Scientific Journal of Impact Factor(SJIF): 4.14
e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue for ICPCDECT 2016,Volume 3 Issue 1

Voice Controlled Image Processing Application With Motion Detection

Rajan Jha,Kalpesh Dixit, PratishKarkera, Saurabh Suryavanshi

Information Technology, St.John College Of Engineering & Technology
Information Technology,St.John College Of Engineering & Technology College
Information Technology,St.John College Of Engineering & Technology College
Information Technology,St.John College Of Engineering & Technology College

Abstract —*A Voice Command device (VCD) is a device controlled by means of the human voice. By removing the need to use buttons, dials and switches, consumers can easily operate appliances with their hands full or while doing othertasks. Voice Commands would prove to be helpful in providing a hands free format for the users. It increases the accessibility range of any application and requires very less training for users. In recent decade, Voice controlled applications have become increasingly available. The key behind voice recognition software is the ability for technology to convert a person's voice into a recognizable data pattern. The image processing can be defined as processing or altering an existing image in desired manner.*

In this project, we have made an Image processing application that is controlled using voice commands and also by physical means to interpret the users command and provide the necessary outputs. This image processing application provides the following functions such as starting a camera, taking pictures, save pictures, open picture from the particular directory, apply various types of effect to the image like grayscale, Invert, Rotate channel, mirror and many more. These functions can be operated by voice and manually. The application also provides the slide show of the images, which it is also controlled by the voice, the speed of the slide show can be changed by track bar or by voice. It also provides the moving image module where the image is moved using voice commands

Finally, the application provides the Motion Detection module. The motion detection module senses the motion using Two frame differential technique and the amount of motion detected is visualized by the graph as well as the area where motion is detected is converted to red color.

Keywords*-Voice Command Application, Speech Recognition and Synthesizer, Image Processing, Motion Detection, Desktop Application*

INTRODUCTION

Image Processing- is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Images are also processed as three-dimensional signals where the third-dimension being time or the z-axis.Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging.Closely related to image processing are computer graphics and computer vision. In computer graphics, images are created from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered a highlevel image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans) [1].

Motion Detector- is the process of detecting a change in the position of an object relative to its surroundings or a change in the surroundings relative to an object. Motion detection can be achieved by either mechanical or electronic methods.When motion is accomplished by natural organism it is called motion perception.

.

Voice Command Application- is a device controlled by means of the human voice. By removing the need to use buttons, dials and switches, consumers can easily operate appliances with their hands full or while doing other tasks.

Desktop Application- An application that runs stand alone in a desktop or laptop computer. Contrast with "Web-based application," which requires the Web browser to run. The term may be used to contrast desktop applications with mobile applications that run in smartphones and tablets. A desktop application is one that runs in the traditional Windows desktop in contrast to a tablet application that runs full screen.

Speech Recognition and Synthesizer –Speech recognition is the inter-disciplinary sub-field of computational linguistics which incorporates knowledge and research in the linguistics, computer science, and electrical engineering fields to develop methodologies and technologies that enables the recognition and translation of spoken language into text by computers and computerized devices such as those categorized as Smart Technologies and robotics. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

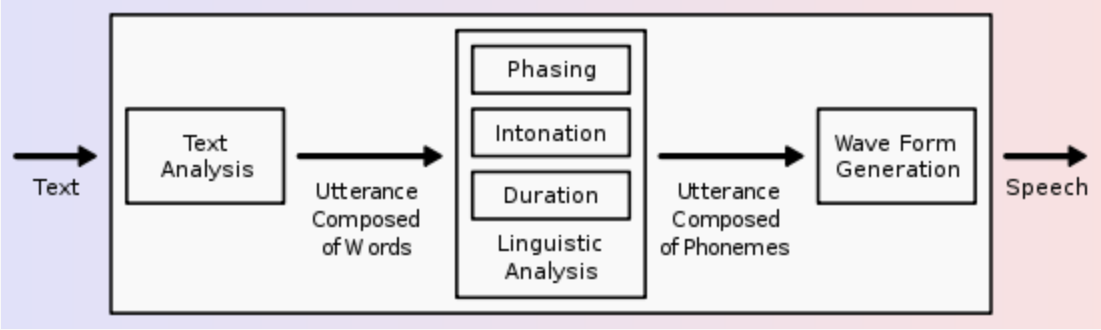
Some SR systems use "training" (also called "enrollment") where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker independent''systems. Systems that use training are called "speaker dependent".

Speech recognition applications include voice user interfaces such as voice dialing (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domotic appliance control, search (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

Speech Synthesizer *and can be implemented in software or hardware products. A* ***text-to-speech (TTS)*** *system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.*

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood clearly. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1990s.



Overview of a typical TTS system

*A text-to-speech system (or "engine") is composed of two parts. a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called*text normalization, pre-processing, *or* tokenization. *The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like* phrases, *clauses, and sentences. The process of assigning phonetic transcriptions to words is called* text-to-phoneme *or* grapheme-to-phoneme *conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end—often referred to as the* synthesizer*—then converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the* target prosody *(pitch contour, phoneme durations), which is then imposed on the output speech*

II. IMAGE PROCESSING TECHNIQUE

Image processing is a method to convert an image into digital form and perform some operations on it in order to get an enhanced image or to extract some useful information from it.

Image processing basically includes the following three steps.

• Importing the image with web cam with speech recognition for start a camera

• Analyzing and manipulating the image that includes visualization of user with the help of Histogram and Motion Detection visualizing with spotting patterns such as graphs

•The last is the output in which result can be altered image with an different techniques use to altered image and visualizing the process done on image with Histogram

III. LITERATURE SURVEY

Yu wang, Qianchen, Baeominzhang et al. has published in Journals and Magazines regarding Image Enhancement using Histogram techniques.

Using charts and graphs can help the audience grasp visually the message that has to be conveyed. Charts and graphs are especially useful if there will be a great deal of details that would normally take up too much time explaining and you need to compact the information to a visual summation. Knowing your audience helps creating the right type of charts and graphs to use. For example, if you are addressing a team of sales representative, you want to search for sales charts and graphs that typically a sales team will understand visually. When utilizing charts and graphs keep in mind that they are to be supplemental to using words. Charts and graphs can be very powerful in visual presentations if done effectively. There are many different formats that can be used in creating charts and graphs that could make it quite difficult or frustrating to choose the correct one to effectively use. Too many charts and graphs can come across to be very confusing, lack clarity or irrelevant.

Bo Cue, Tang Xue et al. has published in Conference Publication regarding Computing Communication *tat the system completes the functions of collecting the voice data, distilling character, specialvoice recognition and voice playing in terms of initializing the system and the identification training. According to the voice recognition arithmetic theory, the pretreatment of voice signal, the character distilling and pattern matching is analyzed. The results of the experiments indicate that the system capability is steady and the identification is effective. The system can be applied in house or small office safety protection.*

Allows user to operate a computer by speaking to it.Free up cognitive working space, Allows dictation of text, commands. Eliminates handwriting, spelling problem.Always spells correctly (doesn't always recognize words correctly);
Requires large amounts of memory to store voice files. Difficult to use in classroom settings, due to noise interference;

IV.SYSTEM METHODOLOGY

Our System works in following Methods. It includes the following ways:

Image Processing-It includes the method of processing images in it. We can start a camera, select image, capture image by speech synthesizer and manually. In this we can add different effects such as grey scaling, saturation, sepia, etc to image by speech synthesizer and also manually. In this we can some effects like slide show of an captured images by increasing the rate of images as maximum or minimum by speech synthesizer and manually.it will lead in to increase the images slide show as fast or slow. Also the Moving Object part is include in it. It will show us the mechanism of an image which can be moved to four different direction such as right, left, up, down. This is also done with speech synthesizer. After some fixed amount of time the image randomly changes. We can also rotate the image by speech recognition and also manually by moving the track bar. We can also increase the brightness by moving the track bar. We also show the visualization of screen in which the image of background screen is capture and it is visualize.

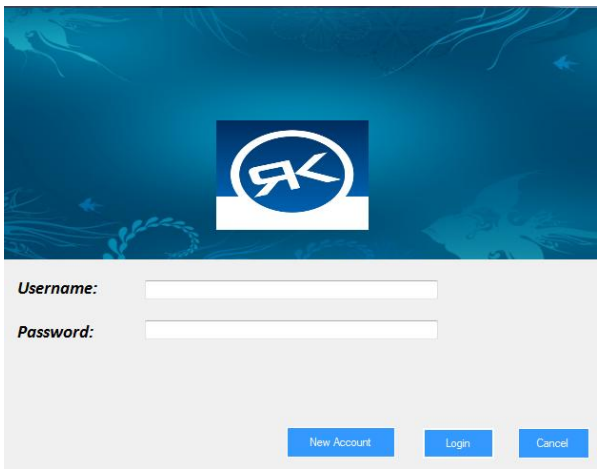


Fig (1): Login page

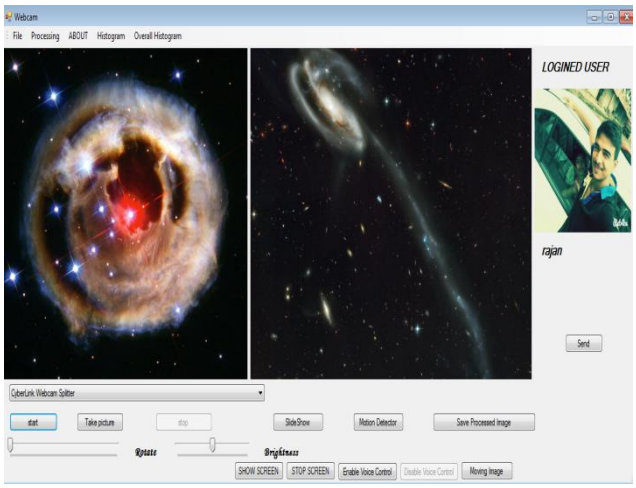


Fig (2): Main screen

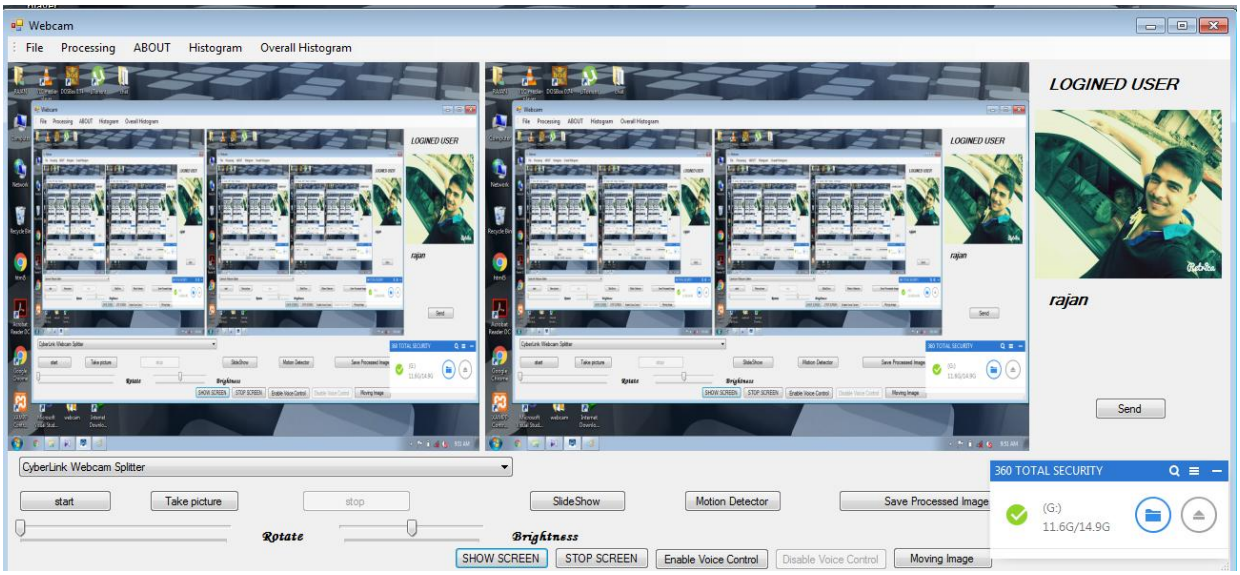


Fig (3): Show Screen or Stop Screen

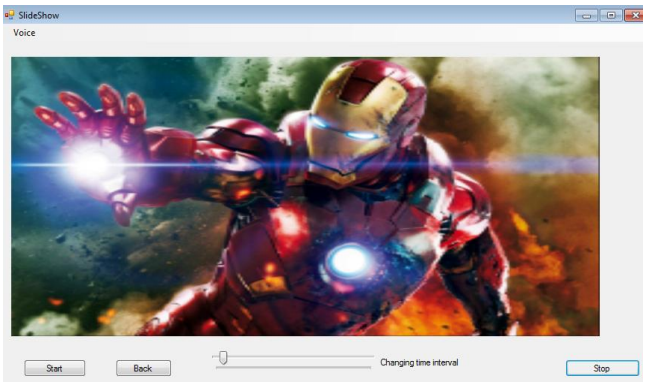


Fig (4): Slide show screen

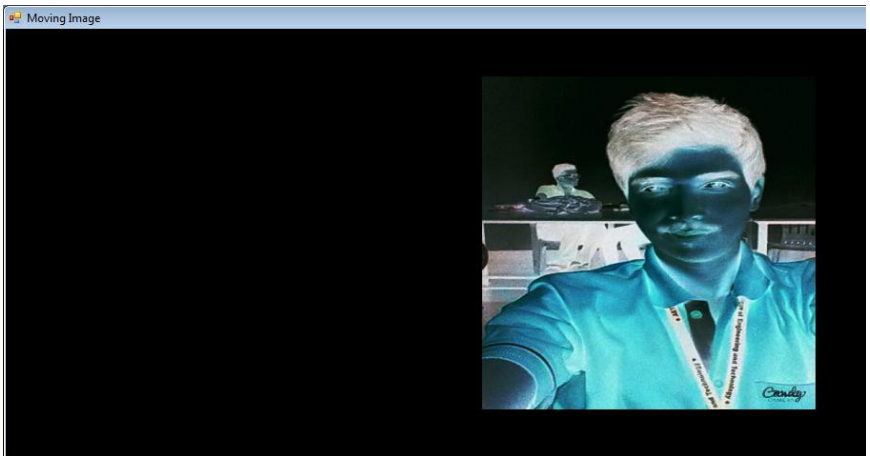


Fig (5) : Moving Object Screen

Motion Detection-we can start this by speech recognition an also manually. This system works in such a way that it first detect the object and the detected object is outlined by the red color. When the detected object perform some motion it recognize the motion performed by an object and the rate of motion performed is visualize by graph on screen [4].

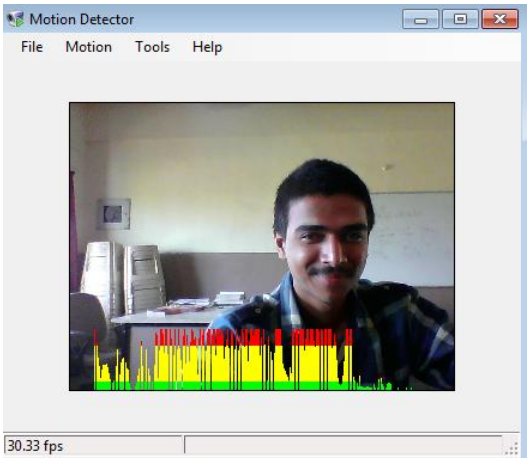


Fig (6) : Motion Detection Screen

This is a web based application which results in following above detailed function. This is also a secured application. It is secured in such a way that when an new user login to it he has to upload his image and when he is added as an user on screen right side he is a user that will be assisted. The new user login and his developed session will be recorded in database also. When user end his session the message is send through Bluetooth to admin Bluetooth device which is an

android phone. The user used function is visualize by using histogram. The visualization is done in two parts first the user session which is used and second the overall session.

This histogram will show us the functions used by user in single session and also in an overall session [5].

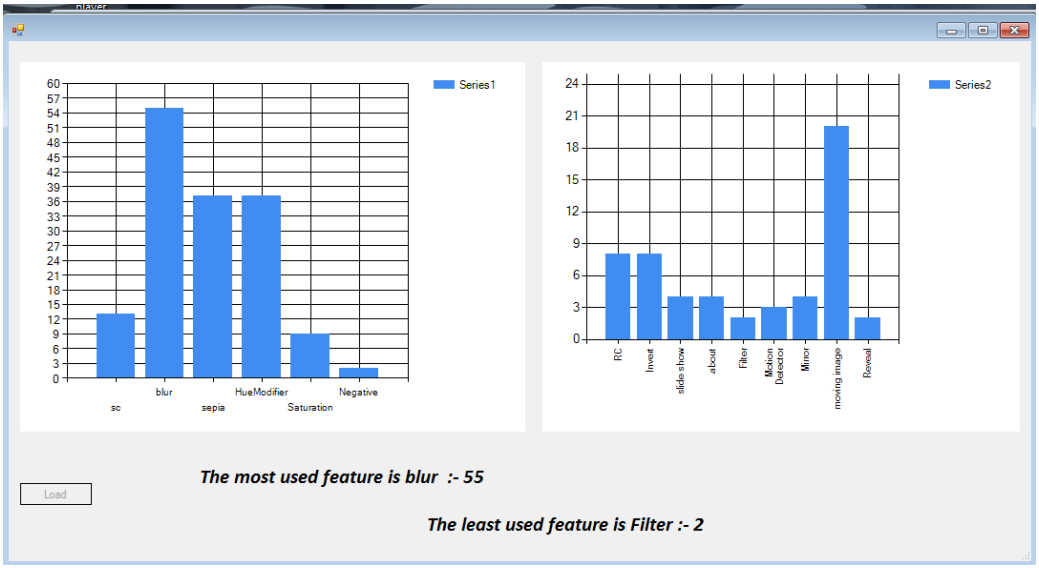


Fig (7): Overall Histogram Screen

We can start a web application through connecting one laptop to another laptop. This is a wirelessly controlling process [2].

REFERENCES

[1] Wang, Yu, Qian Chen, and Baomin Zhang. "Image enhancement based on equal area dualistic sub-image histogram equalization method." Consumer Electronics, IEEE Transactions on 45.1 (1999): 68-75.

[2] Ahonen, Timo, AbdenourHadid, and MattiPietikainen. "Face description with local binary patterns: Application to face recognition." Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.12 (2006): 2037-2041.

[3] Bell, John R. "Group call for a wireless mobile communication device using bluetooth." U.S. Patent No. 6,405,027. 11 Jun. 2002.

[4] Borst, Alexander, and Martin Egelhaaf. "Principles of visual motion detection." Trends in neurosciences 12.8 (1989): 297-306.

[5] Pisano, Etta D., et al. "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated speculations in dense mammograms." Journal of Digital imaging 11.4 (1998): 193-200.

