Object recognition and detection by shape and color pattern recognition utilizing Artificial Neural Networks

Conference Paper · March 2013

DOI: 10.1109/ICoICT.2013.6574562

Object Recognition and Detection by Shape and Color Pattern Recognition Utilizing Artificial Neural Networks

Jerome Paul N. Cruz*, Ma. Lourdes Dimaala*, Laurene Gaile L. Francisco*, Erica Joanna S. Franco*, Argel A. Bandala**, Elmer P. Dadios**

*Department of Computer Science
Polytechnic University of the Philippines
Manila, Philippines

** Gokongwei College of Engineering
De La Salle University
Manila, Philippines

Abstract **— A robust and accurate object recognition tool is presented in this paper. The paper introduced the use of Artificial Neural Networks in evaluating a frame shot of the target image. The system utilizes three major steps in object recognition, namely image processing, ANN processing and interpretation. In image processing stage a frame shot or an image go through a process of extracting numerical values of object's shape and object's color. These values are then fed to the Artificial Neural Network stage, wherein the recognition of the object is done. Since the output of the ANN stage is in numerical form the third process is indispensable for human understanding. This stage simply converts a given value to its equivalent linguistic term. All three components are integrated in an interface for ease of use. Upon the conclusion of the system's development, experimentation and testing procedures are initiated. The study proved that the optimum lighting condition opted for the system is at 674 lumens with an accuracy of 99.99996072%. Another finding that the paper presented is that the optimum distance for recognition is at 40cm with an accuracy of 99.99996072%. Lastly the system contains a very high tolerance in the variations in the objects position or orientation, with the optimum accuracy at upward position with 99.99940181% accuracy rate.**

Keywords- **object recognition; artificial neural networks; vision system; (ANN, pattern recognition, object detection)**

I. INTRODUCTION

Humans recognize a multitude of objects and images with little effort, despite the fact that the image of the objects may vary in different viewpoints, sizes or even when they are rotated. For instance, when a human sees a hand regardless whether it's open or not, he still recognizes it as a hand. Object recognition has been man's daily lifestyle. The more often a person sees an object the more he gets familiar with it [1].

In object recognition, it was known that the human brain processes visual information in semantic space mainly, that is, extracting the semantically meaningful features such as line-segments, boundaries, shape and so on [1]. As for computers, it is a challenge than what humans regard as the simple process of recognition [2].

Numerous attempts are done to realize object recognition such as in [3] that used the principle of visual vocabulary. Another approach in [4] is introduced, wherein fuzzy logic is used to detect an object based on its color. Finally in [5] a combination of artificial neural networks and fuzzy logic is utilized to evaluate an object based on a visual and audio information.

In this study, the researchers developed a system which explored the possibility of creating a vision system that recognizes an object regardless of its viewpoint and illumination through artificial neural networks. Furthermore the study can serve as the stepping stone to more sophisticated object detection systems enabling them to consider Artificial Neural Networks.

II. ARCHITECTURE OF THE SYSTEM

Figure 1 shows the block diagram of the designed system. As can be seen there are three main blocks used in the system. The image processing block deals with extracting numerical characteristics of an object from a frame of the input video information. The neural network block is in charge of dispensation of the numeric characteristics from the image processing block. Lastly, the interpretation of data block translates the numeric output of the neural network block to its equivalent linguistic meaning.

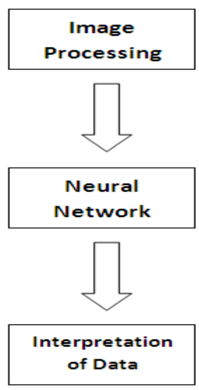

Figure 1. The General Block Diagram of the System

A. *IMAGE PROCESSING BLOCK*

Figure 2 shows a sample grabbed frame from an input video stream. As shown, the image is placed under a uniformly white background. The image is then processed to extract numerical values. The first step is to transform the original image into its equivalent grayscale image then to a binary image shown in figure 3.

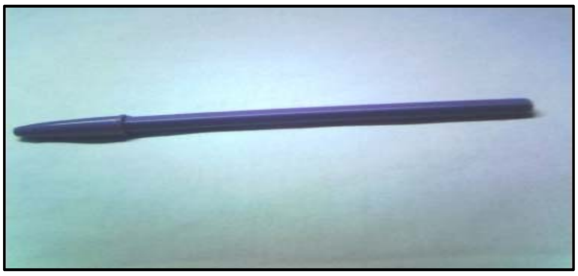

Figure 2. Image Snapshot from Input Stream

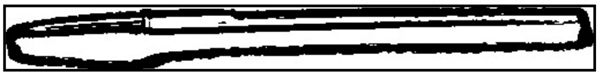

Figure 3. Binary Equivalent of the Input Image

As showed in figure 3 the image is cropped to fit the size of the object from top to bottom and left to right. Since the position of the object is arbitrary which depends on the placement of the user, the system must determine where the object is. This is commonly known as object detection. This was realized when dynamic cropping process was implemented.

$$\boldsymbol{S_i} = \sum_{i=n=1}^{\substack{i=480\\n=640}} \boldsymbol{P}_{[i,n]} \begin{cases} \boldsymbol{r_t} = \min\limits_{0<S_i<\infty} \boldsymbol{i} \\ \boldsymbol{r_b} = \max\limits_{0<S_i<\infty} \boldsymbol{i} \end{cases} \qquad (\boldsymbol{1a})$$

$$\boldsymbol{S}_{\boldsymbol{i}+\boldsymbol{480}} = \sum_{i=n=1}^{\substack{i=480\\n=640}} \boldsymbol{P}_{[n,i]} \begin{cases} \boldsymbol{r_l} = \min\limits_{0<S_i<\infty} \boldsymbol{i} \\ \boldsymbol{r_r} = \max\limits_{0<S_i<\infty} \boldsymbol{i} \end{cases} \qquad (\boldsymbol{1b})$$

The boundaries of cropping to obtain figure 3 is calculated by equations 1a and 1b. These equations are applied to the binary image drawn from the raw image. $\boldsymbol{P}_{[i,n]}$ is the original image matrix. $\boldsymbol{S_i}$ is the sum of the pixel values of each rows. $\boldsymbol{S}_{\boldsymbol{i}+\boldsymbol{480}}$ is the sum of the pixel values of each columns. $\boldsymbol{r_t}$ is the top boundary of the image to be cropped and the first vertical pixel of the object. $\boldsymbol{r_b}$ is the bottom boundary of the image to be cropped and the last vertical pixel. $\boldsymbol{r_l}$ is the left most boundary of the image to be cropped and the first horizontal pixel. $\boldsymbol{r_r}$ is the right most boundary of the image to be cropped and the last horizontal pixel.

Generally, equations 1a and 1b calculates the sum of each rows of the image pixel matrix and the sum of each column of the image pixel matrix. Thus, a row or a column not included in the image would generate a sum of zero since the background is white.

Figure 4. Binary Equivalent of the Input Image

A similar image in figure 3 with color characteristics can be produced using the same boundaries calculated from the equations 1a and 1b. Figure 4 illustrates the said image.

$$C_a = \frac{\sum_{j=1}^{r_b-r_t}\sum_{\boldsymbol{i}=\boldsymbol{1}}^{r_l-r_r} \boldsymbol{I}_{[\boldsymbol{i},\boldsymbol{j}]}}{(\boldsymbol{r_b}-\boldsymbol{r_t})(\boldsymbol{r_l}-\boldsymbol{r_r})} \qquad (\boldsymbol{2})$$

$$C_v = \frac{\sum_{j=1}^{r_b-r_t}\sum_{\boldsymbol{i}=\boldsymbol{1}}^{r_l-r_r} (\boldsymbol{I}_{[\boldsymbol{i},\boldsymbol{j}]} - C_a)^{\boldsymbol{2}}}{(\boldsymbol{r_b}-\boldsymbol{r_t})(\boldsymbol{r_l}-\boldsymbol{r_r})} \qquad (\boldsymbol{3})$$

The first output of this block is calculated by equation 2. The distribution of colors within the cropped image is characterized. $\boldsymbol{r_b}, \boldsymbol{r_t}, \boldsymbol{r_l}, \boldsymbol{r_r}$ are the same values derived from equation 1a and 1b. $\boldsymbol{I}_{[\boldsymbol{i},\boldsymbol{j}]}$ is the cropped image matrix from the boundaries generated from equation 1a and 1b.

The second output of this block is given by equation 3. The identity of each pixel against the overall distribution of color in the image is identified.

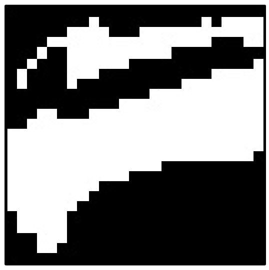

Figure 5. Resized Binary Image

Furthermore, figure 3 is resized into a 25x25 image matrix resulting into an image shown in figure 5. Each row and each column of the said image is summed resulting into 50 distinguished numbers. These will be the set of output of the image processing block that will characterize the shape of the target image and characterized in equations 4a and 4b.

$$\boldsymbol{R_i} = \sum_{i=n=1}^{\substack{i=25\\n=25}} \boldsymbol{B}_{[i,n]} \qquad (\boldsymbol{4a})$$

$$\boldsymbol{C_n} = \sum_{i=n=1}^{\substack{i=25\\n=25}} \boldsymbol{B}_{[n,i]} \qquad (\boldsymbol{4b})$$

B. Artificial Neural Networks Block

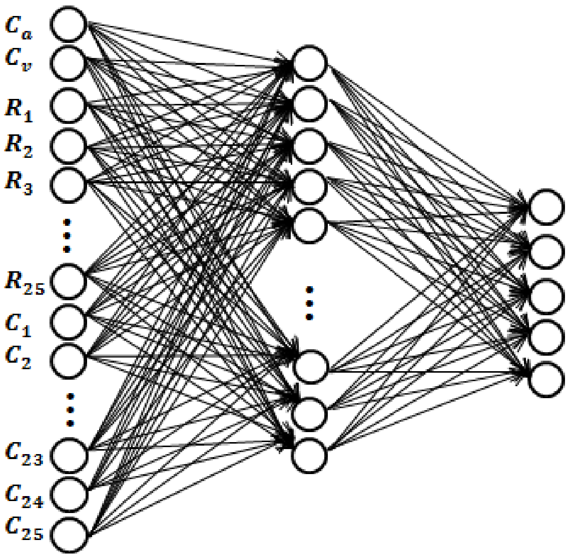

Figure 6. Architecture of the Artificial Neural Network

The artificial neural network used in this study utilizes the feed forward architecture. As can be seen in figure 6 the flow of information is from the neurons in the left towards the neurons on the right. There are 52 input neurons 20 hidden neurons and five output neurons.

The 52 input neurons are composed of two neurons for color characteristics of the image to be detected namely C_a and C_v. Furthermore the remaining 50 neurons is subdivided into two 25 of which is for the horizontal shape characteristics R_1 to R_{25} and vertical shape characteristics C_1 to C_{25}.

There are five output neurons which are defined to generate binary numbers. Each combination of binary numbers will represent a certain object. The assignment of binary output in each input is associated during the training stage.

Each neuron has a construction characterized by figure 7. The neuron is composed of a weight, which is a numerical element multiplied to the input numerical value, a bias which is added to the product of the input and the weight. The sum will be evaluated in an activation function that will generate an output numerical value.

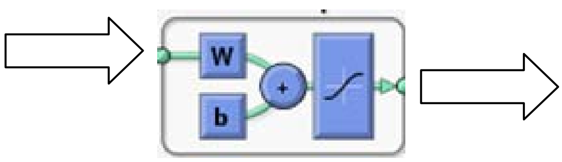

Figure 7. Neuron Architecture

$$\boldsymbol{b} = \frac{\sum_i \boldsymbol{x_i}^2 \sum_i \boldsymbol{d_i} - \sum_i \boldsymbol{x_i} \sum_i \boldsymbol{x_i d_i}}{\boldsymbol{N}(\sum_i (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2)} \quad (5)$$

$$\boldsymbol{w} = \frac{\sum_i (\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{d_i} - \overline{\boldsymbol{d}})}{\sum_i (\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2} \quad (6)$$

Where:
b is the bias of a neuron
w is the weight of a neuron
x_i, input vector of i^{th} example
d_i is the desired (target) response of i^{th} example
N is the sample size

The value of the weight and bias varies in every neuron. The process of determining the value of weight and bias is called learning or training. The algorithm used for learning is called back propagation algorithm. In this learning method, a desired output or target is given with a corresponding set of inputs. In the architecture of the artificial neural networks, back propagation algorithm requires 52 input elements and five output or target elements per set. The calculation of bias is given in equation 5 while the weight is defined in equation 6.

C. Interpretation Block

The output of the neural network is in numeric form assigned in the training stage. Table 1 shows a tabulation of eight objects and its corresponding output value from the artificial neural network.

In the event that a bit is less than 1 or greater than zero the interpretation block rounds up or down the output value to attain a value in table 1 closest to the output of the artificial neural network.

Table1. Association of Output Values to its Equivalent Object

Binary Output	Equivalent Object
00001	Writing instrument
00010	Coin
00011	Headband
00100	Key
00101	Razor
00110	Tweezers
00111	Flash drive
01000	Whistle

III. Graphical User Interface

The researchers developed a graphical user interface which is capable to accommodate all of the processing and calculation for object recognition and also caters ease of use in the user's viewpoint. The graphical user interface is designed and implemented in C# language. Calculations and some specific processes are done in Matlab. The study integrated these two languages to produce a tool which accommodates all of these objectives.

Figure 8 presents the design of the interface. As can be seen there are three options available. First is the "connect button", this option enables the user to operate or turn on the display window by evaluating which camera is available. The user can also choose what camera to use by altering the options in the "camera devices" option. The second option is the "disconnect" button, this is used for terminating an existing camera connection. Lastly, the analyze option is used to trigger the recognition processes in the view window.

Figure 8. The Graphical User Interfaced of the System

IV. Experiment Results

The system's performance is evaluated with several experimental setups that will test certain parameters which are considered to be vital in producing accurate recognition. There are three parameters tested. The first parameter tested is the intensity of illumination in the object to be detected. Another parameter is the distance of the object to the camera. Lastly is the orientation or rotation of the object. The experiments for every parameters are composed of a setup drawing 200 experiment data from 200 trials.

Figure 9 shows the performance of the system when varying the light intensity in the object. As can be seen three distances are used while varying the light intensity on a given object. For a lesser value of light intensity at 318 lumens the highest accuracy rate is found at 30cm distance which yield an accuracy of 99.90847% and the lowest accuracy rate is at distance 50cm. For light intensities 451 lumens, 674 lumens and 906 lumens an average of 99.99% accuracy emerged for all distances. Lastly, for light intensity of 1098 lumens, the highest accuracy rate was given by the distance of 40 cm with an accuracy rate of 99.96284%. On the other hand in this light intensity the lowest accuracy rate was recorded at 50cm distance with an accuracy rate of 97.93216%. It can be seen that the lesser the light the accuracy of the recognition is lesser because ample light intensity is not achieved. However, too much light may cause glair. Thus, the object may reflect too much light resulting to loss of color characteristics.

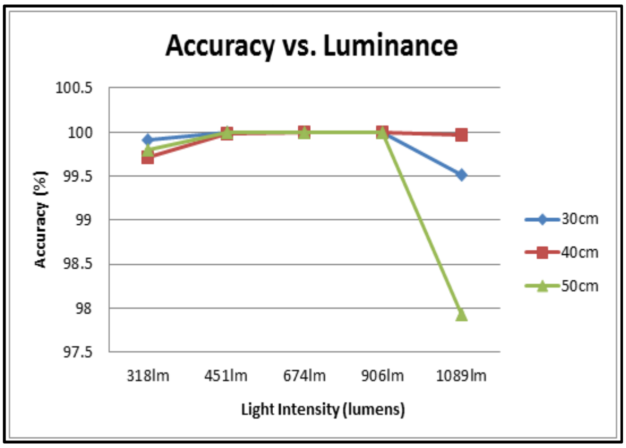

Figure 9. Experiment Results Accuracy against Light Intensity

Figure 10 exhibits the experiment result in varying the distance of the target image to the camera. There are three distances used in this experiment: 30cm, 40cm, and 50cm. As can be seen, all of these distances yield an average accuracy of 99.99% however the least accurate distance is 30cm with an accuracy of 99.9997%. When the distance is increased at 40cm, the accuracy increased to 99.99989%. Further, increase in distance at 50 cm decreased the accuracy at 99.99995%. These results can foretell that having a closer camera would decrease the accuracy of recognition rate but allocating too much distance would also result in decrease in accuracy of recognition rate.

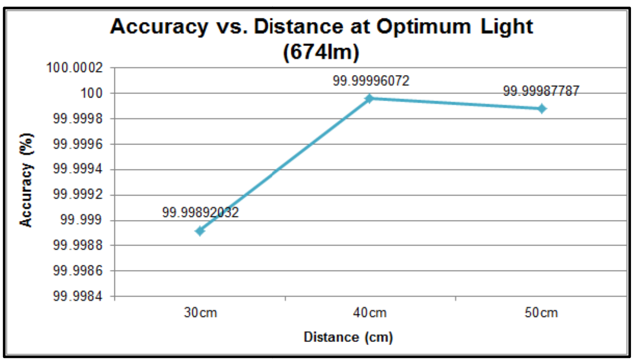

Figure 10. Experiment Results Accuracy against Distance of Object

Finally figure 11 shows the experiment results of varying the orientation of the object against the accuracy of recognition rate. As can be seen in all angles tested, the average recognition rate is above 99.95% with the highest accuracy rate at zero degrees or upright position of the object being evaluated. Also, the orientation of the object has no significant effect on the recognition process. It is because the orientation characteristics of the image are included in the learning or training stage in the study.

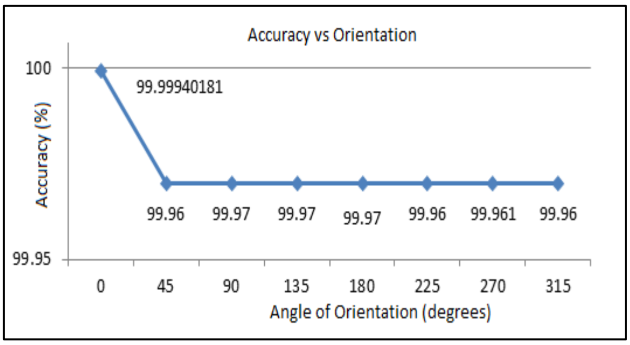

Figure 11. Experiment Results Accuracy against Orientation of the Object

V. CONCLUSIONS

Upon accomplishing experimentations, the paper presents the following generalizations. The system's performance varies with the lighting condition with a recommended lighting range of 318 lumens with 99.71139% accuracy to 1089 lumens with 97.93216% accuracy. The optimum lighting condition is at 674 lumens with an accuracy of 99.99996072%. The distance of the camera has a substantial contribution to the recognition of an image which ranges from a minimum of 30cm with a registered accuracy of 99.99892032% and a maximum of 50 cm with a registered accuracy of 99.99987787%. The optimum distance for recognition is found to be at 40cm with an accuracy of 99.99996072%. Lastly the system contains a very high tolerance in the variations in the objects position or orientation, with the optimum accuracy at upward position with 99.99940181% accuracy rate.

REFERENCES

[1] Bo Zhang, "Computer Vision vs. Human Vision," in *9th IEEE International Conference on Cognitive Informatics (ICCI)*, 2010, p. 3.

[2] Nicolas Pinto, David D. Cox, and James J. DiCarlo, "Why is Real-World Visual Object Recognition Hard?," *PLoS Computatioinal Biology*, pp. 151-156, January 2008.

[3] Su Yu-Chi et al., "A 52 mW Full HD 160-Degree Object Viewpoint Recognition SoC With Visual Vocabulary Processor for Wearable Vision Applications," *IEEE Journal of Solid-State Circuits*, vol. XLVII, no. 4, pp. 797 - 809 , 2012.

[4] Napoleon H Reyes and Elmer P Dadios, "A Fuzzy Approach in Color Object Detection," in *IEEE International Conference on Industrial Technology IEEE ICIT* , Bangkok, 2002, pp. 232 - 237.

[5] Shahnaz Shahbazova, Manfred Grauer, and Musa Suleymanov, "The Development of an Algorithmic Model for Object Recognition From Visual and Sound Information – Based on Neuro-Fuzzy Logic," in *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, El Paso, Texas, 2011, pp. 1 - 6.