Obstacle Detection for Visually Challenged Using Surf

Monisha S\textsuperscript{a}, Saranya M\textsuperscript{b}, Sulochana V\textsuperscript{c}, Hemalatha B\textsuperscript{d}

\textsuperscript{a}\textsuperscript{b}\textsuperscript{c}\textsuperscript{d}Department of Computer Science and Engineering, Velammal Engineering College.

VDGOOD use only: Received 14-Mar-2020; Revised 30-Mar-2020; accepted 20-Apr-2020

Abstract

According to WHO, it is estimated that approximately 1.3 billion people live with some form of vision impairment where India is being the home of world’s largest number of visually challenged people. Visually challenged perform everyday tasks with certain amount of restrictions in mobility due to obstacles in path. To improve the quality of life many embedded systems exists to assist them for obstacle detection using techniques like Stereo Vision, Scale- Invariant Feature Transform(SIFT) and Fast library for approximate Nearest Neighbors. Further supports in mobility to visually impaired people is proposed using image processing technique in obstacle detection to provide obstacle notification with voice assistant. The detection of obstacle is implemented through a KNN classifier and feature extraction is done by SURF algorithm. SURF extracts features of obstacle images captured from different perspectives to estimate segmental key points. These key points are used for categorizing object with the help of KNN classifier. The detected object is intimated through voice assistant to visually impaired people that obstruct their path, thereby warning them for safe mobility. © 2020 VDGOD Professional Association. All rights reserved

Keywords: Features; Obstacle Detection; Descriptor; Matching; Interest Points Surf; KNN.

1. Introduction

Globally, it is estimated that at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion have a vision impairment that could have been prevented or has yet to be addressed. Visually challenged (blind) people has been facing a lot of problems in life by obstacles which are surrounded them. They have to get knowledge about those obstacles to protect themselves from the obstacle. If they know about the nearby obstacle and its type, then they can easily protect from them. For that, they need to get information about nearby obstacles which cause damage to them soon. Object detection and recognition are very important for visually challenged people and that are most researched and used fields under computer vision. They have applications in video surveillance, image retrieval, and video object co-segmentation. Basically, image is a two dimensional matrix that can be shown as coordinates(r,c) with a piece of information called intensity. When this context wanted to be represented in a computer screen, it has to be converted to digital data. This digital data can be used to extract features of
specific objects in order to detect and classify them by using feature descriptors such as Scale Invariant Feature Transformation (SIFT) and Histogram of Gradients. Good object detection and recognition system must be able to detect an obstacle in a good time which can be done by Speeded up robust features (SURF)[1].

The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. SURF descriptors have been used to locate and recognize objects and to extract points of interest from images. And SURF is mainly used to perform object detection and recognition by feature extraction.

2. Related Work

In existing, feature detection and object recognition were done by SIFT in the area of computer vision that is used to detect the features of an image. [7] SIFT is an important object recognition algorithm and it may be used in real-time performance.

This algorithm includes construction of a scale-space, key point detection, outliers removal, assigning orientation to the key points and SIFT features generation for feature extraction process from image. A large collection of can be extracted from images. Each feature should be highly distinctive, so a single feature can be identified from a large database. For object recognition, these SIFT features are acquired from a given image. Even though SIFT gives good performance, it has high dimensionality and not good at illumination changes. But SURF is mainly based on the approximated Hessian Matrix. So it is faster than SIFT. It is widely used in the computer vision applications. The SURF has been high repeatability and distinctiveness [7]. FLANN is used for matching the test features and the database features for object detection and alerting the visually challenged using voice module. [7] FLANN has set of algorithms that are used for nearest neighbor search which is faster in large datasets and also for features which are high dimensional. But KNN classifier in image processing performs better in MATLAB not FLANN.

3. Methodology

3.1. A System Architecture

This paper deals with obstacle detection for visually challenged people. This System Architecture includes SURF algorithm for feature extraction and KNN as classifier. The System architecture consists of training module, obstacle detection module and obstacle notification module. In order to do, image datasets were trained using SURF algorithm. The features are extracted through interest points at the borders and stored in the database. The features are compared with the captured

![Fig.1. System Architecture](image)

3.2. Detection of Object

Detection of object can be done by using SURF Algorithm.

SURF Algorithm

The SURF method (Speeded up Robust Features) is fast and robust algorithm for invariant representation and comparison of images. It can be used for tasks such as recognition, image or 3D reconstruction [1]. The SURF provides fast computation for real-time applications such as tracking and visual perception. The algorithm has three main parts: interest point detection, descriptor and matching.
Feature Extractions

Feature Extraction is to extract interest points from image. Hessian matrix is employed to get intersect points as in fig1.live images from laptop cameras and the results are given through voice message using speaker to the user.

3.3. Integral Images

Integral Image is for calculating the sum of pixel values in an exceedingly given image. It’s used for calculating the typical intensity within a given image. The entry of an integral image at a location $x = (x, y)$ represents the sum of all pixels within the input image $I$ within an oblong region formed by the origin and $x$

$$I_{Σ}(x) = \sum_{i=x}^{i<x} \sum_{j=y}^{j<y} I(i, j)$$

Hessian matrix-based interest points

Surf uses the Hessian matrix for its good performance in computation time and accuracy. Rather than employing a different measure for choosing the situation and therefore the scale (Hessian- Laplace detector), surf relies on the determinant of the Hessian matrix for both[1].

$$H(f(x, y)) = \begin{bmatrix} \frac{∂^2 f}{∂x^2} & \frac{∂^2 f}{∂x∂y} \\ \frac{∂^2 f}{∂y∂x} & \frac{∂^2 f}{∂y^2} \end{bmatrix}$$

For adapt to any scale, the image is filtered by Gaussian kernel, hence for a degree $X = (x, y)$, the Hessian matrix $H(x, σ)$ can be represented by the following:

$$H(x, σ) = \begin{bmatrix} L_{xx}(x, σ) & L_{xy}(x, σ) \\ L_{yx}(x, σ) & L_{yy}(x, σ) \end{bmatrix}$$

Where $L_{xx}(x, σ)$, $L_{xy}(x, σ)$ and $L_{yy}(x, σ)$ is that the convolution of the Gaussian second order derivative with the image $I$ in point. The determinant of the Hessian matrix is calculated by applying convolution with Gaussian kernel and second-order derivative. It pushes the approximation for convolution and box filters. The approximation of second-order Gaussian derivatives are often evaluated at an awfully low computational cost using integral images and independently of size. This leads to Surf algorithm to faster.
The $9 \times 9$ box filters within the above images are the approximations for Gaussian second order derivatives with $\sigma = 1.2$. These approximations are denoted by $D_{xx}$, $D_{yy}$, and $D_{xy}$. Determinant of the Hessian are often represented:

$$\det(H_{\text{approx}}) = D_{xx}D_{yy} - (\omega D_{xy})^2.$$  

Scale-Space Representation

Interest points are found at different scales. Since the hunt for interest points requires comparison of images at different scales. The dimensions space is realized as a picture pyramid in some Feature detection algorithms. Gaussian filter are accustomed smoothen the photographs repeatedly. Smoothened Images are subsampled to induce the upper level of the pyramid. Therefore, various measures of the masks are calculated by the following:

$$\sigma_{\text{appox}} = \text{current filter size} \times \frac{\text{base filter scale}}{\text{base filter size}}.$$  

The scale space is split into numerous octaves. An octave regards to series of maps for covering a doubling of scale. The output for very cheap level of the dimensions space is $9 \times 9$ filters.

3.4. Feature Description

The SURF descriptor falls in two steps. The primary step involves orientation supported information from a circular region round the key point. The second step involves constructing a square region aligned to the chosen orientation and extract the descriptor from it.

Orientation Assignment

In order to be invariant to rotation, surf tries to spot an orientation for the interest points. For achieving this:

1. Surf calculates the Haar-wavelet responses in $x$ and $y$-direction, and this in an exceedingly circular neighborhood of radius $6s$ round the key point, with $s$ dimensions at which the key point was detected. The sampling step $s$ is scale dependent. Hence the wavelet responses are calculated at that scale. The scale of the wavelets is big at high scales. Hence integral images are used for fast filtering.

2. Calculate the sum of vertical and horizontal wavelet responses in an exceedingly scanning area (fig4), then change the scanning orientation (add $\pi/3$), and re-calculate, until the orientation with largest sum value is obtained. This value is that the orientation for the feature descriptor.

![Fig.5. Orientation Assignment](image)

4. Descriptor Components

Extraction of descriptor points are often done by:

1. Construction of a square region round the key point and oriented along the orientation. The scale of this window is $20s$.

2. Then the region is separate into smaller $4 \times 4$ square sub-regions. For simplicity, we represent $dx$ as the Haar wavelet response within the horizontal direction and $dy$ as the Haar wavelet response within the vertical direction. The responses $dx$ and $dy$ are weighted with $\sigma = 3.3s$ so as to extend the robustness of deformation and localization errors.

3. Then, the responses $dx$ and $dy$ are summed up over each sub region and form a primary set of entries to the feature vector. For intensity change within the polarity we extract the sum of absolutely the values of the responses, $|dx|$ and $|dy|$. Therefore, each sub-region will have four- dimensional descriptor vector $v$ for its underlying intensity structure $V = (\Sigma dx, \Sigma dy, \Sigma|dx|, \Sigma|dy|)$. This leads to a descriptor vector of 64 bit. Hence SURF is quicker than SIFT.
4.1. Features Matching

Matching compares the extracted descriptor with the query image within the database. Feature Matching finds the gap between two or more descriptors.

4.2. Object Classification

Classification of object is done by KNN classifier.

4.3. KNN classifier

K-Nearest Neighbor algorithm is that the most effective method used for classification. KNN may be a non-parametric algorithm used for regression problem and classification of objects. KNN is additionally used for feature similarity to predict the cluster. \( K \) is the value that's accustomed identify similar neighbors for brand new point. The new information location is set by \( K \) nearest neighbors and hence the decision is predicted on feature similarity. Choice of \( K \) includes a drastic impact on the results we obtain from KNN (fig 7).

Take the test set and plot the F1 score against different values of \( K \). Error rate is high for test set when \( K=1 \). Therefore, that model overfits when \( k=1 \). When a worth of \( K \) increases, the F1 score starts to drop. Error rate is minimum for test set when \( K=5 \) as represented in the below graph (fig 8). The optimal value of \( K \) is obtained from the worth of \( K \) at elbow of test error rate.

Step 1: Pick the worth of \( K \) that ought to be an odd number.
Step 2: Find the gap of the new point for every of the training data sets.
Step 3: Find the \( K \) nearest neighbors to the new information.
Step 4: To classify the article, count the amount of information points in each category among the \( k \)
neighbors. New point will belong to class which has the foremost neighbors. For regression, the worth of the new info are going to be the typical of the k nearest neighbors.

4.4. Distance Calculation

Distance is calculated by the following metrics:
1. Euclidean Distance.
2. Manhattan Distance.
3. Chebyshev Distance.

4.5. Euclidean Distance

The Euclidean distance or metric is the line distance between two points in Euclidean space (fig 9).

4.6. Manhattan Distance

The Manhattan distance is the sum of the absolutely values of differences \( \sum|a_i - b_i| \) between two points(fig 10).

4.7. Results and Discussion

In this study, image datasets were trained using SURF algorithm (fig.13,14,15, 16). Predefined objects were detected and results were displayed as both visual (fig.21) and auditory to compare the correctness of auditory findings. It was concluded that auditory results were mostly matched with visual results.

When an object was recognized the interest points (fig.19,20) on the border of the image were calculated and represented as green circle (fig.18). Before calculating the interest points the input image was converted into gray scale image (fig.17) to be most able to work with SURF algorithm. The interest points were compared with the interest points of pre-trained model and according to results with KNN classifier the signal was sent to audino and so the auditory output was intimated to the user.
The captured RGB images are converted to Grayscale images for processing SURF algorithm.

Feature Extraction is to extract interest points from image. Hessian matrix is employed to get intersect points as in fig18.
The pixel values of the grayscale images are plotted in the above Fig.19.

The above Fig.20 represents the mean value of the trained images.

The above Fig.21 represents the name of the detected obstacle.

5. Conclusion

In this paper, we proposed a system which is an integration between obstacle detector and system that helps visually challenged from obstacles when they are moving. Our system is designed to perform obstacle detection and alert through voice message. The proposed algorithms take the onboard camera to achieve complex tasks and safe obstacle detection tasks.

To increase detection performance, Fuzzy logic, and other advanced AI approaches can be implemented as future work.

References


