SMART NAVIGATION TO ASSIST BLIND PEOPLE FOR OBJECT DETECTION

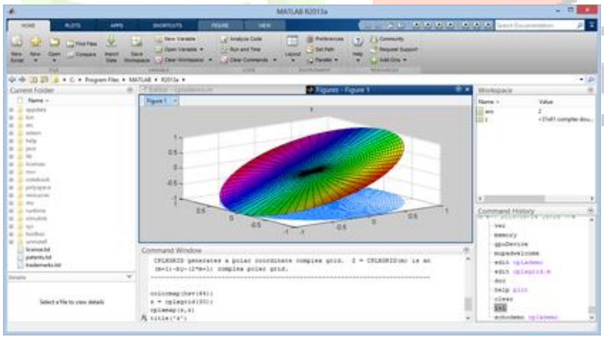
[1] Pachhaiammal@Priya M
[1]Assistant Professor
[1]Computer Science and Engineering,
[1]Sri Sai Ram Institute of Technology, Chennai, India

Abstract: **Vision is one of the very essential human senses and it plays the major role in human perception aboutsurrounding environment. The traditional and oldest mobility aids for persons with visual impairments are the walking cane (also called white cane or stick) and guide dogs. The most important drawbacks of these aids are necessary skills and training phase, range of motion and very little information conveyed. With the rapid advances of modern technology, both in hardware and software front have brought potential to provide intelligent navigation and language capabilities. In proposed system we use camera to capture the image then the image will be processed to extract the features of the object using two algorithms. The simulation results using SIFT algorithm and the SOBEL algorithm is used to extract edge key points matching showed good accuracy for detecting objects. Object detection deals with extracting embedded details, patterns and their relationship in images. The captured image will be compared with theimages that are already stored in the datasets. The images that are matched with the dataset images are converted into text message and then the text is converted into voice notes.All the processing will be done by using MATLAB.**

Keywords:Camera capture,Intelligent navigation,Different language processing,MATLAB.

I. INTRODUCTION

According to the WHO (World Health Organization), there are nearly 285 million people, who are visual impairments,39 million of them are blind and 246 million have a decrease of Visual acuity. Almost 90% who are visually impaired areliving in low-income countries. In this context, Tunisia has identified 30,000 people with visual impairments; including13.3% of them are blind. The daily life of the blind people will not be that much easy as we think. With the help of the improvement in the technology, many devices were helping them to lead their life. The existing system is not that much efficient for detecting the seriousness of the lively object. It can predict the object with 60% efficiency, but the proposed system will predict the object with 95% efficiency. Here we are using two algorithms namely, SIFT and SOBEL. By merging these two algorithms in MATLAB, the features of the object are extracted. **MATLAB** (**mat**rix **lab**oratory) is a multi-paradigm computing environment and fourth-generation programming language. A proprietary programming language developed by Math Works,



II. OVERVIEW

Image segmentation is the foundation of the object recognition and computer vision. In general, image noise should be eliminated through image preprocessing. And there is some specifically given work (such as region extraction and image marking) to do after the main operation image segmentation for the sake of getting better visual effect. Two major computer vision problems, image segmentation and object recognition, have been traditionally dealt with using a strict, bottom-up ordering. Image segmentation is the process of partitioning/subdividing a digital image into multiple meaningful regions or sets of pixels regions with respect to a particular application. The segmentation is based on measurements taken from the image and might be grey level, color, texture, depth or motion. The result of image segmentation is a set of segments that collectively cover the entire image. All the pixels in region are similar with respect to some characteristic or computedproperty, such as color , intensity, or texture. Adjacent regions differ with respect to same characteristics. Edge detection is one of the frequently used techniques in digital image processing. Object recognition is the task of finding a given object in an image or video sequence. For any object in an image, there are many ‘feature’ description of the object. This description extracted from a training image can then be used to identify the object when attempting to locate the object in a test image containing many other objects.

1. IMAGE

An image is an array or a matrix of square pixels (picture elements) arranged in columns and rows. An image (from Latin: imago) is an artifact, for example a two-dimensional picture, that has a similar appearance to some subject usually a physical object

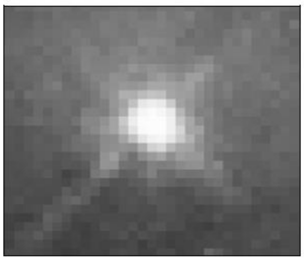
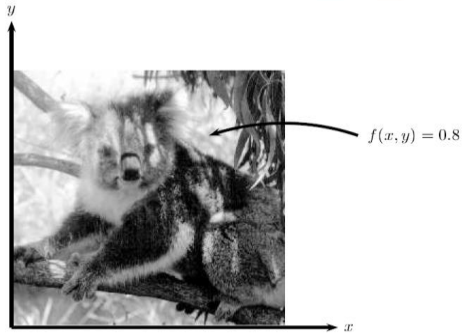


Figure 1.1: An image – array or a matrix of pixels arranged in columns and rows.

2. PIXEL

Image processing is a subset of the electronic domain where in the image is converted to an array of small integers, called pixels, representing a physical quantity such as scene radiance, stored in a digital memory and processed by computer or other digital hardware.



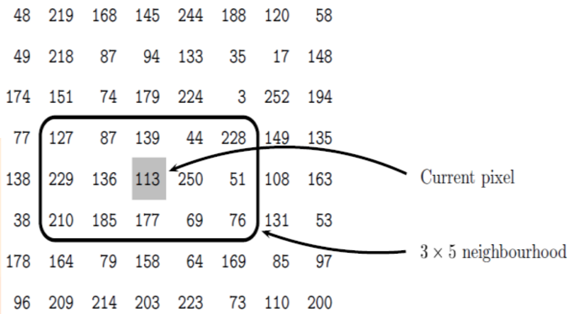


Figure 1.1.1: A grey scale image **Figure1.1.2:** Pixel representation

3. IMAGE RESOLUTION:

Pixels transform into inches through what is called "resolution", -- the number of pixel per square inch on a computer. Resolution allows you to transform pixel into inches and back again.

Two resolution definitions are often used in place of one another. Pixel resolution is the size (in bytes) of your image or its appearance on a computer screen. This number is tied directly to how big your image is on your hard drive. The byte –size of the image file is directly proportional to the pixel count and its size on your computer screen, which simply displays all the pixels in a fixed one-to-one grid.

4. IMAGE BRIGHTNESS:

Brightness actually describes how we experience light and not 'how it is'. If we are going to describe light and 'brightness' properly, there are two essential terms:

1. Luminance
2. IL luminance

Luminance:

Luminance is the light we see: reflected or radiating from objects. It is measured in candela per square metre.

IL luminance:

IL luminance is light we can't see directly – it is so-called ambient light. It is measured in lux.

5.CONTRAST:

The Image contrast ratio refers to the difference between the luminance of the white part of an image, divided by the black part. So if the white part is one hundred times brighter than the black part, it will be 100:1, and so on.

6.IMAGE PROCESSING:

Image processing is any form of single processing for which the input is an image, such as a photograph or video frame. The output of the image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a dimensional signal and applying standard signal-processing techniques to it.

The example in figure 2.1 operate on 256 gray-scale images. This means that each pixel in the image is stored as a number between 0 to 255, where 0 represents a black pixel, 255 represents a white pixel and values in-between represent shades of gray.

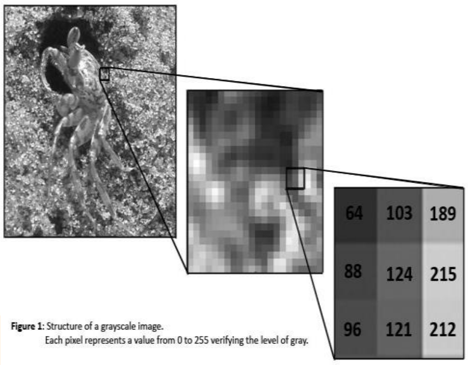


Figure 1.2.1: Structure of grey-scale image

In figure1.2.1 each pixel represent a value from 0 to 255 verifying the level of gray. These operations can be extended to colour images too.

III. PROPOSED SYSTEM

Most artificial vision systems are based on rules for converting images into data sensory substitution tactile or auditory stimuli. Indeed, a large part dedicated to the blind apparatus attaches directly to the cane as a complement for detection. Sensory substitution systems are very encouraging because today they can locate and recognize objects from images with very simple conversion's rules. However, they can only be used to localize simple patterns and cannot be used as tools of substitution in natural environments.These systems were efficient for mobility and localization of objects which is sometimes with a lower precision, thus they have some limitations that we want to surmount them. In fact, they do not identify objects (e.g whether it is a table or chair) and they have in some cases a late detection of small objects. In addition, some seek additional auditory, others require a sufficiently long learning period. The importance of designing a system based on the recognition and localization of objects to meet the major challenges of the blind is in three main categories of needs: displacement, orientation and object recognition.

In our approach, we propose to analyze the frame of images using the features selection's algorithm in order to translate detected object.Then for each frame we have to use the mentioned algorithm. Features extracted for each object are saved in the database. We compute keypoints in the target frame. We make correspondence between features extracted from the frame and those saved in the database. If frames are the same, the list of detected objects,that was stored, will appear, else if they are distinct, features from the target frame will be matched with those of objects saved on the database, in order to identify the new list of detected objects.If the matching between different features shows the detectionof some objects, we save the list of detected objects and extracted features from the frame into the database. Then the text is converted into voice message to indicate the seriousness for the blind people.

1. SIFT

The cost of extracting features of objects is minimized by taking a cascade filtering approach, in which the more expensive operations are applied only at locations that pass an initial test. Following are the major stages of computation used to generate the set of image features:

1. Scale-space extrema detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
2. Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.
3. Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.
4. Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint.

These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features.An important aspect of this approach is that it generates large numbers of features that densely cover the image over the full range of scales and locations. A typical image of size 500x500 pixels will give rise to about 2000 stable features (although this number depends on both image content and choices for various parameters). The quantity of features is particularly important for object recognition, where the ability to detect small objects in cluttered backgrounds requires that at least 3 features be correctly matched from each object for reliable identification.

For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors.The keypoint descriptors are highly distinctive, which allows a single feature to find its correct match with good probability in a large database of features. However, in a cluttered image, many features from the background will not have any correct match in the database, giving rise to many false matches in addition to the correct ones. The correct matches can be filtered from the full set of matches by identifying subsets of keypoints that agree on the object and its location, scale, and orientation in the new image. The probability that several features will agree on these parameters by chance is much lower than the probability that any individual feature match will be in error. The determination of these consistent clusters can be performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed verification. First, a least-squared estimate is made for an affine approximation to the object pose. Any other image features consistent with this pose are identified, and outliers are discarded. Finally, a detailed computation is made of the probability that a particular set of features indicates the presence of an object, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

IV. SOBEL

Sobel operator is used in image processing and computer vision,particularly within edge detection algorithms where it creates an image emphassingedges.It is based on convolving the image with a small,separable, and integer-valued filter in the horizontal in terms of computations. It is a discrete differentiation operator,computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel-Feldman operator is either the corresponding gradient vector or the norm of this vector.The kernals are designed to respond maximally to edges running vertically and horizontally relative to the pixel grid, one kernal for each of the two perpendicular orientations.The hernals can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation.These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient.

The Sobel operator is slower to compute than the Roberts Cross operator, but its larger convolution kernel smooths the input image to a greater extent and so makes the operator less sensitive to noise. The operator also generally produces considerably higher output values for similar edges, compared with the Roberts Cross.As with the Roberts Cross operator, output values from the operator can easily overflow the maximum allowed pixel value for image types that only support smallish integer pixel values (e.g. 8-bit integer images). When this happens the standard practice is to simply set overflowing output pixels to the maximum allowed value. The problem can be avoided by using an image type that supports pixel values with a larger range.The image shows a simpler scene containing just a single flat dark object against a lighter background. Applying the Sobel operator produces.All edges in the image have been detected and can be nicely separated from the background using a threshold of 150.Although the Sobel operator is not as sensitive to noise as the Roberts Cross operator, it still amplifies high frequencies. The image is the result of adding Gaussian noise with a standard deviation of 15 to the original image. Applying the Sobel operator yields and thresholding the result at a value of 150 produces the edge points.

A. IMAGE PROCESSING

Recognition and classification of objects are subjects of very active research. They are performed on several digitalmedia such as images or videos. In a visual scene of real life, object recognition task is often focus on the overalltask of the semantic interpretation. Image and video data can be perceived differently making it very difficult tomaintain complete content information of images or video.There are two broad categories of general object recognition, namely instance recognition and class recognition. The former involves re-recognizing a known 2D or 3D rigid object, potentially being viewed from a novel viewpoint, against a cluttered background, and with partial occlusions. The category-level or generic object recognition is the challenging problem of recognizing any instance of a particular general class such as cat, car, or bicycle. Searching for objects in a video scene can be also accomplishedusing algorithm of features extraction and descriptor for matching signatures of objects. Feature-based systemoperates much faster than a pixel-based system. To analyze the content of videos, the first step is to extract informationthat characterizes each frame. The main purpose of using features instead of raw pixel values as the input to a learning algorithm, is to reduce/increase the in-class/out-of class variability compared to the raw input data, and thus making classification easier.

B. FEATURES EXTRACTION

Features extraction is the process by which certain features of interest within an image are detected and representedfor further processing. It is a critical step in most computer vision and image processing solutions because it marks the transition from pictorial to non pictorial (alphanumerical,usually quantitative) data representation. Types of features that can be extracted from image depend on the type of image (e.g. binary, gray-level, or color), the level of granularity (entire image or individual regions)

desired, and the context of the application. Once the features have been extracted, they are usually represented in an alphanumerical way for further processing. The actual representation depends on the technique used. The features extraction process should be precise, so that the same features are extracted ontwo images showing the same object. It consists on five steps:
1) Find a set of distinctive keypoints.
2) Define a region around each keypoint in a scale- or
affine-invariant manner.
3) Extract and normalize the region content.
4) Compute a descriptor from the normalized region.
5) Match the local descript.

Feature extraction can be obtained by 6 factors:

1) **Mean**:
The mean of a data set is simply the arithmetic average of the values in the set, obtained by summing the values and dividing by the number of values.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} f_i x_i = \sum_{i=1}^{n} p_i x_i$$

2) **Variance:**
The variance of the data set is the arithmetic average of the squared differences between the values and the mean.

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - X_{avg})^2}{n-1}$$

3) **Standard Deviation:**
The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

4) **Skewness:**
It is ameasure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

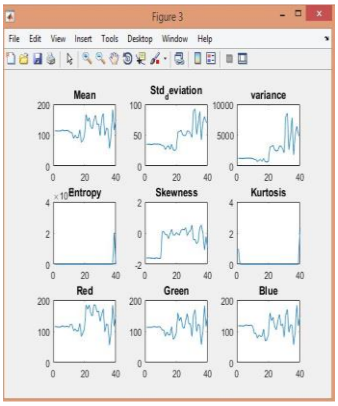
$$g_1 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^3 / N}{s^3}$$

5) **Kurtosis:**
It is measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.

$$\text{kurtosis} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^4 / N}{s^4} - 3$$

6) **Entropy:**
Entropy is defined as the quantitative measure of disorder or randomness in a system.

delta-S = Q/T



C. DESCRIPTORS

Extracted features represent the interesting points found in the image to compare them with other interesting points.

Descriptors are used to describe these features. They are generally based around points of interest of the image andoften associated with a detector of keypoints .The descriptors can be global, local or semi-local:

• Global image descriptor: features overall image are usually based on color indices and the most famous global color descriptor is the color histogram.

• Local image descriptors: Local features are ones that have received the most attention in recent years. The main idea is to focus on the areas containing the most discriminated information.

• Semi-local image descriptor: most shape descriptors fall into this category. This descriptor is based on extracting accurate contours of shapes in the image or in the region of interest. In this case, image segmentation is generally useful as a preprocessing step. Several methods for features localization and description have been proposed in the literature. In this paper, we aimto use features extraction algorithm SIFT.

IV. CONCLUSION

Visual substitution systems are generally focusing on the mobility, the redirection and the detection of obstacles. Ourgoal is to provide a robust and easy system allowing the blind people or visual impaired to describe their environmentsby objects identifiers. This system is based on the analysis of aimages that are stored in dataset. Then it will undergo an image processing to identify different objects detected.In this paper, we have evaluated SIFT and SOBEL, the famous methods for features extraction for objects detection. In order to ameliorate and increase the object recognition rate, we will consider the color while identifying objects and add an auditory translation for identified objects.

V. REFERENCES

[1] Iasonas Kokkinos, and Petros Maragos (2009),”Synergy between Object Recognition and image segmentation using Expectation and Maximization Algorithm”., IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol. 31(8), pp. 1486-1501, 2009.

[2] Wen-Xiong Kang, Qing-Qiang Yang, Run-Peng Liang (2009), “The Comparative Research on Image Segmentation Algorithms,” First International Workshop on Education Technology and Computer Science.

[3] V. Ferrari, T. Tuytelaars, and L.V. Gool(2004), “Simultaneous Object Recognition and Segmentation by Image Exploration,” Proc. Eighth European Conf. Computer Vision, 2004.

[4] B. Leibe, A. Leonardis, and B. Schiele(2004), “Combined Object Categorization and Segmentation with an Implicit Shape Model,” Proc. ECCV Workshop Statistical Learning in Computer Vision, 2004.

[5] Y.Ramadevi, B.Kalyani, T.Sridevi(2010), “ Synergy between Object Recognition and Image Segmentation”, International Journal on Computer Science and Engineering, Vol. 02, No. 08, 2010, 2767-2772.

[6] N.Senthilkumarn, R.Rajesh(2009), ”Edge Detection Techniques for Image Segmentation- A Survey of Soft Computing Approaches”, IJRTE, vol1,No2, 2009 250-254.

[7]. A. Petrosino, "Moving Object Detection for RealTime Applications," in Proc. IEEE 14th Int. Conf. Image Analysis and Processing, Modena, Italy, Sep. 2007, pp. 542-547.

[8] G. Zhang, J. Jia, W. Xiong, T. T. Wong, P. A. Heng, and H. Bao, “Moving Object [3]L. Maddalena, and A

[9] Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching’Karan Gupta1, Anjali V. Kulkarni2 Indian Institute of Technology, Kanpur, India

[10]‘MOTION OBJECT DETECTION OF VIDEO BASED ON PRINCIPAL COMPONENT ANALYSIS ‘Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.

[11] Moving object detection tracking system : a real time implemented, SEIZIÈME COLLOQUE GRETSI — 15-19 SEPTEMBER 1997 — GRENOBLE

[12] Segmentation and tracking of multiple video objects Received 7 June 2005; received in revised form 23 February 2006; accepted 14 July 2006`