

# Continuity Report

Grocery Shopping Assistance

# Grozi

---

Faculty Advisor: Serge Belongie  
T.A.: Brent Lee

---

Ted Chu  
Jae Shin  
David Ing  
Robert Tran  
Cankut Guven  
Yu-Tseng Chou  
Brandon Lehrman  
Jose Melendez Castro  
Mehran Takapoomanesh Baghaei

UCSD      ENG 100L  
Global Ties      Winter 2010

# Introduction

## Goal and Quarter Approach

GroZi's goal is to create a system to aid the visually impaired in shopping for groceries. The scope of GroZi is currently limited to the shelved section of the store. This project has been developing for some quarters now, but in this quarter it was decided that we would start with a new approach that would allow the project to improve the communication and assistance for the blind user.

This quarter we focused on user interface development using sonification. Sonification is the process of transforming information or data into sounds. For this quarter, there was a division from within the team in order to present better and faster results. The team was divided in 3 sub-teams that were mainly identified by the kind of information they processed and received. These were the 3 teams:

- Image Team
- Sonification Team
- Interface Team

All of the sub-teams needed to have constant communication and feedback in order to process efficiently the information that is being extracted from the user and its environment. Each one of the teams used different platforms and programs to process the information.

As an additional work, a member of our team worked during the quarter with gestural interfaces to provide the general team a better understanding of better ways for a person to interact with a device, like the one we attempt to create. This work is resumed in the last part of this continuity report to provide an insight of what will be studied in depth during the next quarter.

# Image Team

Ted Chu, Brandon Lehrman, Mehran Baghaei

For the winter quarter of 2011, the image team conducted distance transforms of real life images of groceries found in a grocery store. The real life image is taken from a camera and then converted into a segmented image by using the Label Me toolbox and **Matlab**. The image team made the **Matlab** code needed to perform the distance transforms more modular for easy construction and flexible arrangement of variables.

The distance transform labels each pixel of the image with the distance to the nearest “boundary pixel” of a binary image. The metric used for the distance transform is the Euclidean distance. The purpose of the distance transform is to indicate to the sonification team where to apply a specific sound at a certain region.

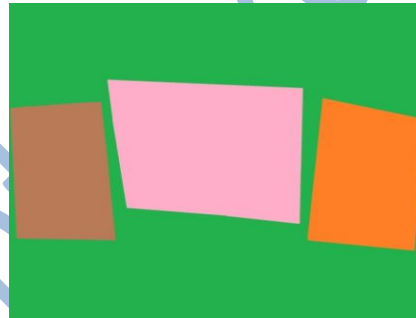


Figure 1 – Real life image of bag of chips      Figure 2 – Segmented image of different brand of chips

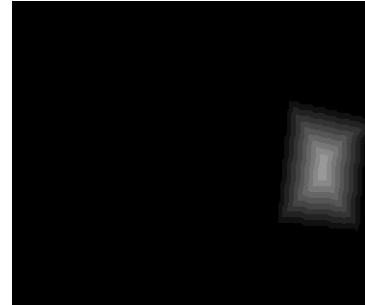
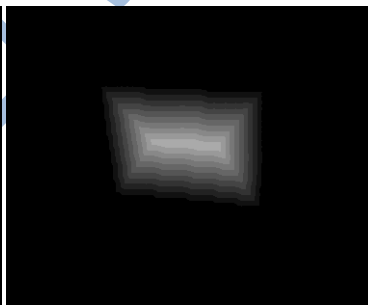
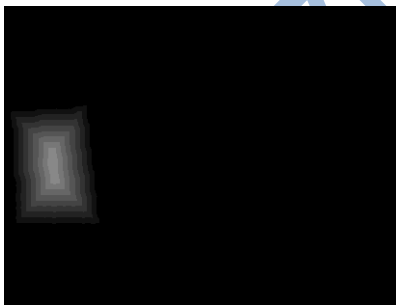


Figure 3 - left chips

Figure 4 – middle chips

Figure 5 – right chips

Figure 1 displays the actual image of the bag of chips within a grocery store. For the winter quarter of 2011, the groceries that are being focused on are the ones that are shelved.

In figure 2, the segmented image shows three different boxes that represents three different regions. The Label-Me toolbox was used to label these different regions in order to segment the image.

Observing Figure 3, 4 and 5, there are three different distance transforms for each region labeled within the segmented image. Figure 3 is for the left region. Figure 4 is for the middle region. Figure 5 is for the right region. Each region signifies a different type of bag of chips. To produce these distance transforms, the **Matlab** code transforms the segmented images of the shelved chips into a binary image where only the item we are looking for is white and everything else black. The code needed to conduct the distance transforms is explained below:

```
Ain = IMRESIZE(Ain,.5);
ImSize = size(Ain);
for i = 1:ImSize(1,1)
    for j = 1:ImSize(1,2,1)
        if((R == Ain(i,j,1))&& (G == Ain(i,j,2))&&(B == Ain(i,j,3)))
            Ain(i,j,1) = 0;
            Ain(i,j,2) = 0;
            Ain(i,j,3) = 0;
        else
            Ain(i,j,1) = 255;
            Ain(i,j,2) = 255;
            Ain(i,j,3) = 255;
        end
    end
end
end
```

```
Bin = Ain(:,:,1);
Object = bwdist(Bin, 'euclidean');
Object= uint8(Object);
imwrite(Object,outFile);
output = imread(outFile);
subplot(2,2,3),imshow(output);
subplot(2,2,3),title('Distance Transform');
```

The code above reads in an image which is passed in and the RGB values, then the code filters out all of the colors except the color that is passed in which represents the color of the object that the distance transform is desired. Then we use **bwdist** **Matlab** library function which calculates the distance of each bright pixel with respect to the dark pixels around it (since the image that is passed in is a binary image), this will result in an image like Figure 3, 4, and 5 above. Sonification team then uses these images and adds sound to it indicating the distance to the center of the object.

# **Sonification Team**

Jae Shin and David Ing

## **Goals:**

- Receive information from image and interface team
- Make Sound Database based on MSRC objects and Grocery Related Items
- Work off old code / fix & debug code

Sonification is defined as the use of nonspeech audio to convey information. A main task of the Sonification team was to work with the process of transforming data or information into sound. This quarter, our primary goal was to make a sound database and to fix and develop old code. So, in the beginning of this quarter, we were working to get familiar with the ChucK programming language using miniAudicle. Once we got familiar with the chuck, we decided move on to the task of making a sound database. We were able to successfully finish the sound database based on MSRC objects and Grocery related items but since we got the old code almost at the end of the quarter, we could not work with the code as much as we thought.

## **Sound Database:**

We downloaded 600 image files that are for research purpose from the MSRC (Microsoft Research in Cambridge) and created almost 1000 corresponding sound files in mp3 format. Since each image file contained more than one image we had to make 2 to 3 sound files for each image file. We've also made a database for grocery related items since our main goal is to assist visually impaired people in grocery shopping. While we were creating the database, we thought that it would be better if we could make sound files containing sound effects. Idea was to create a sound file based on characteristics of the item along with saying out the name of the item. This technique of showing the sound effects would be part of 'auditory icon' which is a very natural sound that represents an object or an action. This is something what we are planning to achieve next quarter.

## Debugging Code:

The previous quarter sent us the code they had worked on for picture sonification. Specifically there were two files/programs meant to simulate data from the other teams and two files/programs meant to receive the data and make sounds.

There were two different processes of making sounds based on the information: XY object and Distance transform. (For further information, refer to Continuity Report of Fall 2010.)

For both these programs there were certain inherent bugs in the code/design that we had to rectify.

Moving cow problem:

The “Moving Cow Problem” was a problem with the program that simulated the sending of Data. When the user would click around to simulate the touch screen, the program’s understanding of the location of the objects in the picture would move around. For example, in this picture, the cows would keep moving around any time the mouse was clicked.

Through examining the algorithm, we found a problem with the recalculation of the positions and we fixed this by counterbalancing the change with an offset. There is still a minor tracking problem that we hope to fix in the future, but the moving cow problem is fixed.



## **X-axis problem**

The code only dealt with changes in the X-axis and completely forsook the Y axis.

We developed code that worked with the Y-axis as well in anticipation of the full integration of the User-Interface, Image, and sonification team

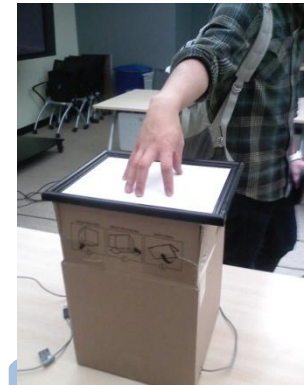
## **Future Ambition:**

One of our future goals is to make a sound database more toward sound effects to give better guidance to those who are visually impaired. We believe adding sound effects to current speech sound could definitely give more specific details to blinds. While we were researching for the sonification techniques, we found out about two techniques called 'Auditory Icons' and 'Earcons'. The definition of Auditory Icon is a very natural sound of the objects or the actions. Good examples would be the sound of animals or the sound when we open the soda can. Earcon is a brief, distinct sound used to represent a specific event. Earcon could helpful when we use distance transformation. Currently, we use a 'bip' sound when the fingers approach the desired object. If we can add a feature of earcon when the user picks up the item it would be very helpful. We hope to develop a system similar to the interface of City Maps.

Another goal is to join all the programs of the different subteams together using Open Sound Control (OSC). There are measures for receiving OSC packets inherent in the programs already

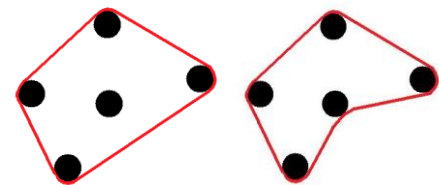
## Interface Team

Last quarter, we had a multi-touch screen that consists of a webcam, a box and a frame and a program in C# that outputs the name of the object each finger is touching on an image (Figure 1). Building off from last quarter, we implemented the code with a new gesture using the idea of a convex hull this quarter.



**Figure 1.** The multi-touch device

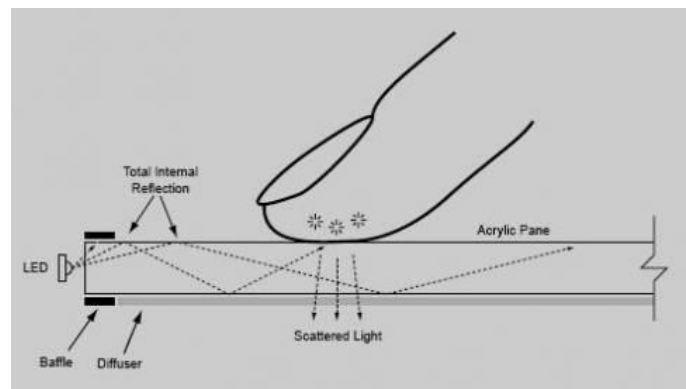
When 3-5 fingers are detected, we assume that the user is attempting to locate or search for some object. We then implemented a “searching mode” which follows the three main steps below: approximates a polygon by the positions of the finger touches, checks the color of each pixel within the polygon and lastly outputs the object with the highest



**Figure 2.** Example of convex and concave hull

percent area. The idea of a convex hull, the boundary of a point set, helps drawing a polygon efficiently since it includes every single point while using the least amount of “string” possible. Two figures on the right illustrate an example of a convex hull (left) and a concave hull (right) and how drawing a convex hull is more efficient (Figure 2). The next possible approach follows a scan line algorithm to mark/fill the region within the polygon and to reduce comparison time.

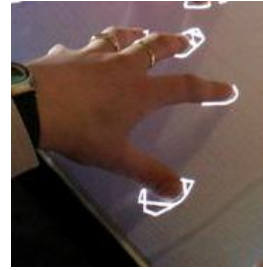
As our project progresses, we decided to build a new multi-touch prototype closer to the final product.



**Figure 3.** A diagram that illustrates the FTIR technology



The new prototype should ideally produce more accurate finger blobs, be more durable while still remain affordable. We will utilize FTIR technology (Frustrated Total Internal Reflection) to track finger touches (Figure 3). This approach involves feeding infrared LED lights through a sheet of acrylic. When the user touches the surface of the acrylic, the IR light will scatter, thus indicating the instance of contact. Furthermore, we will utilize a PS3 eye camera equipped with a 850nm bandpass filter to specifically track IR light and ignore all else. This will essentially yield finger blobs with higher contrast (Figure 4).



**Figure 4.** Ideal image of a new multi-touch prototype with high

The code for the project can be accessed at <http://code.google.com/p/grozi/>.

GROZI Winter 2010

## **Future plans:**

- Sub-team integration.
- Making and designing a new touchpad that is robust and cheap.
- Test the whole system as one.
- Exploring other options.

GROZ Winter 2010

## **Acknowledgements**

- Dr. Serge Belongie
  - Our Faculty advisor
  
- Dr. Mandy Bratton
  - Program Director
  
- Barb Donovan
  - Program Specialist
  
- Brent Lee
  - Teacher Assistant
  
- Zach Salin
  - Global TIES Fellow

GROZI Winter 2010