

# Image Recognition by Sound Pattern Generated by the Image

Zubair Khan  
Invertis Institute of  
Engineering and technology,  
Bareilly, India

Saurabh Singh  
Invertis Institute of  
Engineering and technology,  
Bareilly, India

Gaurav Agarwal  
Invertis Institute of  
Engineering and technology,  
Bareilly, India

## ABSTRACT

In this paper we are proposing algorithm for image recognition by sound patterns generated by the image. There are five senses in human sight, smell, taste, touch, and hearing. Sight is probably the most developed sense in humans, followed closely by hearing. Blindness is the condition of lacking visual perception due to physiological or neurological factors. For blind persons their ears work as their eyes. This paper presents a novel idea of generating an audio signature from an image which can be used to visualize the image. After some training sessions the user (may be blind person) will be able to recognize the image. This paper is divided in seven parts the first part of paper give the introductory second part deals with related work the important part of the paper is methodology and propose algorithm the last part is conclusion.

**Keywords:** Image to sound , Listen the image, image processing, sound generator, Sensory replacement.

## 1. INTRODUCTION

Some blind people navigate within their environment by listening echoes from the environment. They create some sounds by tapping their sticks, this ability is termed as echolocation. A person can be trained to navigate the environment by echolocation. Various characteristics like location and size of the nearby object can also be gathered by echolocation. various spatial information can be conveyed by echoes and other sounds. A blind person can use sound to recognize an object better than a person having sight because he or she can better concentrate towards sound as his brain does not waste its energy in vision. So if somehow we generate sound pattern from a given image then that pattern can be used to recognize the image. In this paper we are generating sound according to the orientation and length of the edges present in the image. This system can be used by a blind person to see (listen) an image and he or she can also navigate in the environment without a stick.

## 2. RELATED WORK

In 1940-1950 a device was developed by Dr. Cooper at Haskins Laboratory which was used to play a tone corresponding to a pattern and the tone could be used to recognize the pattern[2][3][4][5]. In daily life we see various examples of identifying characteristics of an object by the sound generated or echoed by the object. Our idea is inspired by this novel human ability of perception [1]. The traditional way of dealing with blindness and deafness has been some form of sensory substitution - allowing a remaining sense to take over the functions lost as the result of the sensory impairment. With visual loss, hearing and touch naturally take over as much as they can, vision and touch do the same for

hearing, and in the rare cases where both vision and hearing are absent (e.g., Keller 1908)[6][7][8][9][10].

## 3. PROPOSED METHODOLOGY

The proposed system basically deals with the production of sound through an image:

**The Method works as follows – the method is divided in two phases**

### A) Phase 1-

- 1- Detection of horizontal, vertical and diagonal edges of the image (using Sobal operator) .
- 2- Find length and orientation (Horizontal, Vertical and diagonal) of each edge and record it to be used in second phase for sound generation.

### B) Phase 2-

- 1- Three audible frequencies are selected and assigned to three kinds of edges ( Horizontal, vertical and diagonal).
- 2- Now read the recorded data (orientation and length of the edge and play the sound by selecting the frequency according to the orientation and duration according to the length of the edge.

- Edge Detection
- Sound Production
- Assignment of sound to the detected edges to generate specific sound signature.

### 3.1 Edge detection in an image

This is the first module. Edge detection is a pre-processing step for our system, to refer to algorithms which aim at identifying points in a digital image at which the image brightness changes sharply or more formally has discontinuities. The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the objects in the image.

- Discontinuities in depth,
- Discontinuities in surface orientation,
- Changes in material properties and

- Variations in scene illumination.

An image is a set of pixels which are arranged in the form of matrix. An image consists of horizontal, vertical and diagonal edges. In our work sobel operator is used to detect the edges of the image. Next section explains the function of sobel operator.

### 3.1.1 Sobel Operator for edge detection

The Sobel operator [fig-1] is used in image processing, particularly within edge detection algorithms. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel operator is either the corresponding gradient vector or the norm of this

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Fig 1. Sobel Operator

The magnitude of the gradient is then calculated using the formula:

$$|G| = \sqrt{Gx^2 + Gy^2}$$

An approximate magnitude can be calculated using:  
 $|G| = |Gx| + |Gy|$

The GX mask highlights the edges in the horizontal direction while the GY mask highlights the edges in the vertical direction. After taking the magnitude of both, the resulting output detects edges in both directions[fig-2][fig-3].

Detect edges using the Sobel method



Fig 2



Fig 3

vector. The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations.

The Sobel operator performs a 2-D spatial gradient measurement on an image. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. The Sobel edge detector uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows). A convolution mask is usually much smaller than the actual image. As a result, the mask is slid over the image, manipulating a square of pixels at a time. The actual Sobel masks are shown below:

The Sobel operator represents a rather inaccurate approximation of the image gradient, but is still of sufficient quality to be of practical use in many applications. More precisely, it uses intensity values only in a 3x3 region around each image point to approximate the corresponding image gradient, and it uses only integer values for the coefficients which weight the image intensities to produce the gradient approximation.

**3.2 Sound production:** In this module we have generated sound using MATLAB's sound function. MATLAB can send data to our computer's speaker, allowing us to visually manipulate our data, and listen to it at the same time. Here we have used a sound function whose syntax is given as follows:

- sound(y,Fs)
- sound(y)
- sound(y,Fs,bits)

**sound(y,Fs)** sends the signal in vector y (with sample frequency Fs) to the speaker on PC and most UNIX platforms. Values in y are assumed to be in the range -1.0<=y<=1.0. Values outside that range are clipped. Stereo sound is played on platforms that support it when y is an n-by-2 matrix. The values in column 1 are assigned to the left channel, and those in column 2 to the right.

**sound(y,Fs,bits)** plays the sound using bits number of bits/sample, if possible. Most platforms support bits are 8 or 16.

To differentiate between horizontal edge and vertical edge we have used two different frequencies. Listening to the sound generated with variations due to different frequencies one can make out the presence of horizontal and vertical edges. If there are more no of horizontal edges in an image then the frequency allotted to horizontal edge is more prominent. Same is the case for vertical edges and mixed sound is produced in case of equal no of horizontal and vertical edges. Efficient working of this system is fulfilled after proper training at user side. In summarized way we can say that – frequency of the sound will be selected according to the orientation of the edge (Horizontal or vertical) and duration of sound will be decided by the length of the edge

#### 4. Automated Training Database Designing:

Here we are performing the automated training and database designing. Seeing is something that most of us expect to do with our eyes. But what if you are born blind or lose your sight later in life? Peter Meijer suggests you consider seeing with your ears instead. The concept of our system is to make easy for a blind person to identify an image on the basis of the sound which is produced after scanning an image. Automated training allows us to represent visual information - to "see" - with sounds. The device used here is a laptop or a PC integrated with webcam, microphone and headphones. In automated training we will design a database of some images such as a tree, a building, furniture, a car etc. That images can be of square, rectangle, circle, rhombus, and triangle or any other shape.

Requirements for automated training :

- Collecting different images of different common shapes and of simple objects.
- Taking output for each binary image and saving it.
- Using microphone, record the category of image in human voice whether the image is classified as:
  - Simple horizontal : containing all horizontal edges with in an image.
  - Horizontal : containing more number of horizontal edges as compared to vertical edges.
  - Complex : containing equal number of horizontal and vertical edges with in an image.
  - Vertical : containing more number of vertical edges as compared to horizontal edges with in an image.
  - Simple vertical : containing all vertical edges with in an image.
- After performing all these actions create database which has three columns<image, img\_output, img\_record> and multiple rows.
  - Image column- we store image.
  - Img\_output column -we store output of this image which comes after running the program on that image .
  - Img\_record column – we store recorded sound for specifying the category of the image.

Store output and corresponding recorded sound for respective image in the database.

- A blind person now sits in front of laptop and wear headphones. Then we people are taking any image and told him/her that this is the image of 'XYZ' and low frequency sound is of horizontal edge & high frequency sound is of vertical edge and now listen the sound for perception its shape and then play recorded sound for telling him that this image lies in which category.

When we provide this training to a blind person more than one time for a single image, then blind person is used to recognize shapes and category of image easily.

#### 5. Code in Mat Lab:

```
sh = 0;
sv=0;
sd=0;
t1=1;
t2=1;
t3=1;
t4=1;
a=imread('c:\im\i1.jpg'); % Read the
Image from system or a digital camera

a=rgb2gray(a);
[M, N]=size(a);

mkH= [-1 -2 -1 ;1 2 1]; % Masking to
detect horizontal edges
mkV= [1 -1;2 -2;3 -3]; % Masking to
detect vertical edges
md= [-1 1; 1 -1];% GIVES +VE VALUE FOR
FWD DIAGONAL, -VE VALUE FOR
a1 =[2,5000];
a2 =[2,5000];
b1=[2,21919];
b2=[2,21919];
a = double(a);
for i=1:M
    for j=1:N
        if a(i,j) >50
            a(i,j)=1; % Converting image
to binary Image
        else
            a(i,j)=0;
        end
    end
end
figure,imshow(a);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% DETECTING HORIZONTAL EDGES
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:M-2
    for j=1:N-3
        sh=0;

        for k=1:2
            for l=1:3
                sh=sh+a(i+k-1,j+l-1) *
mkH(k,l);
                %sv=sv+a(i+k-1,j+l-1) *
mkV(k,l);

            end
        end

        if (sh~=0)
            a1(1,t1)=i;
            a1(2,t1)=j;
            t1=t1+1;

        end
    end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% DETECTING VERTICAL EDGES
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```

for j=1:N-3
    for i=1:M-2

        sv=0;
        for k=1:3
            for l=1:2
                sv=sv+a(i+k-1,j+l-1) *
mkV(k,l);

            end

        end

        if(sv~=0)
            a2(1,t2)=i;
            a2(2,t2)=j;
            t2=t2+1;
        end

    end

    end

    k1=1;
    k2=1;
    len1=0;
    len2=0;

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    STORE EDGE LENGTH AND ORIENTATION
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

    % ENTRY INTO B1
    for i=1:t1-2
        if(a1(1,i) == (a1(1,i+1)))
            len1=len1+1;

        else
            b1(1,k1)=len1;
            len1=0;
            b1(2,k1)=a1(1,i); % the value of i
            k1=k1+1;
        end
    end

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% FOR LAST
    ENTRY%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        b1(1,k1)=len1;
        len1=0;
        b1(2,k1)=a1(1,i); % the value of i
        k1=k1+1;

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    for i=1:t2-2
        if(a2(2,i) == (a2(2,i+1)))
            len2=len2+1;

        else
            b2(1,k2)=len2;
            len2=0;
            b2(2,k2)=a2(2,i); % the value of J
            k2=k2+1;
        end

    end

    end

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%FOR LAST
    ENTRY%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
        b2(1,k2)=len2;
        len2=0;
        b2(2,k2)=a2(2,i); % the value of i
        k2=k2+1;

    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

GENERATE SOUND ACCORDING TO EDGE LENGTH AND
ORIENTATION
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

b1
b2
t1=0;

if (k1>k2)
    k=k1;
else
    k=k2;
end

for i=1:k-1

    fs = 8192; % sampling
    rate
    T = 1/fs;
    if ( i<k1 && b1(1,i)>30 )
        t1=0.005*b1(1,i);

        f1 = 1000; % frequency
        in Hertz
        t = (0:T:t1); % time
        vector
        omegal = 2*pi*f1; % angular
        frequency in radians
        phi = 2*pi*0.75; % arbitrary
        phase offset = 3/4 cycle
        x1 = sin(omegal*t + phi); % sinusoidal
        signal, amplitude = 1
        %figure,plot(t, x1); %
        plot the signal
        xlabel('Time (seconds)');
        ylabel('x1');
        title('Simple Sinusoid');
        sound(0.9*x1, fs);
    end

    if (i<k2 && b2(1,i)>30 )
        t1=0.005*b2(1,i);


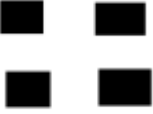


        f1 = 2000; % frequency
        in Hertz
        t = (0:T:t1); % time
        vector
        omegal = 2*pi*f1; % angular
        frequency in radians
        phi = 2*pi*0.75; % arbitrary
        phase offset = 3/4 cycle
        x1 = sin(omegal*t + phi); % sinusoidal
        signal, amplitude = 1
        %figure,plot(t, x1); %
        plot the signal
        xlabel('Time (seconds)');
        ylabel('x1');
        title('Simple Sinusoid');
        sound(0.9*x1, fs);
    end
end

```

## 6. RESULT

We worked on small images as our input as describe in following table no 1. We found that by listening the sound generated by the system we can identify the type of an image like whether it is a house, tree, car or building etc. We run the system with different images to generate the sound signatures, 10 persons were first trained 5 times for each image and then made to listen the signatures. Following table[Table 1] shows the results-

Image	Right Guess	Wrong Guess	Can't Say
-------	-------------	-------------	-----------

 <b>Verical Bar</b>	<b>8</b>	<b>2</b>	<b>0</b>
 <b>Four Squeres</b>	<b>6</b>	<b>2</b>	<b>2</b>
 <b>Circle</b>	<b>7</b>	<b>2</b>	<b>1</b>
 <b>House</b>	<b>2</b>	<b>3</b>	<b>5</b>

**Table 1**

## 7. CONCLUSION

We can see system works great for simple images, the performance decreases with the complexity in image. With automated training session we can overcome this problem. Above results shows the potential in the technology. This system can be a boon for blind persons if they are given proper training they will be able to identify the type of object which is in front of them and it produces very good results.

## 6. REFERENCES

- [1]. Abravanel, E. 1971. Active detection of solid-shape Information by touch and vision. *Perception & Psychophysics*, 10, 358-360.
- [2.] Bach-y-Rita, P. 1967. Sensory plasticity: Applications to a vision substitution system. *Acta Neurologica Scandanavica*, 43,417-426.
- [3.] Bach-y-Rita, P. 1972. *Brain mechanisms in sensory substitution*. New York: Academic Press. Bliss, J.C., M.H. Katcher, C.H. Rogers, and R.P. Shepard. 1970. Optical-to- tactile image conversion for the blind. *IEEE Transactions on Man-Machine Systems*, MMS- 11.58-65.
- [5]. Brabyn, J.A. 1985. A review of mobility aids and means of assessment. In *Electronic spatial* plasticity in blind humans. *Nature*, 389: 180-183.
- [6]. D. N. Jiang, L. Lu, H.J. Zhang, et. al. "Music type classification by spectral contrast features", Int. Conf. Multimedia Expo (ICME02), Vol. 1,pp. 113-116, 2002.
- [7]. G. Tzanetakis, and P.Cook, "Music genre classification of audio signals", IEEE Trans. Speech Audi Processing, 10 (5), 293-302, 2002
- [8] . M. Grey, and J. W. Gordon. "Perceptual effects of spectral modifications on musical timbres." Journal of the Acoustical Society of America, Vol. 63(5), pp. 1493-1500, 1978.
- [9]. J. Smola, and B. Scholkopf, "A Tutorial on Support Vector Regression",NeuroCOLT2 Technical Report Series, NC2-TR-1998-30
- [10] artino, G. and Marks, L. E., ``Synesthesia: Strong and Weak," *Current Directions in Psychological Science*, Vol. 10, No. 2, April 2001, pp. 61-65