Friedrich-Alexander-Universität Erlangen-Nürnberg

Lehrstuhl für Multimediakommunikation und Signalverarbeitung

Prof. Dr.-Ing. André Kaup

Master Thesis

Text Recognition Algorithms for Screen Content Quality Assessment

Sebastian Hirt

July 2023

Supervisors: Prof. Dr.-Ing. André Kaup
M. Sc. Hannah Och

Lehrstuhl für Multimediakommunikation
und Signalverarbeitung
Prof. Dr.-Ing. André Kaup

Friedrich-Alexander-Universität
Erlangen-Nürnberg

Master Thesis

for

Mr. Sebastian Hirt

Text Recognition Algorithms for Screen Content Quality Assessment

Texterkennungsalgorithmen für die Qualitätsbewertung von Bildschirminhalten

Screen content (SC) refers to images and videos as they can be found on screens during office work or similar. Consequently, next to buttons, icons, and computer graphics, they contain textual information in most cases. For images containing text, human viewers place a high importance on readability of text when assessing image quality. However, conventional objective image quality assessment algorithms, such as PSNR, MS-SSIM or even screen content specific quality metrics, such as ESIM or GFM, only consider text readability indirectly, e.g. by taking into account gradient distortions, or not at all.

Since text readability plays an important role in the subjective quality assessment of screen content images, evaluating text recognition rates can be a useful addition to conventional quality metrics. However, subjective tests are not feasible, since they are too expensive and time-consuming. Instead, current Deep Learning-based text detection and recognition algorithms have shown very high text recognition rates and can be utilized to simulate the human reader.

In this thesis, Mr. Hirt will explore the application of such algorithms for the assessment of screen content image quality. First, Mr. Hirt will research state-of-the-art text recognition and detection methods. Since ground truth textual information is not included in available screen content datasets, Mr. Hirt will generate a labeled dataset to evaluate the efficiency of the researched algorithms on screen content data. Available datasets with subjective quality scores will be utilized to investigate the correlation between text recognition rates and human judgement. Since most datasets do not contain textual ground truth information, in a further step, Mr Hirt will investigate the feasibility of using recognized text from pristine images as ground truth instead. A structured implementation and detailed documentation of the framework and the performed experiments is part of the work.

Start: 01.02.2023
End: 31.07.2023

A. Kaup

Prof. Dr.-Ing. A. Kaup

Declaration

I confirm that I have written this thesis unaided and without using sources other than those listed and that this thesis has never been submitted to another examination authority and accepted as part of an examination achievement, neither in this form nor in a similar form. All content that was taken from a third party either verbatim or in substance has been acknowledged as such.

Place, Date	<full name> <address>

Contents

Kurzfassung

Abstract in German...

Abstract

Abstract in English...

Symbols and Notations

In order of appearance.

$+$	Addition

Abbreviations and Acronyms

MOS Mean opinion score. 3

TER Text error rate. 3

Chapter 1

Introduction

The introduction may also include acronyms, and references like [1].

Chapter 2

Methods

2.1 Optical Character Recognition

2.2 Metrics

The Text error rate (TER) is defined as follows:

$$\text{TER} = \frac{S + D + I}{N} \tag{2.1}$$

with S being the number of substitutions, D the number of deletions, I the number of insertions, and N the total number of characters of the text prediction. The TER ranges from 0 to 1, where 0 means perfect recognition and 1 means no correct characters. Because the Mean opinion score (MOS) is defined in the range 0 to 100 and 100 represents a high subjective quality, the two metrics are unintuitive to compare. Therefore, transform the TER by subtracting it from 1 and multiplying it by 100 to get a MOS-like value, see Equation 2.2.

$$\text{TER} = (1 - \text{TER}_{raw}) \cdot 100. \tag{2.2}$$

Chapter 3

Dataset

This chapter discusses the dataset used in our work.

3.1 Overview

An overview of the dataset can be seen in Figure 3.1.

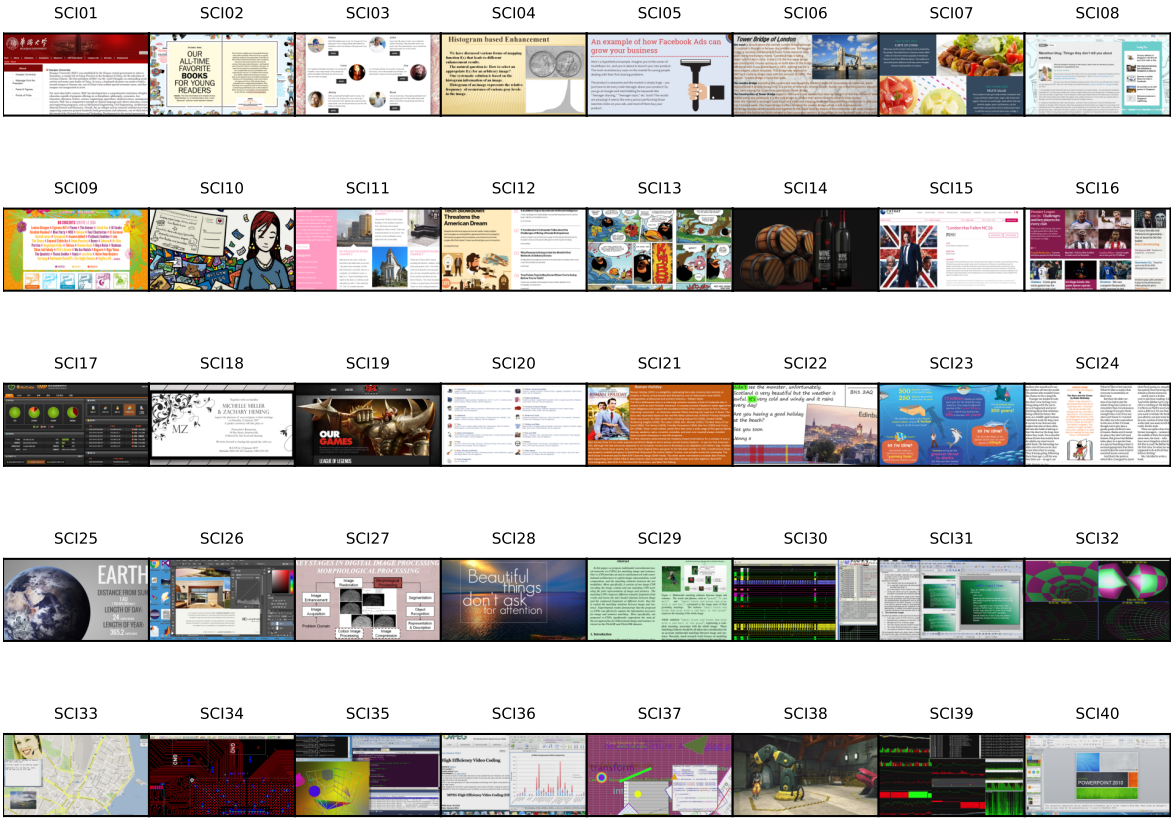


Figure 3.1: Overview of the dataset.

3.2 Analysis

In Figure 3.2 the TER is plotted against the MOS. It shows the TER and MOS of all 1800 distorted images compared to their reference image.

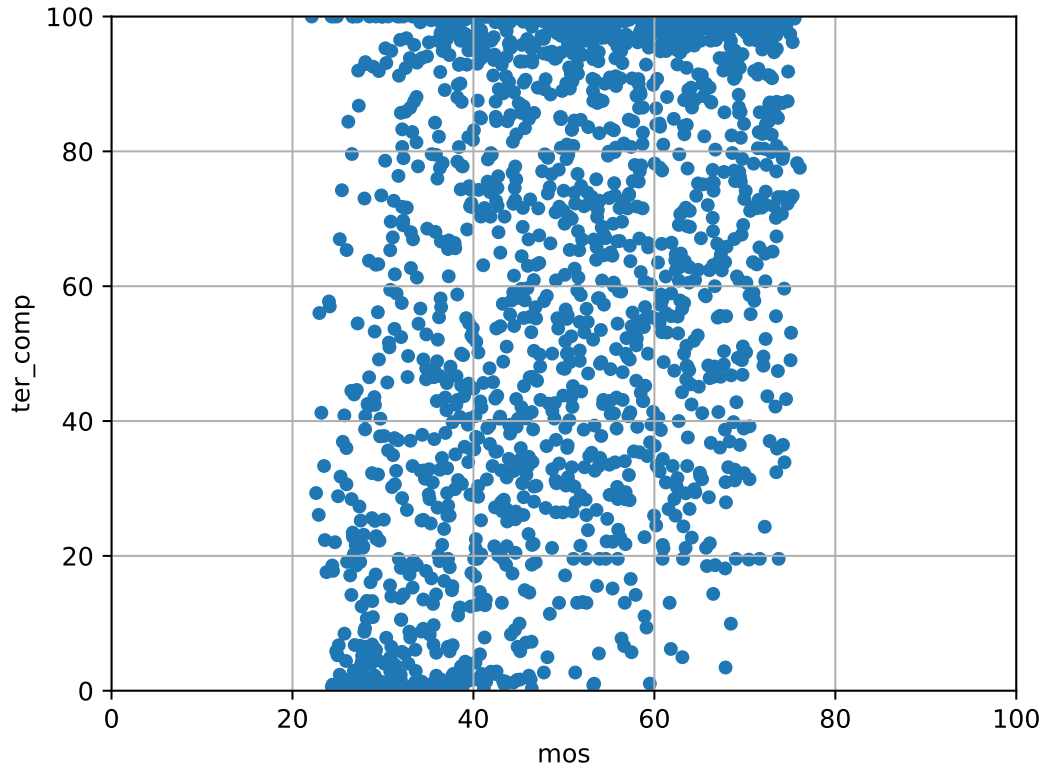


Figure 3.2: TER of all distorted images compared to their reference image plotted against the MOS.

In Figure 3.3 the TER is plotted against the MOS. We are only using a selection of images in this plot. These images (SCI4, SCI5, SCI6, SCI22 and SCI29) have their main focus on text, and have simple text structure.

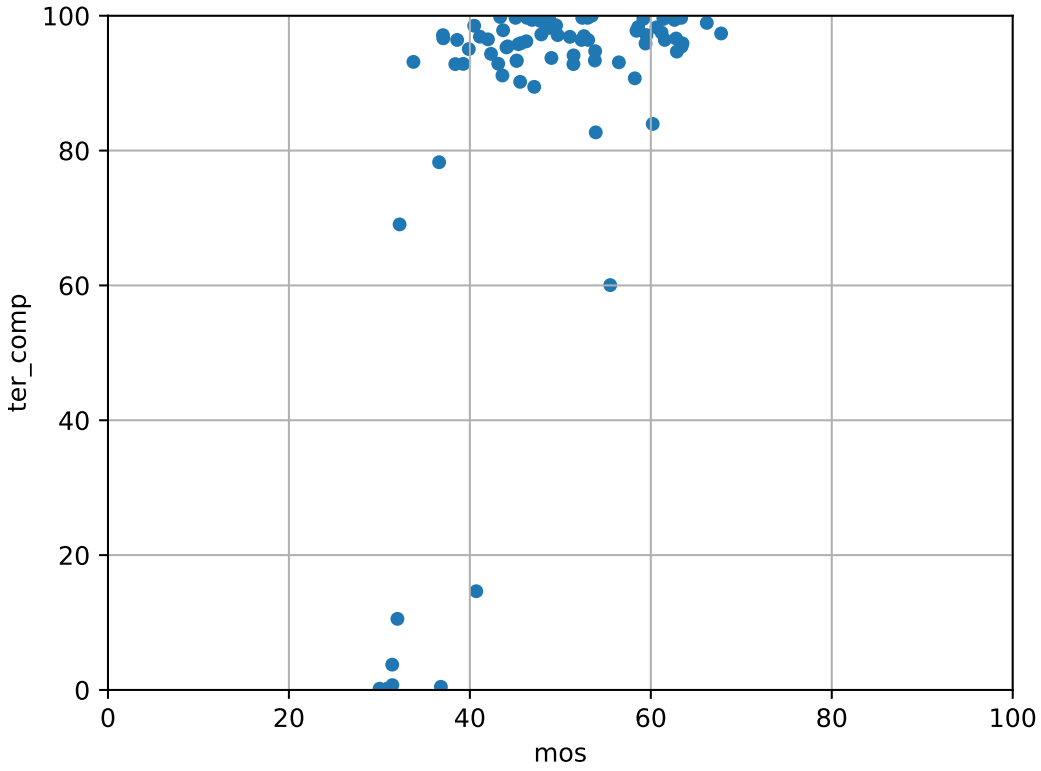


Figure 3.3: TER of a selction of distorted images compared to their reference image.

Chapter 4

Evaluation

This is the Evaluation.

Chapter 5

Conclusion

Conclusion...

Appendix A

First Appendix

Appendix (optional).

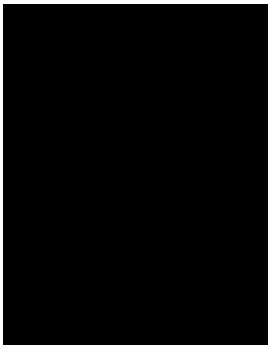
List of Figures

List of Tables

Bibliography

[1] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge Similarity for Screen Content Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4818–4831, Oct. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7954714/ ↑1

Curriculum Vitae



Short CV of the author. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.