



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكرو فيلم

بسم الله الرحمن الرحيم



HANAA ALY



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكروفيلم



شبكة المعلومات الجامعية التوثيق الإلكتروني والميكروفيلم



HANAA ALY



شبكة المعلومات الجامعية
التوثيق الإلكتروني والميكروفيلم

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغييرات



يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



HANAA ALY



Cairo University

TEXT EXTRACTION AND ENHANCEMENT FROM IMAGERY FILMS AND NEWS

By

Hossam Ahmed Fadel Elshahaby

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

TEXT EXTRACTION AND ENHANCEMENT FROM IMAGERY FILMS AND NEWS

By
Hossam Ahmed Fadel Elshahaby

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Under the Supervision of

Prof. Dr. Mohsen Rashwan

.....
Professor of Electronics and
Communications Department,
Faculty of Engineering, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

TEXT EXTRACTION AND ENHANCEMENT FROM IMAGERY FILMS AND NEWS

By
Hossam Ahmed Fadel Elshahaby

A Thesis Submitted to the
Faculty of Engineering at Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Electronics and Communications Engineering

Approved by the
Examining Committee

Prof. Dr. Mohsen A. Rashwan,

Thesis Main Advisor

Assistant Prof. Dr. Omar A. Nasr,

Internal Examiner

Prof. Dr. Khalid M. El Sayed,

External Examiner

Faculty of Computers and Information, Cairo University

FACULTY OF ENGINEERING, CAIRO UNIVERSITY
GIZA, EGYPT
2021

Engineer's Name: Hossam Ahmed Fadel Elshahaby
Date of Birth: 03/02/1988
Nationality: Egyptian
E-mail: hossam.elshahaby@gmail.com
Phone: +20 111 654 22 55
Address: Mohandessen, Agouza, Giza
Registration Date: 01/10/2014
Awarding Date: 2021
Degree: Master of Science
Department: Electronics and Communications Engineering



Supervisor:
Prof. Mohsen Rashwan

Examiners:
Prof. Mohsen A. Rashwan (Thesis Main Advisor)
Dr. Omar A. Nasr (Internal Examiner)
Prof. Khalid M. El Sayed (External Examiner)
Faculty of Computers and Information, Cairo University

Title of Thesis:

Text Extraction And Enhancement From Imagery Films And News.

Key Words:

Text Detection and Text Recognition; Edge Features; Multiple Frames Integration; Films videos; Computer Vision.

Summary:

This research solves problems of text detection, verification, segmentation, and enhancement in text imagery applications like news and films. Recent approaches are applied in an efficient way. In news videos, locating multiple captions is done using edge detection by grayscale-based and color-based techniques. Stationary as well as moving captions across frames are automatically classified as horizontal or vertical motion using combinatory techniques of recurrent neural network and correlation-based technique. The Convolutional Neural Nets (CNNs) is used to verify the caption as a caption containing text for further processing. In films, several CNNs are implemented to detect frames containing text with high accuracy. Error handling and correction algorithm are applied to resolve classification problems. Multiple frames integration technique is used to extract inserted text in graphics and enhance it. The Correctly Detected Characters (CDC) overall average weighted accuracy for news text recognition using Autoencoder Neural Network (ANN) is 96.07% while the CDC average weighted accuracy for films text translation is 97.79%.

Disclaimer

I hereby declare that this thesis is my own original work and that no part of it has been submitted for a degree qualification at any other university or institute.
I further declare that I have appropriately acknowledged all sources used and have cited them in the references section.

Name: Hossam Ahmed Fadel Elshahaby

Date: / / 2021

Signature:

Dedication

I would like to dedicate my thesis to my parents and sister for supporting and encouraging me during my Master's degree.

Acknowledgments

Firstly, I would like to thank God for helping me and facilitating all obstacles out of my hand until this research is finally finished successfully. Next, I would like to thank the RDI team for supporting me with their experience and data set used in my research. Special thanks to Professor Dr. Mohsen Rashwan for his contribution with several ideas which I used to solve the problem of caption detection and enhancement from images as well as empowering me during the whole period of research.

Also, I would like to thank Professor Dr. Mohsen Rashwan for his patience during the period of the research. Indeed, I apologize from the bottom of my heart for any misunderstanding or mistakes. Really, I respect and do love you Dr. Mohsen. Special thanks from the bottom of my heart goes to the examining committee for their valuable comments and corrections during discussion of thesis.

Finally, this work would not be carried out without the support coming from the electronics and communications department members in the Cairo University Faculty of Engineering (CUFE).

Table of Contents

DISCLAIMER.....	I
DEDICATION.....	II
TABLE OF CONTENTS.....	IV
LIST OF TABLES	VII
LIST OF FIGURES	X
NOMENCLATURE	XII
ABSTRACT	XIV
CHAPTER 1 : INTRODUCTION	1
1.1 MOTIVATION.....	1
1.2 SCOPE	4
1.3 PROBLEM DEFINITION.....	4
1.4 WHAT IS IMAGERY TEXT?	5
1.5 HOW CAN TEXT DIFFER IN ITS NATURE?	7
1.6 APPLICATIONS	7
1.6.1 News Captions Text.....	7
1.6.2 Films Graphics Text.....	8
1.7 THESIS ORGANIZATION	9
CHAPTER 2 : LITERATURE REVIEW.....	11
2.1 INTRODUCTION	11
2.2 RELATED PUBLISHED WORK.....	12
2.2.1 Text Detection And Extraction	12
2.2.1.1 Caption text Image Extraction.....	12
2.2.1.2 Graphical Text Image Extraction	14
2.2.1.3 Document Text Image Extraction	15
2.2.1.4 Scene Text Image Extraction.....	15
2.2.1.5 Heterogeneous Text Image Extraction	17
2.2.2 Text Image Enhancement	19
2.3 TEXT DETECTION AND RECOGNITION FUNDAMENTALS	20
2.3.1 Text Detection.....	20
2.3.1.1 Text Localization	21
2.3.1.2 Text Verification	21
2.3.2 Text Recognition.....	22
2.3.2.1 Text Segmentation	23
2.3.2.2 Word Recognition	24
2.4 SUMMARY AND DISCUSSION	24
CHAPTER 3 GRAPHICAL TEXT IMAGES IN FILMS.....	25

3.1	CHALLENGES	25
3.2	TEXT EXTRACTION METHODOLOGIES	25
3.2.1	Text Localization	25
3.2.2	Text Verification	26
3.2.2.1	Geometrical Method	26
3.2.2.2.	Deep Learning Neural Networks Method.....	26
3.2.3	Feature Extraction	27
3.2.3.1.	Geometrical Method	27
3.2.3.2.	Deep Learning Neural Networks Method.....	28
3.3.	SELECTED SYSTEM.....	38
3.3.1.	Flow Chart For Films Multimedia application.....	40
3.3.2.	Applied Technique.....	41
3.3.2.1.	Adopted Algorithm	42
3.3.2.2.	Text Image Enhancement	44
3.3.2.3.	Post Processing.....	44
3.3.3.	Error Handling And Correction	45
3.3.3.1.	Handling Classification Error	45
3.3.3.2.	Handling Translation Issues	46
3.4.	DATASET.....	46
3.5.	EVALUATION CRITERIA.....	48
3.5.1.	Execution Performance	48
3.5.2.	Text Extraction Performance	48
3.6.	SUMMARY AND DISCUSSION	50
CHAPTER 4 : CAPTION TEXT IMAGES IN NEWS		51
4.1.	CHALLENGES	51
4.2.	CAPTION EXTRACTION METHODOLOGIES	55
4.2.1	Hough Transform.....	55
4.2.2.	Color-Based Edges Technique	58
4.3	PROPOSED SYSTEM	59
4.3.1	Flow Chart For News Multimedia Application	59
4.3.2.	Applied Technique.....	60
4.3.2.1.	Text Captions Preprocessing	60
4.3.2.2.	Text Captions Localization	61
4.3.2.3.	Text Captions Classification	61
4.3.2.4.	Text Caption Verification.....	64
4.3.2.5.	Point Feature Matching	64
4.3.2.6.	Text Caption Enhancement	65
4.3.2.7.	Post Processing.....	66
4.4.	DATASET.....	66
4.5.	NEWS MULTIMEDIA APPLICATION RESULTS	69
4.6.	SUMMARY AND DISCUSSION	73
CHAPTER 5 : SUPER-RESOLUTION FOR TEXT ENHANCEMENT.....		74
5.1.	WHAT IS SUPER-RESOLUTION?.....	74
WE PERFORMED A TEXT RECOGNITION EVALUATION FOR BOTH FILMS AND NEWS MULTIMEDIA APPLICATIONS.		75
a)	Films Multimedia Application.....	75
b)	News Multimedia Application.....	75

5.2.	SUPER-RESOLUTION USING MATHEMATICAL METHOD	77
5.2.1.	Interpolation Process	79
5.2.2.	Image Super-Resolution Reconstruction Process	80
5.2.3.	Least-Squares Estimation.....	80
5.2.4.	Experiments	82
5.2.5.	Results.....	83
a)	Manual Calculations	83
b)	Automatic Calculations	83
5.3.	SUPER-RESOLUTION USING AUTOENCODER NEURAL NETWORK.....	86
5.3.1.	Experiments	87
5.3.2.	Results.....	95
5.4.	SUMMARY AND DISCUSSION.....	98
CHAPTER 6 CONCLUSION AND FUTURE WORK		100
REFERENCES		102

List of Tables

Table.3.2 Film application training dataset	46
Table.3.3 Film application testing dataset for text classification.	47
Table.3.4 Film application graphical caption extraction evaluation dataset containing data with multiple languages for text.	47
Table.3.5 System performance comparison for video using either geometrical or neural network method for text or non-text frame classification.	48
Table.3.7 Text extraction performance comparison for video using either geometrical or neural network method for text frame detection.	49
Table.3.8 System performance using the system evaluation dataset.....	49
Table.4.1 Components controlling the state of the layer [82].	62
Table.4.2 Formula for each component [82].	64
Table.4.3 News application training model dataset.....	67
Table.4.4 News application testing model dataset	67
The AcTiV-D videos are considered extrinsic inputs to the system while the other self- gathered videos are considered intrinsic inputs.....	67
Table.4.5 News application system evaluation dataset.	68
Table.4.6 Feature extraction performance using Hough transformation and color-based methods.	70
Table.4.7 Classification performance for the RNN network.....	71
Table.4.8 News stories detection performance for the system.	72
Table.5.1 Abbyy OCR results before super-resolution for films.	75
Table.5.2 Abbyy OCR results before super-resolution for news.	76
Table.5.5 Abbyy OCR Configuration 1 results before and after super-resolution using the mathematical model for Low-Resolution (LR) Arabic news videos.	83
Table.5.6 Abbyy OCR metrics for Configuration 1 using manual calculations for Mathematical Method.	83
The following results are that I got from the automatic calculations for the configuration 1 experiment in the thesis before we adjust the parameters and use Configuration 2 and Configuration 3.	83
Table.5.7 Character Error Analysis of Configuration 1 for the Mathematical Method using Automatic Calculations.	84

Table.5.8 Word Error Analysis of Configuration 1 for the Mathematical Method using Automatic Calculations.	84
Table.5.9 Analysis of Results using both Manual and Automatic Methods.	84
As seen in the table, there was a great difference between errors calculated by hand and those calculated automatically. This is due to the wrong calculations with a possible human bias. These wrong results mislead me and caused me almost one and half year delay to discover and correct. The results are then enhanced using configuration 2 and 3.	84
Table.5.10 Abbyy OCR Configuration 2 Wrong Recognitions (WR) before and after super resolution using the mathematical model for LR Arabic news videos...	85
Table.5.11 Abbyy OCR metrics after super-resolution using the mathematical model Configuration 2 for Low-Resolution (LR) Arabic news videos.	85
Table.5.12 Abbyy OCR Configuration 3 Wrong Recognitions (WR) before and after super-resolution using the mathematical model for Low-Resolution (LR) Arabic news videos.	85
Table.5.13 Abbyy OCR metrics after super-resolution using the mathematical model Configuration 3 for Low-Resolution (LR) Arabic news videos.	85
Table.5.14 Comparison between the different Autoencoder Networks.	89
Table.5.16 Abbyy OCR Wrong Recognitions (WR) before and after super-resolution using Autoencoder Model for Low-Resolution (LR) Arabic news videos of Experiment 1.	94
Table.5.17 Abbyy OCR metrics using manual calculations for Autoencoder method of Experiment 1.	94
Table.5.18 Character Error Analysis of Experiment 1 for the Autoencoder Method using Automatic Calculations.	94
Table.5.19 Word Error Analysis of Experiment 1 for the Autoencoder Method using Automatic Calculations.	95
Table.5.20 Analysis of Results using both Manual and Automatic Methods.	95
As seen in the table, there was a great difference between errors calculated by hand and those calculated automatically. This is due to the wrong calculations with a possible human bias. These wrong results mislead me and caused me almost one and half year delay to discover and correct. The results are then enhanced using Experiments 2, 3, 4, and 5.	95