IJESC

PixelTone: A Multimodal Interface for Image Editing

Satish.S.Bhojannawar[1], Akshata M.H [2], Pratibha S.K[3], Pratibha M.R[4], Trupti .A.H[5]
Associate Professor & HOD[1], Student[2, 3, 4, 5]
Department of CSE
KLE College of Engineering and Technology, Chikodi, Karnataka, India

Abstract:
Image editing can be a challenging task, and it becomes even more difficult on the small, portable screens of mobile devices that are now frequently used to capture and edit images. To address this problem we present PIXELTONE, a multimodal image editing interface that combines speech and direct manipulation. We observe existing image editing practices and derive a set of principles that guide our design. In particular, we use natural language for expressing desired changes to an image, and sketching to localize these changes to specific regions. To support the language commonly used in image editing we develop a customized natural language interpreter that maps user phrases to specific image processing operations.

I.INTRODUCTION

Before going on to the main points, what is image editing? Everyone is familiar with this, but not everyone has a broader knowledge about this topic; Image editing is a process wherein a person is to transform an image into something that is even more creative and expressive, by conveying what the artist wants, rather than settling with the original image. Image editing can be a daunting task with a steep learning curve. Not only interfaces for image editing often complex, but they also expect the user to learn the language of image processing. Users must understand image properties such as hue, saturation levels, and cropping, and learn how they are changed and combined to achieve a desired effect. To add complexity, effective image edits are often localized to a specific region, e.g., to brighten a portion of an image, this task usually requires sophisticated direct manipulation. Such manipulations are easier b with the large displays available in most desktop environments. However, the reality is that image editing is now frequently performed on small, portable devices such ascameraphones, which makes complex interactions even more challenging. Speech interfaces can make complex tasks more accessible because they allow users to simply state goals without first learning an interface. Research on integrating speech interfaces into software applications starting in the 1980s gave rise to today's systems. Popular speech interfaces allow users to efficiently perform complex operations (e.g., sending text messages and getting navigation information). However, image editing is hard to perform with speech alone, since people are not good at describing spatial locations; visual tasks benefit from a combination of speech and direct manipulation interfaces. If we look at how professionals communicate desired image edits, we find a combination of symbols, shapes, and text; a type of communication that we believe would lend itself well to a multimodal interface that combines speech and direct manipulation. Notably, even professional shave very limited shorthand for describing the changes they want to make and frequently resort to more natural instructions. We found that the multi modalinter face more naturally captures both existing work practice and desired functionality. We explore the effectiveness of a multimodal interface that combines speech and direct manipulation for image editing. To guide our exploration, we distill a set of design principles for multimodal image editing applications by observing existing editing practices. We implement these principles in the PIXELTONE, a multimodal image editing system for a computer or mobile device that allows end users to express desired changes in natural language and localize those changes by sketching. PIXELTONE provides users with a powerful mechanism to obtain desired results.

Scope

Image editing can be a daunting task with a steep learning curve. Not only interfaces for image editing often complex, but they also expect the user to learn the language of image processing. Users must understand image properties such as hue, saturation levels, and cropping, and learn how they are changed and combined to achieve a desired effect. To add complexity, effective image edits are often localized to a specific region. These problems can be solved by developing Pixeltone (A multimodal interface for image editing).This system's main goal is to ease the image editing task through voice. The user can tell his desired changes in natural language. The system receives that voice input and converts it into text and then produce the desired results.

II.LITERATURE SURVEY

While natural language interfaces have been used extensively in a number of domains, natural language interfaces for image editing have receive dlittleattention. are techniques that natural language to correct colors in images. However, they focused on global operations and limited the user interaction to natural language only. Comparing speech only, a multimodal interface is more effective for visualtasks.Customer mobile applications mostly support global image edits and none of them include speech interface as an interaction component. Computer vision techniques may make certain types of image editing operations more difficult. The main difficulty with the existing system is the problem of how to translate image editing operations through voice based.

III.METHODOLOGY

We are going to implement PIXELTONE (A multimodal interface for image editing) on a mobile device, using a client-server architecture.
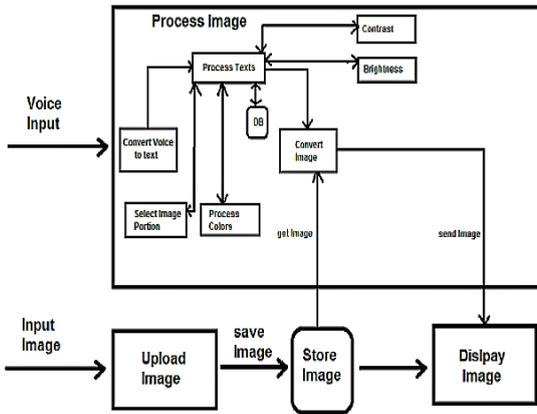
Figure.1. PIXELTONE Architecture.

1.Voice Input
Pixel tone requires a clear voice signal as input to it. All browsers support for MIKE option in it. By enabling MIKE extension in browser we can give voice as input.

2. Image Input
For editing any image we have UPLOAD IMAGE option. Upload image check right extension of image and store it in local folder so that further processing can be done using this file.

3. Convert Voice to text
Google's Web kit Speech Recognition is used to convert input voice signal to text format so that we can process input signals and use it as required.

4. Process texts
Process texts matches converted input signal to available/expected input and send it to different functions to process it further.

5. Select Image portion
Select the co-ordinates of image and store it in database with specific key.

6. Process Colors
Checks input signal and break it four-color changes.

7. Contrast and Brightness
We have specific functions to increase or decrease contrast and brightness of an image.

8. Display image
We use Display image to show final converted image.

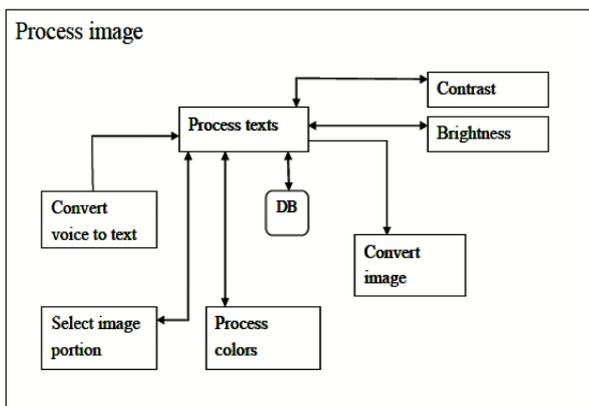## Functionality of the system

Process image

Figure.2. Process image
- *Process texts* matches converted input signal to available/expected input and send it to different functions to process it further.

- *Select image portion* component select the co-ordinates of image and store it in database with specific key.
- *Process colors* checks input signal and break it for color changes.
- We have specific functions to increase or decrease contrast and brightness of an image.

Convert voice to text
The new JavaScript Webkit Speech Recognition API makes it easy to add speech recognition to our web pages. This API allows fine control and flexibility over the speech recognition capabilities. Whenever user speaks something for the purpose of image editing, then Webkit Speech Recognition API receives user's voice input and converts that into text.
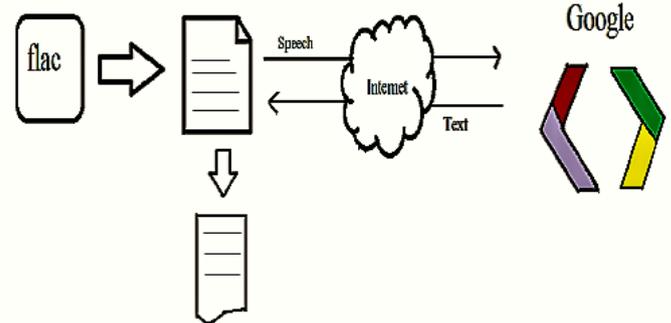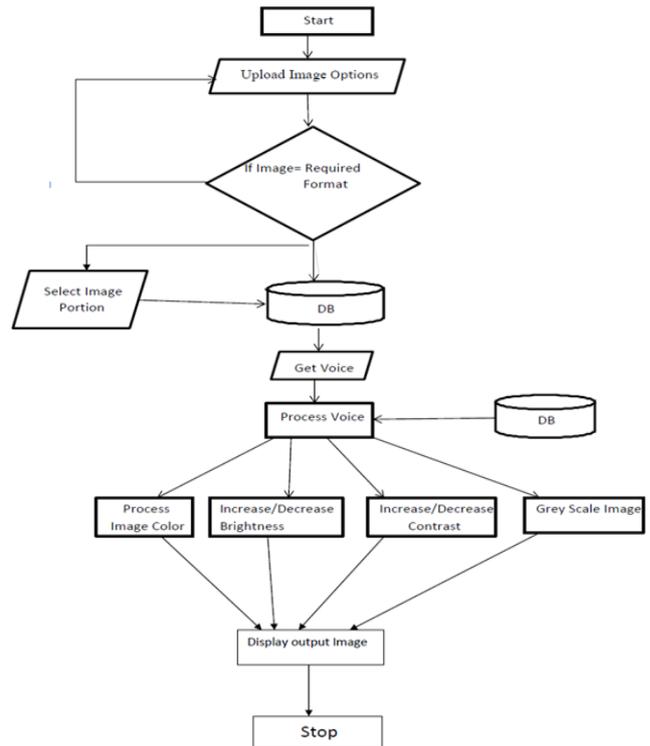

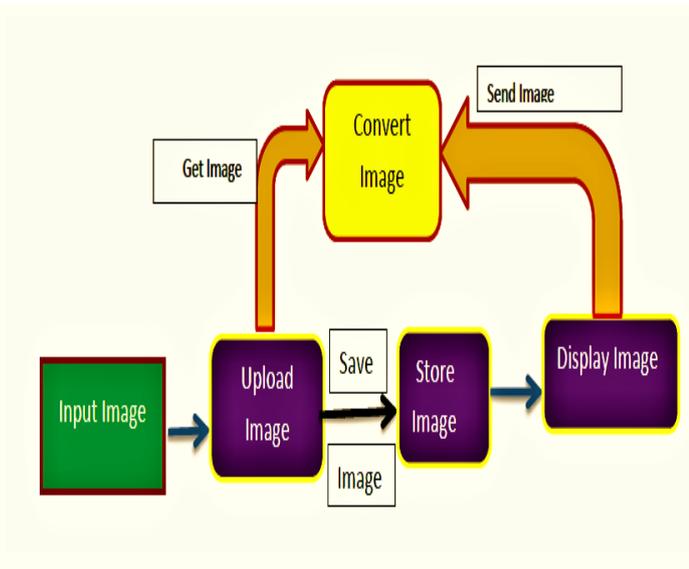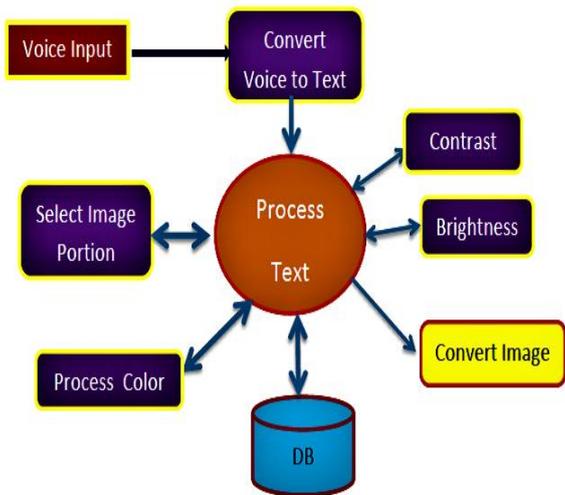**Figure.3. Web kit Speech Recognition**

Data Flow Diagram

System Overview
Pixel tone combines speech and direct manipulation for image editing. Pixeltone allow end-users to express their desired changes in natural language and localize those changes by selecting a portion of the image. This system takes voice signal as input to it. Then it coverts user's speech into text by using Google's "Webkit Speech Recognition" API. Process texts component matches converted input signal to available/ expected input and   send it to different functions to process it further. This system uses Eclipse IDE.

Data Dictionary
Data Dictionary of Process Image

Components

WebkitSpeechRecognition
Takes voice input from user and converts that into text.
JQuery
It retrives the co-ordinates of the selected portion in the image
Database
Selected co-ordinates of image are stored in database with specific key.
Contrast and brightness
It has specific functions to increase or decrease contrast and brightness of an image
Display image
It is used to show final converted image.

Overview of User Interface

First user need to upload the image. Then user can select the entire or portion of the image, then user need to save that image or portion of the image through voice input for example "save image" by clicking on the mike symbol. Then that image will be saved. User can then select the entire or portion of the image and can edit the image through voice. That means user can tell the desired changes by clicking on the mike symbol, for example: " Change the color of the frontal lobe from red to green". The sentence spoken by the user will be displayed on the screen and then user gets the desired result.

Advantages
1.Adding voice to an image editing task reduces time to complete image editing task.

2.Speech interfaces can make complex tasks more accessible because they allow users to simply state goals without first learning an interface

III.CONCLUSION

In this paper we introduce a multimodal system to support image editing tasks through speech and direct manipulation. The system is motivated by a number off or mative studies on how both professional and novice users make use of natural language to indicate the changes they would like to make. Found that the multi modal interface more naturally captures both existing work practice and desired functionality Image editing is an incredibly difficult task, and the shift to mobile devices makes this task harder. Additionally, the language of image manipulation is varied, ambiguous, and subjective. The multimodal provides users with a powerful mechanism to obtain desired results.

IV.REFERENCES

[1].GieradLaput , Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar: PixelTone A Multimodal Interface for Image Editing April 27,2013

[2].Apple.Siri.http://www.apple.com/ios/siri/.Online;accessed Sept 18, 2012.

[3].Androutsopoulos, L. Natural language interfaces to databases - an introduction. Journal of Natural Language Engineering 1 (1995), 29–81.

[4].iSpeech.http://www.ispeech.org/api. accessed Sept 18, 2012.

[5].Pajari, M. Color mark up terminology. Tech. rep., Widen Enterprises Inc., 2009.

[6]. Riley, E. How to talk to a retoucher, 2012.

[7].Milota, A. D. Modality fusion for graphic design applications. In Proc. international conference on Multimodal interfaces

[8].Oviatt, S. Multimodal interactive maps: designing for human performance. Human Computer Interaction 12,1 (Mar. 1997).

[9].Gelfand, N., Adams, A., Park, S. H., and Pulli, K. Multi-exposure imaging on mobile devices. In Proceedings of the international conference on Multimedia, MM '10 (2010).

[10].Pausch, R., and Leatherby, J. H. An empirical study: Adding voice input to a graphical editor. J. of the American Voice Input/Output Society 9 (1991), 2–55.

[11].Samad, T., and Director, S. W. Towards a natural language interface for cad. In Proc. of the 22nd ACM/IEEE Design Automation Conference, ACM/IEEE DAC   (1985).

VI. AUTHORS

Prof.Satish S. Bhojannawar, is Associate Professor & Head of the CSE Dept. at KLECET, Chikodi, Karnataka. His area of interest is Wireless Networks. He can be reached at satish@klecet.edu.in

Ms. AkshataM.H(01373365) is studying   in 4th year of B. E (CSE) at KLECET, Chikodi, Karnataka.

Ms. Pratibha.Kaddu(01373392) is studying   in 4th year of B. E (CSE) at KLECET, Chikodi, Karnataka

Ms. Pratibha.Rachannavar(01373393) is studying   in 4th year of B. E (CSE) at KLECET, Chikodi, Karnataka

Ms. Trupti.Hoogar(01373381) is studying   in 4th year of B. E (CSE) at KLECET, Chikodi, Karnataka