Intelligent Adaptive Object Recognition

Safa Almazmome[1] Li Zhang[2]
Newcastle University, School of Computing Science, Newcastle Upon Tyne.[1]
Northumbria University, Department of Computer Science and Digital Technologies, Faculty of Engineering and Environment, Newcastle Upon Tyne. [2]

Introduction

Computer vision and machine learning techniques have played an important role in human robot/computer interaction. Object recognition is one of the most important and challenging computer vision tasks. It involves methods of real-time object tracking, feature extraction and recognition. In order to produce an object recognition system with reasonable performance, important features of objects, such as textures, colors, shapes and sizes, are required to be extracted. This project involves an object recognition system that recognizes 7 objects using an artificial intelligence approach, namely a neural network. We used a two layer network to work as a classifier to classify objects in images as being in one of several known classes. The reason for choosing a neural network as the main method for the system was that a neural network has the ability to extract patterns that are too complex even for humans to notice. A neural network is also capable of categorizing the data that have been trained.

Method:

A. System Implementation

For the development of the object recognition system, a static scene with a fixed white background and an object in the foreground was used. This made it easier to identify and isolate the foreground object without facing a collision problem. Different objects were used as subjects to train the neural network. For this project we used image pixel features extraction and color features extraction methods in order to extract information on the object's shape and color features.

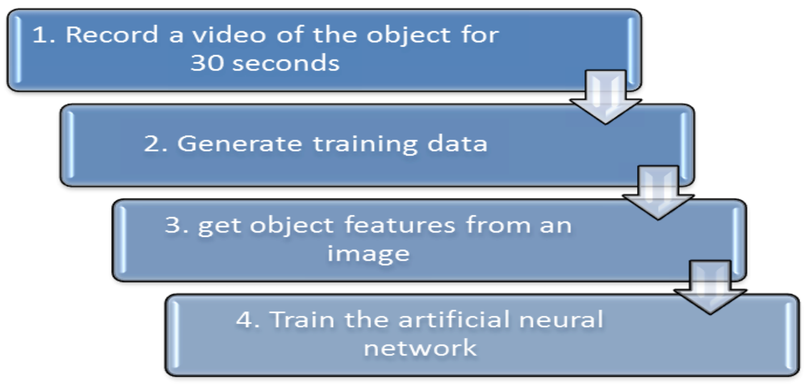

Figure 1: The process of implementing the system

B. Image processing

The following image processing procedures are applied to the objects collected from a real-time video stream in our system:

1. Converting the recorded videos from YUY2 to RGB and HSV format to extract the object shape and color easily.
2. Converting the images to grey-scale representation, which includes only the brightness information of every pixel in the image (for example, see Figure 2).
3. Converting grey-scale images to black and white (see Figure 3). Grey-scale images still contain too much shape information. Therefore, brightness information is removed as well. This is done by converting the image into a black and white (binary) array of pixels. Each pixel in this representation carries either a 1 or a 0 to indicate whether the pixel belongs to the foreground of the scene or the background, respectively.
4. Filling holes using a hole-filling algorithm, because the existence of holes within the body of foreground pixels complicates the task of object shape identification.
5. Image rescaling. Large images require much larger data sets and a longer time to train the network.

Figure 2: Converting to grey-scale

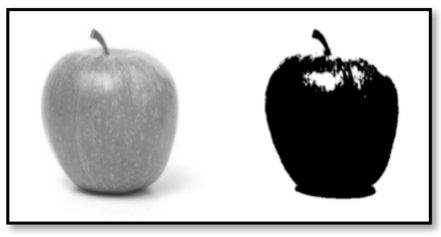

Figure 3: Convert to black and white

C. Feature extraction

• Color features:

Color features are simply a measure of the relative frequency of pixels that carry certain colors in the image. For example, if the colors red, green and blue were given the numerical values 0, 1 and 2, we can arrange the relative frequency count in the matrix format [Cr Cg Cb]; where Cr, Cg and Cb represent the relative counts of red, green and blue pixels respectively. Therefore, every element in the matrix represents the relative count of the color of its index (e.g. Cg has an index of 1 and so it is the count of green pixels). In our development of the system we have used a color (hue) scale ranging from 0 to 1, representing the different colors in the visible spectrum from red to blue. We have divided this spectrum into 11 units (approximately 0.1 units in length) and counted the relative frequency of pixels in each of these color groups.

• Pixel features:

We have chosen to use a smaller copy of the image (5x5 pixels) as a set of pixel features. This array is formed by rescaling the grey-scale copy of the image into a 5x5 size. The total pixel count in this case will be 25 and this is used to create the pixel features of the image.

D. The design of artificial neural networks

In our investigation, we have used a set of 36 features (11 of which are color-based features and the rest are pixel intensity features extracted from object images (see Figure 4)) as inputs to the neural network. In our experiments, different network topologies with a different number of hidden layers and neurons attached to the hidden layers are used to test the performance of the network. Since we aim to distinguish 8 objects with different shapes and colors, the network has 8 nodes, representing the 8 objects in the output layer (see Figure 5).

Figure 4: Test Images

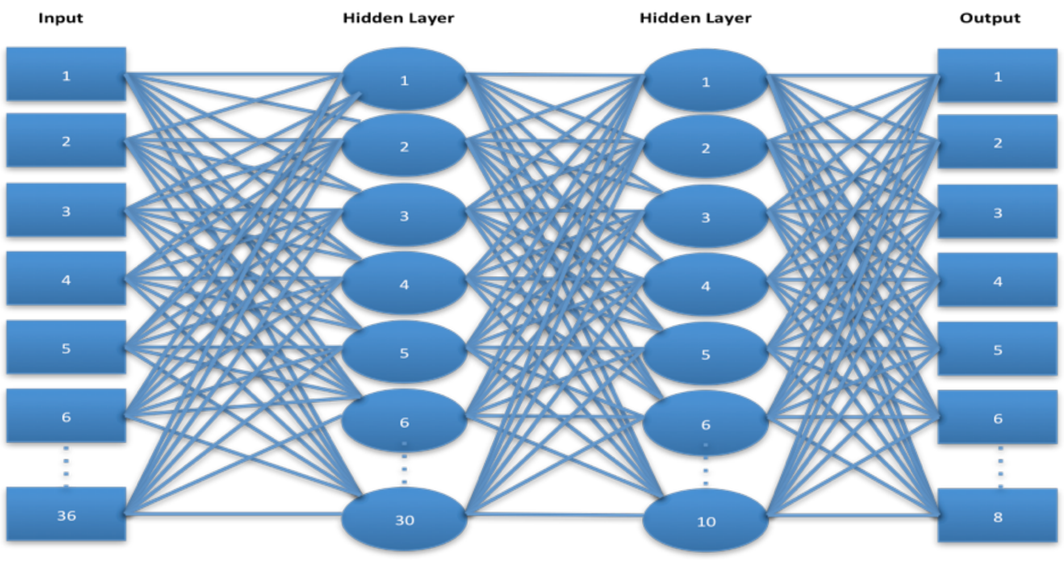

Figure 5: Artificial Neural Network

Output:

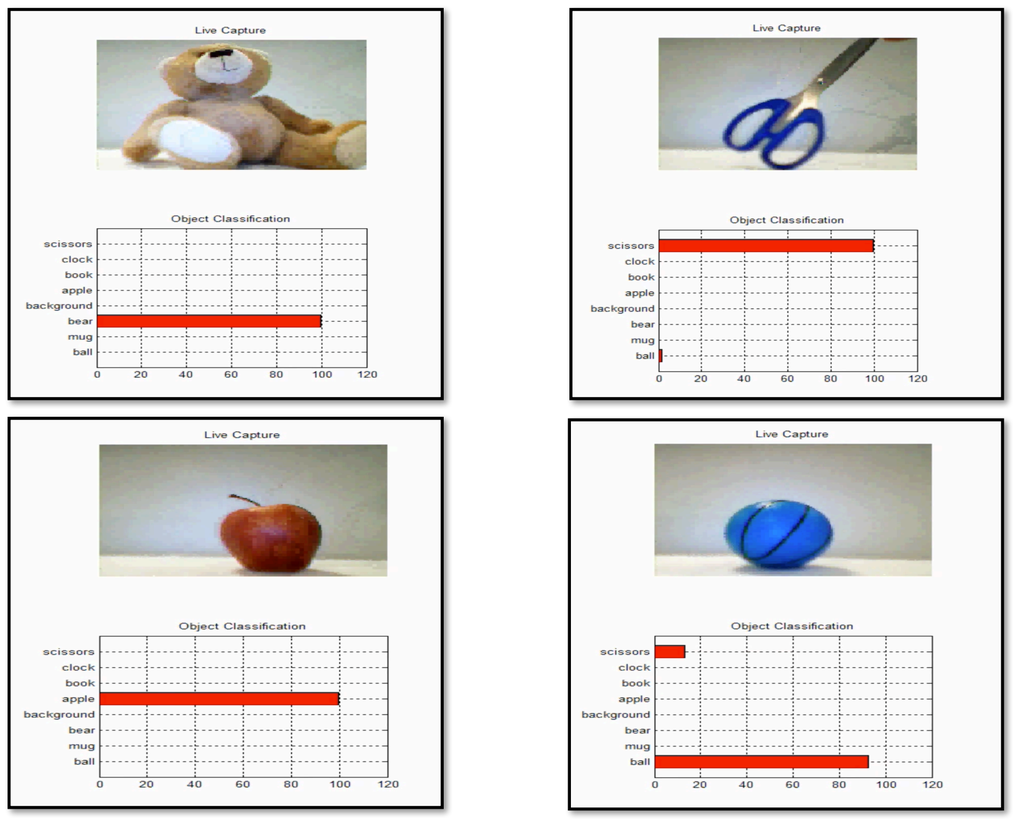

Figure 6: Recognition Output

CONCLUSION AND FUTURE WORK

In this paper, we have investigated the suitability and performance of artificial neural networks as classifiers for object recognition in live video streams. Several areas for further investigation were identified in this project. It would be interesting to investigate how the performance of the system can be improved with additional sets of classifiers. Some of these might be extracted from a video camera, one or multiple additional video cameras or even by other sensors (such as force sensors used to measure the object's weight). Such additional information might require more complex networks for processing and so a thorough investigation of their effectiveness and usefulness must be performed.