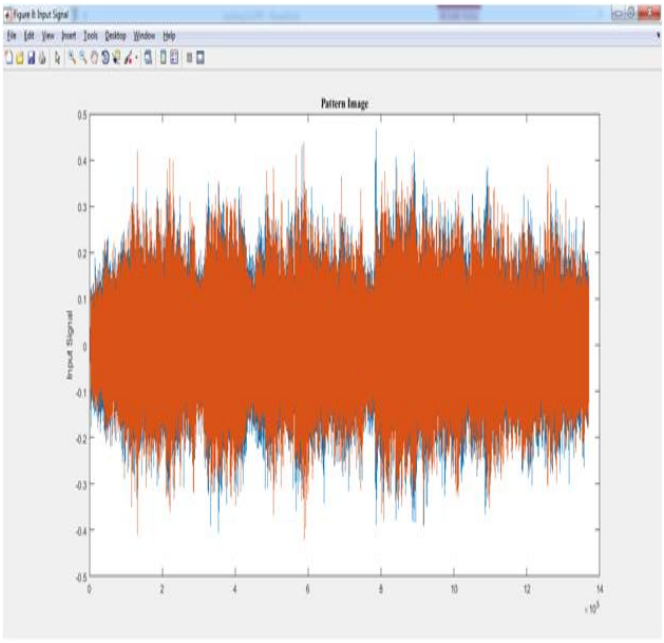

Fig.10 Audio signal after feature extraction

VI. CLASSIFICATION

Classification is used to process both the audio and visual data to get the emotion of the customer by comparing the classified results of both data. K-NN classification algorithm is used to classify the data.

K-NN Classification

K-NN is an algorithm used for classifying the results.In pattern recognition, *k*-nearest neighbor algorithm (*k*-NN), is used for classification and regression. The *k* closest training examples in the feature space are the input in both cases. The output is based on whether *k*-NN is used for classification or regression:

In this classification, the result is a class membership. The objects are classified by a plurality vote of its neighbor, with the object being assigned to the class which is most common among its k nearest neighbors (k is a positive integer, which is small). If k=1, then the object is assigned to the class of the nearest neighbor. K-NN regression, the output is the property value for the object. This value is calculated by the average of the values of its k nearest neighbors.

Properties

K-NN is considered as one of the special case of a variable-bandwidth, kernel density "balloon" estimator of the uniform kernel. The naive version of the algorithm is easy to implement by computing the distances from the test example to all the stored examples, but it is computationally intensive for large training sets. Using an approximate nearest neighbour search algorithm, we can make K-NN computationally tractable even for large data sets. This nearest neighbor search algorithms generally seek to reduce the number of distance evaluations that is actually performed.

K-NN is having some strong consistency results. As the amount of data approaches infinity, the two-class K-NN algorithm is guaranteed to yield an error rate no worse than twice the Bayes error rate (i.e, the minimum achievable error rate given in the distribution of the data).The speed of K-NN can be improved by using proximity graphs.

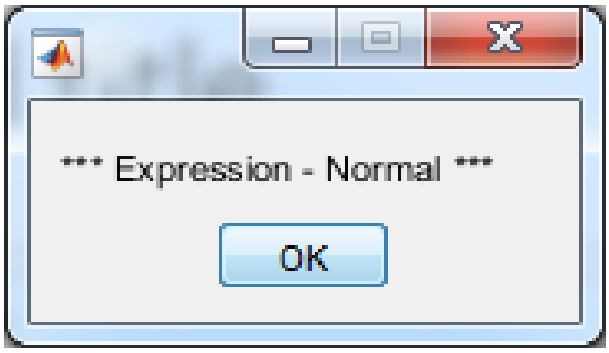

Fig 11. Expression Estimation

VII. RESULT AND DISCUSSION

The outcome of this video analysis is the different emotions of people giving reviews about the product or a movie. The emotions varies from happy to sad or fear. We have analyzed these basic emotions and further working on more emotions and the proper format to display the result (emotions). We have achieved an accuracy of about 98.5% as of now.

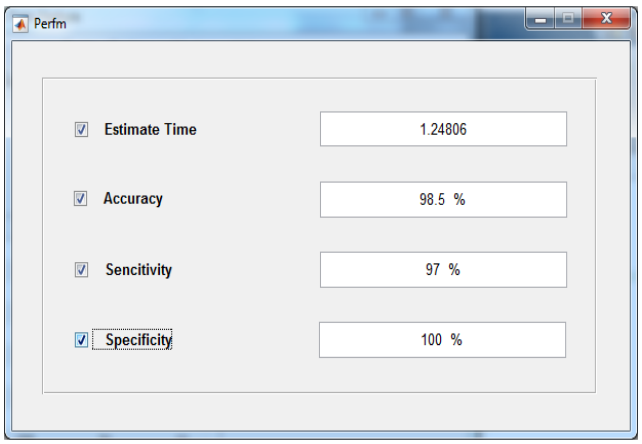

Fig12. Performance Estimation

VIII. CONCLUSION AND FUTURE WORK

Customer feedback about a product is very essential to improve the sales and business of any product. Our work provides an effective system for analyzing the customer emotion when using a product from the feedback videos. The dataset consists of videos of participants giving reviews of various brands of mobile phones. The video is split into visuals and audio for processing. After preprocessing of video the video frames are analyzed and emotion is detected. Further the audio alone is extracted from the video and processed and analyzed to detect the emotion of the reviewer. We have collected minimum dataset as now and analysis of emotion has been done. Further in future dataset consisting of more number of videos can be analyzed and emotion can be detected. This analysis of emotion is helpful to understand a huge number of folk's emotion regarding a product which helps in analyzing their intentions in a digital manner further rising to improvement of business strategy.

IX. REFERENCES

[1] K. A. Richards and E. Jones, "Customer relationship management: Finding value drivers," Ind. Market. Manage., vol. 37, pp. 120–130, 2008.

[2] http://www.academia.edu/266817/Online_Multi-Person_Tracking-by-Detection_From_a_Single_Uncalibrated_Camera

[3] L. Vidrascu and L. Devillers, "Detection of real-life emotions in call

[4] centers," in Proc. Annual Conf. Int. Speech Commun. Assoc., 2005,

[5] pp. 1841–1844.

[6] D. Pappas, I. Androutsopoulos, and H. Papageorgiou, "Anger detection in

[7] call center dialogues," in Proc. IEEE Int. Conf. Cognitive Infocommun.,

[8] 2015, pp. 139–144.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1. Jun. 2005, pp. 886–893.

[10] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 9, pp. 1806–1819, Sep. 2011.

[11] A. A. Butt and R. T. Collins, "Multi-target tracking by Lagrangian relaxation to min-cost network flow," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 1846–1853.

[12] L. Leal-Taixe, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese, "Learning an image-based motion context for multiple people tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 3542–3549.

[13] A. Gilmore, L. Moreland, "Call centers: How can service quality be managed?", Irish Market. Rev, vol. 13, no. 1, pp. 3-11, 2000.

[14] A. Feinberg, I. S. Kim, L. Hokama, K. De Ruyter, C. Keen, "Operational determinants of caller satisfaction in the call center", Int. J. Service Ind. Manage., vol. 11, no. 2, pp. 131-141, 2000.

[15] S. Ben-David, A. Roytman, R. Hoory, Z. Sivan, "Using voice servers for speech analytics", Proc. Int. Conf. Digit. Telecommun, Aug. 29–31, 2006.

[16] D. Melamed, M. Gilbert, "Samsa: Speech analytics", Spoken Language Understanding: Systems for Extracting Semantic Information from Speech, pp. 397-416, 2011.

[17] R. Polig et al., "Giving text analytics a boost", IEEE Micro, vol. 34, no. 4, pp. 6-14, Jul./Aug. 2014.

[18] R., Polig, K. Atasu, C. Hagleitner, "Token-based dictionary pattern matching for text analytics",

Proc. 23rd Int. Conf. Field Programmable Logic Appl., pp. 1-6, 2013.

[19] Priyanka.G, "Prediction of Airline Delays using K Nearest Neighbor Algorithm", International Journal of Emerging Technology and Innovation Engineering (IJETIE), ISSN: 2394 – 6598, Vol. 4, No. 5, pp.87-90, 2018.

[20] Priyanka.G,"Data Analytics in Agriculture", International Journal of Emerging Technology and Innovation Engineering (IJETIE), ISSN: 2394 – 6598, Vol. 4, No. 5, pp. 91-94, 2018.

Cite this article as :

Bharathi E, Bagyalakshmi M, Ambika G, Priyanka G, "Analysis of Customer Emotion from Video based Feedback of a Product", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 44-52, March-April 2019. Available at doi : https://doi.org/10.32628/CSEIT19527
Journal URL : http://ijsrcseit.com/CSEIT19527