

# SAAnnot-C3Pap: Ground Truth Collection Technique of Playing Posture Using Semi Automatic Annotation Method

So-Hyun Park<sup>†</sup> · Seo-Yeon Kim<sup>††</sup> · Young-Ho Park<sup>†††</sup>

## ABSTRACT

In this paper, we propose SAAnnot-C3Pap, a semi-automatic annotation method for obtaining ground truth of a player's posture. In order to obtain ground truth about the two-dimensional joint position in the existing music domain, openpose, a two-dimensional posture estimation method, was used or manually labeled. However, automatic annotation methods such as the existing openpose have the disadvantages of showing inaccurate results even though they are fast. Therefore, this paper proposes SAAnnot-C3Pap, a semi-automated annotation method that is a compromise between the two. The proposed approach consists of three main steps: extracting postures using openpose, correcting the parts with errors among the extracted parts using supervisory, and then analyzing the results of openpose and supervisory. Perform the synchronization process. Through the proposed method, it was possible to correct the incorrect 2D joint position detection result that occurred in the openpose, solve the problem of detecting two or more people, and obtain the ground truth in the playing posture. In the experiment, we compare and analyze the results of the semi-automated annotation method openpose and the SAAnnot-C3Pap proposed in this paper. As a result of comparison, the proposed method showed improvement of posture information incorrectly collected through openpose.

Keywords : Playing Posture, Ground-Truth, Semi-Automatic Annotation Method, Artificial Intelligence, Computer Vision

## SAAnnot-C3Pap: 반자동 주석화 방법을 적용한 연주 자세의 그라운드 트루스 수집 기법

박 소 현<sup>†</sup> · 김 서 연<sup>††</sup> · 박 영 호<sup>†††</sup>

## 요 약

본 논문에서는 연주자 자세의 그라운드 트루스 획득을 위한 반자동 주석 방법인 SAAnnot-C3Pap를 제안한다. 기존 음악 도메인에서 2차원 관절 위치에 대한 그라운드 트루스를 획득하기 위하여 2차원 자세 추정 방법인 오픈포즈를 활용하거나 수작업으로 라벨링 하였다. 하지만 기존의 오픈포즈와 같은 자동 주석 방법은 빠르지만 부정확한 결과를 보인다는 단점이 있고, 사용자가 직접 주석을 생성하는 수작업 주석화의 경우 많은 노동력이 필요하다는 한계점이 있다. 따라서 본 논문에서는 그 둘의 절충 방안인 반자동 주석화 방법인 SAAnnot-C3Pap을 제안한다. 제안하는 SAAnnot-C3Pap은 크게 3가지 과정으로 오픈포즈를 사용하여 자세를 추출하고, 추출된 부분 중 오류가 있는 부분을 슈퍼바이즐리를 사용하여 수정한 뒤, 오픈포즈와 슈퍼바이즐리의 결과값을 동기화하는 과정을 수행한다. 제안하는 방법을 통하여 오픈포즈에서 발생하는 잘못된 2차원 관절 위치 검출 결과를 교정할 수 있었고, 2명 이상의 사람을 검출하는 문제를 해결하였으며, 연주 자세 그라운드 트루스 획득이 가능하였다. 실험에서는 반자동 주석 방법인 오픈포즈와 본 논문에서 제안하는 SAAnnot-C3Pap의 결과를 비교·분석한다. 비교 결과, 제안하는 SAAnnot-C3Pap은 오픈포즈로 잘못 수집된 자세 정보를 개선한 결과를 보였다.

키워드 : 연주 자세, 그라운드 트루스, 반자동 주석화 방법, 인공지능, 컴퓨터 비전

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1C1C2004282).

† 준 회원 : 숙명여자대학교 빅데이터 활용 연구센터 책임연구원

†† 비 회원 : 메가존클라우드 데이터팀 매니저

††† 종신회원 : 숙명여자대학교 IT공학과 교수

Manuscript Received : November 18, 2021

First Revision : May 23, 2022

Second Revision : June 14, 2022

Accepted : June 15, 2022

\* Corresponding Author : Young-Ho Park(yhpark@sm.ac.kr)

## 1. 서 론

최근 딥러닝 기반의 자세 추정 연구가 활발히 진행되고 있다. 또한, 자세 추정 훈련 모델을 위해 데이터를 라벨링 하는 데이터 주석화 연구가 이루어지고 있다. 운동, 발레 등 다양한 도메인에서 등에서 주석화 연구들은 많이 진행되었지만, 음악 도메인에서의 주석화 연구는 부족한 실정이다. 본 논문

에서는 음악 도메인의 연주 자세 추정 훈련을 위한 반자동 주석화 방법을 연구한다.

기존 음악 도메인에서 자세 추정 훈련을 위한 주석화 연구로는 자동 주석화 방법이 주로 제안되었다[1]. 자동 주석화 방법으로 2차원 관절 위치를 추출하는 오픈포즈(openpose)가 있다. 오픈포즈란 이미지나 영상에서 사람의 2차원 관절 위치를 추출하는 방법이다[2]. 오픈포즈는 2차원 관절 위치를 추출하기 위해 사용되기도 하지만 몇몇 연구들에서는 그라운드 트루스(ground-truth) 수집을 위해 사용된다[3]. 그라운드 트루스란 알고리즘의 예측 정확도를 평가하기 위해 예측 값과 비교되는 정답 값을 의미한다. 하지만 기존 오픈포즈를 이용한 자동 주석화 연구는 겹침, 가림 등의 문제로 잘못된 2차원 관절 위치를 추출하는 경우가 존재한다. 부정확한 데이터의 경우 해당 데이터는 그라운드 트루스로 사용하지 못한다. 잘못된 그라운드 트루스를 사용하면 연주 자세 생성, 자세 분석 등의 2차 분석을 진행할 때 정확한 결과를 얻기 어렵다. 예를 들어 [3]은 연주자 자세 예측 모델 구축 훈련 시 전체 연주자 자세 데이터에서 부정확한 부분을 제외하고 정확한 부분만 사용하였다. 부정확한 부분을 제외하고 모델을 훈련할 경우 데이터에 공백이 생기기 때문에 정확한 자세 분석 모델을 구축하는 데 어려움이 있다.

이와 같이 연주자 자세 정보를 활용한 2차 연구의 정확도 개선을 위한 연주 자세의 그라운드 트루스 수집 연구가 필요하다. 본 논문에서 제안하는 SAAnnot-C3Pap 방법을 위해 다음과 같은 연구를 수행한다. 첫 번째로 연주자 자세의 그라운드 트루스 획득을 위한 최초의 반자동 데이터 셋 수집 방안인 SAAnnot-C3Pap 제안한다. 제안하는 방법에서는 첫 번째 단계에서는 오픈포즈로 2차원 관절 위치를 추출한다. 두 번째 단계에서는 오픈포즈로 추출한 2차원 관절 위치 중 잘못 추출된 이미지에 대해서 슈퍼바이저리(supervisely) [4]로 다시 수작업 주석화를 진행한다. 세 번째 단계에서는 오픈포즈로 추출한 결과와 슈퍼바이저리로 다시 수작업 주석 작업을 한 결과를 동기화한다. 이를 위해 먼저 오픈포즈로 추출한 결과에서 2명 이상의 사람이 검출될 경우 연주자를 검출한다. 후에 오픈포즈 결과물과 슈퍼바이저리 결과물을 동기화한다.

또한 본 논문에서는 연주자 자세 분석, 연주자 자세 자동 생성 분야에 활용될 수 있는 연주 자세의 그라운드 트루스 제공한다. 연주자 자세 분석, 연주자 자세 자동 생성 등 다양한 음악 도메인 분야에서 연주자 자세 정보를 활용한다 [3]. 자동 주석화 방법인 오픈포즈로 추출한 자세는 부정확한 데이터를 포함하고 있으므로 정확한 연주자 자세 분석 모델이나 연주 자세 자동 생성 모델 구축에 어려움이 있다. 본 논문에서 제공하는 그라운드 트루스를 활용하면 보다 나은 성능의 모델을 구축에 활용될 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터 주석화와 관련된 연구와 슈퍼바이저리를 소개한다. 3장에서는 본 논문에서 제안하는 데이터 반자동 주석화 방법을 소개한다. 4장에서는 본 논문에서 제안하는 데이터 반자동 주석화 방법 적용 전후 결과를 보인다. 5장에서는 결론을 내리고 향후 연구를 소개한다.

## 2. 관련 연구

본 장에서는 데이터 주석화에 관련된 연구와 슈퍼바이저리를 소개한다. 데이터 주석화는 크게 3종류로 사용자가 직접 주석을 생성하는 수작업 주석화, 자동으로 주석을 생성한 뒤 사람이 다시 확인하여 수정하는 반자동 주석화, 주석을 완전히 자동으로 생성하는 자동 주석화가 있다.

### 2.1 수작업 주석화

본 절에서는 데이터 주석화 중 수작업 주석화와 관련된 연구를 소개한다. 최근 수작업 주석화와 관련된 많은 연구가 수행 중이다.

비디오 데이터 검색 연구에서는 특징 기반과 주석 기반 비디오 검색 기반 시스템을 혼합하여 사용한다[5]. 특징 기반 검색 수행 시 주석자는 키 프레임에 대해 요약한 내용을 수작업으로 입력한다.

협업적 주석 시스템에서는 다수의 사용자가 링크드 데이터를 이용하여 협업적으로 주석을 처리할 수 있는 시스템을 제안하였다[6]. 주석 처리 시에는 사용자가 직접 동영상에서 객체의 영역을 마우스로 설정하고, 객체에 대한 정보를 저장한다.

객체 시공간 정보 관련 시스템에서는 수동 및 자동 주석 생성 도구를 복합적으로 사용하여 객체들의 시공간 정보를 효율적으로 저장, 색인, 관리하는 시스템을 제안하였다[7]. 사용자는 수동으로 비디오 객체와 이벤트 발생과 관련된 주석을 작성하고 씬을 구분할 수 있다.

멀티미디어 데이터베이스 시스템은 대용량 비디오 데이터에 대해 사용자의 다양한 의미검색을 지원하는 시스템을 제안하였다[8]. 주석자는 자동으로 추출된 키 프레임에 대해 가장 기본적인 하나의 단어로 주석을 입력한다.

자유형 주석 처리 시스템은 사용자가 작성한 자유형 주석의 범위를 보다 정확하게 인식하는 방법을 제안하였다[9]. 사용자가 직접 자유로운 형태의 주석을 작성하면, 이는 수평 및 수직 요소의 모호성 등을 처리하는 과정을 거쳐 보다 정확한 범위의 주석으로 변환된다.

### 2.2 반자동 주석화

본 절에서는 데이터 주석화 중 반자동 주석화와 관련된 연구를 소개한다. 반자동 주석화와 관련된 연구는 다음과 같다.

멀티미디어 검색 시스템에서는 다양한 의미 정보를 가지고 있는 멀티미디어 데이터를 자유롭게 표현하고 정확하게 검색한다[10]. 제안하는 방법에서는 MPEG-7 기반의 메타데이터 형식을 활용하고 있으며, 의미 엔티티 생성 시 주석이 자동 및 반자동으로 처리된다.

또한, 비디오 주석 시스템은 비디오 감시 데이터로부터 사람 관련 정보를 빠르게 찾을 수 있도록 사람의 출현과 퇴장을 기준으로 키 프레임을 정의하고, 해당하는 키 프레임에서 사람 관련 주요 정보를 자동으로 추출하여 데이터베이스에 저장하는 연구이다[11]. 키 프레임 추출 및 객체 인식은 자동으로 처리되고, 이후 객체와 관련된 특성을 사용자가 편집할 수 있다.

토지 정보 분류 연구를 위한 이미지 데이터 셋 개발에 필요한 반자동 주석화 도구 연구에서는 지피에스 정보에 기반하여 1차 레이블링 결과를 자동으로 생성한 뒤, 이를 사람이 실제 결과와 비교하여 수동으로 수정하는 과정을 거친다 [12].

### 2.3 자동 주석화

본 절에서는 데이터 주석화 중 자동 주석화와 관련된 연구를 소개한다. 다양한 도메인에서 자동 주석화와 관련된 연구가 수행 중이다.

엑스레이 영상의 효과적인 분류를 위한 자동 주석화 연구에서는 영상으로부터 색 특징과 질감 특징을 추출하고 추출된 두 개의 특징 벡터를 서포트 벡터 머신에 적용한 뒤, 마지막으로 주석 코드 배열을 사용하여 자동으로 분류된 의료 영상을 얻는다[13].

음악 신호 분리 연구에서는 음악 신호에서 보컬과 비보컬 신호를 자동으로 분리하는 방법을 제안하였다[14]. 음악 신호 분리 시 동적 시간 워핑 알고리즘과 은닉 마르코프 모형을 활용하여 보컬, 비보컬 신호의 경계를 자동으로 분리한다.

노인 행동 판별을 위한 학습 데이터베이스를 생성하는 주석 프로그램에서는 웨어러블 센서와 카메라 등을 활용하여 고령자의 행동, 자세, 시간 등을 자동으로 파악하고 기록한다[15].

물체 인식 딥러닝 모델을 구성을 위한 자동 주석화 방법은 필요한 데이터 레이블링 과정을 하나의 프로그램에서 사용할 수 있는 주석화 도구를 제안한다[16]. 파이썬 기본 그래픽 유저 인터페이스와 물체 인식 딥러닝 모델인 레티나넷을 활용해 영상 내에서 80개의 사물 범주를 자동으로 인식하며, 해당하는 객체의 범위를 표시해준다.

비주얼 의류 검색 시스템은 사용자가 의류 이미지를 통해 의류의 종류와 가격, 판매처, 이미지 등을 효율적으로 검색할 수 있게 한다[17]. 소매길이와 잠금장치를 속성으로 의류 이미지를 분석하며, 해당 의류가 어떤 카테고리에 속할 확률이 높은지 자동으로 파악한다.

AnsNGS는 질병과 관련된 변이를 파악하고 기능적 결과를 조사하는 도구를 제안한다[18]. 사용자가 텍스트 기반의

파일을 입력하면 기존에 존재하는 주석 데이터베이스에서 관련된 정보를 확인하고, 각종 질병과 관련된 변이가 있으면 이에 관한 주석을 자동으로 달아서 데이터를 출력해준다.

수중환경에서 유용하게 활용될 수 있는 데이터 셋을 생성하고, 해당 데이터 셋에서 주석을 자동으로 처리하는 방법이 있다[19]. 3D 카드 모델을 사용하여 이미지를 합성하고, 이후 알고리즘을 통해 필요한 라벨을 자동으로 추출한다.

WWW 영상 주석 방법은 저수준 특징, 페이지 태그, 전체적인 단어 빈도수와 지역적 단어 빈도수를 기반으로 한 주석화 방법이다[20]. 위의 4가지 속성을 기반으로 자동화된 주석 알고리즘을 사용하며, 이를 통해 웹페이지에 있는 영상의 주석을 자동으로 처리한다.

현재 얼굴인식(FR) 솔루션에 상황 정보를 활용한 새로운 얼굴 주석 프레임워크가 있다[21]. 클러스터링과 일반 학습을 기반으로 하며, 의복 및 얼굴 이미지에 자동으로 레이블을 달아준다.

### 2.4 슈퍼바이즐리란

슈퍼바이즐리는 이미지, 비디오 등 다양한 데이터에 라벨링 작업을 수행할 수 있게 하는 웹 기반의 플랫폼이다. 전문 지식이 없는 사람도 편리하게 사용할 수 있으며, 관절 수를 사용자가 원하는 개수로 세팅할 수 있고, 무료로 사용할 수 있다는 것이 장점이다.

Fig. 1은 슈퍼바이즐리에서 주석화를 진행하는 예시 화면이다. 이미지에 좌표를 찍기 위해서는 이미지를 업로드한다. 이미지 업로드 후에는 업로드한 이미지에 좌표를 추가하고 수정한다. 이미지에 좌표를 추가하기 위해서는 기존에 정의해둔 2차원 관절 위치 클래스인 포인트 클래스를 불러온다. 포인트 클래스란 자세 라벨링을 위해 라벨링 전 정한 관절의 종류를 의미한다. 포인트 클래스를 불러온 후 이미지 크기에 맞게 포인트 클래스의 크기를 조절한다. Fig. 1의 노란색 박스로 표시된 좌측 상단과 우측 하단의 빨간 점을 조정하여 어노테이션 자체의 크기를 조절할 수 있으며, 빨간색 박스로 표시된 화살표 아이콘을 클릭해 개별 2차원 관절 위치를 조정할 수 있다. 또한, 파란색 박스로 표시된 좌측의 두꺼운 십자가 화살표 아이콘을 클릭해 2차원 관절 위치 전체를 이동시킬 수 있으며, 검은색 박스로 표시된 커서 모양 아이콘을 클릭하고 지우고 싶은 2차원 관절 위치를 선택해 원하는 관절을 삭제할 수 있다.

## 3. C3Pap를 위한 반자동 주석 방법

본 장에서는 연주 자세의 그라운드 트루스 수집을 위한 데이터 반자동 주석화 방법인 SAAnnot-C3Pap(Semi Automatic Annotation for C3Pap)를 제안한다. 제안하는 방법은 크게

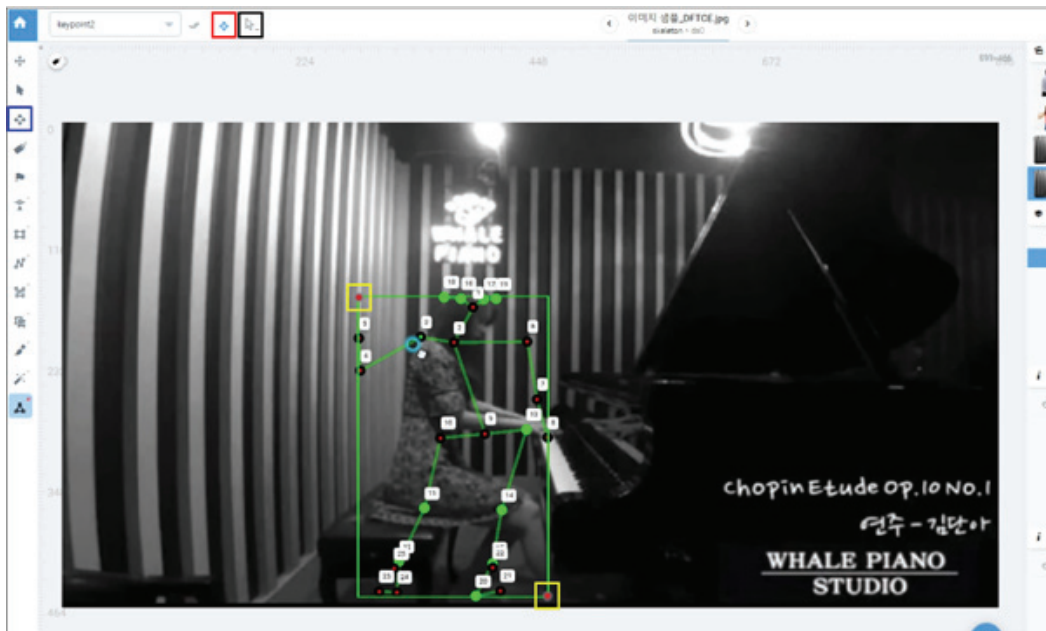


Fig. 1. The Process of Adding and Editing Coordinates to Images Uploaded to Supervisely

3가지 과정으로 오픈포즈로 자세를 추출하고, 추출한 것 중 잘못된 부분을 슈퍼바이즐리로 수정한 뒤, 오픈포즈와 슈퍼바이즐리의 결과값을 동기화하는 과정을 거친다.

### 3.1 전체 프로세스

본 절에서는 본 연구의 전체 프로세스를 설명한다. Fig. 2는 제안하는 기법의 전체 프로세스를 보여준다. 첫 번째 단계는 오픈포즈를 사용하여 피아노 연주 영상에서 2차원 관절 위치를 추출하는 단계이다. 이를 위하여 피아노 연주 영상을 1초 단위로 나누고 1초 단위의 영상을 다시 30개의 프레임으

로 나누어 각 프레임 이미지를 생성한다. 이후에 오픈포즈로 전체 프레임 이미지에 대한 2차원 관절 위치를 추출한다. 두 번째 단계는 추출된 2차원 관절 위치 중 관절 위치가 잘못 추출된 이미지를 선택하고 슈퍼바이즐리를 사용해 수동으로 2차원 관절 위치를 추출한다. 세 번째 단계에서는 첫 번째 단계에서 오픈포즈로 추출한 2차원 관절 위치와 두 번째 단계의 결과물인 슈퍼바이즐리로 수동으로 라벨링한 2차원 관절 위치를 동기화하는 단계이다. 사람이 2명 이상 감지 될 경우 연주자를 추려내고 이후에 동기화를 진행한다.

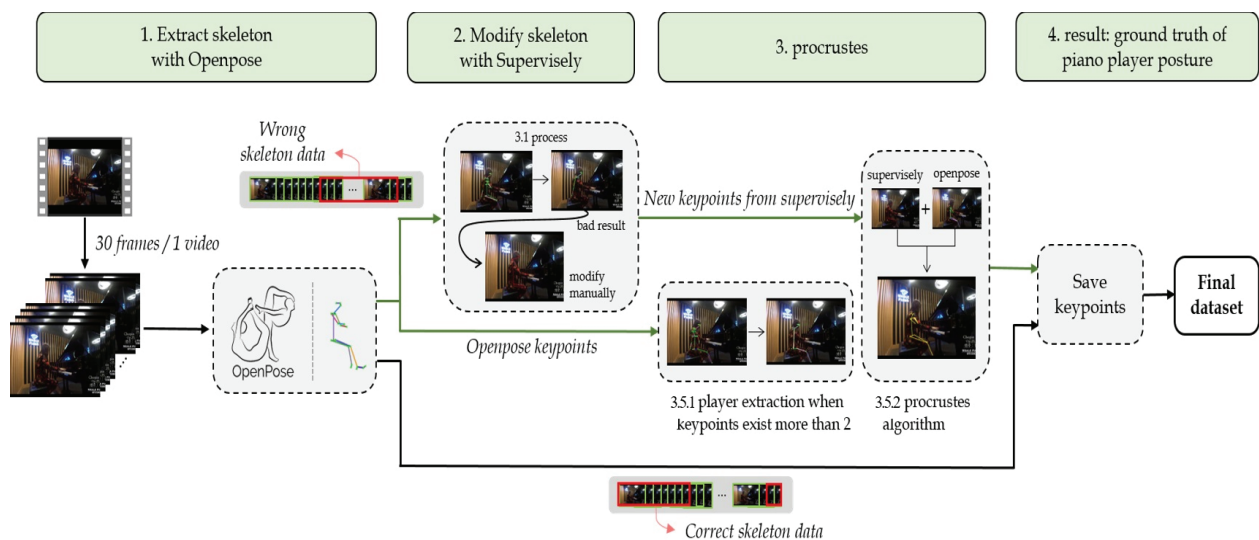


Fig. 2. Proposed Semi-automatic Data Set Collection Method

### 3.2 C3Pap

본 절에서는 본 연구에서 사용한 데이터셋을 소개한다. 본 연구에서는 프로페셔널 피아니스트와 아마추어 피아니스트의 연주 영상 데이터인 C3Pap(classic piano performance posture version amateur & pro) 데이터 셋을 사용한다 [1].

다음은 C3Pap 요약 정보를 보여준다. 총 395개의 영상을 포함한다. 연주자 수는 13명이며 이 중 프로페셔널 피아니스트는 8명, 아마추어 피아니스트는 5명이다. 전체 영상 길이는 1506초며 곡의 개수는 17개이다. 영상 평균 길이는 4.18초이고 최대 길이는 12초, 최소 길이는 1초이다. 프레임 레이트는 29.97fps이고 각 영상은 오디오를 포함한다. 본 논문에서는 영상을 이미지로 나누어 사용한다.

### 3.3 오픈포즈로 자세 추출

본 절에서는 반자동 주석화 방법 중 오픈포즈로 2차원 관절 위치를 추출하는 방법을 소개한다.

오픈포즈는 하나의 이미지에서 여러 사람의 2차원 관절 위치를 실시간으로 추출할 수 있는 소프트웨어로 본 논문에서 피아노 연주 영상에서 연주자의 2차원 관절 위치를 자동으로 추출하는 방법으로 사용한다. 오픈포즈는 사람의 손, 얼굴, 발을 포함한 신체의 135개 지점에 대한 사람의 골격을 생성한다. 오픈포즈는 2차원뿐만 아니라 3차원 자세 관절 좌표 추출 기능도 제공한다. 2차원 자세 관절 좌표 추출 기능은 실시간으로 여러 사람의 2차원 관절 위치를 추출할 수 있고, 3차원 자세 관절 좌표 추출 기능은 실시간으로 한 사람의 2차원 관절 위치를 추출한다.

본 논문에서는 피아노 연주 영상을 초당 30프레임으로 나누고 각 프레임에서 오픈포즈로 2차원 관절 위치를 추출한다. 오픈포즈로 2차원 관절 위치를 추출할 때 여러 옵션을 적용할 수 있는데, 이번 실험에서는 2차원 관절 위치 스케일 옵션(keypoint scale option)을 변경하며 데이터에 가장 적합한 해상도 옵션을 찾아 적용하였다. 오픈포즈의 스케일 옵션은 2차원 관절 위치 (x,y) 좌표의 스케일을 조정하는 옵션으로 0부터 4까지 총 5개의 옵션이 있다. 0번 옵션은 원본 이미지와 동일한 해상도로 스케일링, 1번 옵션은 순 출력 사이즈(net output size)로 스케일링, 2번 옵션은 최종 출력 크기(final output size)로 스케일링, 3번 옵션은 (0,0)이 왼쪽 상단 모서리이고 (1,1)이 오른쪽 하단 모서리인 경우에 [0,1] 사이의 값으로 스케일링, 4번 옵션은 왼쪽 상단 모서리가 (-1,-1), 오른쪽 하단 모서리가 (1,1)인 이미지에 [-1,1] 사이의 값으로 스케일링한다. 일부 데이터에 대해 옵션을 변경하여 결과를 얻어내는 추가적인 실험을 통해 0번 옵션과 2번 옵션이 슈퍼바이저리로 얻은 2차원 관절 위치와 가장 비슷한 결과를 도출하는 것을 알아낼 수 있었고 본 연구의 실험에서는 0번 옵션을 적용하였다.

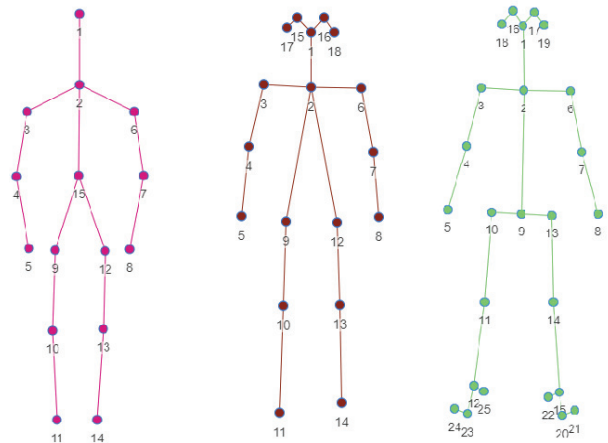


Fig. 3. Point Example  
(Left : MPI, Middle : COCO 18, Right : BODY 25)

본 연구에서는 신체의 25개의 지점에 해당하는 2차원 관절 위치를 추출하였으며 추출한 지점은 코, 왼쪽 눈, 오른쪽 눈, 왼쪽 귀, 오른쪽 귀, 목, 왼쪽 어깨, 오른쪽 어깨, 왼쪽 팔꿈치, 오른쪽 팔꿈치, 왼쪽 팔목, 오른쪽 팔목, 중앙 골반, 왼쪽 골반, 오른쪽 골반, 왼쪽 무릎, 오른쪽 무릎, 왼쪽 발목, 오른쪽 발목, 왼쪽 뒤꿈치, 오른쪽 뒤꿈치, 왼쪽 발, 오른쪽 발, 왼쪽 발가락, 오른쪽 발가락이다.

영상의 해상도가 좋지 않거나 사람 신체의 일부만 보이는 경우에는 영상의 배경이나 피아노를 사람으로 인식하는 경우의 좋지 않은 결과를 보이는 경우가 있었다. 따라서 본 논문에서는 이런 아웃라이어(outlier)에 대한 처리를 진행한다.

### 3.4 포인트 클래스

본 절에서는 본 논문에서 사용한 포인트 클래스를 설명한다. 포인트 클래스는 Fig. 3처럼 다양한 종류가 있지만 본 논문에서는 오픈포즈 논문에서 새로 제안한 BODY 25를 사용하여 오픈포즈로 뽑은 2차원 자세 좌표 중 잘못된 부분에 대해서만 다시 라벨링을 진행한다. 오픈포즈는 기존 MPI, COCO 포인트 클래스에 왼/오른 엄지 발가락, 왼/오른 새끼 발가락, 왼/오른 발꿈치를 추가하였다. 또한, 상체가 왜곡(Occlusion)되거나 이미지 밖으로 나갔더라도 양쪽 다리와 골반을 연결하기 위해 양쪽 골반에 중간값을 추가하였다.

### 3.5 오픈포즈와 슈퍼바이저리 결과값 동기화

본 절에서는 반자동 주석화 방법 중 오픈포즈와 슈퍼바이저리의 결과값을 동기화하는 방법을 소개한다.

오픈포즈로 추출한 데이터와 슈퍼바이저리로 수정한 데이터를 결합하여 하나의 2차원 관절 위치로 만드는 과정을 수행한다. 결과값 동기화 과정은 두 가지 단계로 이루어진다. 첫 번째 단계에서는 오픈포즈에서 2명 이상의 사람이 검출될 경우 연주자를 검출한다. 두 번째 단계에서는 프로크러스



티즈 알고리즘을 사용하여 두 관절의 2차원 관절 위치를 비교하고 결합하는 과정을 진행한다. 비교를 위해 사용한 알고리즘은 두 도형의 유사도를 도출하는 프로크루스티즈 (procrustes) 기반의 형상 분석(shape analysis) 알고리즘으로, 일반적으로 도형이 가지는 위치, 크기, 회전의 특성들을 동일하게 맞추어 각 도형이 가지는 특성을 고려하지 않고 비교할 수 있도록 하였다 [22].

### 1) 아웃라이어 검출 알고리즘

본 절에서는 프로크루스티즈 알고리즘을 수행하기 전에 거친 오픈포즈에서 추출한 데이터를 정제하는 과정을 소개한다. 실험에 사용한 모든 피아노 연주 영상에 한 사람의 연주자만 있음에도 오픈포즈가 추출한 2차원 관절 위치에는 피아노에 비친 사람 혹은 이미지의 배경, 피아노 다리와 같이 사람이 아닌 부분을 사람으로 인식하여 2명 이상의 사람을 추출하는 경우가 있어 올바른 한 명의 연주자만 추출하는 데이터 정제 과정이 필요하였다.

데이터의 정제는 다음과 같은 과정으로 진행하였다. 먼저 두 개 이상의 2차원 관절 위치를 가지는 데이터의 특징을 찾기 위해 데이터를 추출하는 데 사용한 이미지 위에 하나의 2차원 관절 위치만 찍어서 만든 여러 결과 이미지를 비교하였다. 그 결과 두 개 이상의 2차원 관절 위치가 추출된 데이터에서 이상치(outlier)와 연주자 2차원 관절 위치(정상 데이터)의 첫 번째 요소(얼굴 좌표)의 위치가 눈에 띄게 다르다는 발견을 하였다.

슈퍼바이즐리로 만들어 신뢰할 수 있는 2차원 관절 위치와 오픈포즈에서 추출된 얼굴 좌표 2차원 관절 위치의 차를 계산하여 지정한 숫자인 임계치(threshold)를 넘는 경우 이상치로 분류하도록 하였다. 본 실험에서는 임계치를 100으로 사용했으나 이 크기는 어떤 데이터를 사용하는지에 따라 달라질 것으로 예상된다. 이 방법은 슈퍼바이즐리와 오픈포즈의 정상 데이터 사이의 가장 큰 차이 값(a)보다 이상치와의 가장 작은 차이 값(b)이 커야만 적용할 수 있다는 문제점이 있다. 만약 a가 b보다 큰 값을 가진다고 가정했을 때, 모든 정상 데이터를 포함하기 위해서는 임계치가 a보다 크게 되는데 그런 경우 일부 이상치가 정상 데이터로 분류된다. 또 모든 이상치를 거르기 위해서는 b보다 작은 임계치를 가지게 되는데 그런 경우 정상 데이터의 일부가 이상치로 분류된다. 그럼에도 비정상적으로 분류되는 비율이 무시할 정도라면 a가 b보다 커도 아래의 방법을 사용할 수는 있다.

이상치를 검출하는데 사용한 알고리즘은 Fig. 4와 같다. 첫 번째 라인인 오픈포즈 결과값을 저장한 폴더에서 각 파일을 읽어서 가져오는 것을 수행한다. 두 번째 라인에서 가져온 데이터가 비어있는 것을 확인하면 세 번째 라인에서 전역 변수에 저장하고 있던 이전 데이터 반환을 수행한다. 네 번째

#### Algorithm Detecting outliers from Openpose keypoints

**Input** : results keypoints extracted from Openpose (.json)  
**Output** : one keypoints array

#### Algorithm :

```

/* Open and read 'pose_keypoints_2d' from 'people' data in Openpose_results.json*/
1 : FOR each file in Openpose results directory
    /* data format: {..., 'people': [{'pose_keypoints_2d':[], ...}, {...}]}*/
2 :     IF 'people' is empty
3 :         THEN RETURN previous keypoints array
4 :     ELSE THEN FirstKeypointsArray ← Save first 'pose_keypoints_2d'
        array from 'people' data
5 :         IF second 'pose_keypoints_2d' exists
6 :             THEN SecondKeypointsArray ← Save second
                'pose_keypoints_2d' array
    /* Read supervisory data for comparison*/
7 :     supervisoryKeypoints ← Read supervisory data
        with the same file name
8 :     difference ← Calculate head element's difference between
        FirstKeypointsArray and supervisoryKeypoints
9 :     IF difference > threshold
10 :        THEN RETURN SecondKeypointsArray
11 :    RETURN FirstKeypointsArray

```

Fig. 4. Algorithm to Get Outliers in Openpose

라인은 데이터가 비어있지 않은 경우 첫 번째 키포인트 데이터를 변수에 저장한다. 다섯 번째 라인에서는 첫 번째 라인에서 읽어 온 데이터에 2차원 관절 위치가 두 개 이상 존재하는지의 여부를 확인하는 과정을 수행한다. 여섯 번째 라인에서는 두 번째 2차원 관절 위치 데이터를 변수에 저장한다. 일곱 번째 라인에서는 슈퍼바이즐리에서 첫 번째 라인에서 읽은 데이터와 동일한 이름을 가진 데이터를 읽어서 변수에 저장한다. 여덟 번째 라인에서는 일곱 번째 줄에서 저장한 슈퍼바이즐리 데이터와 네 번째 줄에서 저장한 데이터의 머리 좌표의 차를 구하여 변수에 저장하는 과정을 수행한다. 아홉 번째 라인에서는 여덟 번째 라인에서 저장한 변수와 임계치를 비교한다. 열 번째 라인에서는 여섯 번째 라인에서 저장한 두 번째 2차원 관절 위치 배열을 반환한다. 열한 번째 라인은 첫 번째 2차원 관절 위치 배열을 반환한다.

### 2) 프로크루스티즈 알고리즘

여러 2차원 관절 위치 중 하나의 2차원 관절 위치를 선택한 후에는 프로크루스티즈 기반의 형상 분석 알고리즘을 적용한다. 과정은

다음과 같다. 먼저 이동(translation)을 통해 두 도형의 중심점을 맞춘다. 2차원 관절 위치는 행렬(matrix)이기 때문에 프로베니우스 놈(Frobenius Norm)으로 행렬의 크기를 구한다. Equation (1)이 이에 해당한다[23]. 프로베니우스 놈은 행렬의 모든 성분의 제곱의 합의 제곱근을 구하는 방법으로 행렬의 크기를 구하기 위해 사용한다[24]. 다음은 앞서 구한 행렬 크기로 각 도형의 크기를 단위 크기(unit scale)로 맞추는 스케일링(scaling)을 수행한다. 그리고 주축(major axis)을 정렬하기 위해 비교하려고 하는 2차원 관절 위치를 비교당하는 2차원 관절 위치에 맞춰 회전(rotation)시킨다. 마치

막으로 앞에서 구한 두 2차원 관절 위치 사이의 차이(disparity)를 계산하면 알고리즘이 마무리된다. 차이가 0에 가까울수록 비교한 두 대상의 유사도가 높다.

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2} \quad (1)$$

본 논문은 유사도를 구하는 것이 아닌 오픈포즈와 슈퍼바이즐리로 추출한 2차원 관절 위치를 통합한 프로크루스티즈 된 하나의 2차원 관절 위치를 얻는 것이 목적이었다. 그래서 프로크루스티즈 기반의 형상 분석 알고리즘에서 유사도를 구하는 마지막 부분 대신 오픈포즈에서 추출한 데이터를 슈퍼바이즐리로 추출한 데이터와 비슷하게 하도록 변환하는 과정을 통해 변환된 행렬 즉, 2차원 관절 위치를 구하였다[25].

본 절에서 제안하는 프로크루스티즈 알고리즘은 Fig. 5와 같다[25]. 첫 번째와 두 번째 라인은 각각 X행렬과 Y행렬의 크기를 구하는 과정을 수행한다. 세 번째 라인은 X, Y 행렬의 중간값을 계산한다. 네 번째 라인에서는 세 번째 라인에서 구한 중간값을 사용해 행렬을 이동시킨다. 다섯 번째 라인에서는 프로베니우스 놈 수식으로 행렬의 크기를 계산한다. 여섯 번째 라인에서는 행렬의 크기를 사용해 행렬 X와 Y를 단위 크기로 스케일링한다. 일곱 번째 라인에서는 Y 행렬의 열의 크기가 X보다 작은 경우 0으로 채우는 작업을 수행한다. 여덟 번째 라인에서는 X행렬의 전치행렬(transposed matrix)과 Y행렬의 내적(dot product)을 수행한다. 아홉 번째 라인에서

#### Algorithm Procrustes

**Input** : two keypoints to be compared and combined (X, Y)  
**Output** : combined keypoints (Z)

#### Algorithm :

```

/* Procrustes to make the combined keypoints*/
/* x is a fixed object(criterion) and y is an object to be compared*/

/*1. translation*/
1 : n, m ← row and column of X
2 : ny, my ← row and column of Y
3 : midX, midY ← center coordinates of the X and Y matrix
4 : X0, Y0 ← translate matrix to center with midX and midY

/*2. scaling*/
5 : normX, normY ← calculate Frobenius norm with X0, Y0
6 : X0, Y0 ← scale to unit size with normX and normY
7 : IF my < m THEN fill column Y with zero

/*3. rotation*/
8 : A ← calculate dot product of two matrix (X0.T(transpose) and Y0)
/* left singular vectors of A and right singular vectors of A in order*/
9 : U, Vt ← calculate SVD(singular value decomposition) and return
orthogonal matrix obtained by eigen decomposition of AAT and ATA
10 : ssvA ← ∑ (singular values of A)
11 : T ← dot product of U.T and Vt.T

/*4. transform Y matrix into combined matrix*/
12 : Z ← transform Y0 matrix with normY(scale), ssvA and T(rotation),
and midX(translation)
13 : RETURN Z

```

Fig. 5. Procrustes Algorithm

는 앞 라인에서 계산한 내적으로 특이값분해(Singular Value Decomposition, SVD)를 한다. 열 번째 라인에서는 특이값의 총합을 계산한다. 열한 번째 라인에서는 아홉 번째 라인에서 구한 left singular vector의 전치행렬과 right singular vector의 전치행렬의 내적을 수행한다. 열두 번째 라인에서는 앞에서 구한 크기, 회전, 이동 값을 Y 행렬에 적용해 X행렬과 병합된 새로운 행렬을 열세 번째 라인에서 반환한다.

#### 4. SAAnnot-C3Pap 적용 결과 분석

오픈포즈와 슈퍼바이즐리의 결과값 동기화는 아래의 두 단계를 거친다. 먼저 첫 번째 단계에서 오픈포즈에서 2명 이상의 사람이 검출된 경우 연주자를 검출하는 알고리즘을 적용한다. 그 다음 두 번째 단계에서 프로크루스티즈 알고리즘을 적용한다. 본 장에서는 각 단계의 적용 전후 결과를 보여준다.

먼저 오픈포즈로 추출한 결과에서 2명 이상의 사람이 검출된 경우 연주자를 검출하는 알고리즘 적용 전후 결과를 분석한다. Fig. 6의 왼쪽(Data extracted with Openpose)은 아웃라이어 검출 알고리즘 적용 전 이미지로 오픈포즈로 연주자의 2차원 관절 위치를 보여준다. 오른쪽(Extract human keypoints)은 아웃라이어 검출 알고리즘 적용 후 이미지로 올바른 한 명의 연주자의 2차원 관절 위치를 보여주는 이미지이다. Fig. 6의 위 이미지에서는 피아노 다리를 사람으로 잘못 인식하였는데 알고리즘 적용 후에는 잘못 인식한 결과가 삭제되었다. Fig. 6의 아래 이미지는 피아노 건반을 사람으로 잘못 인식하였는데 알고리즘 적용 후에는 잘못 인식한 결과가 삭제되었다.

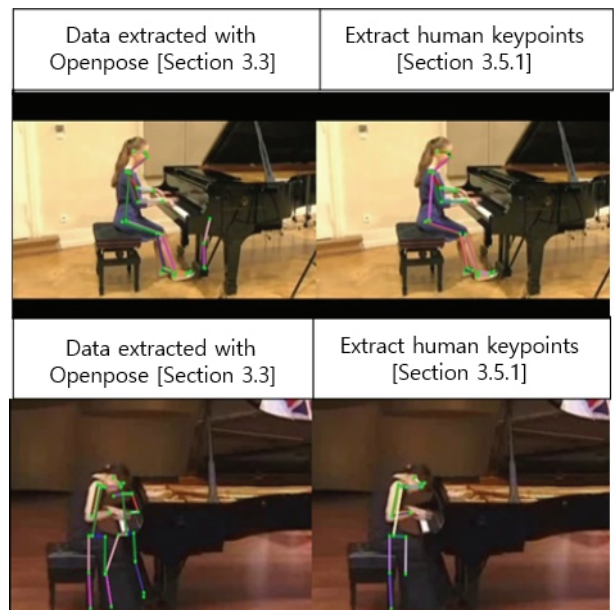


Fig. 6. Effect of Extract Human Keypoints

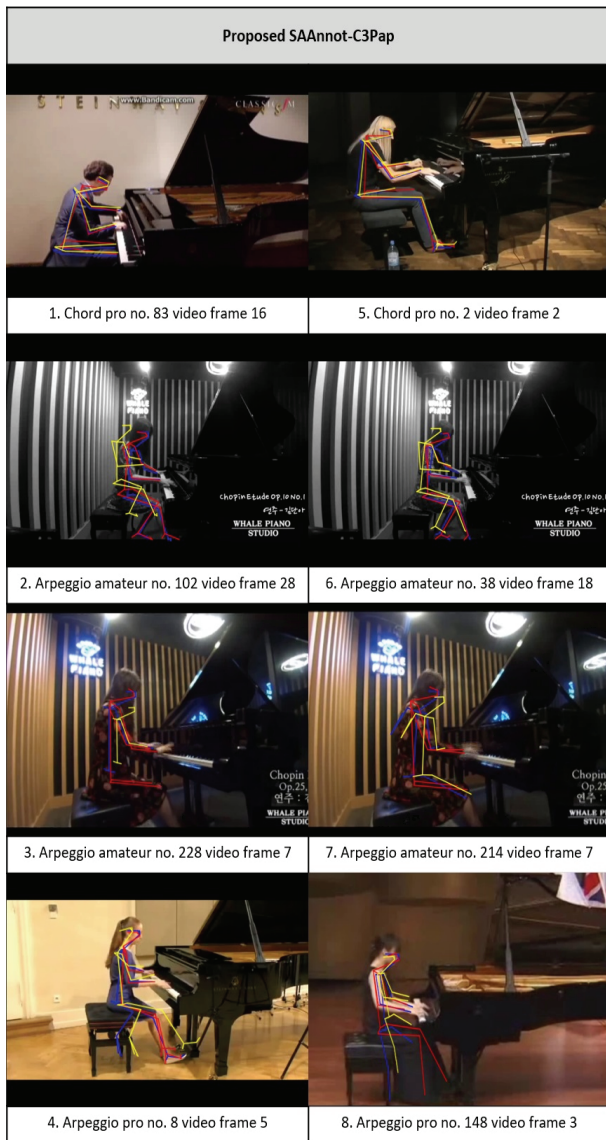


Fig. 7. After Applying the Procrustes Algorithm

다음으로는 프로크러스티즈 알고리즘 적용 전후 결과를 분석한다. 비교하기 쉽도록 Fig. 7에 프로크러스티즈 알고리즘으로 새로 추출한 2차원 관절 위치를 오픈포즈, 슈퍼바이즐리 데이터와 함께 나타내었다. 오픈포즈는 파란색, 슈퍼바이즐리는 빨간색으로 나타내었고, 프로크러스티즈된 결과는 노란색으로 나타내었다. 자동으로 만들어진 오픈포즈 2차원 관절 위치와 수동으로 만든 슈퍼바이즐리 2차원 관절 위치를 프로크러스티즈하여 오픈포즈로 추출한 2차원 관절 위치를 슈퍼바이즐리 2차원 관절 위치와 비슷하게 조정하는 과정이다. Fig. 7의 네 번째 예시의 경우에는 프로크러스티즈 하기 전 오픈포즈 데이터에 오류가 있어 프로크러스티즈한 결과에 피아노 의자의 다리가 사람의 다리로 잘못 인식되는 오류가 나타난다. 이런 문제를 해결하기 위해서는 프로크러스티즈 하는 기준을 오픈포즈 2차원 관절 위치가 아닌 슈퍼바이즐리

2차원 관절 위치로 한다. 오류가 있는 오픈포즈 2차원 관절 위치를 슈퍼바이즐리로 수정했음에도 이런 오류가 발생하는 경우가 있다. 프로크러스티즈하는 기준을 슈퍼바이즐리 2차원 관절 위치로 하는 경우를 다음과 같은 두 가지 경우로 한정하였기 때문이다. 첫 번째 경우는 오픈포즈가 영상의 사람의 관절 위치를 전혀 검출하지 못한 경우이고, 두 번째 경우는 검출된 관절 위치가 그라운드 트루스와 차이가 크게 나는 경우이다.

### 5. 결론

본 논문에서는 연주자 자세의 그라운드 트루스 획득을 위한 반자동 주석 방법인 SAAnnot-C3Pap를 제안하였다. 제안하는 SAAnnot-C3Pap는 기존 음악 도메인에서 사용하는 자동 주석화 방법인 오픈포즈의 잘못된 2차원 관절 위치 추출 문제를 해결하기 위한 방법이다.

본 논문에서 제안하는 SAAnnot-C3Pap는 슈퍼바이즐리를 사용하여 오픈포즈에서 발생하는 잘못된 2차원 관절 위치 검출 결과를 교정하였다. 또한, 프로크러스티즈 알고리즘을 활용하여 2명 이상의 사람이 검출되는 문제를 해결하고 오픈포즈와 슈퍼바이즐리로 얻은 결과를 동기화시켰다.

실험에서는 반자동 주석 방법인 오픈포즈와 본 논문에서 제안하는 SAAnnot-C3Pap의 결과를 비교·분석한다. 비교 결과 제안하는 SAAnnot-C3Pap는 오픈포즈로 잘못 수집된 자세 정보를 개선한 결과를 보였다. 향후 연구는 본 연구를 통해 수집된 그라운드 트루스 기반의 연주 자세 자동 생성, 연주 자세 분석 연구를 진행할 예정이다.

### References

- [1] S. H. Park and Y. H. Park, "Audio-visual tensor fusion network for piano player posture classification," *Applied Sciences*, Vol.2020, No.10, pp.6857-6871, 2020.
- [2] Github-CMU-Perceptual-Computing-Lab/openpose [Internet], <https://github.com/CMU-Perceptual-Computing-Lab/openpose>.
- [3] E. Shlizerman, L. Dery, H. Schoen, and I. Kemelmacher-Shlizerman, "Audio to Body Dynamics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Utah, pp.7574-7583, 2018.
- [4] Supervisely - Web platform for computer vision. Annotation ... [Internet], <https://supvisely.ly>.
- [5] K. W. Lee, "Video Data Retrieval System using Annotation and Feture Information," *Journal of Korea Academia-Industrial Cooperation Society*, Vol.7, No.6, pp.1129-1133, 2006.



- [6] Y. H. Lee, K. J. Oh, H. M. Deok, S. B. Park, and G. S. Jo, "A collaborative video annotation and browsing system using linked data," *Journal of Intelligence and Information Systems*, No.5, pp.455-460, 2011.
- [7] J. S. Kim, T. E. Kim, and M. H. Kim, "A video annotation system that automatically generates spatio-temporal relationships among objects," *KIISE Transactions on Computing Practices*, Vol.18, No.10, pp.701-710, 2012.
- [8] K. W. Lee, T. H. Oh, and K. M. Cho, "A multimedia database system using method of automatic annotation update," *Journal of Korea Academia-Industrial cooperation Society*, Vol.11, No.6, pp.417-420, 2006.
- [9] W. S. Sohn, J. K. Kim, S. B. Lim, and Y. C. Choy, "A method of context based free-form annotation in XML documents," *Journal of KIISE: Software and Applications*, Vol.30, No.9-10, pp.850-861, 2003.
- [10] H. G. An and J. J. Koh, "Development of multimedia annotation and retrieval system using MPEG-7 based semantic metadata model," *The KIPS Transactions: Part D*, Vol.14, No.6, pp.573-584, 2007.
- [11] J. S. Kim, K. Y. Kim, H. I. Kim, and Y. S. Kim, "A video annotation system with automatic human detection from video surveillance data," *KIISE Transactions on Computing Practices*, Vol.18, No.11, pp.808-812, 2012.
- [12] D. W. Jang, J. W. Lee, and J. S. Lee, "Development of semi-automatic annotation tool for building land cover image data set," *Journal of Broadcast Engineering*, Vol.2019, No.11, pp.69-70, 2019
- [13] K. H. Park, B. C. Ko, and J. Y. Nam, "Medical image automatic annotation using multi-class SVM and annotation code array," *The KIPS Transactions: Part B*, Vol.16, No.4, pp.281-288, 2009.
- [14] K. G. Lee and M. Cremer, "Automatic labeling of training data for vocal/non-vocal discrimination," *The Journal of the Korean Electro-Acoustic Music Society*, No.7, pp.29-41, 2009.
- [15] G. Won et al., "Development of multi-modal sensor data annotation system for elderly behavior analysis," *IEIE Transactions on Smart Processing & Computing*, Vol.2019, No.11, pp.781-783, 2019.
- [16] S. W. Lim and G. M. Park, "Development of python-based annotation tool program for constructing object recognition deep-learning model," *Journal of Broadcast Engineering*, Vol.25, No.3, pp.386-398, 2020.
- [17] E. K. Lee, Y. W. Kim, and S. S. Kim, "Research: Annotation technique development based on apparel attributes for visual apparel search technology," *Fashion & Textile Research Journal*, Vol.17, No.5, pp.731-740, 2015.
- [18] Y. J. Na, Y. L. Cho, and J. H. Kim, "AnsNGS: An annotation system to sequence variations of next generation sequencing data for disease-related phenotypes," *Healthcare Informatics Research*, Vol.19, No.1, pp.50-55, 2013.
- [19] M. H. Jeon, Y. J. Lee, Y. S. Shin, H. S. Jang, T. K. Yeu, and A. Y. Kim, "Annotation Tool = synthesizing image and automated annotation tool for CNN based under water object detection," *Journal of Korea Robotics Society*, Vol.14, No.2, pp.139-149, 2019.
- [20] S. S. Yoon, H. Y. Moon, and Y. W. Rhee, "A WWW images automatic annotation based on multi-cues integration," *Journal of the Korea Society of Computer and Information*, Vol.13, No.4, pp.79-86, 2008.
- [21] J. Y. Choi, W. De Neve, Y. M. Ro, and K. N. Plataniotis, "Automatic face annotation in personal photo collections using context-based unsupervised clustering and face information fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.20, No.10, pp.1292-1309, 2010.
- [22] F. Meng et al., "Procrustes: A python library to find transformations that maximize the similarity between matrice," *Computer Physics Communications*, Vol.276, pp.108334, 2022.
- [23] G. H. Golub and C. F. Van Loan, "Matrix Computations," 3rd ed. Baltimore, MD: Johns Hopkins, 1996.
- [24] Eric W. Weisstein, "Frobenius Norm," From MathWorld--A Wolfram Web Resource.
- [25] "Simple examples of multidimensional scaling (MDS, Multi-dimensional Scaling) and Procrustes Analysis on the joint points of human posture (python)", ProgrammerSought [Internet], <https://www.programmersought.com/article/74856266632>.



### 박 소 현

<https://orcid.org/0000-0003-1022-1790>

e-mail : shpark@sm.ac.kr

2016년 숙명여자대학교 멀티미디어학과  
(석사)

2020년 숙명여자대학교 IT공학과(박사)

2020년 ~ 현 재 숙명여자대학교 빅데이터  
활용 연구센터 책임연구원

관심분야 : 인공지능, 빅데이터, 영상처리



**김 서 연**

<https://orcid.org/0000-0001-9833-4823>  
e-mail : sykim2048@gmail.com  
2021년 숙명여자대학교 컴퓨터과학전공  
(학사)  
2021년~현 재 메가존클라우드 데이터팀  
매니저

관심분야: 데이터베이스, 시각화, 데이터 분석



**박 영 호**

<https://orcid.org/0000-0002-5284-9589>  
e-mail : yhpark@sm.ac.kr  
1992년 동국대학교 컴퓨터공학과(석사)  
2005년 한국과학기술원 전산학과(박사)  
1993년~1999년 한국전자통신연구원  
교환전송연구단 선임연구원

2005년~2006년 한국과학기술원 첨단정보기술연구센터 연구원  
2005년~2006년 동국대학교 컴퓨터멀티미디어학과 겸임교수  
2006년~현 재 숙명여자대학교 IT공학과 교수

관심분야: 데이터베이스, XML, IR(정보검색), 멀티미디어 데이터베이스,  
Bio정보공학, 영상미디어, 예술&공학인터페이스,  
데이터베이스 관리시스템, 머신러닝, 빅데이터, 데이터분석,  
Telecommunication System