Multilingual Artificial Text Extraction and Script Identification from Video Images

Azra Batool

Enrollment #: 01-244112-003

Supervised By

Dr. Imran Ahmed Siddiqui

Department of Computer and Software Engineering

Bahria University, Islamabad Campus

2011-2014

Multilingual Artificial Text Extraction and Script Identification from Video Images

A THESIS SUBMITTED TO THE BAHRIA UNIVERSITY IN THE PARTIAL FULFILLMENT OF REQUIREMENTS FOR THE DEGREE OF MS SOFTWARE ENGINEERING

Azra Batool

Enrollment #: 01-244112-023

Supervised By

Dr. Imran Ahmed Siddiqui

Department of Computer and Software Engineering

Bahria University Islamabad

2011-2014

CERTIFICATE OF ORIGINALITY

I certify that the intellectual contents of the thesis

"Multilingual Artificial Text Extraction and Script Identification from Video Images"

is the product of my own research work except, as cited properly and accurately in the acknowledgment and references, the material taken from any source such as research papers, research journals, books, internet, etc. solely to support, elaborate, compare and extend the earlier work, Further, this work has not been submitted previously for a degree at this or any other University.

The incorrectness if the above information, if proved at any stage, shall authorize the university to cancel my degree.

Signature: _____. Dated: January 31st , 2014 .

Name of the Research student: Azra Batool .

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, Most Beneficent and the Most Merciful.

Alhamdulillah, all praise to Allah Almighty for the strength and blessing in completing this research work. I would like to gratefully and sincerely thank my supervisor Dr. Imran Ahmed Siddiqui for his patient guidance, understanding, encouragement and excellent supervision throughout my work. I have been lucky to have a great supervisor who cared a lot about my work and extremely honored to work in his supervision. I express my gratitude to him for his continuous and quick response to my questions and quires. I am extremely thankful to him, without his constant support and help, it would have been impossible to complete this research.

I am extremely thankful to my friend Safia Shabbir, who is always there to help me and give her best suggestions in every situation for all affairs including studies. She was always there cheering me up and stood by me through good times and bad. I would also like to thank my Uncle Mr. Shabbir Hussain and Aunt Mrs. Shabbir Hussain for their affection, love and realization of being a part of their family.

Finally, I would like to express my deepest gratitude to my beloved parents, Mrs. Sakina Raza (my Amman) and specially Mrs. Muhammad Raza (my Baba) for their love, prayers, support and encouragements throughout my life and especially during my research. I would also like to thanks to my siblings for their constant support and love.

DEDICATION

This thesis is dedicated to my adorable parents Mr. Muhammad Raza and Sakina Raza for their love, endless support and encouragement throughout my life.

Also this thesis is dedicated to the great teacher; I have ever met, Dr. Imran Ahmed Siddiqui.

ABSTRACT

With the tremendous growth in the amount of multimedia data, especially videos, has increased the need for efficient indexing and retrieval techniques. In addition to the audio-visual content itself, a power tool that be employed for indexing of videos is the caption text appearing in them. An important component of textual content based video indexing and retrieval systems is the detection and extraction of text from video frames. Most of the existing text extraction system target textual occurrences in a particular script or language. We have proposed a generic multilingual text extraction system that relies on a combination of unsupervised and supervised techniques. The unsupervised approach is based on application of image analysis techniques which exploit the contrast, alignment and geometrical properties of text and identify candidate text regions in an image. Potential text regions are then validated by an Artificial Neural Network (ANN) using a set of features computed from Gray Level Co-occurrence Matrices (GLCM). Detected text regions are then binarized to segment text from the background. The script of the extracted text is finally identified using texture based features based on Local Binary Patterns (LBP). The proposed system was evaluated on video images containing textual occurrences in five different languages including English, Urdu, Hindi, Chinese and Arabic. The promising results of the experimental evaluations validate the effectiveness of the proposed system for text extraction and script identification.

TABLE OF CONTENTS

List of Figures

List of Tables