

C  
001.642  
LAK

**FEATURE EXTRACTION AND CLASSIFICATION  
OF  
SINHALA FONTS**

**A PROJECT REPORT PRESENTED BY**

**By  
Venura Lakshman  
PGIS/SC/MSC/CSC/05/41**

**to the Board of Study in Computer Science & Statistics of the  
POSTGRADUATE INSTITUTE OF SCIENCE**

*in partial fulfilment of the requirement*

*for the award of the degree of*

**MASTER OF SCIENCE IN COMPUTER SCIENCE**

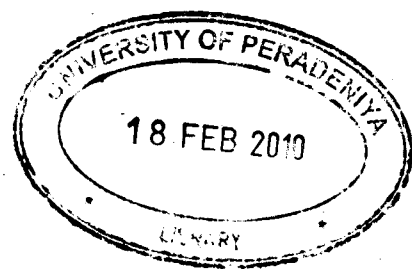
of the

**UNIVERSITY OF PERADENIYA**

**SRI LANKA**



**2009**



**628208**

# **FEATURE EXTRACTION AND CLASSIFICATION OF SINHALA FONTS**

**A PROJECT REPORT PRESENTED BY**

**By  
Venura Lakshman  
PGIS/SC/MS/CSC/05/41**

Optical character recognition (OCR) is varied from language to language .It is an important area in pattern recognition and image processing. The basic things in OCR are extracting the features & classification of relevant characters. In addition to that, I further developed to map classified characters with Unicode.

In this project, I used Artificial Neural network (ANN) to achieve above tasks. Artificial neural networks are a method of computation that tries to achieve human-like performance in the field of image and character recognition.

In this case, this application is used ANN with two layers. Also the application is limited for several Sinhala characters as Sinhala alphabet is very complex than other alphabets. For easy identification we can categorize Sinhala fonts into three main classes. But identification of some complex characters require further developments.

Since this is Matlab implemented, this application has many features than other OCRs. Since image processing part has embed to the application it is not required to use separate image processor. Therefore it is easy to use & trouble free.