User Guide for
Preprocessing Engine of the Optical Character Recognition System for Sinhala Scripts version 0.1

30th April 2005

PAN Localization Project
Language Technology Research Laboratory
University of Colombo School of Computing
Sri Lanka

Table of Contents

1   Introduction

This document explains how to use OCR Preprocessor v0.1. It is an application that consists of basic image processing techniques such as image segmentation and noise removal. The current version of OCR Preprocessor works only for the images in bitmap format.

2   Running the Application

After the successful installation of OCR preprocessor v0.1 (Refer installation guide for OCR Preprocessor v0.1 to install the application) and all dependencies, follow the steps below to start the application.

I.   Run the application from Windows start menu at location **Start -> All Programs -> OCR Preprocessor -> OCR Preprocessor** illustrated in Figure 2.1. If the link for "OCR preprocessor v0.1" is missing in the start menu please reinstall the application.
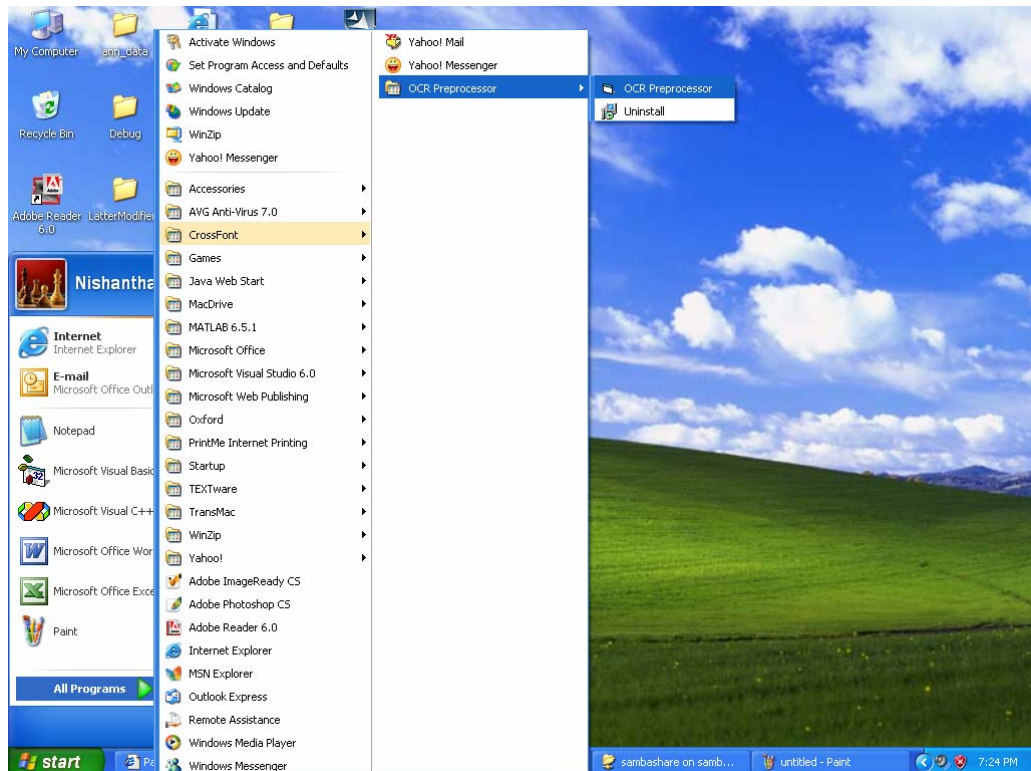

Figure 2.1: **Running the "OCR Preprocessor v0.1" from Windows start menu**

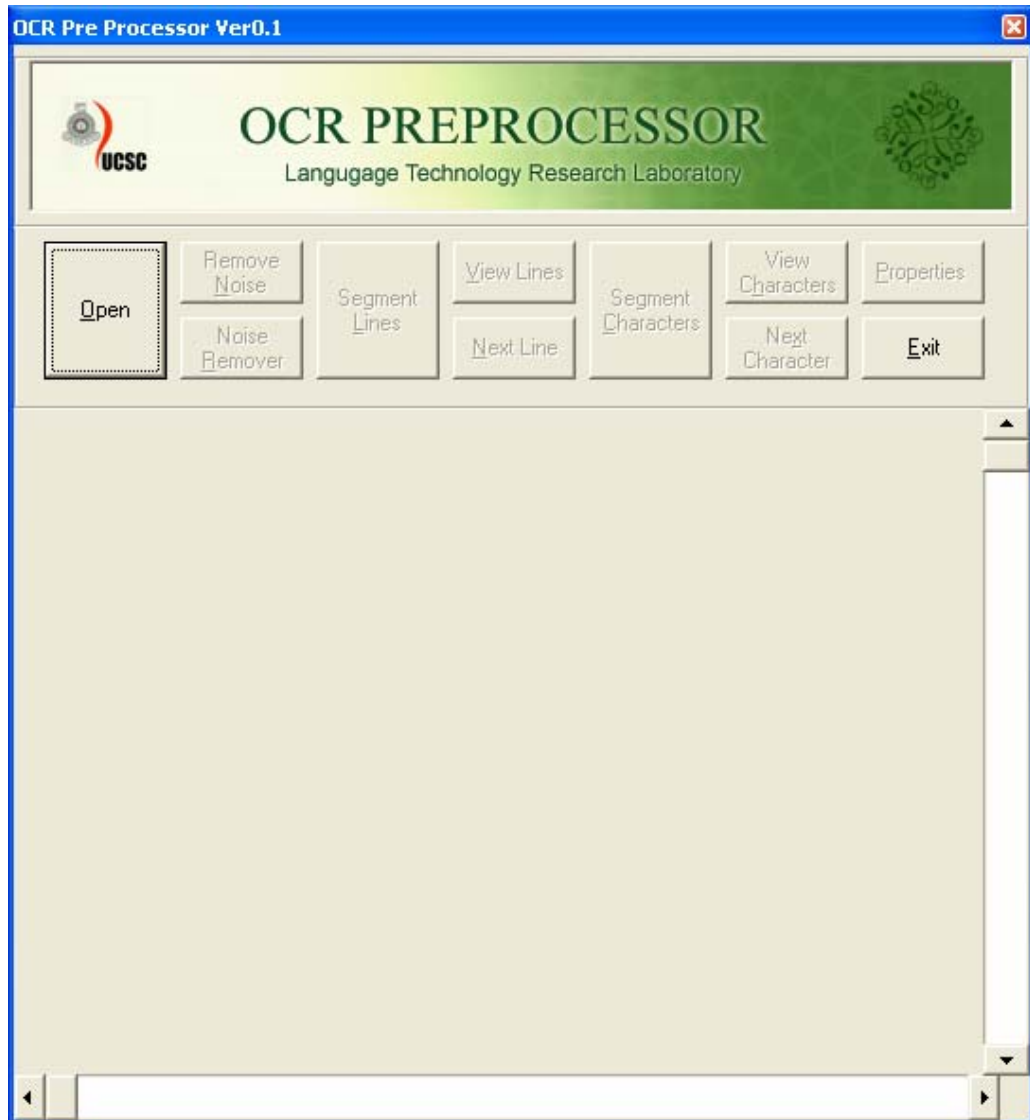II.   As the application starts, it shows main window as shown in Figure 2.2.

Figure 2.2: OCR Preprocessor v0.1, at start up

3   Using Application

3.1 Open the source image

In order to open the image follows the steps given below.

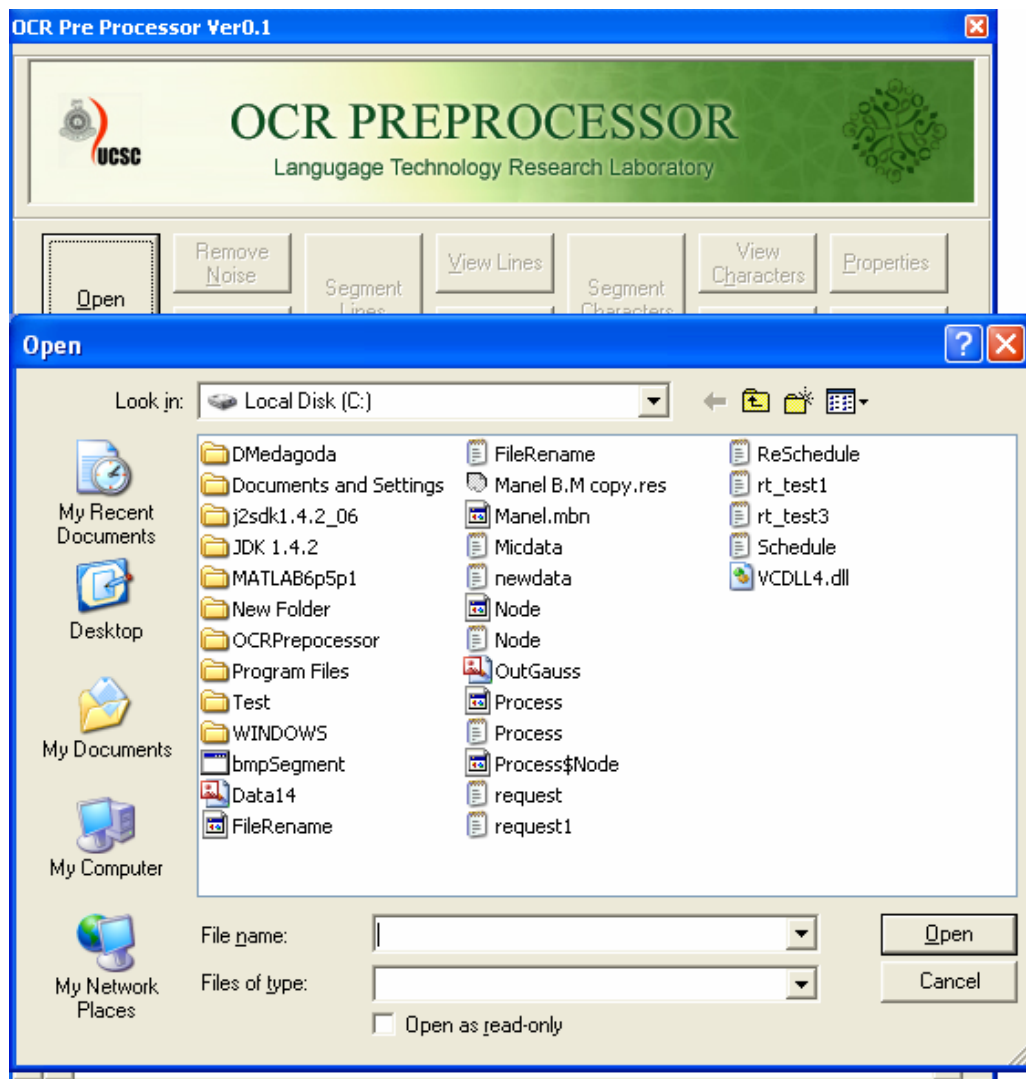I.   Click the "Open" button. Then the open window will appear as shown in Figure 3.1.1

Figure 3.1.1: Preprocessor v0.1, open the image

II.   Browse the image file to be opened.
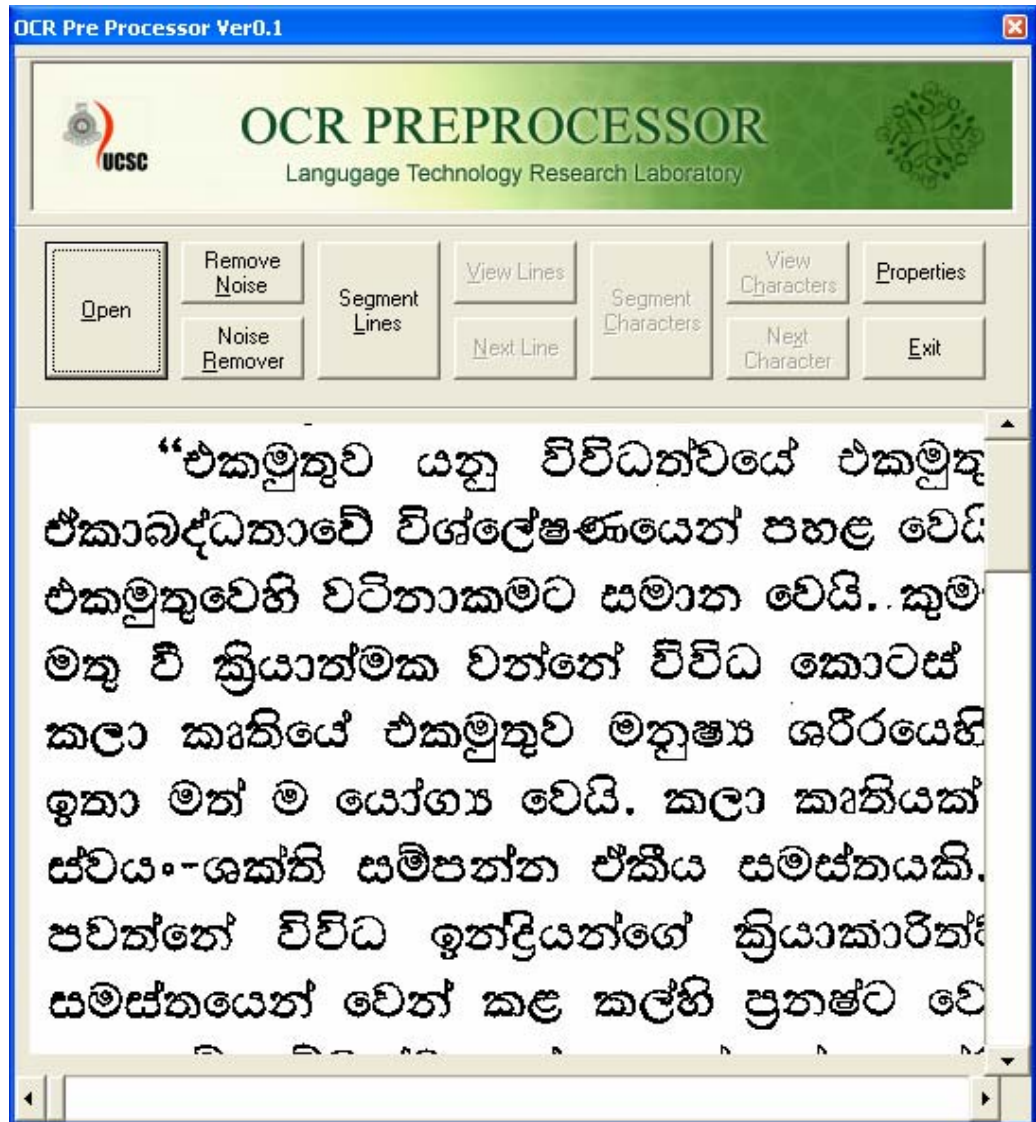
III. Then image file will appear as shown in Figure 3.1.2

Figure 3.1.2: Image file

3.2 Removing Noise in the image

In order to remove noise in the image follows the steps given below.

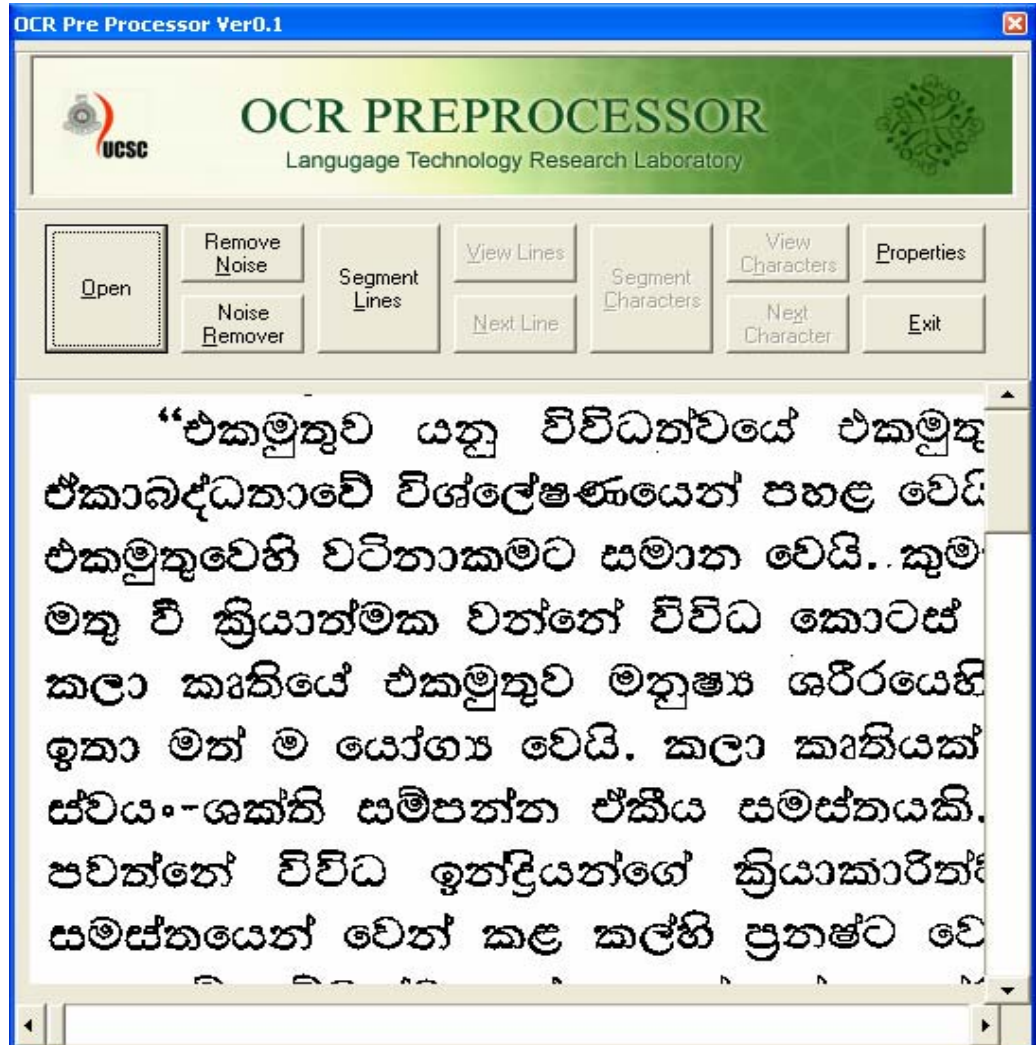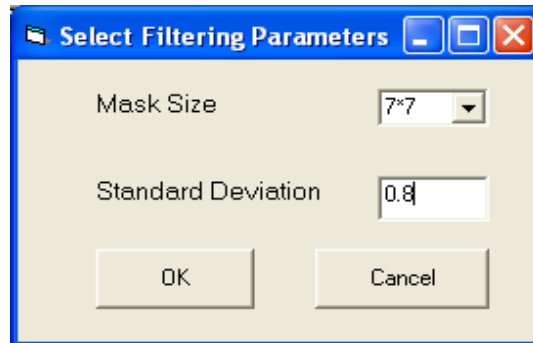I. Click the "Remove Noise " button, then filtered image appears as shown in the Figure 3.1.3

Figure 3.1.3: Filtered Image

II.   Click "Noise Remover" to select the parameters for customize filtering. Enter the values of the parameters as shown in Figure 3.1.4.

Figure 3.1.4: Parameters for filtering

III.   Click "OK" to remove the noise.

3.3 Segment the image into lines

To segment the image into lines follow the steps given below.

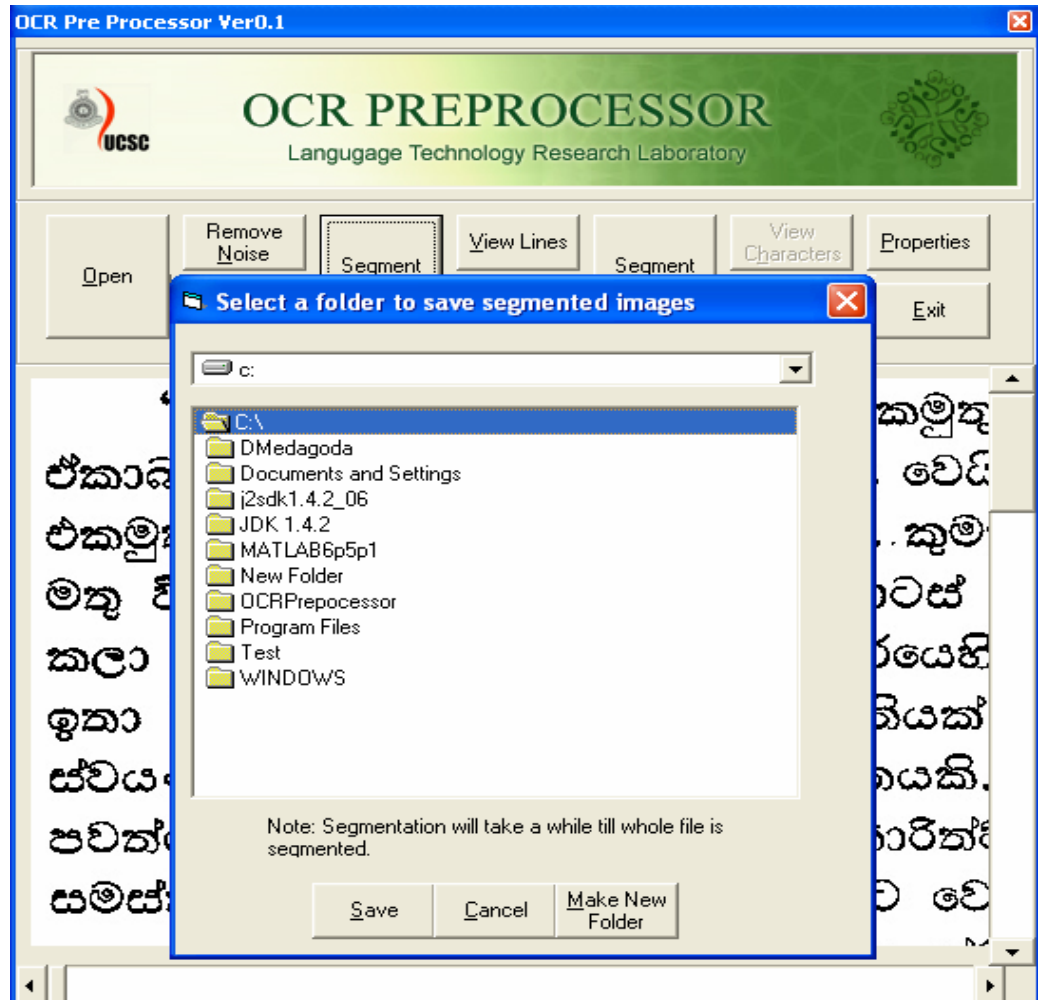   I. Click "Segment Lines" then a dialog box appears as the Figure 3.3.1

Figure 3.3.1: Select a folder to save segmented images

II. Select a folder to save the segmented image or you can create a new folder. Click "OK" to segment the image into lines. Then it gives a massage as in the Figure 3.3.2
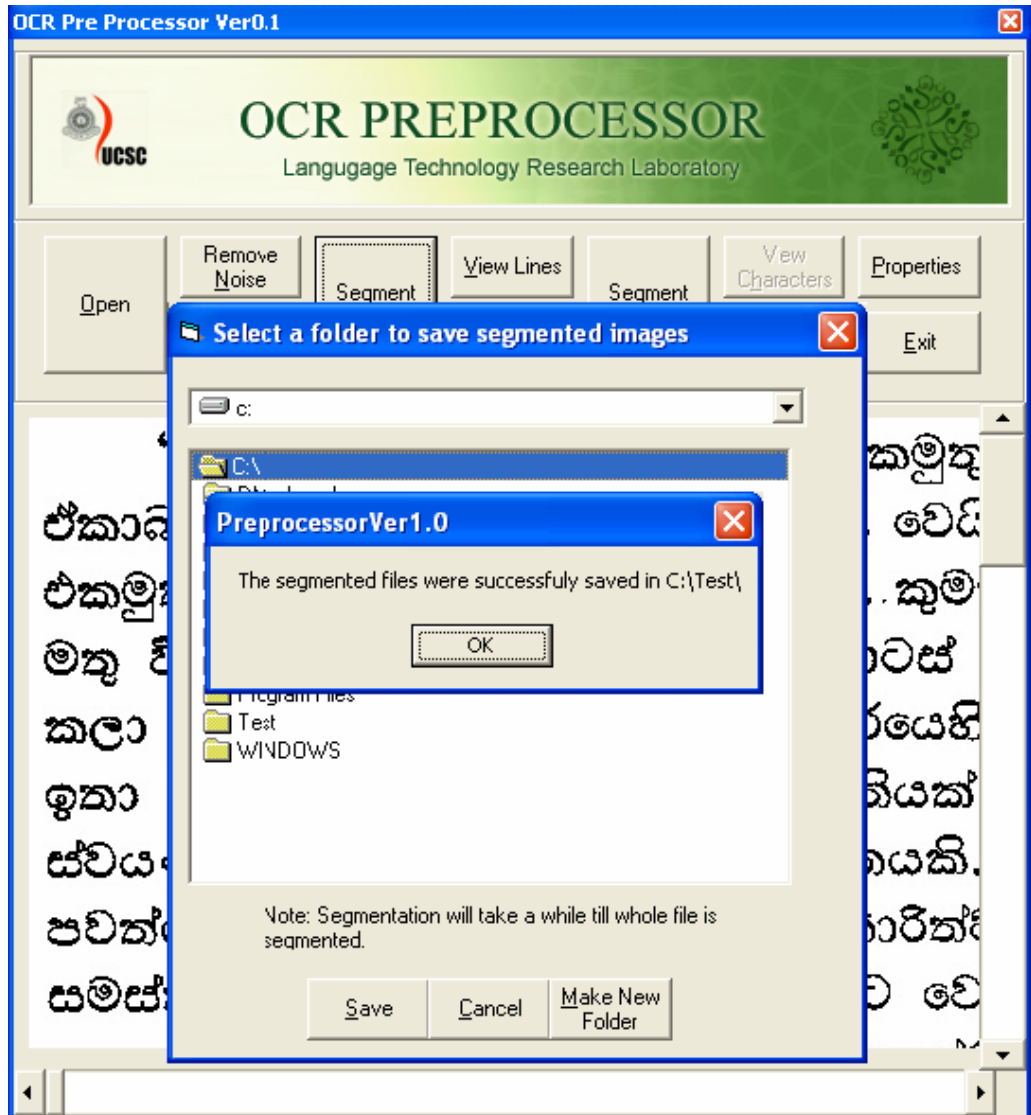

Figure 3.3.2: Image is segmented into lines

3.4 View Lines

Click "View Lines" button then the line number 0 appears in Figure 3.4.1
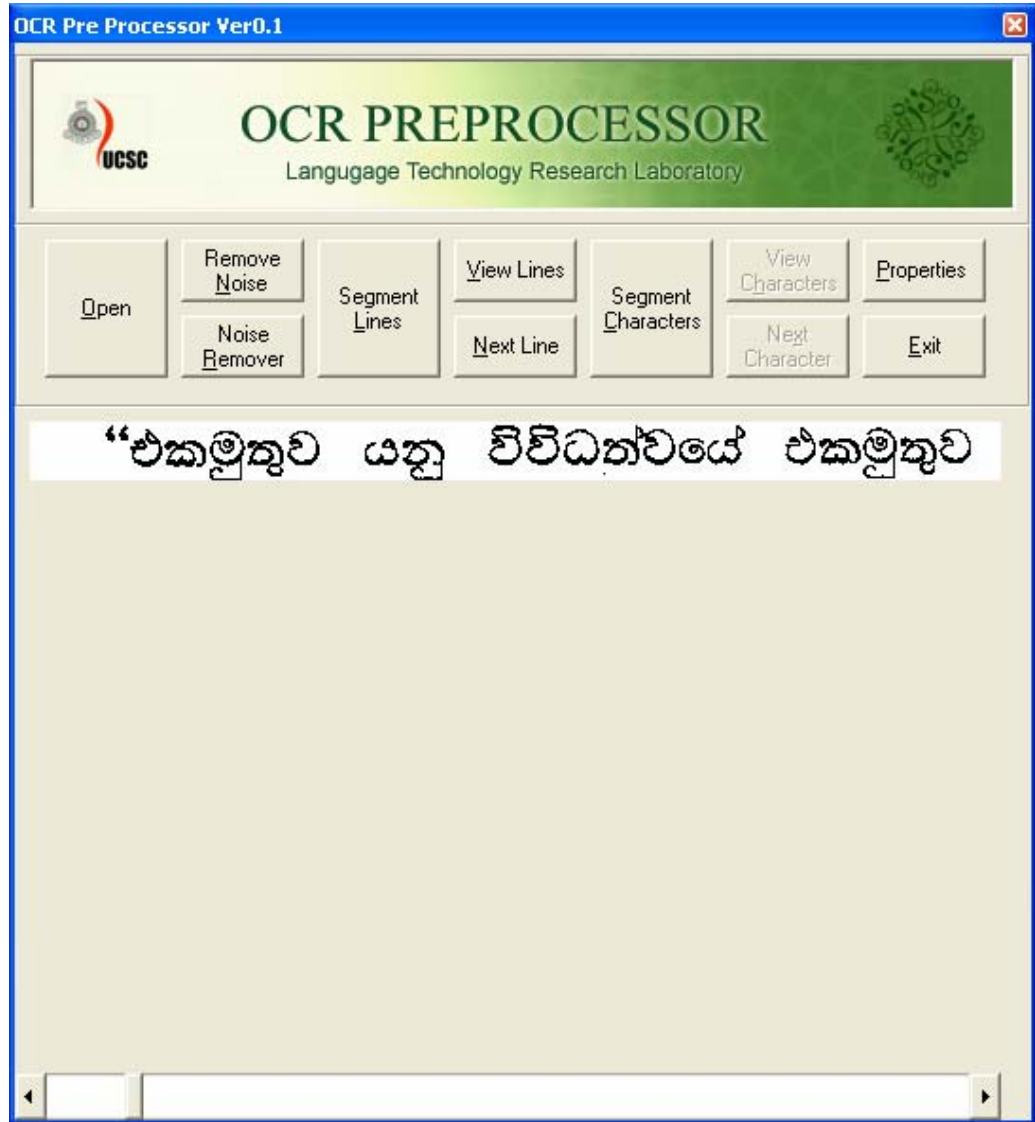

Figure 3.4.1: View Lines

3.5 Segment Lines into Characters

In order to segment the lines into characters, follow the steps given below.

I.   Click the "Segment Characters" button

II. Select a folder to save the segmented image or create a new folder.

III. Click "OK" button to segment the image to characters. When the segmentation is completed a massage appears as shown in Figure 3.5.1
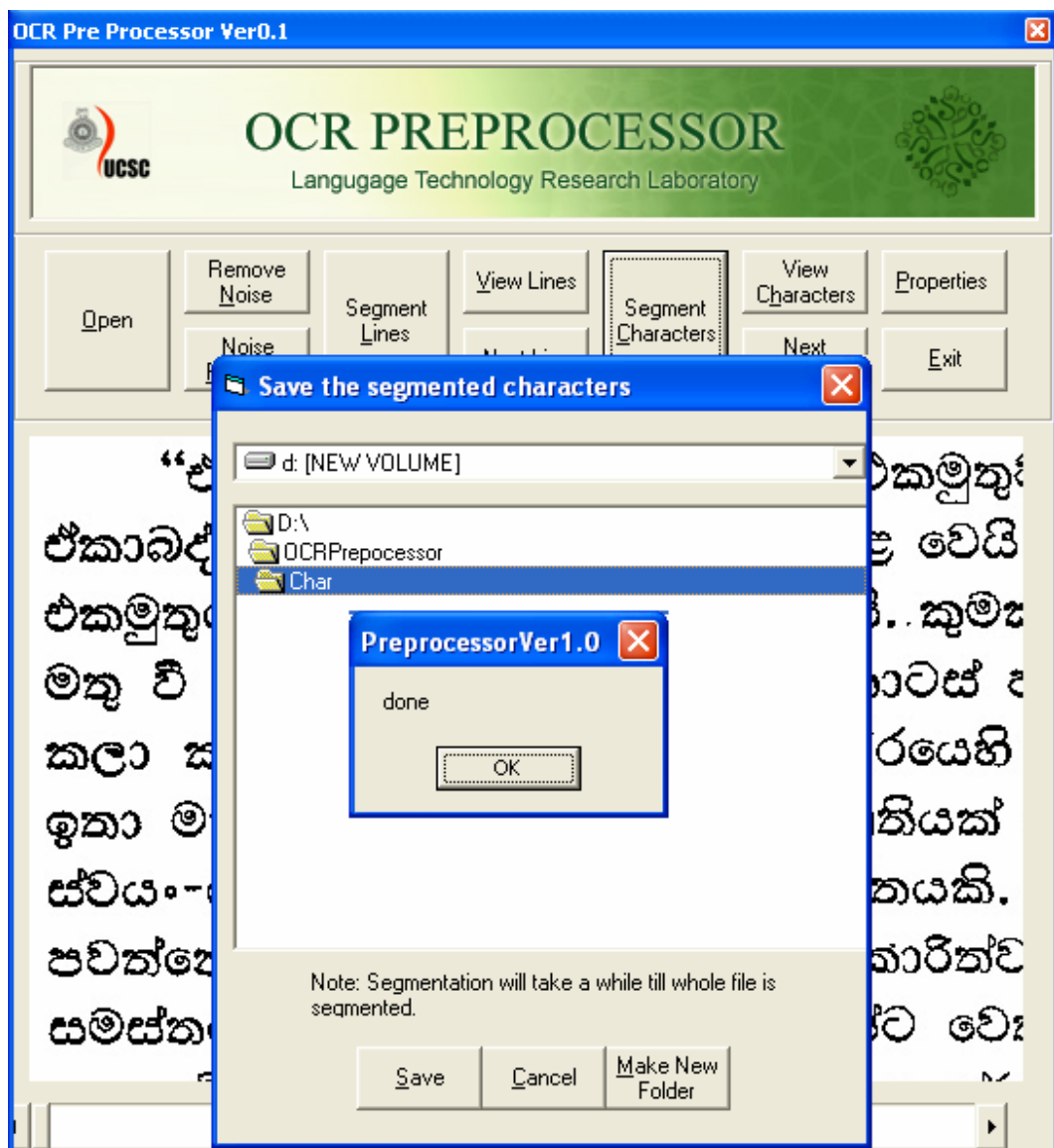

Figure 3.5.1: Segment Characters

IV. Click "Segmented Image" then the segmented characters appear as shown in Figure 3.5.2
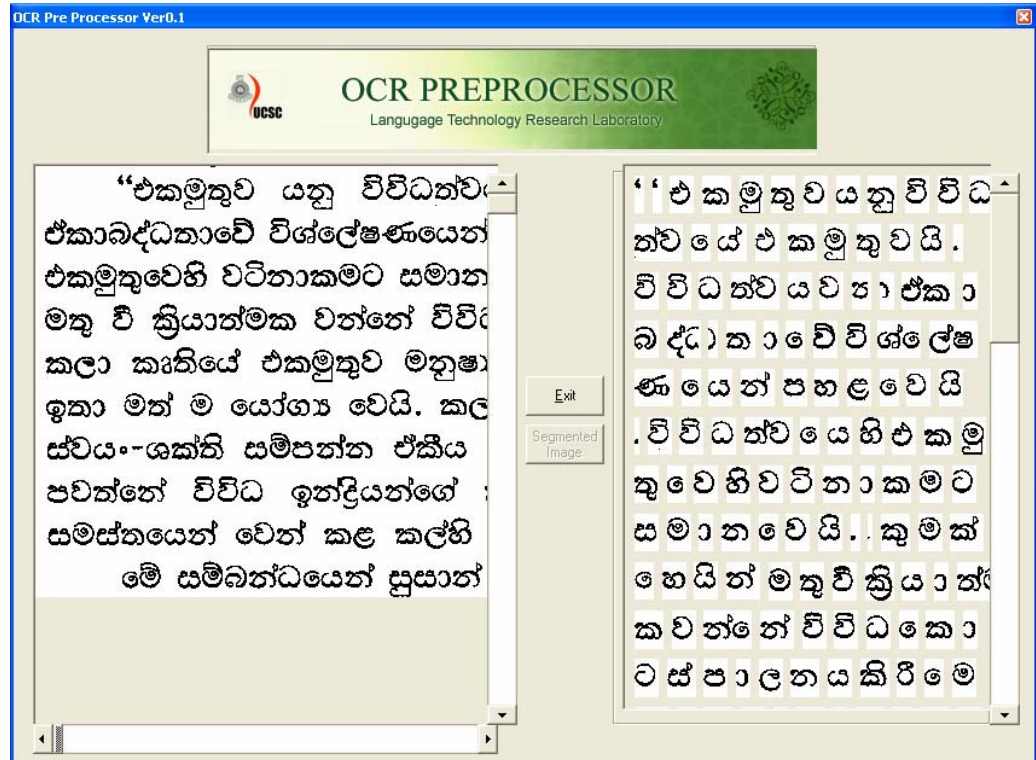

Figure 3.5.2

3.6 View Characters

In order to view the segmented characters click the "View Characters" button. Then the input box appears and inserts the line number, as shown in Figure 3.6.1
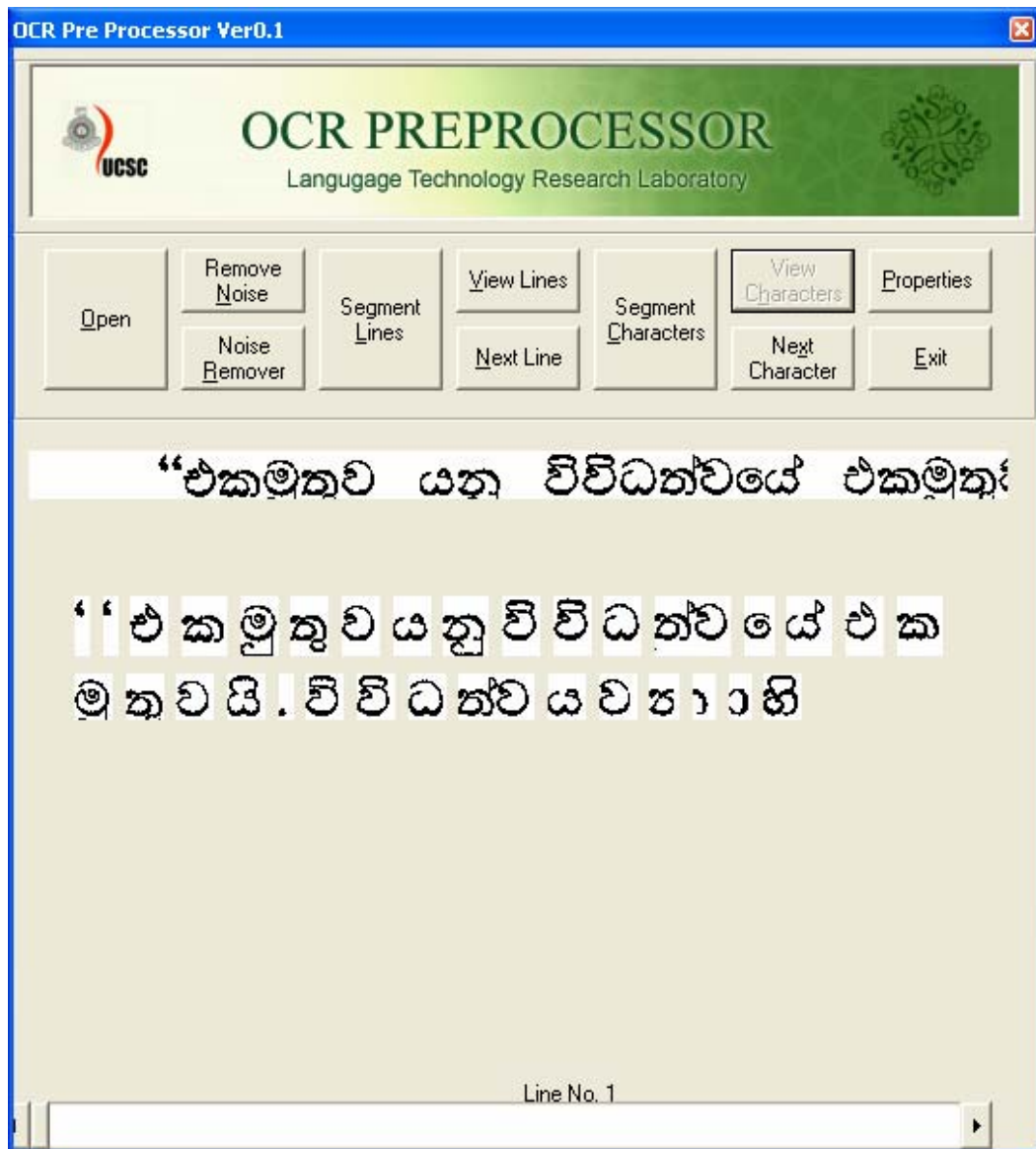

Figure 3.6.1: View Characters

3.7 Finding the properties of an Image

Click "Properties" button. Then the properties of the image appear in a massage box
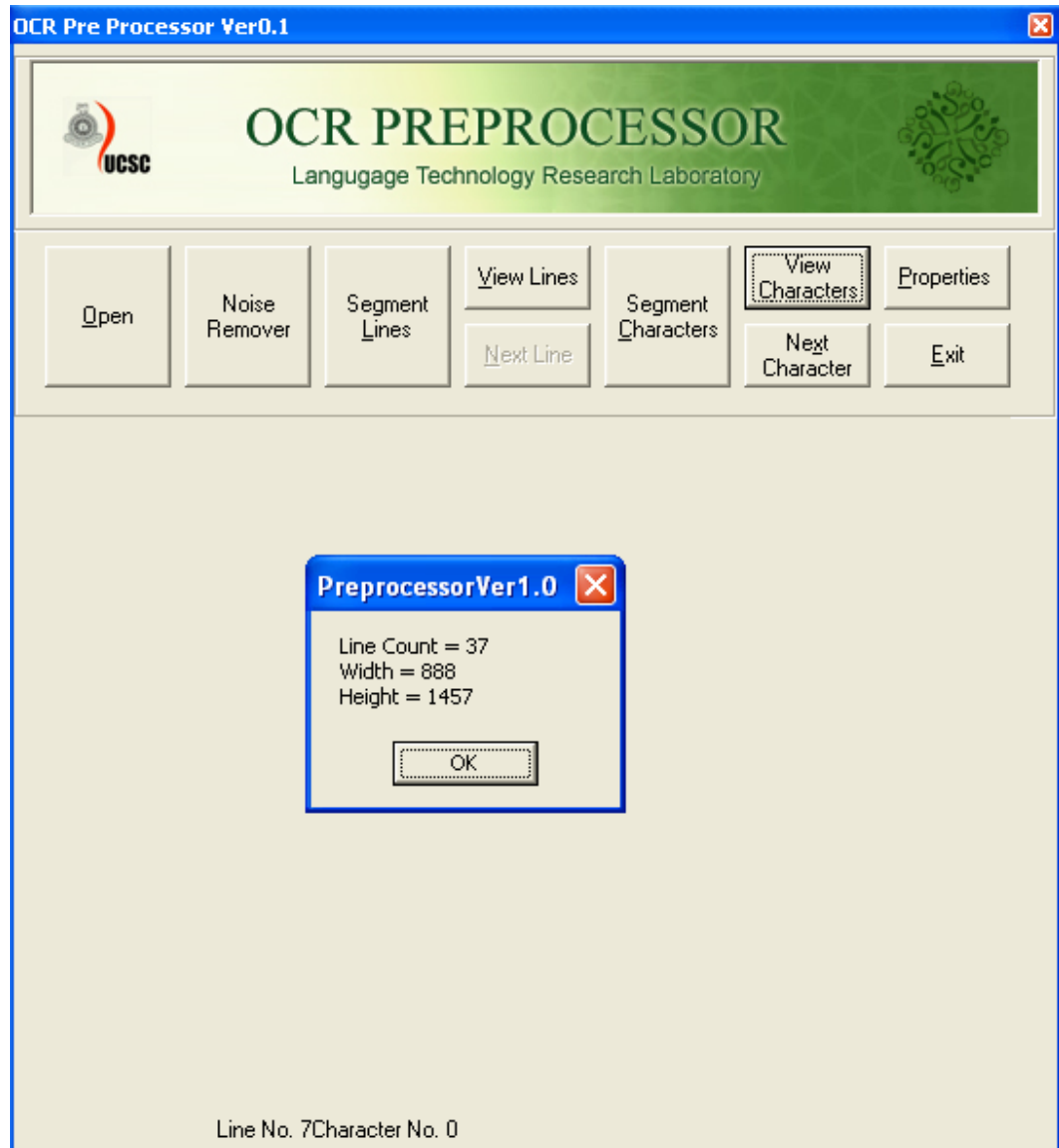
as shown in the Figure 3.7.1.

Figure 3.7.1: Properties of the image

3.8 Close the program

Click "Exit" button to close the application