International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-6, September 2015
ISSN: 2395-3470
www.ijseas.com

OPTICAL SENSE PERCEPTOR FOR VISUALLY CHALLENGED

Mr. A. Muthu Rathinam[#1],Mr. P.Suresh kannan[#2],Mr. M.Venkatesh[#3], Mr. E. Ganessh Kumar[#4],

Mr. M. Ashok kumar[*5]

[#]*Student,*[*]*Assistant Professor*
Department of Electronics and Communication Engineering
Chandy College of Engineering
Tutucorin(T.N) India

Abstract

This paper talks about a simple and economical solution to satisfy the basic need of visually challenged, by making them more independent. This technology works with the theory of robotics that helps them access any book and it also translates arbitrary video images from a camera into sound. We can also includes GPS that helps them to point out their location at any instance.

Keywords: *Author Guide, Article, Camera-Ready Format, Paper Specifications, Paper Submission.*

1. Introduction

Sense of sight- God's greatest gift to his most beloved invention– 'men'! It's an irony that some of our contemporaries are deprived of this wonderful ability. This paper aims to bring a new life to the people so called 'blind'. The Opto-Sono Converter (OPTICAL SENSE PERCEPTOR) is a robotic device that consists of camera which is attached to the glasses which the person wears and is placed near forehead. This in turn is connected to a portable device comprising of embedded systems that processes the necessary information, which is converted to sound that can be heard through the stereo earplugs. The Opto-Sono Converter (OSC) works in 2 modes. The ANALYSIS and READ mode. In the analysis mode, the processor converts the image to sound thus leading to synthetic vision with sensations by exploiting the neural plasticity of the brain through training. In the Read mode, the camera takes picture of the page that is to be read and converts the text to speech with the help of embedded technology.GPS can also be attached, to help them find out where they are and feedback is given in the form of audio. It gives them a sense of distance, direction, evaluation, size and visual texture, not just for a single object but also for multiple objects. Thus the OSC helps the visually impaired to read, and navigate without human intervention most visually challenged people use a cane that is used to extend the user's range of touch sensation. I-Canes are sophisticated canes that direct the person based on the obstacles found in the path. It sends out ultrasonic waves and based on the reflected wave the cane directs the person. One main disadvantage of these canes is that it blindly picks a direction to avoid an obstacle which in turn might mislead a person. The primary advantage of Braille is that it allows users to read in their preferred manner. However, their sheer bulk makes Braille books too cumbersome to store. Another problem is the limited number of books available in Braille. Cassette tapes provide a means for storing and accessing audio recordings of information in the first place, tapes provide access to the user in a strictly linear manner. Neural implants are a breakthrough among the technologies invented for visually challenged, but there are many medical constraints and is very expensive. It also involves a lot of risk as the life expectancy of the individual is at stake. Thus none of these technologies provide an efficient solution to the problems faced by the visually impaired, and it causes more problems than it solves.

1.1 OUR IDEA:

Our idea is to propose a solution that is both cost effective and will solve most of the hurdles faced by the visually challenged people. It will make them more independent and self-secured. Using the OSC they can read any book provided the camera captures the page that is to be read. The text from the image is extracted and the corresponding ASCII codes are generated which is converted to audio with help of inbuilt vocabulary to speech converter. To access the text beyond the existing linear manner, the OSC is attached with attributes that help them to play, pause, move forward, and back. Other than reading the text, it also

works in another mode that is known as the analysis mode in which any picture is converted into its corresponding frequency range that is played in the earplug. The person is trained to these sounds and thus can figure out what his surroundings look like. In short, the processor converts the images to sound thus leading to synthetic vision, by exploring the neural plasticity of the human brain without training. It scans each camera snapshot from left to right, while associating height with pitch and brightness with loudness. The brighter an object, the louder is its sound. With the employed stereo output, objects on the left or right, sounds on the left or right, respectively. For a given column, every pixel in this column is used to excite an associated oscillator in the audible frequency range. This robotic device can be muted at any moment, to hear external sounds. GPS can also be attached along with this to help the person to find out exactly where they are and, depending on their location feed back is given in the form of audio.

1.2 BRAIN FEATURES:

Various sections in the human brain govern each organ in the body. The occipital cortex governs the sense of sight. The iris of the eye is connected to this occipital cortex through the optic nerves. Any damage or deformation to any part might lead to irreversible and undesired effects like partial or total blindness. The retina of the eye also plays a vital role as it is the retina that captures and recognizes the image seen by the person. Under normal conditions, the occipital cortex receives predominantly visual inputs but perception is also highly influenced by cross-modal sensory information. (b) Following visual deprivation, neuroplastic changes occur such that the visual cortex is recruited to process sensory information from other sense. (c) After neuroplastic changes associated with vision loss have occurred, the visual cortex is fundamentally altered in terms of its sensory processing, so that simple re-introduction of visual input (by a visual prosthesis; orange arrow) is not sufficient to create meaningful vision. (d) To create meaningful visual percepts, a blind person can incorporate concordant information from remaining sensory sources.

The various sensory organs receive information at a high rate, which is sent to the brain. The brain of the visually impaired people does not respond or detect the signals from the eye. The frame rate of the eye is high and the decoding process is complicated. The visually challenged people are more sensitive to sound and touch, as the brain does not have to spend time to the signals received by the eye and the pay more attention to these other sensory organs.

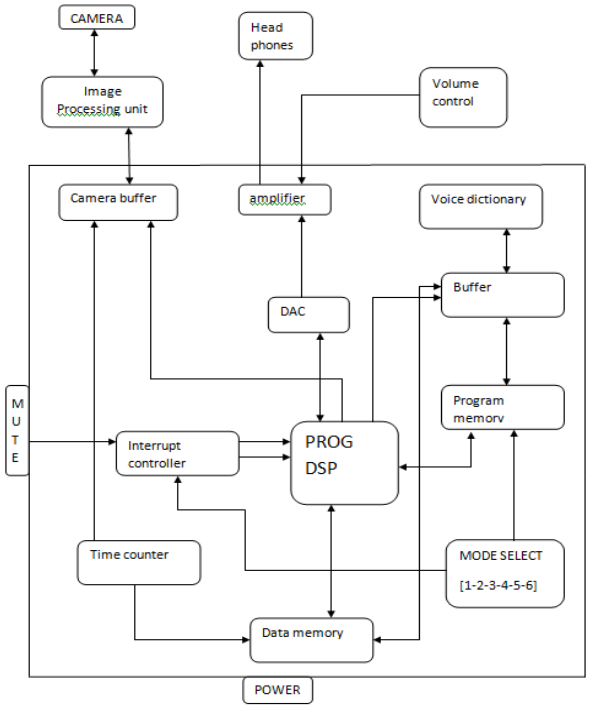
1.3 OUTLINE MODEL:

As shown in the figure the camera is attached to the glasses and high power LEDs are also attached to enhance the images captured in dim light. This in turn is connected to a portable processor comprising of embedded systems that processes the necessary information. The processed data is converted to sound, which can be heard through the stereo earplugs. The processor consists of 7 buttons to activate the following, which are READ mode, ANALYSIS mode, GPS activation, FORWARD, PLAY/PAUSE, BACK, and MUTE. These buttons have projections, for the visually challenged to distinguish them.

1.4 INTERNAL ARCHITECTURE

The main blocks of the OSC are as shown in the figure.The camera is used to capture the surroundings, which has special provisions for opening and closing the aperture to protect the C-MOS sensors. This is in turn connected to the image processing unit, which is used to control the aperture opening and histogram equalization, so that the reflective surface errors can be avoided. The images are stored in the camera buffer and are sent to the data memory at n interval of 0.1 seconds by the time counter. The programmable Digital Signal Processor takes the input from the data memory, processes it and sends it to the respective blocks depending on the mode selected. The DSP is interfaced with the program and data memory, in which the routine for the modes are loaded beforehand. The DAC unit converts the digital signal sent by the processor into corresponding analog signals, which are amplified and played in the stereo earplugs. The external switch controls the volume of the amplifier.

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-6, September 2015
ISSN: 2395-3470
www.ijseas.com



In the read mode the images to be extracted is stored in the data memory and the images are transferred to the processor, which in turn processes the data with the help of the routines loaded in the program memory. The text is extracted and the corresponding ASCII values are stored in the data memory with the help of queue. When the play button is pressed, the signal is sent to the DSP, which in turn extracts the data from the data memory and sends it to the buffer. This buffer sends the data to the voice dictionary, which compares the ASCII values with the inbuilt vocabulary and plays the note if the code matches; else the word is spelt out. The mode selection unit is used to select the mode of operation of the OSC. The interrupt controller is used to control the working of the processor in read, analysis mode, or dormant mode (mute). The power section supplies the power for the whole circuit with the help of batteries.

1.5 MATHEMATICAL MODEL OF ANALYSIS MODE:

Any image can be represented by a M x N matrix comprising of 'm' rows and 'n' columns. The OSC takes the gray scale image for analysis purpose. In gray scale images 255 represents a white pixel and 0 represents a black pixel and intermediate values represent the percentage of black and white. For analysis purpose, the resolution is taken as 320 x 240.

Each row is assigned with a frequency of 500 to 10,000 Hz linearly.

$$\alpha_1 + (t * x) = 500 \; at \; x = 1$$

$$\alpha_1 + (t * x) = 10000 \; at \; x = 240$$

By solving these simultaneous equations, we get,

$$\alpha_1 = 460.25$$

$$t = 39.748$$

By substituting the values on 'n' in place of 'x', we get the corresponding frequency for each row using this formula

$$f = 460.25 + (39.748 * x)$$

Assuming that the images stored is 'x' the value of each pixel is represented by x (m,n).
If x (m,n)=0, the corresponding amplitude is 0. If x (m,n)=255, then the amplitude is 15.

$$A = \beta * X$$

By solving this we get

$$\beta = 0.058823$$

Thus the amplitude of the sine wave can be determined by the equation,

$$A = (0.058823 * X)$$

The frequency and amplitude which is found from equations 1 and 2 is substituted in

$$S = A * \sin(2 * \pi * f * t)$$

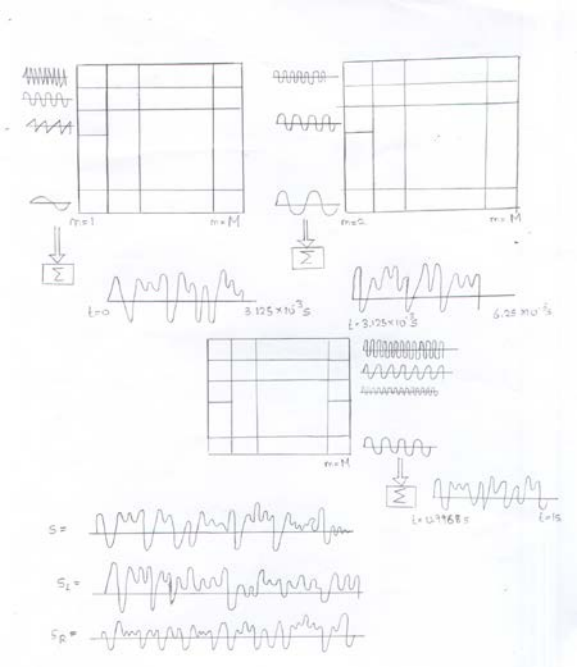
$$s = (0.058823 * x) * \sin(2 * \pi * (460.25 + (39.748 * x)) * t)$$

And thus a sine wave is plotted. And this is done to obtain sine wave for every pixel. Each column has a particular time period, and the whole image is processed in 1 second. Thus for accessing 1 column we get 3.125*10^-3 seconds. Each column has a sample of 138 values.

1.6 ANALYSIS MODE ALGORITHM:

- The image is stored in the data memory and the counter determines the frame rate of the camera.
- The image is analyzed from bottom to top along the column and from left to right.
- The last pixel in the first column is analyzed and converted to sine wave whose frequency and amplitude are determined by the position and value of the pixel.

- Similarly a wave for each column is generated and summed to form a vector. The line period for each column is 3.125*10^-3 seconds and thus for the whole picture it is 1 second.
- Using the left and right channels, a virtual 3-dimensional image is provided through sounds.
- Both the channels have to coordinate simultaneously. Thus when the amplitude increases on the left earplug the corresponding amplitude decreases on the right earplug, and vice versa.



1.7 NECESSITY FOR READING MODE IN OSC:

Though we seldom stop to think about it, sighted individuals are continually bombarded every day by the printed world. Some of the sources of this abundance of printed media in our environment include transportation, advertising, news and commercial signs. However, this is a phenomenon that people who are blind currently do not experiment. Most of their vision troubles prevent them from having access to textual information. Even the process of eating out is complicated by the fact that few restaurants have menus in Braille. All of these facts underscore the necessity for a portable, autonomous, small-size and easy to use automatic text reader. With the emergence of multimedia technology and powerful mobile devices it is now possible to imagine an inexpensive system able to capture images in real time and transform image into speech transformation.
Optical Character Recognition (OCR) is currently developing algorithms to characterize the visual content of images and to recognize text. Our objective is to make these algorithms working in real time into a device devoted to helping people who are blind or visually impaired. While several devices have been developed in the past to assist the reading of printed text, they have all fallen short of the user expectations. Most have been too cumbersome or not readily available to be practical and truly portable. Sometimes they even create more problems than they solve.

1.8 TEXT DETECTION:

Traditionally, document images are scanned with a flatbed,sheet-fed or mounted device. However, digital cameras have shown their potential as an alternative imaging deivce. But camera-based images require specific processing. The first is detection and localization of the text regions. The idea is to locate the text elements without necessarily recognizing them, cut them out of the image,determine the reading order and finally correct their perspective. The read mode uses the principles of OCR, which deals with the recognition of the printed text and storing it in the coded standards the most intuitive characteristics of text are its regularity. Printed text consists of characters with approximately the same size and line thickness that are located at a regular distance from each other. Such regularities can also be observed from edges being detected on textual boundaries,

1. Text possesses cetain frequency and orientation information.
2. Text shows spatial cohesion –characters of the same text string(a word, or words in the same line) are of similar heights, orientation, and spacing. Characters contrast with their background since they are designed to be read easily. Characters appear in clusters at a limited distance aligned to a virtual line. Most of the time the orientation of these virtual lines is horizontal since that is the conventional way of writing.

Text detection techniques can broadly be classified as edge[5][6],color[7][8],or texture-based[9][10]. Edge-based techniques use edges information in order to characterize text areas. Edges of text symbols are typically stronger than those of noise or background areas. These methods operate essentially in grayscale format and do not require much processing time. Nevertheless, they do not cope with complex text images like pictures of magazines or scene images where edge in information alone is not sufficient to separate text from a noisy background. The use of color information allows the image to be segmented into connected components of uniform color. A reduction of the color palette is often required. The main drawbacks of this approach are the high color processing time and the high sensibility to uneven lighting and sensor noise.

Texture-based techniques try to capture texture aspects of text. In our approach, the document image consists of several different types of textured regions, one of which from the text-content in the image. Thus, we pose the problem of locating text in images as a texture discrimination problem. Our method for texture characterization is based on.
Gabor filter which have been used earlier for a variety of texture classification and segmentation tasks. We use a subset of Gabor filters proposed by Jain and Farokhnia associated with an edge density measure. These features are designed to identify text paragraphs. Each individual filter will still confuse text with non-text regions but an association of filters will complement each other and allow text regions to be identified unambiguously. We use a reduced K-means clustering to cluster feature vectors. In order to reduce computational time. We apply the standard K-means clustering to a reduced number of pixels and a minimum distance classification is used to categorize all surrounding non-clustered pixels. Emprically, the number of clusters (value of K) was set to three , a value that works well with all test images. The cluster whose centre is closest to the origin of feature vector space is labeled as background while the furthest one is labeled as text.

1.9 PERSPECTIVE CORRECTION:

As previously suggested the user disabilities introduce specific constraints for text recognition task. A common problem relates to the position between the" text object" and the camera. The user cannot be certain that the document and the camera are placed facing each other. Documents that are not frontal parallel to the camera's image plane will undergo a perspective distortion. In general supposing that the document itself is on plane , the projective transformation from the document plane to the image plane can be modeled by a 3 by 3 matrix in which eight coefficients are unknown and one is a normalization factor. The removal of perspective can be achieved once the eight unknown's have been found. The early optical character recognition (OCR) systems were often tested against document's under ideal conditions. Usually, images were electronically converted from paper with a scanner, so the surface texture was fairly even, lighting was well distributed and the image captured at an overhead angle. If any of these variables were to be slightly altered, however many of the assumptions that made these systems successful would not apply. For better character recognition the following sections talks about the steps to be undertaken.

1.10 CHARACTER SEGMENTATION:

The following steps talks about character segmentation which is used to separated each character from the other. Other possibilities based on segmentation free recognition can be applied . Moreover they are often more accurate with many assumptions, heavy and

not convenient for our embedded platform. Therefore character segmentation still has to be improvised. The traditional connected component algorithm devised by Rosenfeld and Platz, takes advantage of divisions between regions by labeling them as distance objects.

Another major point in character segmentation is broken and touching characters. In normally seen images the first category seldom of course because recognizable characters are thick enough. Nevertheless touching character are often present in these kinds of images due to perspectives like size, thickness,blur,etc…

We have already minimized the number of touching characters. For the ones left, a very common tool is the caliper distance between the uppermost and bottom most pixels in each column to find the place where to separate the components.

1.11 GLOBAL POSITIONING SYSTEM:

The global positioning system(GPS) is a worldwide radio-navigation system formed from a constellation of 27 satellites (24 working and 3 substitutes in case of emergency) and their ground stations. The orbits are arranged so that at any time anywhere on earth, there are atleast four satellites tracking the receiver.

A GPS receiver's job is to locate four or more of these satellites figure out of the distance to each, and use this information to deduce its own location. This operation is based on a simple mathematical principle called trilateration. All GPS is a distance system. This means that the only thing that the user is trying to do is determine how far they are from any given satellite. There is no inherent vector information which implies azimuth and elevation , in the GPS signal. All that the GPS satellite does is shoot out a signal in all directions although there is preferential orientation towards the earth/

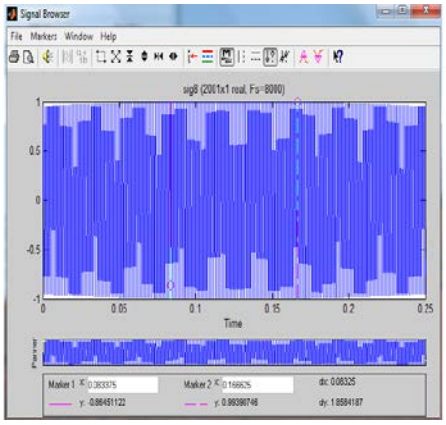
In essence the GPS operated on the principle of trilateration. In trilateration the position of an unknown point is determined by measuring the lengths of the sides of a triangle between the unknown the unknown point and two or more known points. This is opposed to the more commonly understood triangulation, where a position is determined by taking angular bearing from two points a known distance apart and computing the unknown point's position from the resultant triangle.

Therefore, the only thing needed by the user to calculate distance from any given satellite is a measurement of the time it took for a radio signal to travel from the satellite to the receiver. By this method the GPS unit get the latitude and longitude location of the users position. Then this coordinates is given to coordinate locater superimposes the coordinates on the map and find the location where the user Is and returns it back to GPS unit. The GPS unit will read the location aloud so that blind people can able to navigate their path.

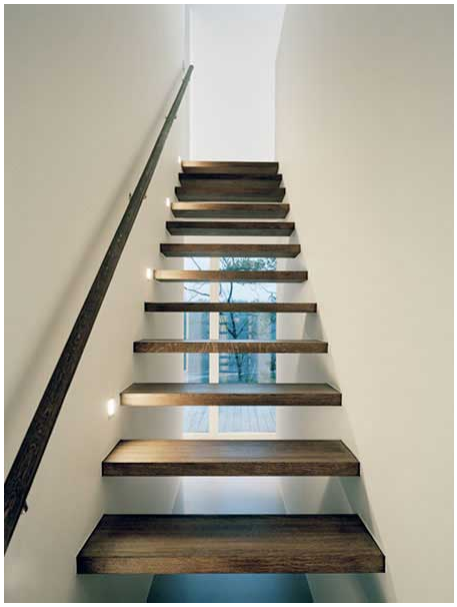
1.12 SIMULATION RESULTS:

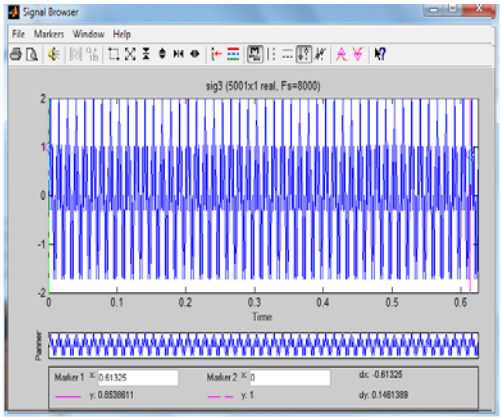
The following waveforms show the simulation results of the pictures analyzed using MATLAB . In this image the variations for an obstacle and a depression can be clearly observed.

The change in the frequency can be noted from the images and without obstructions. The first image shows an image with car present.



The following is the waveform of a image consisting of a staircase





1.13 CODING FOR CONVERTING IMAGE TO TEXT:

```
function
image2text(input_image,output_text);

y=input_image;
img=imread(y,'240','320'); %getting the
gray scale value for each pixel

 [m n r]=size(img);

char2=[];
for(i=1:m)
for(j=1:n)
char1=[];
for(k=1:r)
    char1=[char1
imgnum2str(img(i,j,k))];
end
char2=[char2;char1];
end
end

char_temp=[];
if(r==1)
    char_sign='.';
end
if(r==3)
    char_sign='/';
end
for(i=1:2*n-2)
    char_temp=[char_temp char_sign];
end
char_temp=[imagenum2strall(m)
char_temp];

char2=[char_temp;char2];

stringing=char2;

for(i=1:length(y)-4)
```

```
    name_file(i)=y(i);
end

filename=output_text;
f = fopen(filename,'wt');
if(f==-1)
    disp('error  writing  to  the  text
file')
end
if(f~=-1)
   for dion=1:3*m+1

fprintf(f,'%s\n',stringing(dion,:));
%\n is similar as in C i-e 'enter' at
the end of one row(see help fprintf)
   end
  fclose(f);              % closing the
file as in C++
end


function strg=imgnum2str(no)

base=16;

a1=rem(no,base);
nom=no;

for(i=1:1)
num1=floor(nom/base);
a1=[a1 rem(num1,base)];
nom=num1;
end

a2=fliplr(a1);

char1=[];
for(i=1:length(a2))
    if(a2(i)<=9)
        char1=[char1                  '0'
int2str(a2(i))];
    end
    if(a2(i)>9)
        char1=[char1 int2str(a2(i))];
    end
end

str1='0123456789';
str2='ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefg
hijklmnopqrstuvwxyz!.@#$%^&*()<>:"{}_+-
=,;'`~';

strg=[];
for(j=1:2)
for(i=1:16)
if(char1(2*j-1)==str1(2*i-1))
    if(char1(2*j)==str1(2*i))
        strg=[strg str2(i)];
    end
end
end

function char=imagenum2strall(num);

strg1='abcdefghijklmnopqrstuvwxyzABCDEF
GHIJKLMNOPQRSTUVWXYZ0123456789
#$%&();@[]{}-`^,"*+_=><|?.:\/!~';
lens=length(strg1);

if(num<=lens)
    y=num;                    %the text
values are ascii coded.
end
if(num>lens)
    x=floor(num/lens);
    y=num-lens*x;
end
char=[strg1(x+1) strg1(y+1)];
```

Gabor filter:

```
                    function
gb=gabor_fn(sigma,theta,lambda,psi,gamm
a)

sigma_x = sigma;
sigma_y = sigma/gamma;

% Bounding box
nstds = 3;
xmax                                  =
max(abs(nstds*sigma_x*cos(theta)),abs(n
stds*sigma_y*sin(theta)));
xmax = ceil(max(1,xmax));
ymax                                  =
max(abs(nstds*sigma_x*sin(theta)),abs(n
stds*sigma_y*cos(theta)));
ymax = ceil(max(1,ymax));
xmin = -xmax; ymin = -ymax;
[x,y] = meshgrid(xmin:xmax,ymin:ymax);

% Rotation
x_theta=x*cos(theta)+y*sin(theta);
y_theta=-x*sin(theta)+y*cos(theta);
```

```
gb=exp(-
.5*(x_theta.^2/sigma_x^2+y_theta.^2/sig
ma_y^2)).*cos(2*pi/lambda*x_theta+psi);
```

1.14 CONCLUSION:

The OSC is hoped that producing sounds from the images, may lead to visual experiences ,which truly have the feel of vision. It can be used to built a more complete mental map of the environment. It gives them a sense of distance, direction, evalution,size and visual texture and not just for a single object but also for multiple objects and landmarks make up the surroundings environment. Thus the practical implications of the OSC would exposed the visually challenged to the real world around them making them more confident and independent.

BIBLIOGRAPHY:

1. **Mr. A. Muthu Rathinam** is pursuing, UG in the discipline of Electronics and Communication Engineering at Chandy College of Engineering, Tuticorin, under AnnaUniversity, Chennai, India.

2. **Mr. P. Suresh Kannan** is pursuing, UG in the discipline of Electronics and Communication Engineering at Chandy College of Engineering, Tuticorin, under AnnaUniversity, Chennai, India.

3. **Mr. M. Venkatesh** is pursuing, UG in the discipline of Electronics and Communication Engineering at Chandy College of Engineering, Tuticorin, under AnnaUniversity, Chennai, India

4. **Mr. E. Ganesh Kumar** is pursuing, UG in the discipline of Electronics and Communication Engineering at Chandy College of Engineering, Tuticorin, under AnnaUniversity, Chennai, India

5. **Mr. M. Ashok Kumar** is working as a Assistant professor of Electronics and Communication Engineering at Chandy College of Engineering, Tuticorin, under AnnaUniversity, Chennai, India