Electronic aid based Assistive Text and Product Label Reading from Hand-Held Objects for Specially Abled Persons

Asha Bai L
PG student, Department of Electronics and Communication Eng., B T Lakshman Institute of Technology, Bangalore, India

Dr. Manju Devi
Prof and Head, Department of Electronics and Communication Eng. B T Lakshman Institute of Technology, Bangalore, India

Nirmala Bai L
Asst. Professor, Department of Electronics and instrumentation Eng., Dr. Ambedkar Institute of Technology, Bangalore , India.

Abstract — **The proposed system is a portable camera based visual assistance prototype for blind people to identify currency notes and also helps them to read printable texts from the handheld objects. To read printable texts, an efficient algorithm that combines an Optical Character Recognition (OCR) with Hierarchical optimization is used. In Pattern Recognition OCR every character is localized and separated then the resulting character image is sent to a pre-processor to reduce noise and to perform normalization. Certain characteristics will be extracted from the character for comparison. After comparison the identified characters are grouped and reconstructed to form original text strings, then the output is given to the speech engine to perform text to voice conversion. For identification of currency notes a novel recognition system is developed using SIFT (Scale Invariant Feature Transform) to improve precision and accuracy. The input image undergoes pre-processing and thereafter the distinct features are extracted and compared with the templates from the database. The resulting outcome is given through Earphone to the blind users.**

Keywords—: hierarchical optimization, OCR, SIFT, text to speech engine.

I. INTRODUCTION

Blind people always wanted to live independently like normal People. But most of the times like while reading texts they need to depend on others. Latest advancements in technology made it possible to provide assistance to these people by designing products that use computer vision and camera with optical character recognition (OCR) system. Reading has become an essential part in the modern world. Texts in printed form are available everywhere in books, bills, cheques, demand Drafts, pamphlets, product labels, newspapers, etc. Different types of software such as screen reader and magnifiers are available to help blind people and people with poor eyesight to use a computer or other devices but there is only less number of products which help them to read texts in the outside world. When blind people are assisted to read printed texts and products, it will improve their confidence and provide independency in this society. Recently, many devices have been developed to provide portability in text reading, but the process is a bit tedious and creates inconvenience for blind people. One such product is the Barcode scanner; the basic concept behind this device is that the products or objects need to be stored separately for each barcode and all the data will be held in a database. Any time the user can scan the barcode and get the information about that object.

For currency recognition, there are many systems available for office usage but no such portable devices are available. It reduces the independency of the blind people to the greater extent. Our paper is mainly focused on addressing these issues for the blind people. To help blind people in reading printable text, we have connected a camera with our processing system, the camera will capture the readable texts and system will perform OCR extraction process to extract the text information. For currency note recognition we are using the efficient SIFT algorithm to extract information and with that information the correct denominations can be identified. After all the above processes the extracted information is given to the text to speech engine and then given to the users through ear phone.

II. TEXT READING

For text reading the video is captured from the camera fitted on the glasses on the user using Matlab. The frames are segregated and pre-processing is done. To identify texts, the captured image is binarized after the gray-scale conversion. Now the novel hierarchical optimization method is applied along with OCR to extract the texts, then the output is given to the speech engine to generate audio output.

A. Hierarchical optimization algorithm

The reason for using of this algorithm is that the basic OCR [5] alone is not suitable to read robust and distorted text. Our algorithm will use pattern character recognition and hierarchical optimization to perform better recognition of the printable texts. Another reason for using this algorithm is because of its high degree of accuracy in recognition along with its speed. In our case the images are taken by a blind

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRET - 2016 Conference Proceedings

user, there is a good possibility for a distorted image. These are requirements that need to be satisfied by our algorithm, stableness to errors in recognized characters, high speed, easily trained and tuned. By taking above requirements into consideration, pattern recognition is best fitted algorithm to produce the expected output.. In pattern recognition an object with needed character is selected out of the original image and compared with all the patterns in the database. While comparing both the patterns, it is possible that one of the patterns should be shifted vertically or horizontally at the least. And also the recognition time depends on size and the rotation of the pattern. The major hindrance in using pattern algorithm for text reading on noise and distorted images is the more distortion of the characters, it makes direct comparisons of pattern impossible.

In order to improve speed of template searching that belongs best to the character on the image hierarchical probabilistic matching is used. The above method consists of following steps: when comparing the templates with character, using definite algorithm only part of the template positions are checked based on the image and point number (resolution) of templates and search area is changed initially one after one and then step after step till it uses all the points

In other words, Powel's method with multidimensional optimization was selected as an optimization algorithm. The following values can be changed without dependency from each other by resolution: 1) Translations along the axes x and y, 2) translation in the direction (x, y), (-x, y) etc. 3) Rotation around the axes x and y 4) Rotation around the z-axis 5) scaling i.e. the Translation along the z-axis. So the optimization by each point number (resolution) is done in 8-dimensionsal space. Based on optimization algorithm, the location of the template for the next iteration is defined based on quality criterion, which is calculated as below:

$$QC1 = \sum_{t=0}^{N} d_t^k$$

$$QC_2 = \sum_{t=0}^{N} ((n_e / n_t),(n_e / n_c))$$

N - Amount of the template points, di - distance from i[th] point of the template to the nearest point of the recognized character, k - ½, 1 or 2. Ne - amount of the template points coincident to the points of the recognized character, Nt - the total amount of the template points, Nc - amount of the points of the image area which is bounded by the applied template.
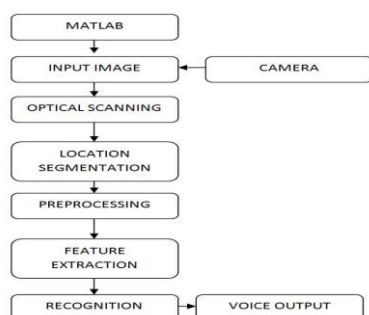
B. Block Diagram

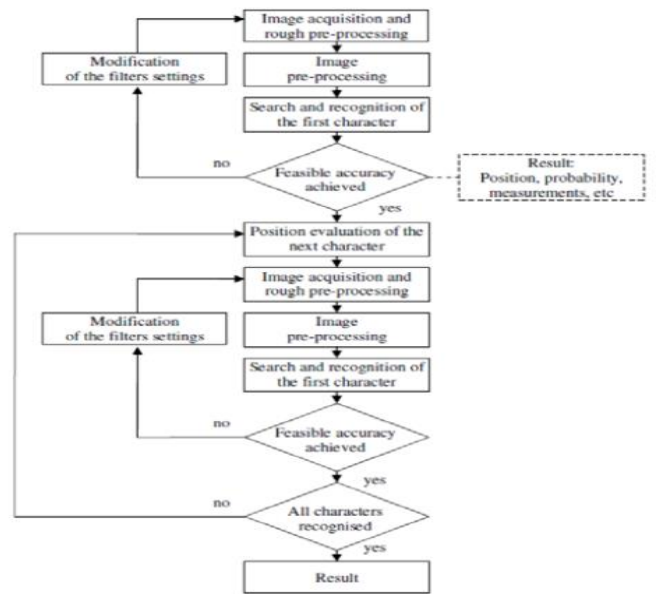FIGURE-1. BLOCK DIAGRAM OF THE TEXT DETECTION USING OCR

C. Flowchart

FIGURE-2 . FLOWCHART FOR COMMON CHARACTER RECOGNITION

D. Speech output generation

The recognized text from OCR is written on a text file. The text file is given as input to the speech engine. Speech engine converts the texts from the file and store it into an array and after that it will be compared to the library and then audio is generated based on the output.

III. CURRENCY NOTE DETECTION

Currency Note Detection is done through SIFT Algorithm. SIFT [Scale Invariant Feature Detection] is a Keypoint feature Descriptor which helps us to identify different types currencies from the given image by Matching its keypoint Features. Precision and accuracy makes this superior than the similar descriptors. We extract the features from the image with distinctive properties which is best suited for image matching process. These features will not vary with respect to scaling or image rotation and also illumination will not show much variation. These points are not disrupted by closure, scramble or noises because these are situated properly in frequency and spatial regions. By applying a cascade filter, complex calculations have been reduced and also it reduces the time required for extraction..

SIFT algorithm

These steps are followed in SIFT for extracting keypoints from the image,

Scale-space Exterma detection

Keypoint Localization

Orientation Assignment

Keypoint Descriptor

A. Scale-space extrema detection

Scale Spaces are created by removing the unnecessary details from an image. While removing those details, the false details should not be added to the image. This process is efficiently done by using Gaussian Blur. In Sift, the scale spaces are produced by applying Gaussian blur continuously and for the next stage image size is reduced half of its original value and blurring is applied again. This process will be continued till acquiring the required scale spaces. Gaussian blur has expression that is applied to each pixel and it results in image blurring.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

L - Output image (Blurred), G - Gaussian operator, I - Input image, x, y - coordinates of the location, σ – Parameter (Scale). Amount of Blurring is based on this value, * - convolution operation to apply Gaussian blurs G to I.

$$G(x, y, \sigma) = (1 \div 2\pi\sigma^2)e^{-(x^2 - y^2)/2\sigma^2}$$

The above is the Original Gaussian Blur Expression.

For a LOG operation, an image is taken and added with a small amount of blur then 2nd order derivative is calculated for it. This will find edges and corners because these are good for locating key points. The above mentioned derivative calculations are very complex and involve lot of computational time, so a different approach is used. To produce LOG, the Gaussian Difference method is used. It is calculated by subtracting two immediate Gaussian Scales.

The Difference of Gaussian (DOG) is equal to the Laplacian of Gaussian Approximately. Now the Complex calculations have been replaced by simple fast and efficient process. Another advantage of using DOG is that it is scale invariant. But LOG depends on the scale because of the σ2 in the Gaussian expression. This will be eliminated by multiplying the result with σ2. While doing subtraction this value is automatically multiplied so it further reduces the computation time and produce scale invariance. To find the maxima and minima, iteration is done for every pixel and all its nearby pixels are checked.

B. Key point localization

After finding the approximate maxima and minima, the exact key points will be localized. Mark the points as show below, in that we need to find the green region i.e. the exact location of extreme key points.

From the acquired data, sub pixel values can be found using the Taylor's formula for expansion near the approximate point. The formula is given below,

$$D(x) = D + (\partial d^2 / \partial t)x + (1/2)x^T(\partial^2 D / \partial x^2)x$$

The extreme's can be found from the above formula by differentiating and equating it to zero. While doing so, it will improve the stableness and matching property of the algorithm. Some of the key points are on the edge or will have low-contrast, either way they are useless features. In order to eliminate, we use two filters. For first filter Taylor's formula is again used to find the intensity at key point areas, if the found magnitude is less than the fixed value than the key point will be eliminated. For edge detection, two perpendicular gradients are needed from the Key point. If both gradients are large then it's a corner and it will be accepted as Key point, otherwise it will be eliminated.

C. Orientation assignment

In this step orientation is assigned to the key points which passed the above two filters. After the previous steps we have stable and scale-invariant key points. For efficient feature matching we need the points to show rotational invariance. This can be obtained by assigning Orientation to the Key points. The basic concept is to group gradient values i.e. both directions and magnitudes from each pixels around every key point. The best suited orientation is figured and assigned for that particular key point and also relative calculations are made to ensure invariance in rotation. The following formulas are used for orientation Assignment process

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1), y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1))/L(x-1, y)))$$

M(x, y) -- Gradient magnitude

Θ(x, y) - Gradient Orientation

For every pixels near the key point both the above gradient are calculated and a histogram is drawn for the obtained values. In the created histogram, the 360° is divided to 36 sections totally with 10° each. Certain regions of the gradients are marked as Orientation Reception Area. For example if it is 15.789, then it will be put in between 10 and 19 degree section. The peak of the histogram will be at some point after plotting all the pixels around the key point. In the below figure, the peak is between 20 and 29 degrees. So the key point will be assigned to orientation three i.e. the third section. If any of the peaks are over 80% then it will be change into new key point with same location of original key point but with orientation equal to peak. The key Concept is that the images are blurred at 1.5*sigma so the size of the windows kernel should be equal to the same.

D. .Keypoint descriptor

Keypoint Descriptor will describe the unique and highly distinctive fingerprint for every keypoint. In this last step we will develop fingerprint for the keypoints obtained till this step which is invariant to scale and rotational aspects. To generate a unique fingerprint, a window of size 16 X 16 is taken around the keypoint. That will be broken into small windows of size 4 X 4. In those windows gradient magnitudes and orientations are calculated and put in an eight section histogram .

IV. RESULTS
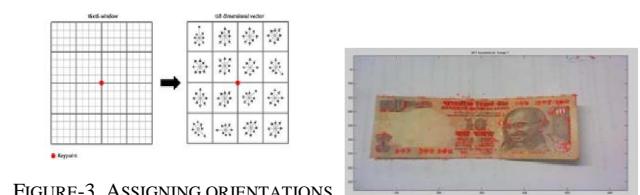

FIGURE-3. ASSIGNING ORIENTATIONS

Any orientation between 0 and 44 is added to the 1st section. Orientation between 45 and 89 is added to 2nd section

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICRET - 2016 Conference Proceedings

and it goes on till last section. Dissimilar to the previous step here the amount of orientation added is also depends on the distance from the keypoint. The whole process is carried out through the weighting function in Gaussian. Its main function is to create a 2D bell curve like gradient and it will be multiplied with magnitude orientations to get a weighted image as shown below
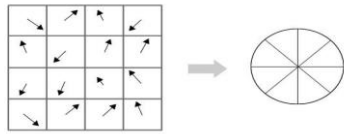

FIGURE-4. WEIGHTED IMAGE.

When the keypoint is at large distance then its magnitude will be small. The same process is continued for the whole 16 pixels and we have fitted 16 completely random orientations into 8 predetermined sections. If the same process is done for all sixteen region we will end up with 4 X 4 X 8 = 128 Nos. After normalization by diving with sum of squares we will get the required feature vectors to uniquely identify a keypoint. Before finalization of features two introduced problems need to addressed. Rotational dependence is adjusted by subtraction of keypoint rotation with each orientations and lighting dependency is adjusted by keeping large thresholds before normalizations. Thus we achieved a illumination and rotationally independent feature vectors for matching. The same process is done for both Templates and Input images then both the keypoints are compared to recognize the correct denomination of the Input Currency Note.

V.   CONCLUSIONS

Here in this paper we have designed a prototype to support blind people for their day to day activities. The proposed prototype reads out the printable text from handheld objects to them and also helps them to identify currency notes with ease. Here we have confined our prototype to identify only Indian currencies but in future this can also be extended to read other country currencies.
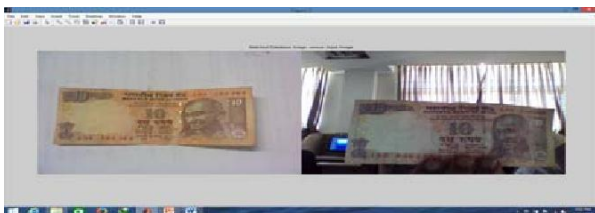

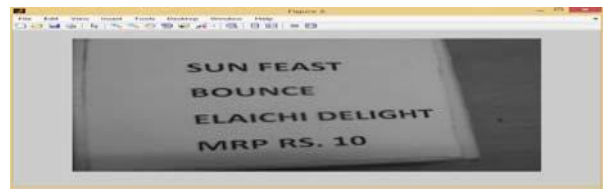FIGURE-5. DETECTED FEATURES IN CURRENCY NOTE.

FIGURE-6. MATCHED OUTPUT (DATABASE IMAGE AND INPUT IMAGE).

FIGURE-7. INPUT TEXT FROM AN OBJECT.

VI.   REFERENCES

[1] Internatinal workshop on camera based document analysis and recognition(cbdar 2005, 2007, 2009, 2011).

[2] X .chen and a. L. Yuille, "detecting and reading text in natural scenes,"vol 2

[3] x.chen,j.yang,j.zhang, and a.waibel , "automatic detection and recognition of signs from natural scences,"ieee trans. Image process, vol 13

[4] D.dakopoulos and n. G. Bourbakis,"wearable obstracle avoidance electronic travel aid for blind" survey," ieee trans. Syst., man, cybern., vol. 40, no. 1, pp. 25–35, jan. 2010.

[5] B. Epshtein, e. Ofek, and y. Wexler, "detecting text in natural scenes with stroke width transform," in proc. Comput. Vision pattern recognit., 2010, pp. 2963–2970.

[6] Y.freundandr.schapire,"experiments with a new boosting algorithm," in proc. Int. Conf. Machine learning, 1996, pp. 148–156.

[7] N. Giudice and g. Legge, "blind navigation and the role of technology," in the engineering handbook of smart technology for aging, disability, and independence, a. A. Helal, m. Mokhtari, and b. Abdulrazak, eds. Hoboken, nj, usa: wiley, 2008.

[8] a. Shahab, f. Shafait, and a. Dengel, "icdar 2011 robust reading competition: icdar robust reading competition challenge 2: reading text in scene images," in proc. Int. Conference .

[9] . K. Kim, k. Jung, and j. Kim, "texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," IEEE transaction.