Volume 6, Issue 12, December 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Voice Vision System

[1]Siddharth Nayak, [2]Srinath R., [3]Sumeet Kripalani, [4]Mehul Padwal, [5]Sujata Khedkar
[1-4] V.E.S.I.T., Mumbai, Maharashtra, India
[5] Associate Professor, (Department of Computer Engineering) V.E.S.I.T., Mumbai, Maharashtra, India

Abstract- ***The paper represents a novel approach for developing a computer based vision for Blind users using single camera provided in smartphones The idea is that the smartphone will detect the objects and obstacles in the user's path and a voice output will direct the user to his/her destination We are focussing on one of the most fundamental challenges in computer vision using monocular cameras The main aim is to use object detection and recognition as well as depth and distance estimation from that object to the camera***

Keywords— SURF, Depth, estimation, smartphone, object

I. INTRODUCTION

Recent advancements in computer vision for blind users use a binocular camera for object detection and tracking. This type of system is expensive and is not available to common public.This paper proposes a cheap and accessible alternative to develop a system that can help blind users detect and track objects in day to day scenarios. Here the cheap alternative is a monocular camera which is used in smartphones. The smartphones processes the stream of videos and in turn gives a feedback to the user. The design of this system uses SURF (Speed up Robust Feature) algorithm for recognition of objects.[1] This objects can be predefined by using a various machine learning algorithms. One the object is detected the main aim is to find the distance to the object and notify the blind user about it. This can be done using floor geometry and point of contact on floor to calculate the in-path distances.[2] The system gives a voice feedback and notifies the user about the objects.[4]

II. DESIGN

The basic design of this model is shown below. The audio video device can be a mobile phone or any other accessory that is easily available in the market. This device can capture video that can be processed using trained models to get the intended result. The processing is done using the surf algorithms for object detection and in-path distance estimation.

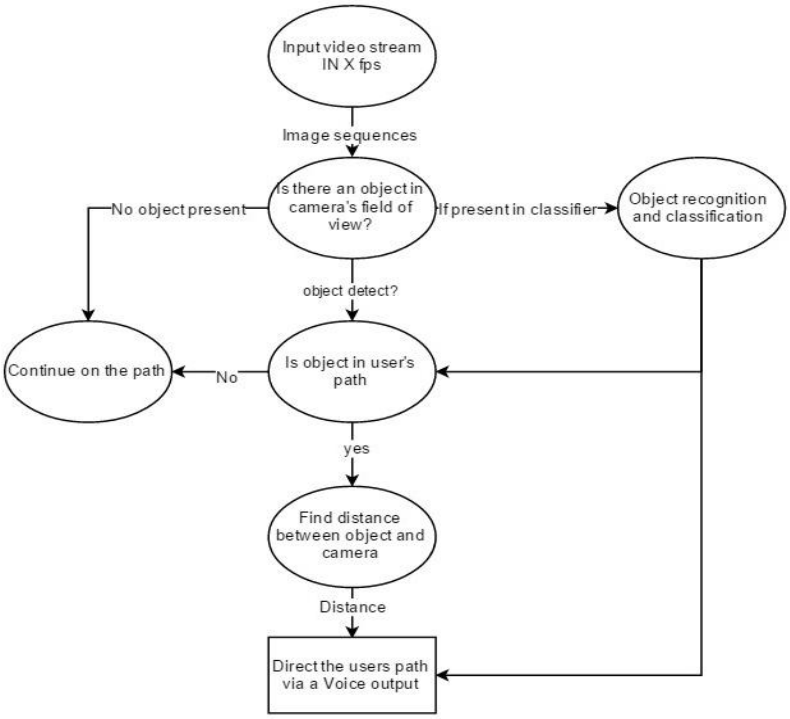

Fig 1. Voice vision system

III. SURF

SURF (Speed Up Robust Features) is the most recent advancement in object recognition .This algorithm utilizes a Hessian based detector and intensity distribution based description feature vector and leverages several proximations, which allow for fast computation without major sacrifices in accuracy and repeatability. From a given video frame, an interest point description feature vector is extracted using the SURF algorithm. The feature points between a training descriptor and the current query descriptor are matched using the k-Nearest Neighbour algorithm and a simple matching rule. If number of matching point is greater than an experimental determined threshold, then the training image is declared as found. The algorithm shows great promise with respect to properties such as repeatability, distinctiveness, and robustness, yet can be computed and compared much faster than any other schemes[1].

IV. DEPTH ESTIMATION

Camera properties include height of camera, focal length, angle of tilt of camera, pixel resolution. These camera parameters are related to respective camera only. The image captured is in the 2-D plane and the relation between 2-D image and the actual 3-D view of the image can be found out to estimate the distance. [2].

The distances in optical axis of the camera are called as the in-path distances. The in-path distances have relatively less horizontal and vertical errors as compared to the oblique distances. Horizontal errors are those which are orthogonal to the optical axis of the camera and increase as the distance from the camera increases. The vertical errors are the errors along the optical axis of the camera and they are maximum when the object is nearer to the camera. The algorithm used to compute the in-path distance is as follows

1. Given an image, identify the point of contact of the object with the ground.
2. Obtain the co-ordinates of this point.
3. Compute the angle between the optical axis and the line joining camera point to the position of the object (Angle β) in Fig.4.
4. Obtain Zcalc which is the intermediate reading for distance, by projecting its corresponding Y on the graph (refer figure 2). An example is shown for the point N (50, 26.92).
5. Obtain the Zact which is the actual in-path distance, by projecting its corresponding Zcalc on the graph (refer figure 3) for the actual in-path distances. An example is shown for point M (26, 22)
6. The actual in-path distance is then divided by cosine of angle β to get oblique distance. But this oblique distance has errors.
7. The horizontal and vertical errors in Zact are then corrected to get corrected oblique distance.

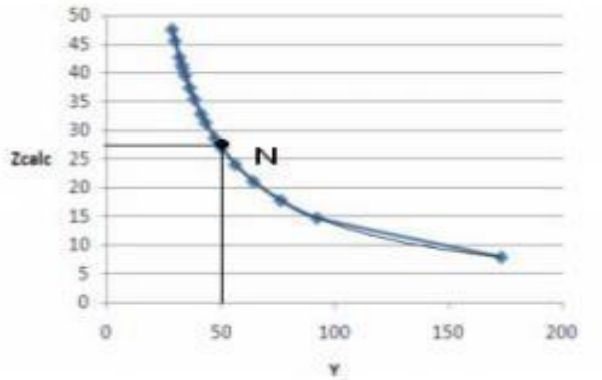

Fig 2. : Graph of Zcalc vs Y Zcalc=1346/Y

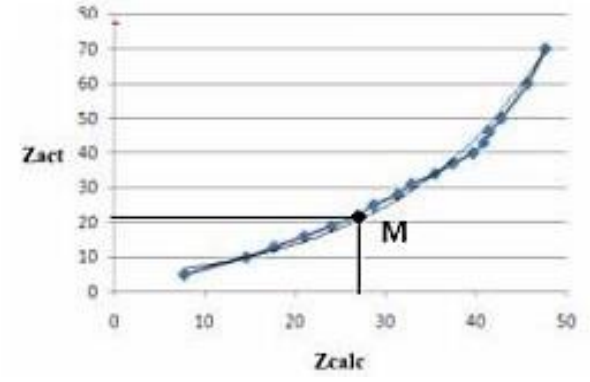

Fig 3. : Graph of Zact vs Zcalc

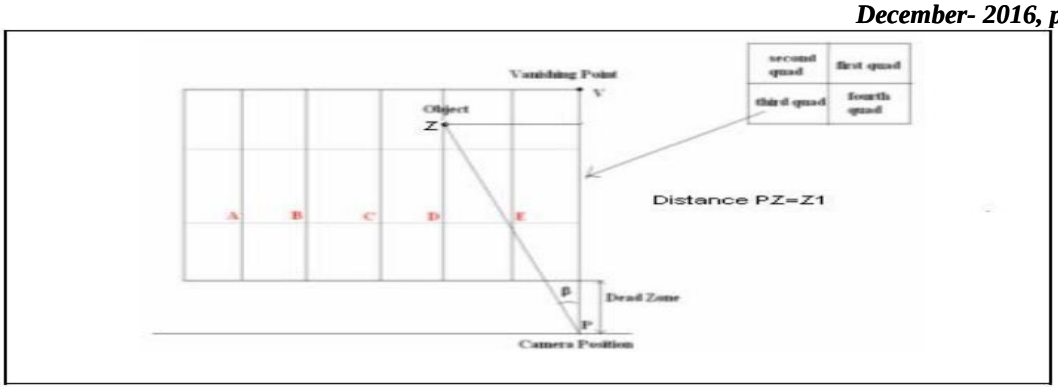

Figure 4. : Figure showing concept used for obliqu e distance calculation.

Zact=4.168e0.059*zcalc

V. DEPTH ESTIMATION

This paper represents a prototype of the system. However, this field of computer science namely computer vision and artificial intelligence is still under intensive development. This prototype requires a lot of computation power. thus, the future work for this project will be to decrease the turnaround time of the whole system which requires less computation and are faster and reliable..

VI. CONCLUSION

The paper represents a cheaper alternative to object recognition and depth estimation. We use monocular cameras instead of binocular ones which are quite expensive. For using this application user doesn't require any other hardware device but rather his/her smartphone. We are currently working on building the prototype for the same. We successfully designed a model hat theoretically defines the problem stated in the abstract.

REFERENCES

[1] SURF: Speeded Up Robust Features Herbert Bay1, Tinne Tuytelaars2, and Luc Van Gool12 1 ETH Zurich {bay, vangool}@vision.ee.ethz.ch 2 Katholieke Universiteit Leuven {Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

[2] Depth Estimation Using Monocular Camera Apoorva Joglekar# , Devika Joshi# , Richa Khemani#, Smita Nair* , Shashikant Sahare# # Dept. of Electronics and Telecommunication, Cummins College of Engineering for Women, Karvenagar, Pune: 411052, India.

[3] https://en.wikipedia.org/wiki/Speeded_up_robust_features

[4] https://support.google.com/accessibility/android/answer/6006983?hl=e

