Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Real Time Detection and Recognition of Hand Held Objects to Assist Blind People

Ankit Dongre[1],Dhruv Purandare[2], Jimil Gandhi[3] , Suchit Adak[4]

[1] *Computer Department ,AISSMS Institute of Information Technology*
[2] *Computer Department ,AISSMS Institute of Information Technology*
[3] *Computer Department ,AISSMS Institute of Information Technology*
[4] *Computer Department ,AISSMS Institute of Information Technology*

Abstract — *There are approximately 315 million people with visual imparity worldwide. The number of these people is also increasing as this generation ages. Recent development in computer vision, digital cameras, and portable computers make it possible to assist these people by developing camera-based products that incorporate computer vision technology. We propose a camera-based object recognition application to help visually impaired people to recognize labeled and unlabelled hand held objects in their daily lives. To differentiate the object from complex backgrounds or other surrounding objects in the camera view, we first propose a motion based method to define a region of interest (ROI) in the image by asking the user to hold the object which is efficient and effective. Reference objects and test images are accompanied by local features (regions, interest points, etc.). If the test image get match with reference image then the output signal created in the form of audio. Output audio is generated and sent towards head-phone through which the blind user could hear the information about the product which is held on hand. In this proposed system we use speeded up robust features (SURF) for tasks such as object recognition, classification or 3D reconstruction, image registration.*

Keywords*-Region of Interest(ROI), computer vision, object recognition, 3D reconstruction, image registration*

II. INTRODUCTION

This proposed system will help blind people to know about product which is held in hand. Reading is very important in today's society. Printed text is everywhere in the form of receipts, reports, statements of home, menus, classroom notes, product packages, instructions etc. Logos are graphical representations that either recall some real world objects, or focuses on a name, or simply display some abstract signs that have strong appeal. Color may have some relevance to assess the logo. But the unique and distinct nature of logos is more often given by a few details carefully drawn out by graphic designers, sociologists and experts of social communication. The graphic layout is equally important to gain attention of the customer and reciprocate the message aptly and permanently. Different logos may have alike schematics with marginally different spatial disposition of the graphic elements, localized differences in the orientation, size and shape. There are many assistive systems available today but they have certain issues reducing the feasibility for the visually challenged persons. For example, portable bar code readers designed to help blind people identify distinguishable products, it helps the users who are blind to acquire information about these products through speech and Braille. But a big drawback is that it is highly difficult for blind users to determine the position of the bar code and to correctly position the barcode reader at the bar code. The input image is acquired by the camera. To assure the hand-held object is visible in the camera view, we use a mobile camera with sufficient clarity. This may often result in other text objects appearing in the cameras view. The generated output audio will be sent to headphone. The headphone will announce the name of the product. This system is helpful for blind person to identify the product which is held on hand. It is a challenge to localize objects and ROIs automatically from captured images with cluttered backgrounds, because details in captured images is most probably surrounded by various background outlining noise and text characters generally appear in multiple fonts, colours, and scales. Multiple algorithms have been made for localization of image areas in scene images.

In this system SURF algorithm is used because it reduces the drawback of SIFT algorithm to some extent. The ability of people who have significant visual impairments to recognize products and product packages will enhance independent living and social self-sufficient by this system.

II. SURVEY DETAILS

Samruddhi Deshpande et. al. [1] in their paper said that camera based system which will help blind person for reading text patterns printed on hand held objects. This is the framework to assist visually impaired persons to read text patterns and convert it into the audio output. To get the object from the background and extract the text pattern from that object, the system first proposes the method that will capture the image from the camera and object region is detected.

Jia Xingteng et. al. [2] in their paper said that SURF (Speeded Up Robust Feature) algorithm has gained great interesting image matching and delocalization or self-navigation of robots noted by its invariant property as well as its low computational complexity. Image matching method based on SURF algorithm has been used in diverse fields such as computer vision, medical diagnosis, and treatment and image mosaic.

Runqing Zhang et. al. [3] in their paper said that Hand gestures recognitions play an important role in human-computer interaction. To facilitate the understanding of computer hand gesture recognition this paper describes a system for human-computer interaction through images local features SURF, and we use threshold segmentation and bag of words algorithms to reduce the feature space dimensions.

Juan et. al. [4] in their paper said that Comparison of SIFT, PCASIFT and SURF summarized the feature detection methods as Scale Invariant Feature Transform (SIFT), Principle Component Analysis (PCA)SIFT and Speeded Up Robust Features (SURF). In this paper the author compares methods with the scale variation, accuracy in matches finding required storage space ,key point localization .They found that SIFT needs more memory space compared to other two methods and the matching speed for SIFT is less although it gives good match findings which improves the accuracy for object detection. Whereas PCA-SIFT show its advantages in rotation and illumination changes and gives less matches compared to SIFT.

Anen Jabnoun et. al. [5] in their paper said that object recognition for the visually impaired based on features extraction provided an overview of various visual alternative systems developed in the recent years. This method is based on interpretation, video analysis and feature extraction .They give the results of comparison of SIFT and SURF in which they concluded that SURF is faster than SIFT ,however SIFT is robust when the matches findings, scale variations are considered. They used video to Audio transformation to provide the object information.

Ricardo Chinchaan et. al. [6] in their paper proposed an object recognition method to assist blind people to find missing items using Speeded-Up Robust Features (SURF). The proposed recognition method starts by matching individual features of the object that is queried by user to a database of features with various personal belongings which are saved in advance. From the experiments the total number of objects detected were 84 out of 100,this shows that their work needs better performance hence to enhance the object recognition SIFT can be used.

Payal Panchal et. al. [7] in their paper said that given the brief idea about various object detection methods. They have also given the comparison of these methods. They concluded that the problem occurs during identification of object if image is not captured properly because the positioning of camera was not proper. Then the object cannot be identified. To solve this problem and improve accuracy they suggested merging multiple methods and making use of it together as per the requirement of the application.

Lukas T et. al.[8] in their paper proposed a system for gaze contingent auditory substitution of spatial vision. This system was designed to be a mobile helper in everyday life of the blind people. The prototype they have developed combines depth measuring and sonification techniques with eye tracking. Eye movements support mental imagery regardless of the presence of visual stimuli. They have developed the Auditory Night Sight as a gaze-contingent system, meaning that the user's gaze immediately determines the direction of perception. In more than nine out of ten trials localization was successful. Size detection seems to be the much more difficult task.

Chucai Yi et. al.[9] in their paper proposed a camera-based text reading assistive system to help blind people read text labels and product packaging from hand-held objects in their daily lives. They proposed an effective and efficient motion based method to derive a region of interest in the video. The performance of the proposed text localization algorithm is numerically evaluated. Then they employed the Microsoft Speech Software Development Kit display the audio output of text information.

Hanen Jabnoun et. al. [10] in this paper proposed a system restores the primary function of the visual system i.e. the identification of surrounding objects. This method is based on the concept of local features extraction. The simulation of results using SIFT algorithm and matching the key points showed good accuracy for detecting objects. They have worked for the key point detection in fast video using transformation in SIFT which is invariant to the changes in luminosity.

IV. PROPOSED SYSTEM

We propose an optical sensor-based object recognition system to help visually impaired persons to read product label from hand held objects in their day to day lives. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, we first propose an efficient and effective motion based method to define a region of interest (ROI) in the image by asking the user to hold the object. In this framework able to match and recognize multiple instances of multiple reference logos in image archives. Reference logos and test images companies by local features (interest points, regions, etc.).If the test image get match with reference image then the output signal created in the form of audio signal. Output audio signal sent toward headphone which will attach to the ear of the blind user so that blind user could identify the product which is held on hand .To make sure the hand-held object is visible in the camera view, we use a mobile camera with sufficiently large optical sensor and good clarity. In this suggested system we use speeded SURF algorithm which stands for **S**peed **U**p **R**obust **F**eatures ,for tasks such as object recognition, classification or image registration .

Since we are implementing the algorithm on a smart phone, we will be using OpenCV for Android SDK.

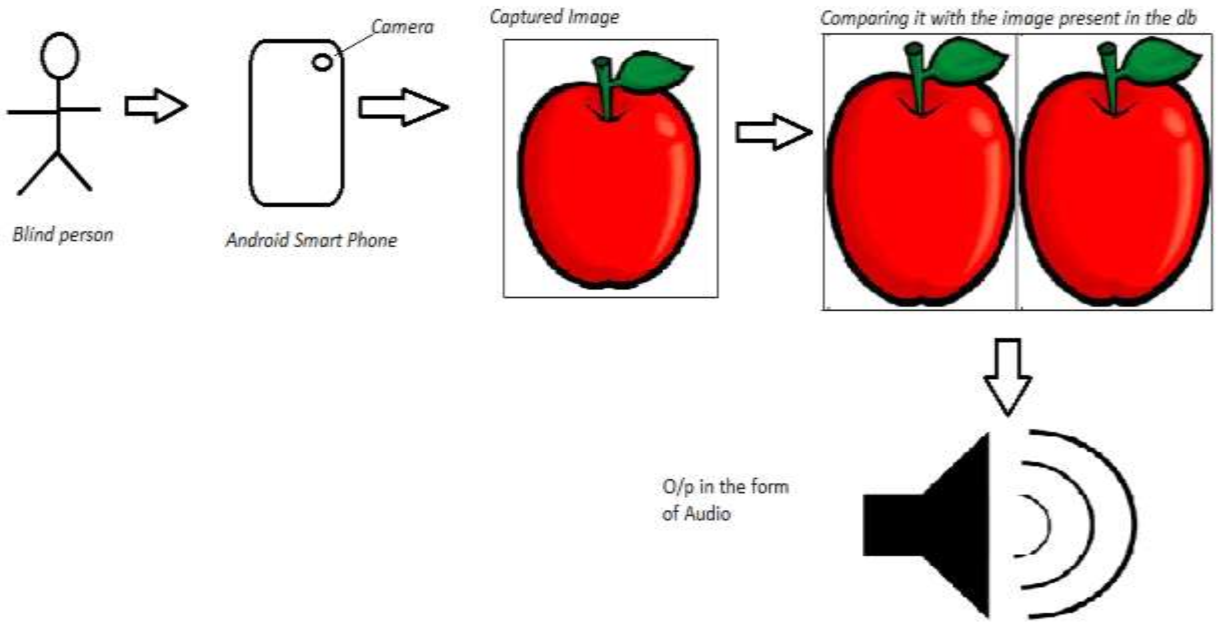

Figure 1:Block Diagram

3.1. Relevant Mathematics

Square-shaped filters are used in SURF as an approximation of Gaussian smoothing. (cascaded filters approach is used by SIFT to detect scale-invariant characteristic points, where the difference of Gaussians (DoG) is calculated on rescaled images progressively.)

A blob detector based on the Hessian matrix to find points of interest is used by SURF. To measure local change around the point determinant of the Hessian matrix is used ,and points are chosen based on where this determinant is maximal. In contrast to the Hessian-Laplacian detector by Mikolajczyk and Schmid, SURF also uses the determinant of the Hessian for selecting the scale, as is also done by Lindeberg.

Given a point p=(x, y) in an image I, the Hessian matrix H(p, σ) at point p and scale σ, is:

$$\mathcal{H}(\mathbf{x},\sigma) = \begin{bmatrix} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{xy}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{bmatrix}$$

where L_{xx}(p,σ) etc. is the convolution of the second-order derivative of gaussian with the image at the point x.

The box filter of size 9×9 is an approximation of a Gaussian with σ=1.2 and represents the least level (highest spatial resolution) for blob-response maps.

3.2. Scale-space representation and location of points of interest

At different scales Interest points can be found , partly because the comparison images where they are seen at different scales are often required to search for correspondences. The scale space is usually realized as an image pyramid in other feature detection algorithms. Images are repeatedly smoothed with a Gaussian filter, then they are sub sampled to get the next higher level of the pyramid. Therefore, several floors or stairs with various measures of the masks are calculated:

$$\sigma_{\text{approx}} = \text{current filter size} \times \left(\frac{\text{base filter scale}}{\text{base filter size}}\right)$$

An octave refers to a series of response maps of covering a doubling of scale ,the scale space is divided into a number of octaves. In SURF, the output of the 9×9 filters helps us to obtain the lowest level of the scale space is obtained from .

Hence, scale spaces in SURF are implemented by applying box filters of different sizes, unlike previous methods. Accordingly, the scale space is analyzed by up-scaling the filter size rather than iteratively reducing the image size. The output of the above 9×9 filter is considered as the initial scale layer at scale s =1.2 (corresponding to Gaussian derivatives with σ = 1.2).

The following layers are obtained by filtering the image with gradually bigger masks, taking into account the discrete nature of integral images and the specific filter structure. This results in filters of size 9×9, 15×15, 21×21, 27×27,.... Non-maximum suppression in a 3×3×3 neighborhood is applied to localize interest points in the image and over scales.

The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space with the method proposed by Brown, et al. Scale space interpolation is especially important in this case, as the difference in scale between the first layers of every octave is relatively large.

IV. CONCLUSION AND FUTURE SCOPE

The ability of people who significant visual impairments or are blind to read printed labels and product packages will boost independent living and they can be socially self-sufficient by this system.

This system can be further improved to recognize multiple objects at once. It could be used to detect larger objects at greater distances. Furthermore the system's accuracy and efficiency can be improved by using better hardware and enhanced versions of SURF algorithm.

REFERENCES

1. Samruddhi Deshpande, Ms. Revati Shriram, "Real Time Text Detection and Recognition of Hand Held Objects to Assist Blind People",2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)International Institute of Information Technology (IIIT),Pune.
2. Jia Xingteng, Wang Xuan, Dong Zhe, "Image Matching Method Based On improved SURF Algorithm", IEEE International Conference on Computer and Communications(ICCC), pp 142-145, 2015.
3. Runqing Zhang, Yue Ming, Juanjuan Sun, "Hand gesture recognition with SURF-BOF based on Gray threshold segmentation",ICSP2016 978-1-5090-1345-6/16/31.00 2016 IEEE.
4. Juan and O. Gwon, "A Comparison of SIFT, PCASIFT and SURF", International Journal of Image Processing(IJIP), 3(4):143152, 2009.
5. Hanen Jabnoun, Faouzi Benzarti, and Hamid Amiri, "Object recognition for blind people based on features extraction",IEEE IPAS14: INTERNATIONAL IMAGE PROCESSING APPLICATIONSAND SYSTEMS CONFERENCE 2014.
6. Ricardo Chincha and Ying Li Tian, "Finding Objects for Blind People Based on SURF Features", 2011 IEEE International Conference on Bio informatics and Biomedicine Workshops.
7. Payal Panchal, Gaurav Prajapati, Savan Patel, Hinal Shah and Jitendra, "A Review on Object Detection and Tracking Methods", INTERNATIONAL JOURNAL FOR RESEARCH IN EMERGING SCIENCE AND TECHNOLOGY, VOLUME-2, ISSUE-1, JANUARY- 2015.
8. Lukas T, Hendrik, Andrea Finke and Helge Ritter CITEC, "Gaze-contingent audio-visual substitution for the blind and visually impaired", 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops.
9. Chucai Yi, Student Member, IEEE, Yingli Tian, Senior Member, IEEE, and Aries Arditi, "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons", IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 19, NO. 3, JUNE 2014.
10. Hanen Jabnoun, Faouzi Benzarti ,Hamid Amiri, "Object Detection and Identification for Blind People in Video Scene", 2015 15th International Conference on Intelligent Systems Design and Applications (ISDA).