BONAFIDE CERTIFICATE

This is to certify that this dissertation report entitled "**A Framework for Document Image Analysis**" submitted to AMRITA VISHWA VIDYAPEETHAM, Mysuru Campus, Mysuru, is a bonafide record of work done by

Pruthvi TR [MY.SC.P2MCA19007]

Aishwarya Govinda Rao [MY.SC.P2MCA19009]

under my supervision from "July 2020" to "June 2021"

Dr. N Shobha Rani

Associate professor

Adwitiya Mukhopadhyay

CHAIRPERSON

Place: Mysuru

Date: June, 2021

Internal Examiner

1……

2……

External Examiner

1……...

ACKNOWLEDGMENTS

First and foremost, I feel deeply indebted to Her Holiness Most Revered Mata Amritanandamayi Devi (Amma) for her inspiration and guidance both in unseen and unconcealed ways.

Wholeheartedly, I thank our respected Director, **Br. Anantaananda Chaitanya** and Correspondent **Br. Muktidamrita Chaitanya** for providing the necessary environment, infrastructure and encouragement for carrying out my dissertation work at Amrita Vishwa Vidyapeetham , Mysuru Campus.

We would like to express our sincere thanks to our beloved Principal **Dr. G Ravindranath** for giving us moral support and continuous encouragement which has been the key for the successful completion of the project.

We are pleased to acknowledge **Mr. Adwitiya Mukhopadhyay**, Chairperson, Department of Computer Science, for his encouragement and support throughout the project.

We would like to express our heart-felt gratitude to our Project Co-ordinator **Ms. Pallavi M S,** Assistant Professor, Department of Computer Science for her valuable suggestions and excellent guidance rendered throughout this project.

We would like to express our heart-felt gratitude to our guide **N. Shobha Rani**, Assistant Professor, Department of Computer Science for her/his valuable suggestions and excellent guidance rendered throughout this project.

Further, we extend our thanks to all the faculty members and technical staff of our department and ICTS for their suggestions, support and providing resources at needed times. Finally, we express our sincere gratitude to our parents and friends who have been the embodiment of love and affection which helped us to carry out the project in a smooth and successful way.

Date: June 2020

PRUTHVI T R [MY.SC.P2MCA19007]

AISHWARYA GOVINDA RAO [MY.SC.P2MCA19009]

DECLARATION BY AUTHORS

This is to declare that this report has been written by me/us. No part of the report is plagiarized from other sources. All information included from other sources has been duly acknowledged. I/We aver that if any part of the report is found to be plagiarized, I/we shall take full responsibility for it.

Pruthvi T R

[MY.SC.P2MCA19007]

Aishwarya Govinda Rao

[MY.SC.P2MCA19009]

Place: Mysuru

Date: June 2020

ABSTRACT

Document Image Processing is one of the widely used technologies today. It has proved to be extremely useful in digitizing physical documents to preserve them for longer times. But to achieve good results for any documented analysis, proper alignment and skew are required. This can be achieved by using skew detection and corrections algorithms. Hence in this work, we propose a skew detection and correction model, along with a line segmentation and table detection algorithm suitable for pre-printed and handwritten documents. For the skew correction model, text images that consist of text in the form of a wave or an arc or diagonal text, etc., are used as a dataset. Hough transform and Radon transform are used for slant correction using the above mentioned transforms dataset. Line segmentation is carried out on the Handwritten South Indian script images using Horizontal Projection Profile. And table detection is carried out on pre-printed documents using edge detection techniques and morphological operations.

Table of Contents

List of Figures

List of Tables