**CHAPTER THREE**

**RESEARCH METHODOLOGY**

**3.1 Chapter Introduction**

This chapter presents the methodology followed in the solution to the problem of wood species classification using Support Vector Machine (SVM) and Artificial Neural Network (ANN). In the development process, the implementation phase involved decomposing the whole process into smaller bits and defining the relationship among them. Top down coding approach was used in the comparative use of the two mentioned algorithms for the wood species classification problem solution Matrix Laboratory (MATLAB) programming language. This involved breaking the development process into subs or what is called modules and each subs being further broken down into even smaller units. This process is repeated until each unit is sufficiently small enough to be conveniently coded from scratch as an independent entity that performs a clearly defined task.

The analysis and comparison of existing pattern recognition systems where different algorithms were used was done. This was actually preceded by the review of related implemented image processing algorithms. The critical analysis led to the adoption of the scheme of requirement specifications that highlighted the nature of the pattern recognition system procedures and algorithms implemented in this project work. The process followed in the implementation of the image recognition system is succinctly written in the following sections of this chapter.

**3.2 Procedure Adopted**

The comparative analysis of SVM and ANN for wood print implementation procedure was accomplished using MATLAB functions. The first stage of the implementation was the data acquisition which involved the gathering of wood species (dataset) images for the database. Local database system was employed, the wood species (dataset) images were divided into two; the training database and the testing database in the ratio 60:40. Once the images in the database were acquired, image pre-processing done using discrete cosine transform (DCT) followed by feature extraction using the two algorithms of choice. SVM and ANN were applied separately and comparatively to the processed images for feature extraction. The implementation is in two sub-parts: verification and identification. Wood species images verification deals with the confirmation of authenticity of a wood image in the local database with one to one matching of the (dataset) database, while identification is used to specify the wood identification with one-to-n matching. Matrix Laboratory software provides the best image processing tools, the digital wood images were analyzed easily using MATLAB. Figure 3.1 shows the flowchart of the overall system architecture. The complete algorithm for the implementation is as follows:

Step 1: Acquisition of wood species image.

Step 2: Conversion of the image into grey scale form.

Step 3: Conversion of the grey scale image into binary form

Step 4: Apply thinning process on the binary image.

Step 5: Unsharp Masking of the binarised image for sharpening of the edges.

Step 6: Feature extraction.

Step 7: feature selection using SVM/ANN

Step 8: Match both selected features for verification; if total numbers of points of two wood print images are same, it is prove that both woods are of the same species or matched else both are not same and related to two separate wood species.

**3.2 Preprocessing**

After the wood images dataset are gotten, the next is the processing of the images. The high resolution wood-prints from digital cameras usually contain unevenly distributed gray values; hence the acquired wood images were subjected to pre-processing. The images used (in the local database needed to be converted to grey scale before normalization began. The pre-processing stage involved normalization (coordination and background removal), binarization and masking. Binarization was used in altering the gray scale images to black and white images (binary images).

In MATLAB, a value of one represents that the pixel is white and value of zero represents that the pixel is black (0 = white, 1 = black). This modification of gray scale image to binary image is executed by using ‘threshold processes to the image. When a threshold process is applied to an image, each pixel values are analyzed to the input threshold. Those pixel values which are smaller than the threshold value is place to zero and those pixel value which are greater than the threshold value is place to one. Figure 3.2 shows the flowchart of the preprocessing stage of the system. At the end of this process each pixel values within the image are either zero or one, and the image has been modifying to binary form. After this conversion the key feature points (the differentiating patterns that distinguish wood species) in the wood print are highlighted with black color while valleys are highlighted with white color. Binarization was done in MATLAB using inbuilt function “im2bw”. Example, b=im2bw “Input Image”; (as shown in the appendix).

After binarization of the wood images were done, then thinning was done; image thinning is the process of decrease the thickness of all key feature points lines into single pixel width. Thinning process converts the original x, y location and angle of direction of the feature points of the image, which assure the true calculation of feature points. Thinning was also used in order to destruct the extra pixel of key feature points till they are just one pixel broad .This is done using MATLAB’s inbuilt morphological thinning function named as “bwmorph”. Example, bwmorph(“Binary image”,‟thin‟, Inf); Bwmorph (shown in the appendix) shows morphological operations on binary image. Unsharp masking follows in order to sharpen the edges of the images.

* 1. **Feature Extraction**

Feature extraction phase involved feature point detection; Preprocessing steps are crucial for further feature extraction. The feature extraction is done by points searching and substructure creation. Points searching is where the Feature Points are detected on the basis of the fact that key feature points are disconnected at arbitrary point (key feature Point) and that points are associated with specific individual wood specie pattern Substructure creation deals with the classification of these key feature points which was done in MATLAB by creating matrix. If the central pixel is one, has only one neighbor pixel that is key feature point. Whereas, if the central pixel is one, has two neighbor pixel that is bi-key feature point.

Resize Image

Import Image into MATLAB

Convert Image to gray scale

Perform single-level decomposition

Reshape Image

Compare reshaped image with the trained image

Classify image

Is there a match?

No match

other images?

Close match?

Report Matched specie

Report Matched specie

No Yes

No Yes

Figure 3.1 Flowchart of the overall system architecture

Import Image into matlab

Reshape Image into Matrix Form

Perform Single-based decomposition

Train Database

Figure 3.2 Wood image preprocessing flowchart

* 1. **Application of SVM for Feature Selection**

The feature selection phase was done in two phases. The first phase was the application of SVM on the pre-processed wood images as shown by the procedure below. An iterative algorithm can be designed which scans through the dataset looking for violators (those that does not match), the violator is made a Support Vector. Blocking points are identified and pruned away. The algorithm stops when all points are classified within an error bound i.e.. The outline of our algorithm is presented in the algorithm below. This was done using MATLAB functions and the codes are to be shown in the appendix.

{

Begin

Acquire wood database

Pre-process image

Normalisation (orientation, background removal), binarization, thinning and masking

}

Apply SVM on the wood image

{

Initialize SVM by defining population size of the images in the training dataset,

Randomly generate initial population of wood images.

Initialize iteration counter.

candidateSV = { closest pair from opposite classes } while there are violating points doFind a violator candidateSV = candidateSV S violator if any *αp <* 0 due to addition of *c* to *S* then

candidateSV = candidateSV \ *p* repeat till all such points are pruned

end if

end while

End

}

**3.5 Application of ANN for Feature Selection**

In the second stage of the feature selection phase, the simulated annealing algorithm performs the following steps of actions in order to select the features of the image and compare with the trained images in the dataset:

The algorithm generates a random trial point on the wood print images. The algorithm chooses the distance of the trial point from the current point by a probability distribution with a scale depending on the current temperature. The trial point distance distribution was set as a function with the NeuralFcn functionality in the MATLAB IDE (Integrated Development Environment). A feed-forward neural network algorithm was used. Image coding using a feed forward neural network consists of the following algorithmic steps on the next page. Database training for the ANN segment of the image recognition is conducted for a representative class of images (training database) using the feed forward ANN algorithm. Once the weight matrices have been appropriately selected, any image can be quickly encoded using the Wh matrix, and then decoded (reconstructed) using the Wy matrix.

It is assumed that the hidden layer of the layer network consists of L neurons each with P synapses, and it is characterized by an appropriately selected weight matrix **Wh**. All N blocks of the original image is passed through the hidden layer to obtain the hidden signals, h(n), which represent encoded input image blocks, x(n) If L<P such coding delivers image compression.

It is assumed that the output layer consists of m=p=rxc neurons, each with L synapses. Let Wy be an appropriately selected output weight matrix. All N hidden vector h(n), representing an encoded image H, are passed through the output layer to obtain the output signal, y(n). The output signals are reassembled into p=rxc image blocks to obtain a reconstructed image, Fr. There are two error matrices that are used to compare the various image recognition techniques. They are Mean Square Error (MSE) and the Peak Signal-toNoise Ratio (PSNR). The MSE is the cumulative squared error between the compressed and the original image whereas PSNR is the measure of the peak error.

Thefeature selection with the application of ANN on the same set of processed wood images. This was done using MATLAB functions and the codes are shown in the appendix. The ANN procedure is as shown below:

}

Begin

Acquire wood print database

Pre-process image

Normalisation – scaling, resizing, orientation, background removal

}

Applying feed forward ANN on the training database of wood images

{

Choose a specie image, F, divided it into rxc blocks of pixels. Each block is then scanned to form a input vector x (n) of size

p= r x c

Assess the match of image coding by the Peak signal-to-noise ratio (PSNR) defined as:

PSNR = 20 log 10 [255/sqrt(MSE)]

End

}

The feed forward ANN algorithm is represented by the flowchart diagram on the next page:

Choose an image

Normalize the chosen image

E’<=E

E’>E

Divide image it into rxc blocks

Check for PSNR

Check for MSE

Scan blocks to form a input vector x (n)

Assess level of image match

Figure 3.1 Feed forward ANN Flowchart

**3.6 Feature Matching**

Image acquisition (the new image that is to be compared) is the first step in feature matching (with the images in the testing database), according to the procedure below:

Image.jpg = Input Image acquisition from dataset (database).

Timage.jpg = Template Image retrieve from database

The point computation step is a very important part of wood species image matching. After the detection of feature points, matching algorithm require to calculate total number of available points in the image separately. To perform this computation two counter variables are used to count both key feature points (that carry specific pattern that actually differentiates wood species from one another) on the wood image. All the points computed are organized and used in plotting the comparative ANN and SVM graphs by MATLAB functions. More of these will be discussed in the next chapter when the desktop app for the comparison of these two algorithms is fully figured out and developed.

Location Detection of feature Point was done because each feature point in the wood image being considered at a time has a specific location. This location information of particular point is significant to store for further matching of wood images. The location of every point in the digital image is given by pixel position, so that it can be taken and stored separately for key feature points. All the required information about points is computed and stored. Now, this is the matching step, here the system compares the computed values with the stored values. This developed system first, compares the combination of both amounts of key feature points with stored data. If the match occurs, the algorithm then compares the location of the key feature points with stored location data. And finally, if all the location matches, the message will be displayed on the screen as “Wood Specie Match Found” instead the message will be displayed as “Wood Specie Match Not Found”. To recap, wood image recognition system is used to analyze two wood images; one is original image and another one is template image stored in the local testing database.