



ICONI



KOREAN SOCIETY FOR INTERNET INFORMATION

The 11th International Conference on Internet

(ICONI 2019)

Dec. 15-18, 2019, Lotte Hotels & Resorts Hanoi, Vietnam

<http://www.iconi.org>

Proceedings of ICONI 2019

| Organized by |

Korean Society for Internet Information (KSII)

| Sponsored by |

Korea Electronics Technology Institute (KETI)

Contents



5-2	Towards a New Ultra-High-Speed Search Algorithm for Convolutional Neural Networks Chakkrit Termritthikun, Yeshe Jamtsho (Naresuan Univ., Thailand), Ivan Lee (Univ. of South Australia, Australia), Roy Morien, Paisarn Muneesawang (Naresuan Univ., Thailand)	113-118
5-3	Differences of Image to Image Translation According to Various U-Nets Sohyun Lim, Junchul Chun (Kyonggi Univ., ROK)	119-121
5-4	Classifying Animal Sounds using Multi-Feature Network Chung-Il Kim, Yongjang Cho, Seungwon Jung, Jehyeok Rew, Eenjun Hwang (Korea Univ., ROK)	122-126
5-5	AVC/HEVC on Hardware and Software for Advancing Video Streaming System Tuan Thanh Le (Gachon Univ., ROK), Eun-Seok Ryu (Sungkyunkwan Univ., ROK)	127-131
5-6	Facial Expression Recognition using Multiple Deep Learning Methods Sin Woo Lee, In Seop Na (Chosun Univ., ROK)	132-135
5-7	A study on smart ordering system for textile industry based on CNN Palash Sontakke, Chang Soo Kim (Pukyong National Univ., ROK), Il-Kyeun Ra (Univ. of Colorado at Denver, USA)	136-138
5-8	Design of Bandwidth-Efficient Transmission Method for User View-oriented Multi-view Video Services Minjae Seo, Jong Ho Paik (Seoul Women's Univ., ROK)	139-142
6-1	OCR Based Webtoon Infringement Site Detection Technology Eui-Jin Kim, Seul-Ki Choi, Jin Kwak (Ajou Univ., ROK)	143-144
6-2	K-mean Clustering Algorithm for Improving Real-time Communication based on Priority Queue Sa Math, Prohim Tam (Soonchunhyang Univ., ROK), Jin Wang, Wenbing Wu, Yongsong Zou (Changsha Univ. of Science & Technology, China), Zilong Jin (Nanjing Univ. of Information Science and Technology, China), Seokhoon Kim (Soonchunhyang Univ., ROK)	145-147

OCR Based Webtoon Infringement Site Detection Technology

Eui-Jin Kim¹, Seul-Ki Choi,² and Jin Kwak³

^{1,3}ISAA Lab, Department of Cyber Security, Ajou University
Suwon, Republic of Korea

²ISAA Lab, Department of Computer Engineering, Ajou University
Suwon, Republic of Korea

[e-mail: dmlwls0403@ajou.ac.kr, skchoi.isaa@gmail.com, security@ajou.ac.kr]

*Corresponding author: Jin Kwak

Abstract

Webtoon is a combination of web and cartoon that originated in South Korea. The size of the webtoon market is increasing with the recent globalization of the South Korean webtoon industry. At the same time, there are increasing instances of copyright infringement in the webtoon industry. To prevent this, 'Naver', one of the webtoon service platforms, is trying to reduce webtoon copyright damage by developing a 'Toon Radar' to detect webtoon copyright infringement sites. However, research on copyright detection technology is not as good as the growth rate of webtoon copyright infringement cases. Therefore, this paper proposes a webtoon infringement site detection technology based on OCR technology.

Keywords: Webtoon Infringement Site, OCR technology, Copyright

1. Introduction

Webtoons are a combination of web and cartoon that originated in South Korea. The size of the webtoon market is increasing with the recent globalization of the South Korean webtoon industry. And the most representative of these, Naver webtoons, is translated into English, Taiwanese, Chinese, Thai, and Indonesian [1].

But with the developing webtoon market, piracy of webtoon sites is increasing. In 2017, the Korean webtoon industry estimated 240 billion won worth of damage caused by webtoon-infringing sites. To prevent these intrusions, webtoon platforms are developing and responding to their webtoon detection programs. As a representative example, Naver is using its own 'Toon Radar' to detect webtoon-infringement sites [2].

Webtoons need different methods of detecting text extraction for text copyrights, such as traditional books because webtoons contain text within images.

In this paper, we analyze OCR as a related study in Section 2, present the detection method for the webtoon infringement site in Section 3, and conclude with Section 4.

2. Related Works

2.1 OCR(Optical Character Recognition)

OCR is a technology for image processing, which converts the shape of letters into a machine-readable form[3]. Since OCRs are affected by unnecessary areas of input images, further character area extraction from images is required for accurate recognition. Tesseract is an engine for OCR technology and it is free software developed and distributed by Google[4].

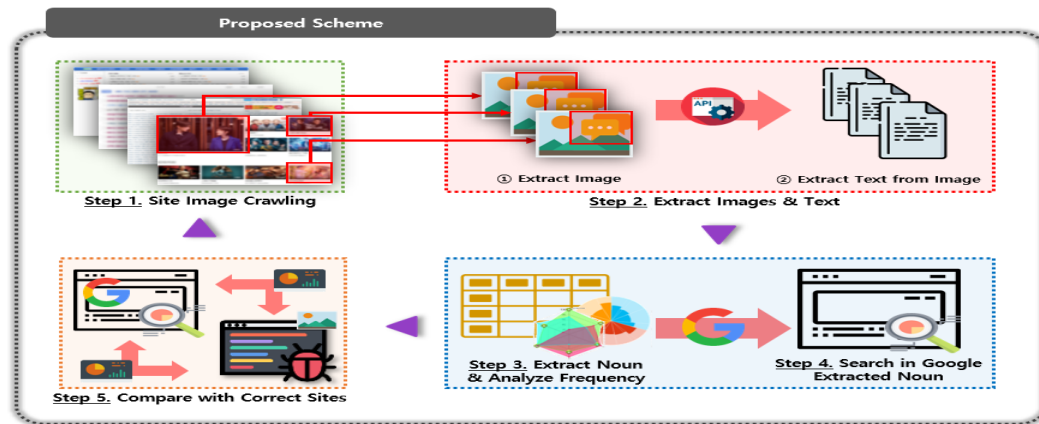


Fig. 1. Webtoon Infringement Site Using OCR detection method

3. Proposed Scheme

In this Section, the detection techniques for webtoon-infringement site detection are as follows. The proposed scheme is based on the sequence diagram shown in Fig. 1.

Step 1. Site Image Crawling

Extract webtoon images through web site crawling.

Step 2. Apply extracted image Tesseract

Extract the text using Tesseract on the extracted image. In this case, To compensate for the low text recognition rate of the Tesseract technology, text recognition is improved through in-image text contour extraction using OpenCV[5].

Step 3. Text-based noun extraction and frequency analysis

Extract nouns based on extracted text, and then proceed with frequency analysis. In this Step, Uses KoNLPy API, a Korean morphology analyzer, for extracting Korean nouns and analyzing frequency[6].

Step 4. Identify the official webtoon site based on high-frequency nouns

Identify the official webtoon site of a webtoon by using a search engine with a high-frequency.

Step 5. Compare to Official webtoon sites

Compare the Web site used in Step 1. and the official webtoon site. If it is an official webtoon site, apply the new site to Step1.~4. If it is not an official webtoon site, it is determined to be a webtoon infringement site.

4. Conclusions

In this paper, we presented a method for detecting webtoons encroaching sites through in-image text recognition based on OCR technology. In order to develop this study, we will increase the recognition rate through machine learning-based handwriting recognition and image learning.

References

- [1] Komacon, "Southeast Asian Comics, Webtoons Market Issues and Market Trends", *Overseas Cartoon Industry Report*, Dec. 2018.
- [2] Seok-Hwan Park, "2018 Comic and Webtoon Industry Issues in Number", *Komacon*, Jan. 2019
- [3] "Optical Character Recognition", https://en.wikipedia.org/wiki/Optical_character_recognition
- [4] "Tesseract OCR", <https://opensource.google.com/projects/tesseract>.
- [5] "OpenCV API Reference", <https://docs.opencv.org/2.4/modules/refman.html>
- [6] "KoNLPy: Korean NLP in Python", <http://konlpy.org/en/latest/>