Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 387 – 394

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet(VisionNet'15)

Vision System with Audio Feedback to Assist Visually Impaired to Grasp Objects

Gautam S[a], K.S.Sivaraman[a], Hariharan Muralidharan[a], A. Baskar[b*]

[a] Student, Dept. of Computer Science and Engineering, Amrita Vsishwa Vidyapeetham (University), Coimbatore, India
[b]Assistant Professor, Dept. of Computer Science and Engineering, Amrita Vsishwa Vidyapeetham (University), Coimbatore, India
*E-Mail: a_baskar@cb.amrita.edu

Abstract

In recent times, there have been many products that cater to the visually challenged, but only few have been readily available for regular use. The reason being, that the technology driving these products are costly, or complex to use. In this work, we propose a "Vision system with audio feedback to assist visually impaired to grasp objects". Our proposed system eliminates these challenges, in its usability, complexity and functionality. The system is designed to serve the following: (1) Finding a desired object in the scene, in which, the object recognition is done using Weighted Matrix Algorithm from the visual input received from the camera; and (2) Assisting the user to the object's proximity, where, the user is guided using audio-feedback, at every step. This approach is represented taking the example of a commonly found object in our household. We have considered a water bottle. The above approach is found to produce suitable results.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Object Recognition, Weighted Matrix, Visually Impaired, Audio Output, Hu Moments;

1. Introduction

The number of visually challenged people is already significant, and is ever-increasing. Owing to their disability, interaction with common household objects is restricted. Technology has assuaged the lifestyle of mankind, whether physically challenged, or visually challenged. Three widely used sensors or methods used to aid the visually impaired are: Audio [10], Visual [3,4,6] and GPS [9]. These have been exploited to the fullest use to ease the condition of the visually impaired. Unfortunately, these endeavours have not been pervasive enough for one or more reasons.

Audio signal has been widely used to facilitate the visually impaired. The following work, Text detection from

1877-0509 © 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.041

natural scene images, uses character extraction to provide a reading model for the visually impaired using text-to-speech [10]. The drawback with text-to-speech is that input in the form of text cannot be used by a visually impaired to interact with objects.

GPS sensors alone are seldom used for indoor navigation. They are generally used along with RFID [9] sensors or tags. There have been many efforts and works on indoor navigation [2]. Obstacle avoidance has also been a well-touched area of research. But as stated before, apart from the navigation, further interaction with the object of interest is restricted. However, the usage of these sensors has not been very common, and do not seem feasible for day-to-day activities.

Only a camera as a visual sensor, can best match the functionalities of a human eye. Furthermore, an audio feedback is the most convenient way to assist a visually impaired user. This system makes use of the visual input scene to generate a vocal output. Not many models have combined visual input [5, 7, 8] and Audio output to assist the visually challenged. A vision system with Audio feedback to assist a visually impaired has been implemented in this paper. Using this system, the user can be brought to the vicinity of the object with ease.

The paper has been divided into 4 sections. Section 2 deals with the elaborate implementation of the system. The result and analysis are discussed in detail in section 3. The final section draws a conclusion to this paper.

2. Proposed Methodology

The proposed system combines two modalities, the vision system and the audio system, to assist the visually challenged. Hu Moments are extracted from the images and used as feature vectors. Hu Moments [11] otherwise called as moment invariants are derived from six absolute orthogonal invariants and one skew orthogonal invariant based upon algebraic invariants, which are not only independent of position, size and orientation but also independent of parallel projection. The moment invariants have been proved to be the adequate measures for tracing image patterns regarding the images translation, scaling and rotation under the assumption of images with continuous functions and noise-free. The proposed methodology can be modularized into three steps, namely: 1) Assist the user to view the region of interest (frame containing the target object), 2) Calibrate the object of interest to the centre of the frame, and 3) Instruct the user to reach and grasp the object. The three steps have been delineated as follows.

2.1 Assist the user to view the region of interest

In this step, the system assists the user in identifying the correct Region of Interest (ROI). The ROI is defined as the frame containing the desired object. The ROI is trained manually prior to this step. To train the ROI, we take 100-150 images of the object along with its surroundings. This takes not more than 5 minutes running a python code. From the images, the Hu Moments are extracted. We use Hu Moments as a feature vector (H_i) from which seven invariant features are obtained as follows:

$$H_i = < h_1, h_2, h_3, h_4, h_5, h_6, h_7 >$$
$$i \in 1 \text{ to } n$$
$$n \rightarrow \text{total number of training images}$$

The feature vector obtained from the trained images [H_1,H_2,H_3…H_N] has been given to the Weighted Matrix (WM) as proposed in our previous paper [1] and the min (m) and max (M) matrices are modelled. The Weighted Matrix algorithm recognizes the object in two phases, the feature Weighted Matrix Model will be constructed using feature vector and is estimated by Hu Moments in the first phase and in the second phase target objects are identified using Diagonal Rank Matrix.

During the testing, a frame is captured. Hu Moments are extracted from this frame as input vector, T.
The matrices 'm' and 'M' and 'T' are all of the order 1 x 7. Deviation from the maximum(R) and deviation from the minimum(r) is calculated as follows:

$$R = M - T. \text{ Hence, R has an order of } 1 \times 7$$
$$r = T - m. \text{ Hence, r has an order of } 1 \times 7$$

Download English Version:

https://daneshyari.com/en/article/487379

Download Persian Version:

https://daneshyari.com/article/487379

Daneshyari.com