Implementation and Evaluation of Object Identification Techniques on Nao Robot Platform

Zeinab Farrokhi1*, Ahmad Kalhor2, Mehdi Tale Masouleh3

1Human and Robot Interaction Laboratory, Faculty of New Sciences and Technologies, University of Tehran
2Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran
3Human and Robot Interaction Laboratory, School of Electrical and Computer Engineering, University of Tehran

*Corresponding Author’s E-mail: z_farrokhi@ut.ac.ir

Abstract

Interaction and communication of children with Autism with a humanoid robot would improve their social and verbal interactions and the quality and the speed of the robot in doing interactions will influence the rate of patient health improvement. In this paper, a system is proposed in which the conversion of image into natural language is deployed on Nao humanoid robot as the performance is improved. Actually in this method an image is shown to the robot and it would recognize the shape of the image straight off and say the quantity and type of them using its voice system. In each step of this system several algorithms would be tested and based on results comparison from the error, speed and accuracy point of view, a combination of most optimal of them will be used in proposed algorithm. At last the algorithm is deployed on the humanoid robot and had better results in accuracy and speed than the robot default mode was obtained. This system operated successfully in 95.7 percent of test images and has 14 percent improvement than the default mode.

Keywords: Robot interaction, Autistic children, Recognize shapes, Image Processing.

1. Introduction

Besides the practical objectives of social humanoid robots design and manufacture, research goals of researching and academic robots is important. In recent years, extensive works and researches in humanoid robots were done in such fields as Cognitive Sciences, Natural Interaction, all body physical Integrations with environment, objects movement and etc. Using robots has significant growth in training, recognition and treatment of children with autism in the world. Various robots, such as Nao Humanoid robots, are currently being used to treat and educate children with autism. Previous researches[6, 9, 4, 19]have shown that this humanoid robot has acted successfully in communicate and interaction with children as a training aid and assistant therapist and in some cases when a robot is in their room, the results show up to 30 percent increase in social Interactions and verbal communications. These studies also demonstrate that these improvements can help the children interactions with their parents and staff of doctors. In future by gathering the information of children with autism, this robot can be implemented as a diagnostic tool in patients treatment. A social humanoid platform can be a proper model for in- vestigating in these field with regards to its
communication- motion abilities and developing its capabilities, inventing the treatments and performance, and approaching it to man would results in more efficiency and growth in these fields. A performance closer to human treats and characteristics such as less error rate and audio output can be added to robot features and improve its efficiency in current applied fields.

The method used in this paper has been developed in response to the robot’s default method, and the ability to express the number of each type of it has been added to it, and it also has faster speed and precision than the default mode. As the system overview is shown in the Fig. 1, this article purpose is improving the conversion of image data into natural language by developing the performance and improving the growth speed and accuracy using humanoid robot visual and audio system. In the order that the robot extracts and processes the information immediately and straight off and tells type and the quantity of each like a human using an audio system. Research is divided into 3 steps: obtaining the environment images, image processing and extracting the intended geometric shapes and telling the achieved information by the robot. The implemented humanoid robot in this research is Nao humanoid robot. The main operating system of this robot is updated version of Gentoo a Linux based operating system, named open Nao.

In this research the data transfer and all the related operations to three parts of this project such as receiving the image, processing it and information extracting and audio informing is based on Linux operating system because of more speed and efficiency. For communicate with Nao robot and access the sources and files of this operating system, can be used the file exchange protocol, included environments or dedicated Nao software.

The shape detection default application is prepared exactly like dedicated Nao software such as choreograph and Monitoring and at the same environment. For Nao robot, its possible to program coding and code-based program development in different languages and environment and just by using two programming language of python and C++ can write all Nao humanoid robot embedded programs. Therefor in this article all the programs are developed in C++ language. In addition, in order to a dynamic communication with user two speakers, four Microphones and 51 optical indicators are prepared for Nao robot.

The geometric patterns subject recognized by humanoid robot of this article can be deployed in various fields of image processing. Such fields as Industry, Medical, Military and Security Sciences, Geology, Astronomy, Urban planning, Cinema and Art, scientific Technologies, politics, psychology, agriculture, Meteorology, Archeology, Economic, Commercials, Traffic and Measurement are just few applicable fields of geometric patterns recognition.

In image processing-based intelligent systems, the data analysis and information extracting is based on accommodating image extracted shapes with predefined geometric algorithms such as line, circle, square, rectangular and etc. In recent years, different algorithms were proposed for edge derivation and image lines [10, 12, 13]. Liow[21] proposed a method of recognizing closed paths and Connected geometric shapes by investigating borders and Kim[22] extracted the Topological features and shapes characteristics directly from the image . Carson did the image edge detection based on maximizing edge signal and wavelet [5].

In machine vision and image processing using some math operations like edge detection with gradient or implementing proper filters, image characteristics like edges, lines, curves, corners and borders can be extracted [5]. In the following we will discuss 3 steps of research in form of several sections respectively. First, methods of receiving image from environment and audio informing of obtained results of image analysis will be expressed, then how this shapes of image will be extracted including several sections. In the next step, the implementation results would be discussed and finally we will work on summery and conclusion of article.
2. MATERIALS AND METHODS

Robot camera images are used in this article. The Nao robot has 2 cameras with 1.2 Mp resolution and 30 frames per second. While all the systemic resources are dedicated to processing, it can provide an image with 1280×960 pixels. But in normal situations, this quality will be reduced to 640 × 480 pixels. As it is shown in the Fig. 4, the Technical Specifications of these cameras are as follows:

- sensor model: MT9M114
- camera output: 1280 × 960 pixels and 30 frames per second and based on YUV422
- visibility (scope): 72.6 degree
- focal distance: 30cm to infinity

The photos of painted pictures and environment that are including geometric shapes is taken by the
robot digital camera. The connection with Nao robot can be with an Ethernet port 48 and Wi-Fi wireless network and the commands can be sent over it. In this research, the data transferring and pictures and commands is done using the Wi-Fi wireless network. Considering that in this research the system operates so that automatic image positioning in front of the camera with high speed and accuracy expression of type and quantity of geometric shapes by the robot, by placing the image in front of the camera, the image taken by robot will be transferred through the wireless system using the humanoid robot libraries in Linux environment to an open source software OpenCV. Thus this image will be processed to extract the features and information. General steps of system is shown in the Fig. 3 of system steps flowchart.

2.1. Geometric patterns detection algorithms

As the shapes extraction from images and the importance of edge detection was mentioned in introduction, image processing and the implementation methods used for this project has been investigated as the first step. The human eye is empirically familiar with certain forms so it would recognize and detect the shapes powerfully with high speed and accuracy but detecting and identifying objects based on shape parameters is one of the most difficult subjects in terms of proposing algorithm. Because there are various components that cause significant changes in the shape representation in the images. So that this article purpose is to bring the performance of the humanoid robot closer to the desired speed and accuracy.

In many cases, geometric features of being circular, triangular, rectangular, or square objects play an important role in describing and identifying objects. Therefore, in some applications, the calculation and measurement of this shapes are used as identification parameters [14]. In many image processing methods, the first step is to extract the feature. Extracting a feature is the process that identifies the characteristics of the data by performing operations on the data. The purpose of extracting features is to make raw data more usable for later processing. Extracting features is a very common process in various types of data processing, such as image processing, audio processing, and so on. Detectable structures in the image are called features and are used to display patterns and objects in various applications such as shapes discovery and identification.

The feature extraction in image processing can also be interpreted as a particular form of problem dimensions reduction, because by defining the useful the inappropriate part of the information is eliminated feature rather than the large amount of information that should be analyzed and processed by an algorithm and thus, some valuable features called feature vector are given as the input to the algorithm. This process is called a feature extraction in the term, and in addition it reduces the complexity of the computations in order to provide a better response to the algorithm [17] and this results in improvement in algorithm speed.
2.1.1. Processing and extracting features from image and identifying and detecting shapes:

Defines an edge in an image as a border or contour in which an important alternation happens in some physical aspects of images. Extracting these features would result in ease of display and analyze of image pages. The edges are determiners of borders and some spots of images that the brightness level changes sharply. The high volume of information contained in an image, including information about the shape of objects, lies at the edges of the latitude lines. Therefore, the change in light intensity, which has a significant effect on the way the image area is displayed, produces much less effect on the edges. So far, due to the importance and variety of edge application, many methods have been introduced in this area. Some of these algorithms are Robert Prewitt, Sobel, and Canny that depending on the advantages and disadvantages, each one is used in its own applications. These methods are measured by comparing the performance of these methods for the application of this paper, We've been

Figure 3: System steps flowchart.
severely numb the effect of these weak points. All edge detection algorithms also have some parameters that should be calculated for each image or each work area. The methods are introduced in follows for comparison. Sobel algorithm: this algorithm detects the edges by using derivative estimates that returns the edges in those points where the gradient of the image is maximal.

- Sobel algorithm takes gradient image in two directions of x and y, so the following operators (Fig. 4) are implemented on the image. Then, using Eq. (5), gradient norm is calculated for each block:

\[ G = \sqrt{G_x^2 + G_y^2} \]  

(1)

The value of G represents the edge strength. Using Eq. (2), edge direction is determined for each block.

\[ \theta = \tan^{-1} \frac{G_y}{G_x} \]  

(2)

It should be noted that in the image, any direction can not be considered as the direction of the edge. Instead, only angles of (horizontal direction) and (positive direction of the diagonal) and (vertical direction) and the (diagonal negative direction) can be considered for the direction of the edge (Fig. 5). Therefore, according to the following equation, the direction of the edge is determined.

\[
\begin{align*}
0 \leq \theta &\leq 22.5 \rightarrow Dir = 0^\circ \\
157.5 \leq \theta &\leq 180 \rightarrow Dir = 0^\circ \\
22.5 \leq \theta &\leq 67.5 \rightarrow Dir = 45^\circ \\
67.5 \leq \theta &\leq 112.5 \rightarrow Dir = 90^\circ \\
112.5 \leq \theta &\leq 157.5 \rightarrow Dir = 135^\circ 
\end{align*}
\]  

(3)

\[
\begin{array}{ccc}
-1 & 0 & +1 \\
-2 & 0 & +2 \\
-1 & 0 & +1 \\
0 & 0 & +2 \\
-1 & -2 & -1 \\
+1 & +2 & +1 \\
0 & 0 & 0 \\
+1 & +2 & +1 \\
\end{array}
\]

Figure 4: The used mask by Sobel operator.

\[
\begin{array}{c}
135^\circ \\
90^\circ \\
45^\circ \\
\end{array}
\]

Figure 5: Edge direction.
• Prewitt algorithm. This algorithm is very similar to Sobel algorithm but the mask coefficients are different. Fig. 6. shows the Prewitt edge detector mask

• Roberts algorithm. This algorithm is very similar to Sobel too. Figure 7 shows the Robert edge detector mask.

• Canny algorithm. The Canny edge detector norms the image to eliminate the noise effect. Then calculates the image gradient to find the high changing areas (High spatial derivations). Then the algorithm moves along these areas to prevent any pixel which gradient is not maximum (finding local maximum). In next step the Hysteresis concept is used. Hysteresis uses two high and low thresholds and if the value and intensity of each pixel was below the lower threshold its value would be zero (is not considered as an edge). If the value was between two thresholds, it would become zero unless there has been one path from this pixel to another with a higher gradient than higher threshold. In other words, there is a connection between this pixel with edge pixels and if the pixel value was more than the higher threshold, that pixel would be considered as an edge pixel.

![Figure 6: The Prewitt edge detector mask for Gx,Gy.](image)

![Figure 7: Used masks by the Robert edge detector.](image)

2.1.2. Processing Identifying Polygon Patterns:

In the topic of identifying the geometric shapes, the straight lines are considered as a very important feature. Also, by extracting the straight lines of the image, many geometric shapes like square, rectangle, triangle, etc. will be displayed. So, as a clue to discovering a shape, you can extract straight lines from the image. A lot of methods are not introduced in the field of extracting straight lines of image. According to research, the Hough conversion is the most common method of detecting straight lines in the image [7, 15].

The Hough conversion is a technique of feature extracting which is used in image analyzing, computer vision and image processing. This conversion is a strong tool for extracting straight lines of an image. So this technique is implemented in this research to extract straight lines feature. The corner finders are a very useful algorithm of image processing which is used in various real time applications like matching images. Corners have important information of shapes and latitudes in the image and is defined as a point with high curvature on the peripheral curve of the object. Corners are the most prominent feature of the peripheral curve due to their stability against transfer, rotation, and scale change.
Accordingly, the exact and accurate extraction of them for next processes a very important issue. Shapes corners are considered as a key feature to find them and due to low volume of computation its vastly been used in various applications. Another positive aspect of the corners is the fact that it is not related to the scale of the image, which is aligned with the deliberate use of the article, which includes a wide range of forms. Commonly used cornering techniques can be divided into two main curve-based groups based on light intensity. In the first group, first extract the edges of the image, and then look at the maximum curvature at these curves. But intensity-based methods, using gradation grades of the image directly, estimate a criterion, which is an indication of the presence of the corner. Two common cornering methods are Harris and SUSAN which place in the second category.

According to studies, in general, Harris’s method is more resistant to noise and better matching between different images than SUSAN method [8]. Harris algorithm is selected for practical implementations of this article because this algorithm is one of the most important for extracting spot features and also has suitable Sustainability. Harris algorithm is a corner detecting method that uses Corner Evaluation function [2, 11, 16, 20]. Hessian is a method of defining the corner Evaluation function that is defined as $C = I_{xx}I_{yy} - I_{xy}^2$. By finding the local maximums of function, c, the corners will be obtained. The other method of corner detecting is Kitchen Rosenfeld algorithm that defines the corner Evaluation function based on the result of multiplication of the local gradient amplitude and gradient direction changes along the edge. This function is defined as follow [11].

$$C = \frac{I_{xx}I_{yy} - 2I_{xy}I_{yx} + I_{yx}I_{yy}}{I_{xx}^2 + I_{yy}^2} \quad (4)$$

The above methods use second derivative that increase sensitivity to noise and have a low stability [3]. But in Harris algorithm for reducing noise effect and more stability, the first derivative is used. assuming use of second gradation grade of image the method of corner detecting based on Harris algorithm is as follow: if the image is shown by I and the image frame with center on $(v,u)$ move as much as $(x, y)$, then:

$$S(x, y) = \sum_u \sum_v w(u, v)(I(u + x, v + y) - I(u, v))^2 \quad (5)$$

$(I(u + x, v + y))$ it would be estimated as follow with Taylor expansion:

$$I(u + x, v + y) \approx I(u, v) + I_x(u, v)x + I_y(u, v)y \quad (6)$$

so $S$ is estimated as the Eq. (7).

$$S(x, y) \approx \sum_u \sum_v w(u, v)(I_x(u, v)x - I_y(u, v)y)^2 \quad (7)$$

$S$ can be in form of Eq. (8) as a matrix.

$$S(x, y) = \begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} \quad (8)$$

That $A$ is defined as Eq. (9).

$$A = \sum_u \sum_v w(u, v) \begin{bmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{bmatrix} = \begin{bmatrix} \langle I_{xx}^2 \rangle & \langle I_{xy} \rangle \\ \langle I_{yx} \rangle & \langle I_{yy}^2 \rangle \end{bmatrix} \quad (9)$$
The $A$ matrix is called Harris matrix. Since calculating special amounts of $A$ is complex, the $M_c$ value can be calculated, instead.

$$M_c = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(A) - k \text{trace}^2(A)$$

(10)

$K$ is the Sensitivity factor of function that usually 0.04 is suggested [1].

3. DISCUSSION AND IMPLEMENTATION RESULTS

The main title (on the first page) should begin 1 3/16 inches (7 picas) from the top edge of the page, centered, and in Times New Roman 14-point, as is shown in the Fig. 9, images taken by robot are processed by common methods of corner detecting and the results were compared together for less error and better performance. A sample test image with 4 ways of corner detection is shown in Fig. 8. The results in the Fig. 8 show that the detected edges using Canny algorithm is superior to others. The Robert algorithm is very similar to Sobel but is more sensitive to noise and uses less pixels for gradient estimation and also has less strength than Canny algorithm. As is illustrated in results the Canny algorithm results better than other algorithms. The most important goal of Canny is low error rate. It is important not to miss the edges in the image and the algorithm does not react to non-edges.

The other goal is to identify the points on the edge, in other words the distance between edge pixels is detected by the detector and its actual state is minimal. The other The next criterion is that the algorithm only has one response to a single edge. This criterion is considered to eliminate the effect of multiple response to a single edge when other two criterions did not have enough ability.

The other problem is noises. The noises have high frequencies like edges. The Canny algorithm emphasizes on noise reduction in the image. So with a tone that Canny algorithm showed a better response and it has been used because of simplicity, High effectiveness, more resistance against noise and its efficiency that matches the considering improvement of speed and accuracy of this project. With presented methods and algorithms, we will be able to identify patterns in accordance with geometric shapes such as squares, circle, rectangle and triangle in a color image. The general format of the algorithm is as follow: 1. The input image is received. 2. preprocessing including image improvement, geometric correcting and noise elimination. 3- color image conversion into a binary image. 4- edge detection of the image. 5- using Hough algorithm to identify straight lines of detected images. 6- corner detection. 7- using geometric shape features (corner, straight lines, area) to identify squares, circle, rectangular and triangles. Square: if we had 4 corners and found 4 equal lines with right angles between them, the result is a square. If the lines were not equal, its a rectangular. Triangle: if we found 3 corners with 3 lines cutting each other, the shape is a triangle.

Then for confirming the correction of shape type detection and more accuracy, the area feature of shapes is used as a confirmation condition in algorithm. Circle: for detecting a circle in color image there are two forms of methods, one is using radiuses with specified value and the other is using Radius with non-specified value that can be anything. In fact, in first form, the circles with specified radius are detected while in the second, any circle with any desired radius is detected. But in second form, the calculation load is more and lasts longer. Also in the article[18] a way of rectangular enclosing around shapes is used to identify the types. In this article we worked on circle recognition using geometric features that eliminated computation complexity and has an effective role in improving speed and accuracy. In this method the area will be calculated then enclose a rectangular around the shape and the rectangular width is divided by 2 and this is assumed as the circle radius. We will investigate following conditions: 1-value (1- rectangular length/ rectangular width) should be less than the threshold value. 2- value (1- shape area) should be less than the threshold value. If above conditions were met, the shape is identified as a circle. The threshold values are obtained practically in this research.
Then results of image analyzes including types and quantity of shapes in the image in the form of a linguistic sentence is sent to the robot through a wireless connection and the robot describes the shown image. The robot performance was investigated by showing various test images and the results recorded in the Table II shows the 95.7 percent of success in robot performance. Also it was compared with default shape detection of robot and the results showing 14 percent of performance improvement are recorded in Table II.

Figure 8: 1. Roberts 2. Sobel 3. Prewitt 4. Canny

Figure 9: The position of the robot.
CONCLUSION

In this paper, software and hardware implementation of the algorithm identifying the type and number of each geometric shapes in the images and its functional improvement and development have been discussed. Identifying the geometric shapes in image processing is very useful due to the use in any field that requires the classification of different objects based on physical appearance, such as robotic artificial intelligence, computer vision, diagnosis of disease, recognition of traffic signs, quality inspection in assembly lines, and so on. Available latitudes in the image are extracted and edited using edge-to-edge algorithms. As shown in the results, Canny algorithm was selected, which shows better results than other algorithms. By combining methods of edge detection, corner detection, straight lines and using shape geometric features, an optimal method for shape, type and quantity detection was achieved and was expressed in natural language by humanoid robot and so the system was completed. The sections of this system was investigated step by step. Among them, optimized algorithms with higher speed and accuracy than other algorithms were considered. In the following of article, after several synthesis of hybrid algorithms and reform and optimizing the codes and software thresholds, a hybrid algorithm architecture for the system was presented and examined and the results show the success of robot performance in real time detection of geometric shapes in test and painted images and also better performance than its default mode. This designed system of image data conversion to natural language can be used for helping training of children with autism and communicating with these children. It also can be used for People with disabilities, who can give their words by showing a geometric shape or a certain number of shapes to a robot or a system.

Table 1: COMPARE THE PERFORMANCE OF THE TWO SYSTEMS.

<table>
<thead>
<tr>
<th>Feature/system</th>
<th>The system implemented in this article</th>
<th>Default system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Accuracy</td>
<td>95.7%</td>
<td>82%</td>
</tr>
<tr>
<td>Expression of the number of any shape</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

Table 2: EXTENDED SYSTEM PERFORMANCE.

<table>
<thead>
<tr>
<th>Type of shapes</th>
<th>Number of shapes in images tested</th>
<th>Number of correctly recognized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle</td>
<td>120</td>
<td>116(96%)</td>
</tr>
<tr>
<td>Rectangle</td>
<td>120</td>
<td>115(95%)</td>
</tr>
<tr>
<td>Square</td>
<td>120</td>
<td>113(94%)</td>
</tr>
<tr>
<td>Triangle</td>
<td>120</td>
<td>118(98%)</td>
</tr>
</tbody>
</table>
References


International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)
Universal Scientific Organization, www.aeus.org
PISSN: 2411-6173, EISSN: 2305-0543