RECOGNIZATION AND RECOMMENDATION OF HANGUL FONTS USING COMPUTER VISION BASED TECHNIQUES

CHUNG SEIJUN 2018-23341 OUM TAEHWOI 2012-11114
KIM HYEONSEO 2015-12377 CHOO WONHO 2014-18868
Computer Vision 2019-1
Department of Computer Science
Seoul National University
(Dated: April 26,2019)

Objective: This project recognizes different Hangul fonts based on Computer Vision Techniques and recommend akin fonts.

1 INTRODUCTION

Fonts are predefined collections of letters, varying in design and size. From the monospaced fonts on a VT100 terminal to the Seoul fonts by Seoul city, the range of available fonts to users have expanded, due to the rise of tools for font creation. Different fonts are utilized based on the purpose of the user, as an intermediary of the user's purposes and emotions.

As more fonts are introduced to the market, the quest for the appropriate font has become an evergoing battle with vast images. Also, finding the appropriate font is not just a matter of choice, as it can spur legal disputes as well.[1]

Compared to Hangul fonts, English fonts consist of 26 upper and 26 lower case letters, a total of 52 characters. The limited number of characters, and their simple form makes English font recognition not as challenging and previous researches have prodcued high recognition rates.

However, Hangul fonts consist of 2,350(KS X 1001) or 11,172(KPS 9566) characters, each composed of onset, nucelus, and codas. Due to this complicated format, research on font recogntion of Hangul characters is not active, compared to other Latin alphabets. In this project, we will use KS X 1001 character set.

The object of this paper is to address these issues, identifying Hangul characters from images to recognize their font and recommend similar fonts.

2 PAPER SURVEY

2.1 Classical Computer Vision

We shall utilize many techniques learned from our class and materials.

2.2 Optical Font Recognition Using Typographical Featurs

This paper[2] aims to identify typeface, weight, slope and the size of text from an image block without any knowledge of the content of the text.

2.3 Large-Scale Visual Font Recognition

This paper[3] addresses the large-scale visual font recognition (VFR) problem, which aims at automatic identification of the typeface.

3 KEY FEATURES

3.1 Separating Hangul character from image

We cannot expect input as an image of a single character. Input image may include many characters, so we should separate it character by character using OCR technique.

3.2 Feature Detection

A Study on Typology for Hangul Fonts Identification and Classification in terms of character's typeface, character size and character slope and etc. based on Computer Vision Techniques.

3.3 Machine Learning(Optional)

Rather than conservative computer theories, we can use machine learning techiniques in detecting characters or fonts. Mass of Hangul character fonts recogntion with CNN, Hidden Markov Model with reference to paper[4] or other materials.

4 ALGORITHM

```
separate characters from input image
for each c in characters:
  find idx such that c == ref_font[idx]
  for each font in fonts:
    compare(c, font[idx])
  select most similar font
select font with most concensus
```

The algorithm above briefly explains how font recogni-

tion is done. First, extract characters from an image. Compare the characters with their equivalents in the reference font sets, and select the best match.

5   EXPECTED RESULTS

We aim for the following goals:

1. Extract distinct features of Korean characters based on Computer Vision Techniques
2. Recognize/Recommend the provided font

6   DATASET

Since there's not enough data set existing, we will make new dataset. We will download .ttf font files offered by Naver Corp. Using the downloaded font files, we will generate 64 * 64 image file for 2,350 Korean letters. Therefore, there will be 2,350 * (# of distinct fonts) image files of size 64 * 64.

References

[1] The Font Bureau, Inc. v. NBC Universal, Inc. et al https://dockets.justia.com/docket/new-york/nyedce/1:2009cv04286/296847

[2] Abdelwahab Zramdini, et al. *Optical Font Recognition Using Typographical Features*. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLI-GENCE, 1988.

[3] Guang Chen, et al. *Large-Scale Visual Font Recognition*. CVPR, 2014.

[4] Chris Tensmeyer, et al. *Convolution Neural Networks for Font Classification*. CVPR, 2017.

[5] D. Ham, D.Lee, I.Jung and I. Oh. *Construction of Printd Hangul Character Database PHD08* Journal of the Korea Contents Association, vol.8, no.11, pp. 33-40, Nov. 2008.