

<http://dx.doi.org/10.17703/JCCT.2023.9.1.565>

JCCT 2023-1-68

텍스트 인식 기법에 기반한 화장품 성분 자동 분석 시스템

An Automatic Cosmetic Ingredient Analysis System based on Text Recognition Techniques

김예원*, 홍선미**, 엄성용***

Ye-Won Kim*, Sun-Mi Hong**, Seong-Yong Ohm***

요약 임신부나 피부질환자 등 화장품 성분에 예민한 사람들이 있다. 또 화장품으로 인한 부작용을 경험하는 사람들이 있다. 이를 피하기 위해, 쇼핑 시에 일일이 화장품에 유해 성분이 있는지 검색하는 것은 번거롭다. 또한 본인에게 잘 맞는 기능성 성분을 알고 기억하는 것은 새로운 화장품을 구매할 때 도움이 된다. 사진 촬영을 통해 현장에서 즉시 화장품 성분을 알 수 있는 시스템이 필요하다. 본 논문은 화장품에 표기된 성분을 촬영해 즉각적으로 화장품 성분을 알 수 있는 스마트폰용 애플리케이션 <화안>을 소개한다. 본 시스템은 효과적인 텍스트 인식 기법을 적용하여, 카메라를 화장품 성분에 비추거나 앨범에서 화장품 성분 사진을 불러올 경우, 해당 화장품의 성분을 자동 인식 및 자동 분류하여 그 화장품의 성분을 현장에서 즉시 제공한다는 점에서 기존의 시스템에 비해 효과적이고 편리하다. 이 시스템이 널리 활용된다면, 일상생활 속 화장품으로 인한 피부 질환을 예방하고 본인에게 맞지 않는 화장품 구매를 줄일 수 있을 것으로 기대된다.

주요어 : 텍스트 인식, 화장품 유해 성분, ML kit, 스마트폰 애플리케이션

Abstract There are people who are sensitive to cosmetic ingredients, such as pregnant women and skin disease patients. There are also people who experience side effects from cosmetics. To avoid this, it is cumbersome to search for harmful ingredients in cosmetics one by one when shopping. In addition, knowing and remembering functional ingredients that suit you is helpful when purchasing new cosmetics. There is a need for a system that allows you to immediately know the cosmetics ingredients in the field through photography. In this paper, we introduce an application for smartphones, <Hwa Ahn>, which allows you to immediately know the cosmetics ingredients by photographing the ingredients displayed in the cosmetics. This system is more effective and convenient than the existing system in that it automatically recognizes and automatically classifies the ingredients of the cosmetic when the camera is illuminated on the cosmetic ingredients or retrieves the photos of the cosmetic ingredients from the album. If the system is widely used, it is expected that it will prevent skin diseases caused by cosmetics in daily life and reduce purchases of cosmetics that are not suitable for you.

Key words : Text Recognition, Harmful Ingredients in Cosmetics, ML Kit, Smartphone Applications

*준회원, 서울여자대학교 소프트웨어융합학과 학사과정
(제1저자)

**준회원, 서울여자대학교 소프트웨어융합학과 학사과정
(참여저자)

***정회원, 서울여자대학교 소프트웨어융합학과 교수 (교신저자)

접수일: 2022년 11월 30일, 수정완료일: 2023년 1월 5일

게재확정일: 2023년 1월 9일

Received: November 30, 2022 / Revised: January 5, 2023

Accepted: January 9, 2023

***Corresponding Author: osy@swu.ac.kr

Dept. of Software Convergence, Seoul Women's Univ, Korea

I. 서 론

국내 화장품 트렌드로 기초화장품에 대한 관심이 두드러진다. 코로나19의 영향으로 기초화장품에 대한 관심도가 이전보다 증가하였고, 특히 천연 화장품에 대한 관심도가 높아지고 있다[1]. 반면, 국내 화장품 시장의 규모가 점점 커지면서 화장품 안전성이 사회적 문제로 대두되고 있지만, <화장품전성분표시제>가 처음 시행된 2008년 이후 현재까지도 소비자의 24.6%만이 <화장품전성분표시제>에 대해 인지하고 있는 실정이다[2, 3]. 또한 성분에 관해 관심이 있더라도 화장품 구매 시에는 실제로 성분을 고려하지 않거나, 표시된 성분을 정확히 구별하지 못해서 성분 확인을 아예 하지 않는 것으로 나타났다. 소비자들은 화장품 성분에 대한 홍보나 교육에 필요성을 느끼고 있으며, 특히 유해 성분에 대한 홍보 또는 교육의 필요성을 매우 높게 인지하고 있다[2-4].

이러한 문제를 해결하기 위해서는 기업마다 화장품 전성분 표기에 신경을 쓰고 제대로 분류하는 것이 가장 좋은 대안이 될 수 있으나, 모든 기업이 이를 수행하지 않는 이상 현실적으로 매우 어려운 상황이다. 따라서 현실점에서 화장품 유해 성분을 파악하는 방법은 소비자가 일일이 화장품을 검색하는 것이 차선의 방법이라고 볼 수 있다[4].

현재 많은 소비자들이 화장품으로 인한 부작용을 경험하고 있으며[5], 화장품에 배합된 유화제, 방부제, 색소, 향료 및 다양한 기능성 성분 등 피부에 부작용을 일으킬 가능성이 있는 성분에 대한 안전성 관리는 매우 중요하다[6].

따라서 화장품 성분을 공유하고 안심하고 사용할 수 있는 환경을 마련하는 것이 필요하다. 이에 본 논문에서는 화장품 유해 성분 즉시 분석을 지원하는 스마트폰용 애플리케이션 <화안> 시스템을 소개한다.

이 시스템은 텍스트 인식 기술을 사용하여 화장품 성분들을 정확하면서도 체계적으로 분류해줌으로써 화장품을 안전하게 사용할 수 있도록 도움을 준다.

본 논문의 구성은 다음과 같다. 2장에서는 화장품 성분과 관련된 기존의 시스템들의 주요 특징과 장단점에 대해 살펴본다. 3장에서는 본 논문에서 제안하는 시스템의 주요 기능 및 디자인, 그리고 내부적으로 사용된 텍스트 인식 모델에 대해 설명하고, 4장에서는 시스템

개발 및 구현 환경에 대해 설명한다. 5장에서는 결론 및 향후 연구 방향을 제시한다.

II. 관련 연구

1. 화장품과 관련된 기존 애플리케이션

글로벌 모바일 시장에서 미용 관련 제품은 인기 쇼핑 품목으로 자리 잡고 있으며, 화장품을 주제로 한 모바일 애플리케이션이 다양하게 개발되고 있다. 이러한 시스템들은 기초부터 색조까지 모든 종류를 포함한 화장품의 특성 및 구성 성분, 사용 후기, 구매까지의 모든 정보를 제공하는 특징이 있다[7].

화장품 성분을 안내하는 대표적인 기존 모바일 애플리케이션으로는 '화해'와 '맘가이드' 그리고 '찍검'을 들 수 있다[8-10].

'화해' 시스템은 사용자의 피부에 따른 맞춤 리뷰를 제공하고, 제품별로 화장품 성분 정보를 찾을 수 있는 애플리케이션으로, 다양한 카테고리별 순위를 알 수 있다. 그러나 화장품의 성분은 검색을 거쳐야 확인할 수 있으며, 등록되지 않은 화장품에 대해서는 담당자의 화장품 등록을 기다려야 하는 단점이 있다[8].

'맘가이드' 시스템은 유아 및 임산부 화장품에 대한 성분을 분석하고 등급과 순위를 확인할 수 있지만, 기초화장품에 대해서만 정보를 제공하고 색조화장품에 대한 성분은 거의 제공하지 않는다는 한계가 있다[9].

한편, '찍검' 시스템은 화장품의 상품명을 카메라로 찍고 검색하면 화장품 성분 및 주의 성분을 확인할 수 있고, 가격 비교 기능 등을 제공한다. 하지만 앨범에서 불러오는 기능을 지원하지 않고, 스토어에 없는 제품의 성분 또한 지원하지 않는다는 단점이 있다. 해당 시스템은 2016년 12월 21일에 출시되었지만, 2022년 11월 30일에 서비스가 중단된다[10].

반면, 본 논문에서 제안한 시스템은 사용자가 스마트폰 카메라로 직접 화장품을 촬영하거나, 앨범에서 기존에 찍어둔 사진을 불러와 그 화장품에 해당하는 성분 정보를 바로 확인할 수 있다. 따라서 화장품의 정확한 명칭을 모르더라도 사용자가 원하는 화장품 성분 정보를 즉시 파악할 수 있다. 또한, 화장품 정보를 다시 확인하고 싶은 경우에도 메인 화면에서 리스트를 통해 성분 명을 다시 확인할 수 있도록 하여 사용자의 편의성을 높였다.

2. 이론적 배경

(1) 머신 러닝 기술

머신 러닝(machine learning) 기술은 인공지능의 한 분야로 컴퓨터가 데이터로 학습하고 경험을 통해 개선하도록 훈련하는 데 중점을 둔다[11].

머신 러닝은 다양한 알고리즘 기법을 적용하는 여러 유형의 머신 러닝 모델로 구성된다. 데이터의 특성과 원하는 결과에 따라, 지도, 비지도, 준지도, 강화 등 네 가지 학습 모델 중 하나를 적용할 수 있다. 사용 중인 데이터 세트와 원하는 결과에 따라 각 모델 내에서 하나 이상의 알고리즘 기법을 적용할 수 있다. 머신 러닝 기술은 기본적으로 사물 분류, 패턴 발견, 결과 예측, 정보 기반 의사결정 등을 수행하도록 설계된다. 알고리즘 기법은 하나씩 사용할 수도 있고 복잡하고 보다 예측 불가능한 데이터가 포함된 경우에는 정확도를 극대화하기 위해 여러 알고리즘을 결합할 수도 있다[12].

(2) OCR

OCR(광학문자인식, Optical Character Recognition)은 사람이 쓰거나 기계로 인쇄한 문자의 영상을 이미지 스캐너로 획득하여 기계가 읽을 수 있는 문자로 변환하는 것이다.

이미지 스캔으로 얻을 수 있는 문서의 활자 영상을 컴퓨터가 편집 가능한 문자코드 등의 형식으로 변환하는 소프트웨어로 인공지능이나 기계 시각(machine vision)의 연구 분야로 시작되었다[13].

OCR의 활용 가능성이 대두되면서 해당 분야의 성능 향상 연구 및 프로그램 개발이 활발히 이루어지고 있다. 그 중 Tesseract OCR은 구글사의 OCR 프로그램으로 현재 가장 많이 사용되고 있다. Tesseract는 다양한 언어의 텍스트 변환을 지원하고 있다. 하지만 모든 언어에서 동일한 변환 일치율을 보이지는 않는다. 라틴문자와 같이 한 음운이 한 글자인 언어는 문자인식이 쉽지만, 그에 비해 한글은 음운 하나가 자음과 모음이 합쳐져 음운의 개수에 비해 글자의 경우의 수가 매우 많다. 따라서 한글의 이런 복잡한 구조 때문에 라틴계열의 언어보다 많은 학습 데이터가 필요해 문자인식의 정확도가 떨어진다[14].

III. 시스템 구성 및 주요 기능

1. 시스템의 주요 기능

본 논문에서 개발한 시스템은 두 가지 방식으로 화장품의 성분을 확인할 수 있다.

첫 번째 방식은 카메라 기능을 사용하는 방식이다. 이 방법에서는 분석하고자 하는 화장품 성분명을 카메라 화면에 비추고 찍으면 사진이 앱 상에서 불러와진다. 분석하기 버튼을 통해 텍스트 인식 기법이 적용되고, 인식 결과에 따라 해당하는 화장품 성분 정보가 분류되어 화면에 출력된다. 이 방식은 쇼핑을 하면서 화장품에 대해 잘 모르는 경우에 활용 가능한 장점이 있다.

두 번째 방식은 앨범에서 가져오는 방식이다. 이미 가지고 있는 화장품 사진을 가져와 분석하기 버튼을 통해 텍스트 인식 기법이 적용되고, 인식 결과에 따라 해당하는 화장품 성분 정보가 분류되어 화면에 출력된다.

분석을 완료한 후에는 저장 기능을 통해 메인 화면에서 리스트 형태로 다시 확인할 수 있다. 기본적으로 화장품 성분 분석은 한글, 영어 유해 성분과 기능 성분을 인식하여 지원한다.

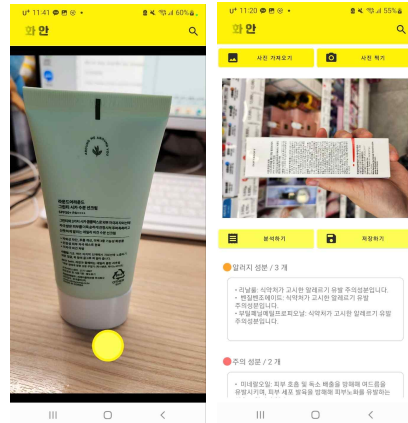


그림 1. 화장품 분석 페이지
Figure 1. Cosmetics Analysis Page

그림 1은 각각 시스템 내부 카메라와 앨범에서 가져온 사진을 통해 화장품 성분이 분석된 화면이다.

2. 텍스트 인식 기능

(1) 적용 모델 소개

본 시스템에서는 정확한 텍스트 인식을 위해 ML(machine learning) kit를 사용하였다. ML Kit는 강력하고

사용하기 쉬운 패키지로 모바일 개발자에게 Google의 머신 러닝 전문 지식을 제공한다. 기기에서 실행하는 데 최적화된 솔루션을 통해 iOS 및 Android 앱을 더 몰입도 높고 유용하게 만들 수 있다[15].

ML Kit는 다양한 기능을 제공하는데 본 논문에서는 Vision API를 사용한다. Vision API는 동영상 및 이미지 분석 API로 이미지에 라벨을 지정하고 바코드, 텍스트, 얼굴, 객체를 감지하는 기능을 제공한다. 이 기술을 통해 애플리케이션에서 머신 러닝 기술을 빠르고 효과적으로 구현할 수 있다[15].

(2) 텍스트 인식 모델

ML kit의 텍스트 인식 API(Text Recognition API)는 라틴어 기반 문자 집합의 텍스트를 인식할 수 있다. 또한, 신용카드, 영수증 및 명함 처리와 같은 데이터 입력 작업을 자동화하는 데 사용될 수도 있다.

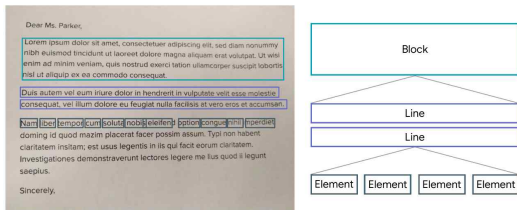


그림 2. 텍스트 구조 분류
Figure 2. Text Structure Classification

텍스트 구조 분석에서는 단어/요소, 줄, 단락 감지를 지원하며 텍스트 언어 식별을 통해 인식된 텍스트의 언어를 확인할 수 있다. 그림 2는 ML kit에서 텍스트 인식 방식을 보여준다. 감지된 모든 블록, 선, 요소, 기호와 관련하여 API는 경계 상자, 모서리 지점, 회전 정보, 신뢰도 점수, 인식된 언어, 인식된 텍스트를 반환한다.

본 시스템에서는 다양한 기기에서 실시간으로 텍스트 인식을 하기 위해서 ML kit의 텍스트 인식 v2 모델을 사용했다. 해당 모델을 통해서 영어와 한국어 스크립트를 인식할 수 있도록 하였고 텍스트 인식 정확도를 높였다[16].

그림 3은 영문 표기 화장품 성분에 대해서 분석하여 분류하는 것을 보여준다. 그림 4는 한글 표기 화장품에 대해서 분석한 것이다. 사각형 모양의 화장품 뿐 아니라, 둥근 모양의 화장품과 빛 반사가 있는 화장품에 대해서도 성분 인식이 되는 것을 확인할 수 있다.

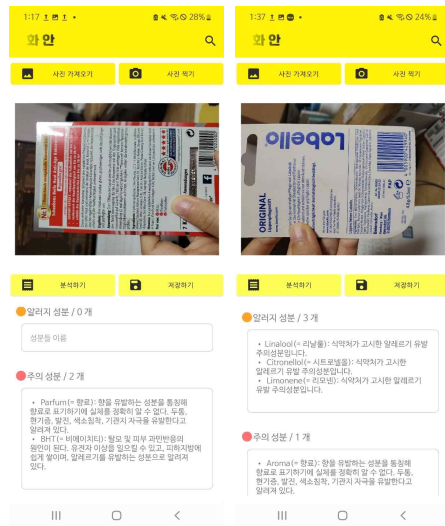


그림 3. 영문 표기 화장품 분석
Figure 3. Analysis of English cosmetics

3. 디자인 색상

본 시스템에서는 사용자의 직관적인 성분 이해를 돕기 위해 분류별 색상 코드를 적용하였다.

먼저, 전체적인 시스템 색상은 밝고 행복함을 상징하는 노란색으로 지정하였다[17]. 또한, 사용의 편리성 및 효율성을 위해 화장품 성분 종류에 따라 각각 다른 색상을 지정하여 사용자가 직관적으로 파악할 수 있도록 하였다.

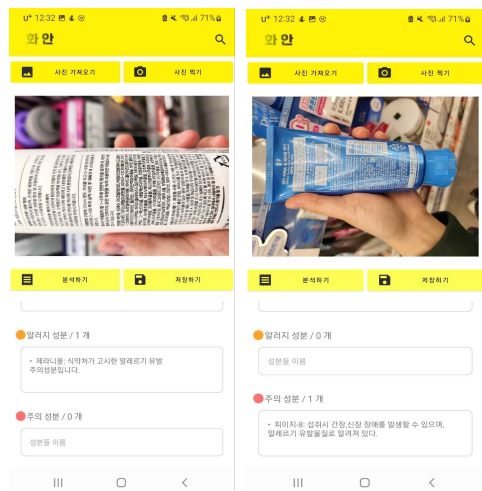


그림 4. 둥근 모양 및 빛 반사가 있는 경우의 분석 예
Figure 4. Analysis of round or light-reflected cosmetics

IV. 시스템 구현

본 시스템은 안드로이드 애플리케이션 개발을 위한 공식 개발 환경(IDE)인 안드로이드 스튜디오를 사용하여 구현하였다. 안드로이드 스튜디오는 IntelliJ IDEA를 기반으로 하여, IntelliJ의 강력한 코드 편집기와 개발자 도구를 사용할 수 있다는 장점이 있다[18]. 또한 안드로이드 스튜디오가 지원하는 언어 중 다양한 참고 자료를 활용하기 위해 JAVA를 선택하였다.

데이터베이스로는 SQLite를 사용하여 사진과 데이터베이스 내의 화장품 성분을 비교하였다. SQLite는 MySQL나 PostgreSQL과 같은 데이터베이스 관리 시스템이지만, 서버가 아니라 응용 프로그램에 넣어 사용하는 비교적 가벼운 데이터베이스이다[19].

시스템 내부의 카메라 기능은 CameraX 라이브러리를 사용한다. CameraX는 더 쉬운 카메라 앱 개발을 위해 빌드된 Jetpack 라이브러리이다. CameraX는 미리보기 기능을 통해 실시간으로 화면에서 이미지를 가져올 수 있으며, 이미지 분석과 캡처 기능을 지원해 고해상도 고화질 사진을 캡처할 수 있도록 지원한다[20]. 또한 한글로 표시된 화장품 성분의 텍스트의 인식률을 높이는 데 도움을 준다[21].

V. 결론 및 향후 연구 방향

본인에게 맞는 화장품을 사용하여 화장품 부작용을 예방하기 위해, 화장품 성분 분석을 지원하는 스마트폰용 애플리케이션 시스템을 개발하였다.

이 시스템에서는 사용자가 애플리케이션 내의 카메라로 직접 화장품을 촬영하거나 앨범에서 화장품 성분 사진을 가져와 그 화장품에 해당하는 성분 정보를 바로 확인할 수 있다. 따라서 화장품의 정확한 명칭을 몰라서 검색을 못하더라도 사용자가 원하는 성분 정보를 즉시 알 수 있다는 장점이 있다.

한글 표기 성분뿐만 아니라 영문 표기의 화장품 성분도 잘 분석되어 해외 화장품 사용으로 인한 부작용도 예방할 수 있다. 더욱이 한번 분석한 화장품 성분은 리스트 형태로 저장하여 메인 화면에서 확인할 수 있도록 지원하여 사용자의 편의성을 높였다.

하지만 화장품 성분 사진을 찍을 때 정확한 성분 인식에 한계가 있다. 화장품 자체가 심하게 등글거나

긴 모양인 경우 한 화면에 성분이 다 담기지 않아 분석하기가 다소 모호하다. 또한, 빛이 반사되어 텍스트가 보이지 않는 경우 역시 정확한 분석이 어려운 한계는 존재한다. 현재로서는 등근 화장품은 사진을 나누어 찍어 분석하고, 빛이 반사되는 경우에는 촬영 각도를 바꾸어 화장품 사진을 찍어야 한다.

향후, 카메라 텍스트 인식 기능이 더 발전한다면 시스템의 편의성과 정확성이 증가할 것이다. 이를 위한 추가 연구가 필요하다.

References

- [1] Hee Suk, Shin, Jae, "A Study on the Trends of Cosmetics through Big Data Analysis -Focusing on text mining and semantic network analysis-", earticle, Vol.66, No.8, pp.85-95, 2021. <https://mearticle.net/Article/A393050>
- [2] Song-I Park, Mee-Ok Choi, "Awareness on the Full Ingredient Labeling System of Cosmetics According to Experiences with Side Effects of Cosmetics", Asian Journal of Beauty & Cosmetology, Vol.13, No.3, pp.301-307, 2015. <https://www.e-ajbc.org/journal/view.php?number=838>
- [3] Su Mi Lee, "Consumer's recognition and actual usage of cosmetic ingredients and functional cosmetics", RISS, 2014. <http://www.riss.kr/link?id=T13419767&outLink=K>
- [4] Min-ah Jang, Jung Min Lee, Seunghee Bae, "A Study of Consumer Choice Behavior Determination Applying the Theory of Planned Behavior (TPB) of Functional Cosmetics", earticle, Vol.71, No.9, pp.95-110, 2022. <https://www.earticle.net/Article/A409445>
- [5] Kim, Taemi, "Purchasing and using behavior according to the experiences of side effects. The Research Journal of the Costume Culture", The Research Journal of the Costume Culture, Vol. 27, No.2, pp.169-180, 2019. <https://doi.org/10.29049/rjcc.2019.27.2.169>
- [6] Lee Eun-Joo, Li Shun-Hua, "Research on Cosmetic Quality Perception and Satisfaction According to Consumer Characteristics", Convergence Society for SMB, Vol.11, No.1, pp.216-224, 2021. <https://doi.org/10.22156/CS4SMB.2021.11.01.216>
- [7] Chun Eunha, Seok HaeMin, Chung Minjee, Ko Eunju, "A Study on the Usage and Preferences of cosmetics branded app and cosmetics app",

- KoreaScience, Vol.19, No.2, pp.180-193, 2017.
<https://doi.org/10.5805/SFTI.2017.19.2.180>
- [8] Goolge Play, Hwahae. <https://play.google.com/store/apps/details?id=kr.co.company.hwahae&hl=ko&gl=US&pli=1>
- [9] Goolge Play, Momguide. <https://play.google.com/store/apps/details?id=kr.co.pushapp.momguide&hl=ko&gl=US>
- [10] Goolge Play, Rocketview <https://play.google.com/store/apps/details?id=io.rocketview.lowcost&hl=ko&gl=US>
- [11] Wikipedia, Machine Learning. https://ko.wikipedia.org/wiki/Machine_Learning.
- [12] SAP, "What is machine learning?" <https://www.sap.com/korea/insights/what-is-machine-learning.html>
- [13] Wikipedia, Optical Character Recognition <https://ko.wikipedia.org/wiki/OCR>.
- [14] Ga-Hyeon Kang, Ji-Hyun Ko, Yong-Jun Kwon, Na-Young Kwon, Seok-Ju Koh, "A Study on Improvement of Korean OCR Accuracy Using Deep Learning", KIICE, Vol.22, No.208, pp.693-695, 2018.
- [15] Google, ML Kit. <https://developers.google.com/ml-kit>
- [16] Google, ML Kit Text Recognition <https://developers.google.com/ml-kit/vision/text-recognition>
- [17] Je Gi-Yeon, Lee Kyoung-Hee, "Color Preference and Color Meaning of University Students", KoreaScience, Vol.13, No.3, pp.346-352, 2011.
<https://doi.org/10.5805/KSCI.2011.13.3.346>
- [18] Android, Introducing Android Studio <https://developer.android.com/studio/intro?hl=ko>
- [19] Wikipedia, SQLite <https://ko.wikipedia.org/wiki/SQLite>
- [20] Android, CameraX Overview <https://developer.android.com/training/camerax>
- [21] Android, Image Capture <https://developer.android.com/training/camerax/take-photo?hl=ko>

※ 본 논문은 2022학년도 서울여자대학교 교내 연구비의 지원을 받았음(2022-0115).