

특집논문 (Special Paper)

방송공학회논문지 제27권 제4호, 2022년 7월 (JBE Vol.27, No.4, July 2022)

<https://doi.org/10.5909/JBE.2022.27.4.499>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

영상분석 기술을 활용한 시니어용 동영상 편집 시스템

장 달 원^{a)†}, 이 재 원^{a)}, 이 종 설^{a)}

Video Content Editing System for Senior Video Creator based on Video Analysis Techniques

Dalwon Jang^{a)†}, Jaewon Lee^{a)}, and JongSeol Lee^{a)}

요 약

본 논문에서는 영상 편집이 익숙하지 않은 시니어 동영상 크리에이터를 위한 동영상 편집 시스템을 설명한다. 영상분석 기술을 이용하여 편집소스 동영상을 분석하여 각종 정보를 제공하고, 자동으로 일부 장면을 삭제한다. 사용자가 다수의 소스 콘텐츠를 입력하였을 때, RNN(Recurrent Neural Network) 기술을 기반으로 샷 단위로 분할하고, 이 중 동영상 편집에서 배제할 부분을 구분한다. 각 샷 별로 중요도를 계산하여 샷 단위로 자동 삭제가 가능하도록 한다. 중요도 계산을 위해서 동영상 초점 정보를 추출하여 활용하는데, 이는 초점이 맞지 않는 영상 또는 흔들린 영상을 배제할 수 있도록 한다. 이후 시스템은 객체 인식을 수행하고, 얼굴이 나온 영상에 대해서 감정, 나이, 성별 등의 정보를 추출하여 사용자에게 제공한다. 사용자는 이런 정보를 활용하여 동영상을 제작한다. 동영상에 자막을 삽입하는 등 동영상을 꾸미기 위한 기능들도 포함되어 있으며, 이런 기능들을 활용할 시, 사용자의 과거 정보를 이용해서 선호 디자인을 쉽게 찾을 수 있도록 앞서 배치하고 있다. 시니어 동영상 크리에이터들이 본 시스템을 통해서 쉽고 빠르게 동영상 콘텐츠를 제작할 수 있다.

Abstract

This paper introduces a video editing system for senior creator who is not familiar to video editing. Based on video analysis techniques, it provide various information and delete unwanted shot. The system detects shot boundaries based on RNN(Recurrent Neural Network), and it determines the deletion of video shots. The shots can be deleted using shot-level significance, which is computed by detecting focused area. It is possible to delete unfocused shots or motion-blurred shots using the significance. The system detects object and face, and extract the information of emotion, age, and gender from face image. Users can create video contents using the information. Decorating tools are also prepared, and in the tools, the preferred design, which is determined from user history, places in the front of the design element list. With the video editing system, senior creators can make their own video contents easily and quickly.

Keyword : Video editing, Video creator, Video shot detection, Video analysis

I. 서론

2020년 기준 우리나라에서 수익을 창출하는 유튜브 채널이 529명당 1개 꼴로 집계되는 등, 국내에서 유튜브의 수는 날로 증가하고 있다. 또한, 최근에는 숏폼 동영상도 많이 제작되고 있다^[1]. 이런 흐름에 발맞추어 디지털 소외계층에서도 동영상 제작에 도전하는 모습이 많이 보이고 있다. 특히 스마트폰의 활용과 유튜브의 시청시간이 급격히 늘어나고 있는 노년층(시니어)에서도 이러한 시도들이 보이고 있다^[2]. 하지만, 동영상 제작에 필수적인 동영상 편집은 기본적인 인터페이스가 다루기 어렵고 가독성이 떨어지는 문제로 인해 반복학습을 하지 않는 이상 금방 잊게 되는 등, 시니어의 특성으로 인한 한계점들이 있다^[12]. 조작의 어려움도 있을 수 있는데, 기존 동영상 편집 구조인 트랙 구조에서 진행 바(Progress bar)를 살펴보고 범위를 지정하며 편집하는 것 또한 촉각 능력이 부족한 시니어 크리에이터에게는 진입장벽이 되고 있다^[12]. 시니어 크리에이터에게는 기존의 편집 소프트웨어와 다른 형태로 동영상 편집을 도와줄 수 있는 방법이 필요하다.

이 논문에서는 동영상 제작에 어려움이 있는 노년층을 대상으로 하는 동영상 편집 시스템의 개발을 다루고 있다. 노년층 중 동영상을 제작하는 시니어 크리에이터를 목적으로 만들었지만, 일반적인 초보 크리에이터들도 활용이 가능하다. 동영상 제작을 전문적으로 수행하기 위해서는 어도비(Adobe)사의 프리미어 프로(Premier Pro)나 애플(Apple)사의 파이널 컷 프로(Final cut pro)와 같은 전문적이고 강력한 기능을 제공하는 동영상 편집 소프트웨어가 주로 사용되고 있다^[13]. 하지만, 초보자가 동영상 제작을 위해 사용하기에는 소프트웨어 비용과 시스템 요구사항 등의 장벽이 존재한다. 쉽게 접근하고 적절한 수준의 동영상 편

집 기능을 빠르게 익혀서 시니어 크리에이터가 보다 쉽게 자신의 동영상을 만들 수 있도록 하는 것인 제안하는 동영상 편집 시스템의 목적이다.

제안하는 동영상 편집 시스템은 인공지능 기술을 이용하여 동영상을 분석하여 정보를 제공하는 등 편의기능을 제공하고 있다. 또한 시니어 크리에이터의 촬영 실수로 모션 블러(motion blur)가 강하게 발생한 영상 부분을 삭제할 수 있도록 도와준다. 일부분 반자동으로 동영상을 만들어 주고 사용자의 의지를 기반으로 반자동으로 만들어진 동영상을 편집하는 구조를 가지고 있다. 사용자 맞춤형으로 디자인 요소 삽입을 지원하는 등, 동영상 제작 과정에 있어서 불편한 요소를 제거하는데 초점을 맞추고 있다.

논문의 나머지 부분은 다음과 같이 채워진다. II장에서는 제안하는 시스템의 구조에 대한 설명을 보이고, III장에서는 시스템의 주요 모듈에 대해서 설명한다. IV장에서는 구현한 시스템의 동작을 보인다. V장에서 사용자가 평가한 결과를 보이고 VI장에서 논문의 결론을 맺는다.

II. 시스템의 구조

그림 1에 시스템의 구조가 간략히 그려져 있다. 사용자가 편집을 수행할 소스 동영상 콘텐츠를 업로드하면, 그것들이 저장되고, 소스 콘텐츠 분석(content analysis), 영상자르기(content cutting), 콘텐츠 꾸미기(content decoration tool) 과정을 통해서 사용자가 순차적으로 동영상 콘텐츠를 편집할 수 있도록 한다. 이중 소스 콘텐츠 분석과 영상자르기 과정은 서버에서 자동적으로 진행되는 자동편집 과정으로 사용자의 선택에 따라서 진행되지 않을 수 있다. 이 경우에는 전체 수동기능만을 사용한 편집과정이 진행된다. 그림 1에 회색 음영으로 나타난 영역이 자동편집 과정으로 콘텐츠를 분석한 결과를 사용자에게 제공한다. 소스 콘텐츠 분석 과정을 거치는 것은 후에 있을 영상자르기 과정을 원활히 진행하고 사용자에게 편집에 필요한 정보를 제공하기 위함이다. 사용자에게 대해서 소스 콘텐츠에서 적절한 내용 부분을 잘라내고 붙이는 과정에 어려움이 있다고 가정한다. 일반적인 동영상 편집기에서는 동영상내 영역을 선택하는 과정을 거쳐야 하지만, 시니어 크리에이터에게는

a) 한국전자기술연구원 정보미디어연구센터(Information & Media Research Center, KETI)

‡ Corresponding Author : 장달원(Dalwon Jang)

E-mail: dalwon@keti.re.kr

Tel: +82-2-6388-6697

ORCID: <https://orcid.org/0000-0002-8131-3978>

※ This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022 (Project Name: Development of intelligent authoring tools for senior creators, Project Number: R2020040068).

· Manuscript May 16, 2022; Revised July 4, 2022; Accepted July 5, 2022.

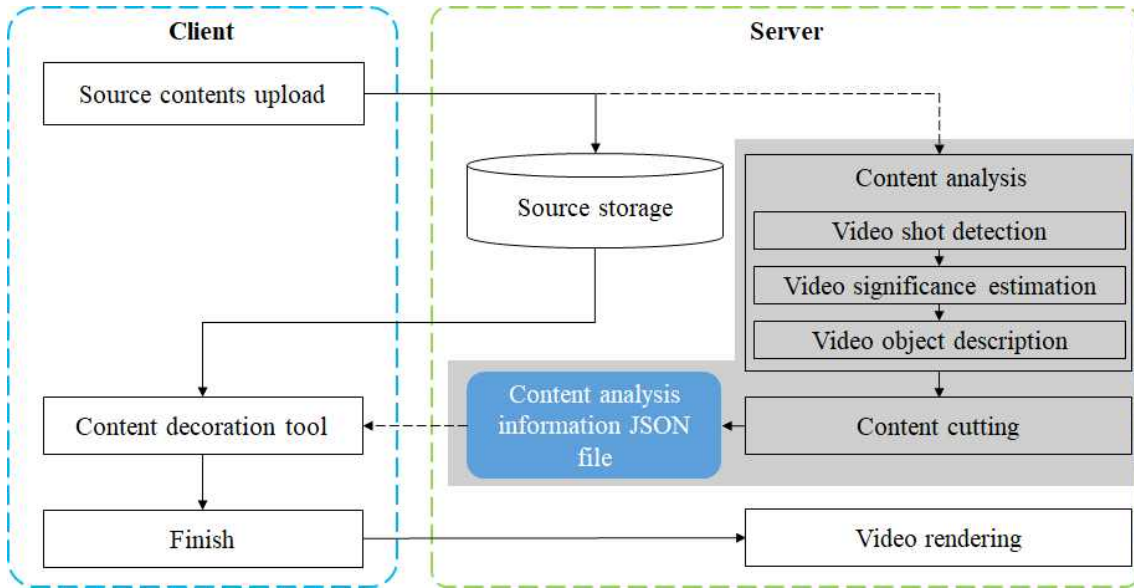


그림 1. 시스템의 구조
 Fig. 1. System structure

시각적으로 어려움이 있을 수 있고, 적절한 영역을 선택하는 인터페이스를 활용함에도 어려움이 있을 수 있다. 이런 손질 과정에 도움을 주는 정보를 제공하기 위해서 콘텐츠 분석 과정을 거친다. 영상 자르기 모듈은 자동으로 진행되어 불필요한 영상 부분을 제거하지만, 사용자가 판단하여 제거된 영상을 되돌릴 수 있다. 콘텐츠 꾸미기 과정은 사용자가 영상효과 적용 모듈과 디자인 요소 삽입 모듈이 포함된다. 아이콘 같은 디자인 요소를 첨가한다든지, 화면 효과를 적용하는 등의 작업이 포함된다. 사용자의 히스토리 기반으로 디자인 요소와 영상효과를 선택하고, 콘텐츠의 객체 정보에 따라서 디자인요소의 위치를 추천하도록 한다. 모든 과정이 마치면 사용자는 종료를 지시하고, 서버에서 소스 콘텐츠와 콘텐츠 분석 정보, 사용자의 편집 정보를 활용하여 비디오 렌더링을 수행한다.

사용자에게 소스 콘텐츠를 편집하기 위한 정보를 제공하는 소스 콘텐츠 분석 과정은 그림 1에 있듯이 크게 3가지 모듈로 구성되어 있다. 소스 콘텐츠 분할을 진행하고(Video shot detection), 분할된 샷 단위로 중요도를 측정하고(Video significance estimation), 콘텐츠에 대한 객체 서술을 진행(Video object description)한다. 위 세 가지 모듈을 포함하여 영상자르기, 콘텐츠 꾸미기 모듈의 자세한 기

능에 대하여 다음 장에서 설명한다.

III. 모듈별 기능

1. 소스 콘텐츠 분할 모듈

소스 콘텐츠에 대해 키프레임을 추출하거나, 샷 경계를 찾아 영상을 요약하는 등의 연구가 많이 이루어지고 있는데^[9-11], 소스 콘텐츠 분할 모듈에서는 이미지 프레임의 특징 추출 및 딥러닝 네트워크 사용을 위해 입력된 소스 콘텐츠를 먼저 짧은 샷 단위로 분할된다. 각각의 소스 콘텐츠는 한번의 촬영으로 만들어진 영상이라는 가정을 하고 있기 때문에, 하나의 샷만으로 이루어져 있을 수 있다. 하지만, 영상의 변화에 기반하여 짧은 단위의 샷으로 분할하는 것이 추후 편집에 도움이 되기 때문에 소스 콘텐츠를 분할하고, 분할된 영상을 기준으로 추후 작업이 이루어진다.

소스 콘텐츠 분할 모듈은 RNN(Recurrent Neural Network)을 기반으로 샷 경계점을 추출한다. 먼저 소스 콘텐츠의 특징 추출을 위해 프레임의 픽셀 평균값 및 표준편차를 특정 크기의 블록으로 나누어 계산하고, 이에 대한 특징 벡

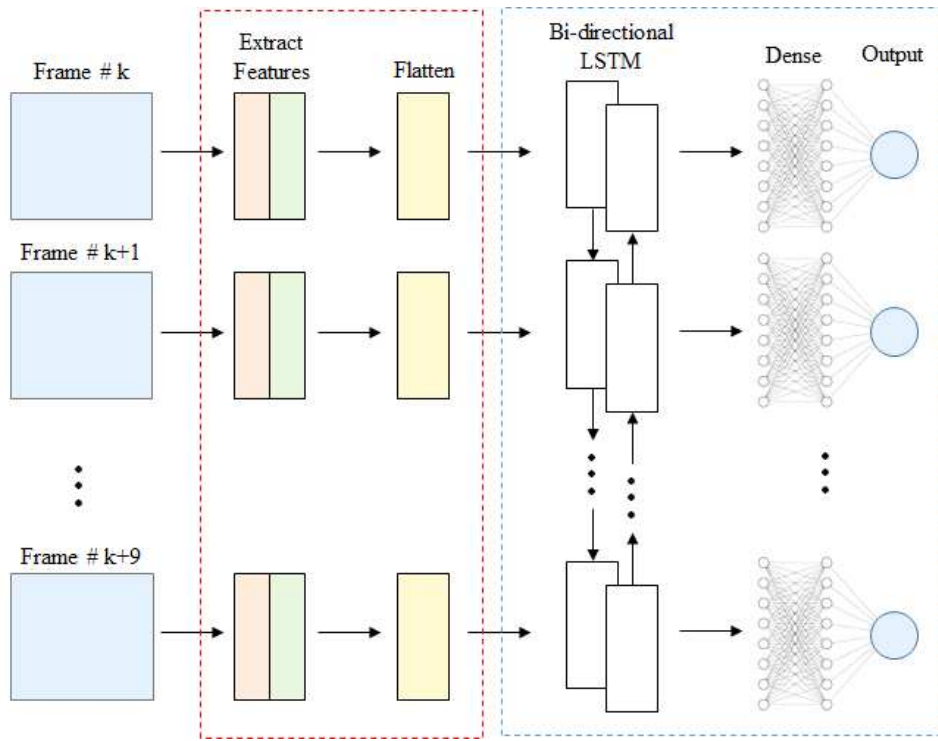


그림 2. 분할 지점을 찾기 위한 딥러닝 네트워크 구조
 Fig. 2. Deep learning network structure for detecting shot boundary

터를 생성한다. 다음 단계의 딥러닝 네트워크 입력으로 전달하기 위해 하나의 이미지 프레임에서 생성된 특징 벡터의 형태를 변환하고, 영상 장면분할을 위해서는 하나의 프레임 이미지를 가지고 분석하기 어렵기 때문에, 연속된 N개의 이미지를 입력으로 하여 분할 지점에 대한 영상 특징을 학습한다. 이를 위해 시계열 데이터 분석에 이용되는 딥러닝 네트워크인 RNN(Recurrent Neural Network)을 사용하며, 위에서 언급한 N개 이미지 프레임에서 추출된 특징 벡터가 RNN의 입력으로 사용된다. 전체 딥러닝 네트워크의 구조는 아래 그림 2와 같다.

특징 추출을 위해 이미지 프레임의 크기를 Width=192, Height=128, Channel=3으로 조정하였으며, RGB채널 이미지와 RGB를 HSV채널 도메인으로 변환한 192x128 크기의 이미지 프레임에 대해 16x16, 32x32, 64x64 크기의 블록으로 나누어, 각 블록에 대해 픽셀 평균값과 표준편차를 계산하여 벡터화하였다. RGB 및 HSV 이미지 프레임에 대해 채널별로 계산한 픽셀 평균값 및 표준편차를 10개의 이미

지 프레임 단위로 딥러닝 네트워크에 입력으로 넣으면, 프레임별 출력값들에 대한 결과가 최종적으로 하나의 값으로 출력되고, 각 프레임별 결과값이 큰 폭으로 변화하는 지점을 통해 영상 분할지점을 추출할 수 있게 된다.

소스 콘텐츠는 실제 물리적으로 분할되어 재가공되는 것이 아니고, 각 콘텐츠별로 분할 지점에 대한 정보를 소스 콘텐츠에 대한 정보를 저장하는 JSON 파일에 저장하여 편집 과정에서 관리하게 한다. 실제 물리적 분할은 최종적인 편집이 종료되어 렌더링이 진행될 시에 진행된다.

2. 소스 콘텐츠 중요도 측정 모듈

이 모듈은 촬영된 소스 콘텐츠의 초점이 맞지 않거나 블러가 심한 영상을 구분하여 제거하는 역할을 수행한다. 시니어 크리에이터를 가정하였기 때문에, 촬영된 영상의 흔들림이 존재할 수 있다. 그리고 촬영된 영상의 초점이 맞지 않을 수 있다. 이런 저품질의 영상들을 제거하기 위해서

Defocus blur detection (DBD)을 수행하고 이 결과를 정리하여 결과를 하나의 수치로 계산하는 과정을 수행한다⁵⁻⁸⁾. 그리하여 특정 값 이하의 영상들은 초점이 맞지 않거나 모션 블러가 심한 것으로 판단하여 편집 과정에서 제거하도록 한다. 앞서 소스 콘텐츠 분할 모듈에서 분할된 샷 단위를 기준으로 하나의 중요도 수치를 계산하고 이를 기준으로 영상의 삭제 여부를 판별한다. 중요도 측정 모듈에 입력된 하나의 샷에 대해서 아래 그림 3과 같은 과정을 거쳐서 중요도를 계산한다.

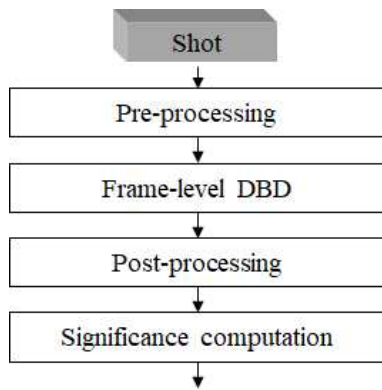


그림 3. 중요도 측정 모듈의 구조
Fig. 3. Structure of significance estimation

여기에서는 Local Binary Patterns (LBP) 알고리즘에 기반을 둔 방법을 통해서 초점 영역을 찾는다⁵⁾. 초점 영역을

찾으며 하나의 프레임에서 각 픽셀별로 값을 가질 수 있으며, 이것을 기준으로 프레임별로 초점영역이 잘 맞았는지를 살펴보고, 그 결과를 모아서 하나의 샷에 대한 결과를 얻는다. 그림 3에 설명된 과정을 하나씩 살펴보면 다음과 같다. 먼저, 전처리 과정으로 영상의 프레임을 추출하고 정해진 사이즈로 리사이즈를 한다. 연속되는 프레임에 중복되는 데이터가 많기 때문에 샘플링을 수행하여 프레임 레이트를 5 fps (frame per second) 이하로 낮추고 작업을 진행한다. 30fps/15fps 동영상 입력에 대해서는 5fps로 낮추고, 24fps 동영상에 대해서는 4fps로 프레임 레이트를 낮추어 작업한다. 전처리 과정 후, LBP 알고리즘을 활용하여 프레임별 초점 영역을 추출한다. 프레임별 초점 영역을 추출한 것에 대해서, 높은 수치로 나오는 값이 일부 영역에 모여 있다던지 이런 경우, 초점이 잘 맞은 영상이라고 볼 수 있다. 프레임 단위 처리를 위해서 후처리로 Gaussian 블러 처리를 하고, 프레임당 하나의 값을 뽑기 위해 Laplacian variance를 계산한다. 이 값을 중요도 기준으로 잡고, 각 샷당 평균치를 계산한다. 중요도 값이 낮은 샷은 초점영역이 잘 맞지 않은 곳으로 결정한다. 그림 4에 이 중요도 측정 과정에 대한 예시가 나타나있다. 그림 4에 나타난 왼쪽 두 그림은 초점이 잘 맞은 영상의 프레임이고, 3번째 그림은 모션 블러가 있는 경우이고 4번째 그림은 초점이 잘 맞지 않는 영상이다. 이 경우, 초점 영역을 추출하고 그것을 프레임당 하나의 숫자로 표현하면 아래 값과 같이 초점 영역이

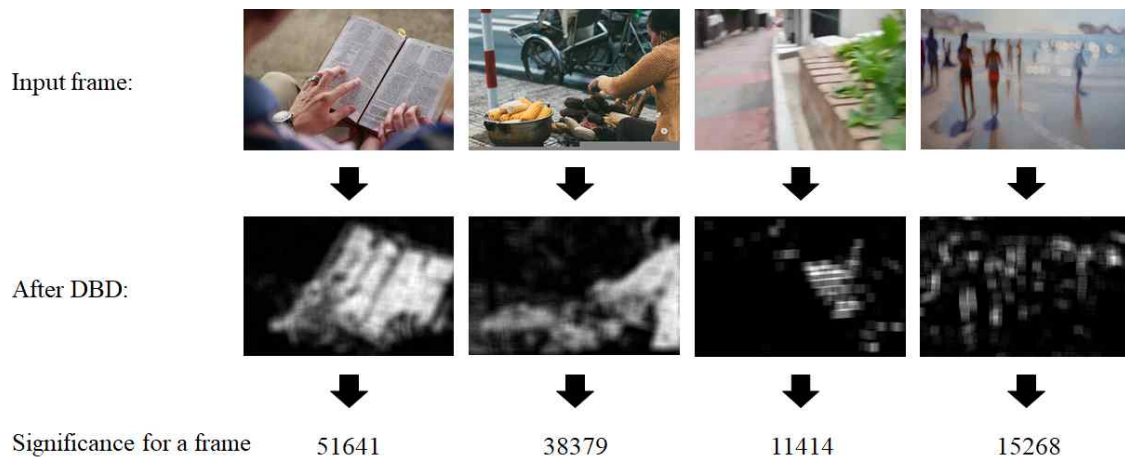


그림 4. 중요도 측정 과정 예시
Fig. 4. Examples of significance estimation

맞는 영상의 경우 높은 값이 나오고, 그렇지 않은 영상의 경우 낮은 값이 나온다.

다양한 DBD 알고리즘 중 LBP 기반의 알고리즘을 선택한 이유는 빠르게 결과를 얻을 수 있기 때문이다^{5,6}. 사용자가 입력하는 영상에 대해서 모든 과정을 거치려면 많은 시간이 걸리며 성능이 높은 작업보다는 빠르게 결과를 뽑을 수 있는 알고리즘을 선택해야 한다.

3. 소스 콘텐츠 객체 서술 모듈

소스 콘텐츠의 내용을 분석하여 사용자에게 정보를 전달하기 위해서 객체 인식 (object detection), 얼굴 검출 (face detection), 감정 인식(emotion recognition), 얼굴 군집화 (face clustering), 성별 및 나이 인식 (gender classification and facial age estimation) 과정을 진행한다. 그림 5와 같은 구조를 거치게 된다. 이 과정을 진행한 결과는 JSON 파일로 저장되어 시스템에서 활용될 수 있고, 소스 콘텐츠의 설명 요소로 사용자에게 전달된다. 그림 6(a)에 결과물 JSON 파일의 예시를 보여주고 있고, 그림 6(b)에 객체 서술 모듈의 결과물을 사용자가 확인하는 화면을 표시하였다. 그림 6(a)의 예시에서는 사람이 검출되어 그곳의 위치를 표시한 내용이다. 소스 콘텐츠 객체 서술 모듈을 통해서 생성된 정보는 사용자에게 각 샷 단위로 간략히 제공되어 편집에 도움이 되도록 한다. 기본적으로 진행되는 객체 인식 과정은 YOLOv5에 기반을 두고 있다³. 객체 인식을 통해서 사

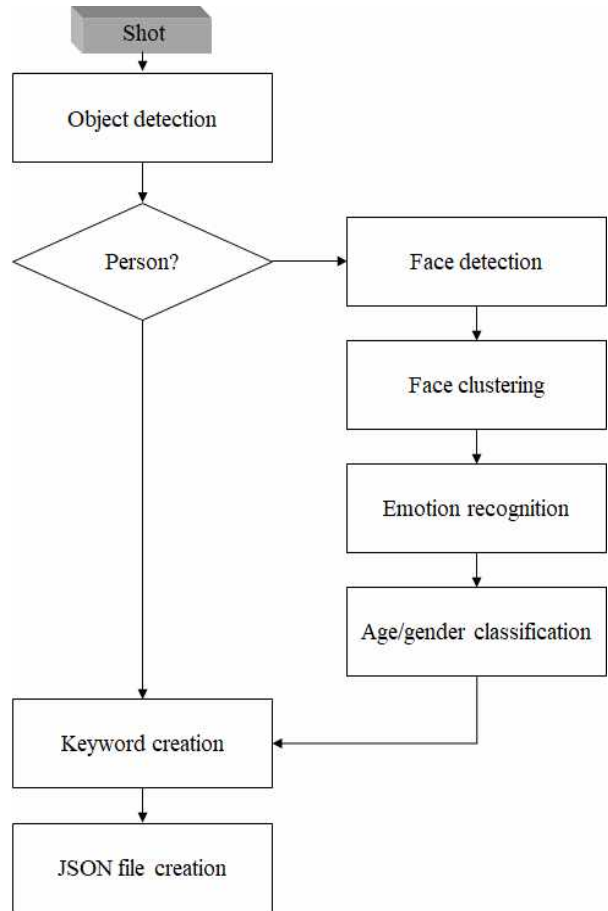
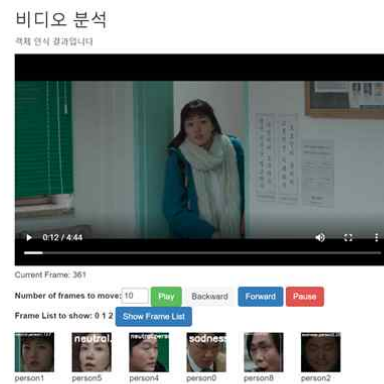


그림 5. 객체 서술 모듈의 구조
Fig. 5. Structure of object description module

```

11  "sourceType": "video",
12  "frames": [
13    {
14      "no": 1,
15      "objects": [
16        {
17          "name": "person",
18          "accuracy": 0.8044747114181519,
19          "xy": [
20            1,
21            524
22          ],
23          "wh": [
24            117,
25            483
26          ]
27        }
28      ]
29    }
30  ]
  
```

(a)



(b)

그림 6. 객체 서술 모듈의 결과물: (a)저장된 JSON 파일, (b) 결과물을 확인하는 화면 예시

Fig. 6. Outputs of object description module: (a) JSON file, (b) Visualization of object description for users

람의 존재를 인식할 수 있지만, YOLOv5는 뒷모습이나 하반신만으로도 사람으로 인식하고, 뒷모습이나 하반신보다는 얼굴에서 많은 정보가 나오기 때문에 얼굴 검출 과정을 별도로 수행해야 한다. 얼굴 검출을 통해서 얼굴의 위치를 파악하고 얼굴 근집화를 위해서 얼굴 이미지를 추출한다⁴⁾. 추출한 얼굴 이미지로 근집화를 수행하고, 각 얼굴 근집에 대하여 썸네일 형태의 대표 이미지를 따로 저장한다. 추출된 얼굴의 정보를 이용해서 감정, 나이, 성별 등의 정보를 추출한다. 이런 정보들은 시니어 크리에이터의 특성을 고려한 것으로 시니어 크리에이터는 가족과 주변 사람 동영상 또는 풍경 위주의 동영상 등이 많을 것으로 판단된다. 인물이 많이 나오는 동영상에서는 얼굴에서 추출한 정보를 많이 활용할 수 있고, 풍경 동영상에서는 반대로 얼굴이 나오지 않는다는 것을 확인할 수 있다.

4. 영상 자르기 모듈

소스 콘텐츠 중요도 측정 모듈에서 계산된 샷별로 정의된 콘텐츠 중요도에 따라서 일부 샷들을 제거하는 역할을 수행한다. 실제 콘텐츠 소스를 물리적으로 분할하는 것은

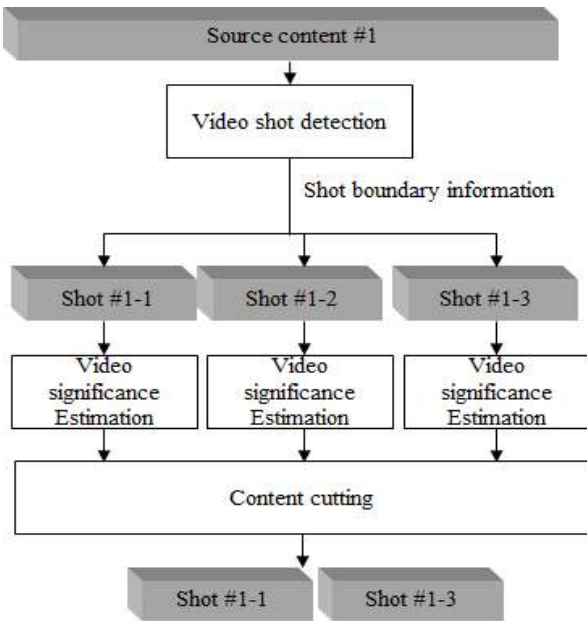


그림 7. 콘텐츠 입력부터 커팅까지 과정의 도식화
Fig. 7. Process from source content input to content cutting

아니기에 소스 콘텐츠의 정보를 저장하는 JSON 파일에 각 샷들의 활용 여부를 작성하여 저장한다. 아래 그림 7.은 소스 콘텐츠가 입력되어 영상 자르기 모듈까지의 과정을 도식화한 것이다. 소스 콘텐츠 분할 모듈을 통해서 소스 콘텐츠의 분할 지점 (shot boundary)를 프레임 단위로 추출하고, 이정보를 기반으로 하나의 소스 콘텐츠를 여러 개의 샷으로 분리한다. 예를 들어 85번째 프레임이 분할 지점이라면 0에서 84번 프레임은 Shot #1-1이 되고 이후 프레임들은 Shot #1-2가 된다. 각각의 샷에 대해서 중요도 측정을 수행한 후 이를 thresholding 하여 일부 샷을 제거하게 된다. 이 과정들이 실제 분할하여 저장되는 것이 아니라 소스 콘텐츠의 정보를 저장하는 JSON 파일에 분할 지점과 각 샷의 중요도, 그에 따른 활용 여부가 저장된다.

5. 콘텐츠 꾸미기

사용자들은 이모티콘 추가, 화면전환 효과 추가 등의 기능을 사용할 수 있다. 많은 수의 영상효과나 디자인 요소를 지원할 수 있고, 이것들은 추가적으로 계속 늘어날 수 있기 때문에, 사용자 편의를 위하여 추천 기능이 필요하다. 사용자들의 히스토리 기반으로 주로 사용했던 영상효과와 디자인 요소를 먼저 추천하는 형태를 가지고 있다. 또한, 디자인 요소를 추가할 시에, 기존의 객체영역을 피하는 기능을 제공한다. 예를 들어 스티커를 삽입하는 경우에, 스티커 사이즈와 스티커를 삽입하려는 샷에 존재하는 객체의 위치 정보를 고려하여 최초의 위치를 선택하여 사용자에게 보여준다. 그림 8에 진행과정이 나타나 있다. 디자인 요소를 추가한 후에 사용자가 위치 수정 등이 가능하지만, 최초 삽입 위치를 정할 시에 객체와 겹치지 않도록 하여 보다 빠른 편집이 가능하도록 한다. 제안하는 시스템에서는 아래의 표 1.과 같은 영상효과와 디자인 요소들을 활용하고 있다.

표 1. 영상효과와 디자인 요소 목록
Table 1. List of Effects and design elements

Effects for video	Color change
	Shot transition effects
	In-shot effects
Design elements	Audio embedding
	Text
	Sticker

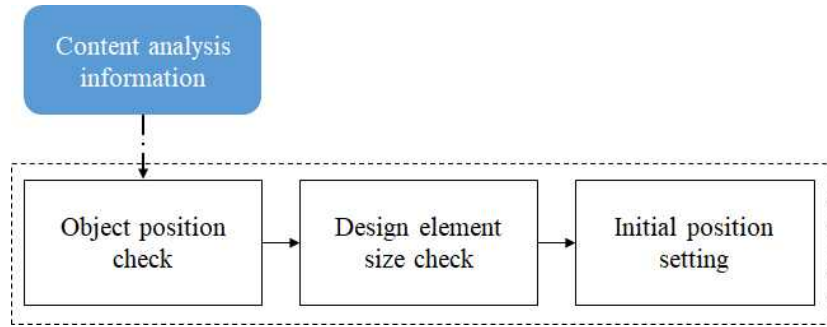


그림 8. 디자인 요소의 위치 지정
Fig. 8. Position setting for design elements

IV. 동영상 편집 시스템 구현

시스템은 현재 웹기반으로 만들어져 있고, 아래 그림 9 (a)와 같이 소스 콘텐츠를 업로드함으로써 동영상 제작을 시작할 수 있다. 소스 콘텐츠 선택이 완료되면 아래 그림 9(b)와 같이 자동편집을 실행할 것인지를 확인한다. 자동편집을 실행할 경우, 소스 콘텐츠 분석과 콘텐츠 커팅 과정을 자동으로 진행하고, 아닐 경우, 일반적인 수동 편집의 과정을 진행한다. 즉, 그림 1의 회색 음영 부분의 모듈을 진행하지 않는 것이다. 소스 콘텐츠 분석 과정을 거치지 않을 경우에는 업로드한 소스 콘텐츠 단위로 편집이 가능하고, 소스 콘텐츠 분석 과정을 거치면 콘텐츠 분할을 진행하기 때문에 샷 단위로 편집이 가능하도록 한다.

위 과정을 거치면, 샷 단위로 분할된 영상들이 아래의 편집 창에 생기고, 이 중 일부는 중요도 측정 모듈에서 삭제 결정되어 안 보이게 처리된다. 즉 사용자가 영상을 올린 후

부터 소스 콘텐츠 분석과 영상 자르기 모듈이 이미 작동하고, 사용자는 그 결과물을 보게 된다. 이 과정이 지난 후 사용자가 콘텐츠 꾸미기 과정을 진행할 수 있고, 편집 그 과정의 메인 화면은 아래 그림 10과 같이 구성된다. 편집의 과정을 정리하여 순서대로 진행할 수 있게 하여 시니어의 쉬운 사용을 유도하였으며, 소스 콘텐츠가 분석된 결과를 보면서 동영상 편집을 진행할 수 있다. 사용자는 왼쪽의 메뉴를 기준으로 동영상을 편집하는 작업을 수행하고, 모든 작업이 종료되었을 때 결과물을 발행할 수 있다.

왼쪽 메뉴 중 영상 편집은 샷 단위로 삭제 또는 프레임 변경 등을 지정할 수 있는 기능을 가지고 있고, 전환효과, 색상조절, 효과 등 영상효과와 오디오, 텍스트, 스티커 삽입 등 디자인요소 삽입을 진행할 수 있다. AI성우 기능은 TTS 기능을 접목하여 목소리를 삽입하는 기능이다.

영상효과나 디자인 요소를 선택함에 있어서는 가장 선호하는 요소 순으로 정렬하여 빠른 편집을 도와준다. 위 그림

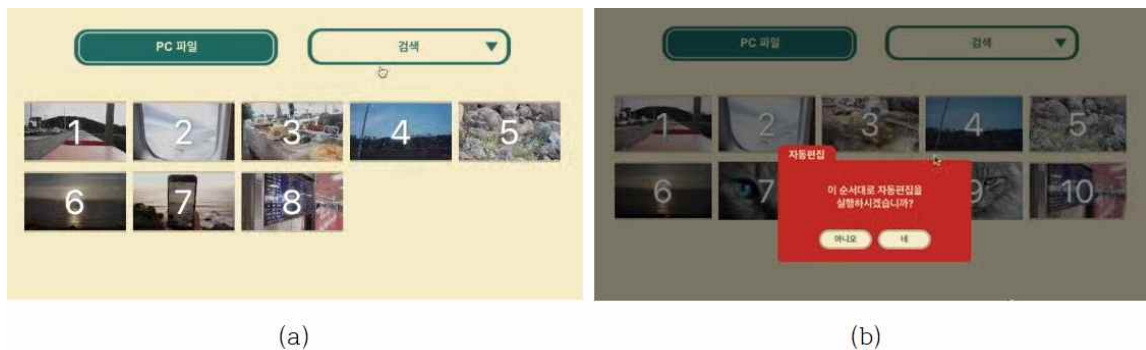


그림 9. 소스 콘텐츠 업로드 과정
Fig. 9. Source contents upload process

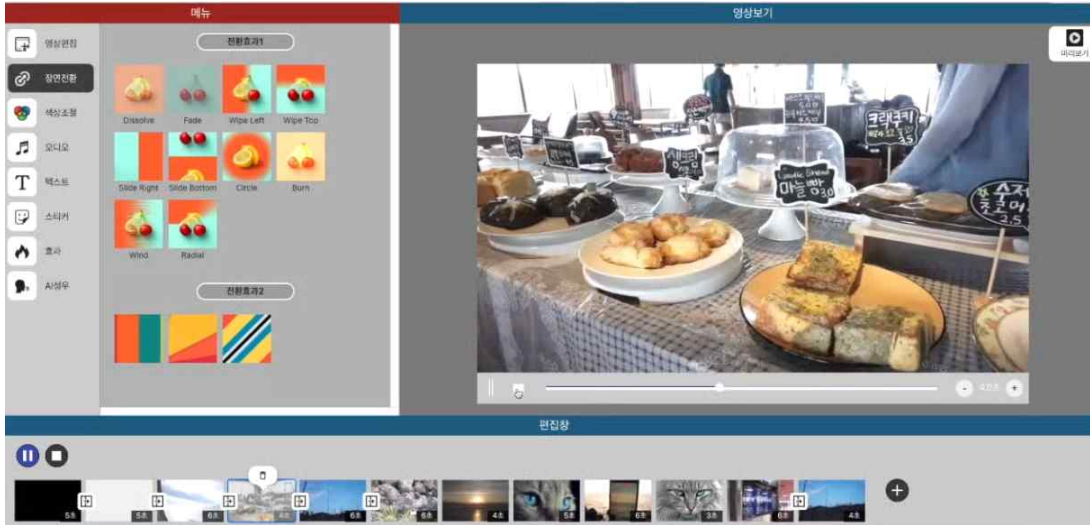


그림 10. 편집 과정 예시
 Fig. 10. Example: editing process

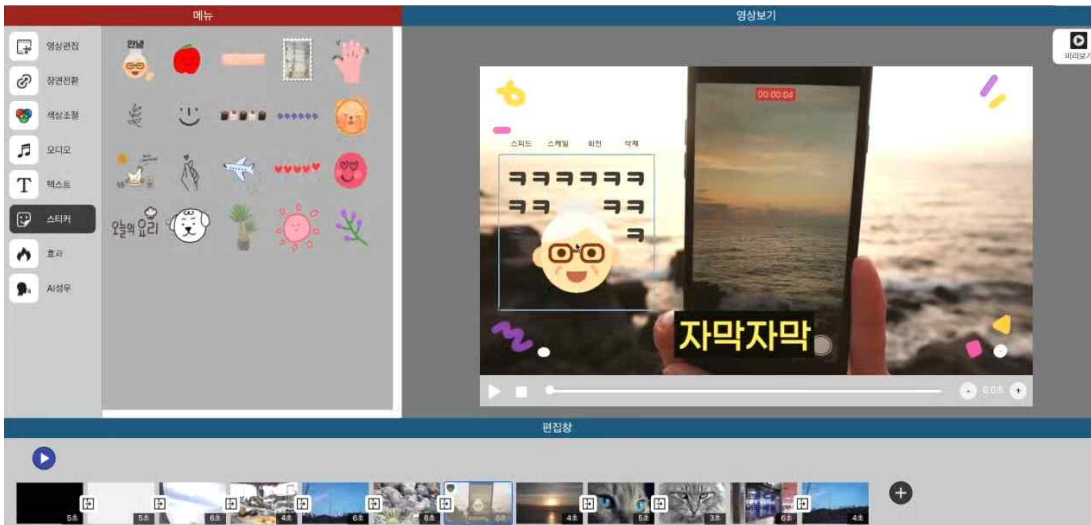


그림 11. 디자인 요소 (스티커) 삽입 예시
 Fig. 11. Example: insertion of design element (sticker)

에서는 전환효과를 선택함에 있어서 다양한 전환효과가 메뉴에 나오는 것을 볼 수 있다. 이 중 기존 사용자의 히스토리 정보를 통해서 효과들이 보이는 순서를 변경하여 보여 준다. 그림 11과 같이 사용하고자 하는 디자인 요소의 리스트를 사용자에게 보여줌에 있어서 사용자의 성향에 맞추어 디자인 요소를 순서대로 보여주게 한다.

V. 시스템 평가

개발된 결과물을 기반으로 만 50세 이상 시니어 중 동영상 편집도구 사용경험이 있는 33인을 선정하여 평가를 진행하였다. 평가참여자는 동영상 편집도구를 사용하는데 있어서 신체적, 인지적 문제가 없는 자로 선정하였다. 평가참

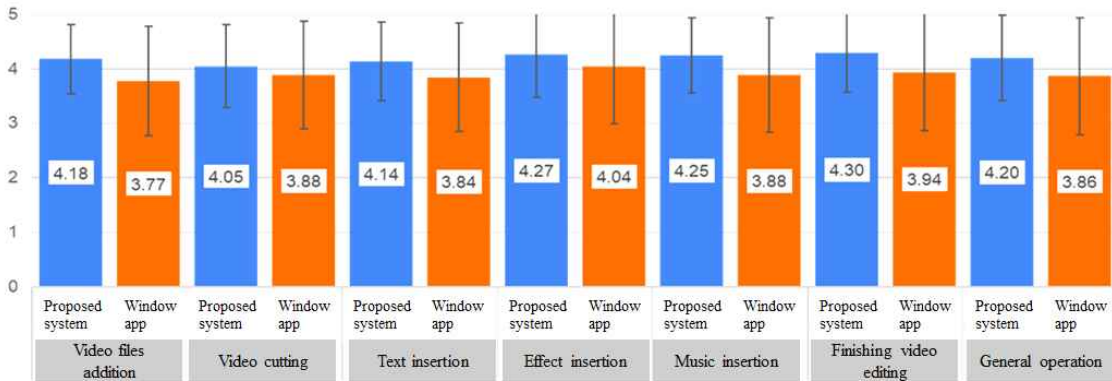


그림 12. 기능별 평가 결과
Fig. 12. Test results of operations

여자는 개발된 동영상 편집 시스템을 활용하여 영상 추가부터 영상 완료까지의 과정을 수행하며 비교 사용성 평가 결과를 도출하였다. 평가 절차 및 저작도구를 소개한 후 1시간 정도의 시간 동안 매뉴얼을 정독하고 프로그램을 사용하고 평가지를 작성하도록 하였다. 제안한 시스템은 윈도우 PC에서 사용 가능한 시스템이기에, 접근성이 좋은 윈도우 비디오 편집기 소프트웨어와 비교를 수행하였다. 영상추가(video file addition), 영상자르기(video cutting), 텍스트 삽입(text insertion), 효과 삽입(effect insertion), 음악 삽입(music insertion), 영상완성(finishing video editing), 일반사항(general operation)에 대해서 평가문항을 구성하고 이를 기반으로 평가를 수행한 결과가 그림 12에 나타나 있다. 전반적으로 평균점이 높은 점수를 받긴 하였지만, 윈도우 비디오 편집기에 비해 많은 항목에서 통계적으로 차이가 있을 만큼 큰 차이는 없었다. 이는 평가에 참여한 시니어들이 동영상 저작의 경험이 있으나, 다양한 기능을 활용해 보았던 것은 아니고, 그러다보니 비교대상으로 삼은 윈도우 비디오 편집기에 대해서도 높은 점수를 주었기 때문이라 판단할 수 있다. 또한 1시간 내에 동영상 편집 기능을 익히며 수행하다 보니 입력동영상들이 단순하고 주요 기능에 대한 활용도도 낮을 수밖에 없었기에 결과의 차이가 크지 않았다고 분석한다. 결과의 편차가 제안한 동영상 편집 시스템에서 훨씬 적게 나타나는데, 사용자들의 최저점이 상대적으로 높게 위치함을 알 수 있다. 평가를 진행한 시니어 사용자들의 주관적 의견으로는 편집단계에 대한 기능이

순서대로 나열되어 있고 사용에 있어 편의성이 좋게 고려되었다는 평가가 있었다.

VI. 결론

본 논문에서는 시니어 크리에이터가 보다 쉽게 동영상 콘텐츠를 만들어서 발행할 수 있도록 도와주는 시스템의 개발에 대해서 설명하였다. 입력된 소스 콘텐츠에 대해서 인공지능 기술을 기반으로 분석을 진행하여 필요없는 부분을 삭제할 수 있고 순서에 맞추어서 콘텐츠를 배치하고, 각종 영상효과를 사용자 히스토리 기반으로 만들어 준다. 일련의 과정에 대해서 수동으로 수정하는 작업을 진행할 수 있도록 정보를 표시하는 기능을 가지고 있다. 동영상 콘텐츠를 편집하는데 어려움이 있는 계층에서 본 시스템을 통해서 쉽고 빠르고 동영상 콘텐츠를 제작할 수 있다. 본 논문에서는 활용 가능한 RNN, YOLOv5, LBP 기반 DFD 기술 등 활용 가능한 인공지능 및 영상처리 기술들을 기반으로 동영상 편집 과정에 도움을 줄 수 있는 기능을 제공하는 동영상 편집 시스템을 제안하였다. 시스템을 구현하고, 실제 시니어 크리에이터를 목표로 하는 일반인을 대상으로 사용성 평가를 수행하였다. 기존의 기본적인 동영상 편집 소프트웨어보다 상대적으로 높은 평가를 받았지만, 큰 차이가 아니기에 시스템과 평가방법에 대해 개선의 필요성이 존재한다. 현재 관련 시스템은 PC에서만 활용 가능하도록

개발되었지만, 현재 많은 모바일 동영상 편집 소프트웨어와 경쟁할 수 있도록 모바일 버전으로 확장을 진행 중이다.

참 고 문 헌 (References)

- [1] J. Lee, "A Study on Types of Short-form Video Contents," *Humanities Contents*, Vol. 58, pp.121-139, 2020.
doi: <https://doi.org/10.18658/humancon.2020.09.121>
- [2] J.-H. Kwon, "A Study on the Planning of a Space for Senior Citizens Using Digital Contents," *Journal of Digital Convergence*, Vol. 18, No. 5, pp. 257-267, 2020.
doi: <https://doi.org/10.14400/JDC.2020.18.5.257>
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE International conference on computer vision and pattern recognition(CVPR)*, Las Vegas, NV, USA, pp. 779 - 788, 2016.
doi: <https://doi.org/10.48550/arXiv.1506.02640>
- [4] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*. Vol. 57, pp. 137 - 154, 2004
doi: <https://doi.org/10.1023/B:VISI.0000013087.49260.fb>
- [5] X. Yi and M. Eramian, "LBP-based segmentation of defocus blur," *IEEE Trans. Image Process*, Vol. 25, no. 4, pp. 1626 - 1638, Apr. 2016.
doi: <https://doi.org/10.1109/TIP.2016.2528042>
- [6] W. Zhao, F. Zhao, D. Wang, and H. Lu., "Defocus blur detection via multi-stream bottom-top-bottom fully convolutional network," *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, Salt Lake City, UT, USA, pp 3080 - 3088, 2018.
doi: <https://doi.org/10.1109/CVPR.2018.00325>
- [7] C. Tang, X. Zhu, X. Liu, L. Wang, and A. Zomaya, "DeFusionNET: Defocus blur detection via recurrently fusing and refining multi-scale deep features," *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, Long Beach, CA, USA, pp. 2700 - 2709, 2019.
doi: <https://doi.org/10.1109/CVPR.2019.00281>
- [8] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition(CVPR)*, Columbus, OH, USA, pp. 2965 - 2972, 2014.
doi: <https://doi.org/10.1109/CVPR.2014.379>
- [9] Y. Gao, Y. Lai and Y.Liu, "Fast Video Shot Boundary Detection Based on Visual Perception", *Proceeding of IEEE International Conference on Consumer Electronics(ICCE)*, Las Vegas, NV, USA, pp.1-4, 2019.
doi: <https://doi.org/10.1109/ICCE.2019.8662083>
- [10] M. Gygli, "Ridiculously fast shot boundary detection with fully convolutional neural networks," *Proceeding of International Conference on Content-Based Multimedia Indexing(CBMI)*, La Rochelle, France, Sep. 2018, pp. 1 - 4.
doi: <https://doi.org/10.1109/CBMI.2018.8516556>
- [11] M. Brindha and R. Amsaveni, "Shot change detection on news videos using color histogram and edge based approaches," *Proceeding of IEEE International Conference on Advances in Computer Applications(ICACA)*, Coimbatore, India, pp.50-54, 2016.
doi: <https://doi.org/10.1109/ICACA.2016.7887922>
- [12] G. Kwon and J. Kwon, "A Study on the Usability of Video Authoring Tool Application for Active Senior," *Journal of Next-generation Convergence Information Services Technology*, 9, no.4 (2020) : 351-361.
doi: <http://doi.org/10.29056/jncist.2020.12.03>
- [13] S.-D. Park, "Education of media by production of image contents - Focusing on Non-Linear Editing," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 23, No. 9, 1096-1103.
doi: <https://doi.org/10.6109/jkiice.2019.23.9.1096>

저 자 소 개



장 달 원

- 2002년 2월 : KAIST 전기 및 전자공학과 졸업(학사)
- 2003년 8월 : KAIST 전기 및 전자공학과 졸업(석사)
- 2010년 2월 : KAIST 전기 및 전자공학과 졸업(박사)
- 2010년 11월 ~ 현재 : 한국전자기술연구원 정보미디어연구센터 책임연구원
- ORCID : <https://orcid.org/0000-0002-8131-3978>
- 주관심분야 : 음향이벤트 인식, 멀티미디어 분석, 음악정보검색, 기계학습

저 자 소 개



이 재 원

- 2015년 2월 : 고려대학교 전기전자전파공학부 졸업(학사)
- 2017년 2월 : 고려대학교 전기전자공학과 졸업(석사)
- 2018년 4월 ~ 현재 : 한국전자기술연구원 선임연구원
- ORCID : <https://orcid.org/0000-0003-4557-1387>
- 주관심분야 : 영상처리, 인공지능, 컴퓨터비전



이 종 설

- 1996년 2월 : 충북대학교 정보통신공학 (공학사)
- 2001년 2월 : 충북대학교 정보통신공학 (공학석사)
- 2013년 2월 : 건국대학교 컴퓨터공학 (공학박사)
- 2000년 10월 ~ 현재 : 한국전자기술연구원 정보미디어연구센터 팀장
- ORCID : <https://orcid.org/0000-0002-4803-4280>
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 머신러닝, 분산처리