

# Data Capturing Hand Pose data

Aug 28th, 2023

To, professor

# Summary

1. 마커 기반 모션 캡처를 통해 고정밀 3D hand pose data capture
2. 표시된 손 이미지를 정렬
3. 마커 제거 네트워크(MR-Net)를 통해 손에 있는 마커 제거되고  
맨손 이미지가 3D hand pose ground truth 상태가 됨

# | Index\_

**1. Build Environment**

**2. Generation of Data**



## **1. Build Environment**

# 1. Build Environment for Data Capturing

## 1. Piano(2대)



Kawai ES-110(전자 피아노)



Kawai GE-20(그랜드 피아노)

# 1. Build Environment(Data Capturing – PianoHand2.5M)

2. Optitrack Prime 13W가 포함된 track1 MoCap 시스템 흑백 IR 카메라 11대 + Optitrack Prime Color FS RGB 카메라 1대  
- 캡처 프레임: 240FPS    노출 시간: 4ms    이미지 해상도: 1920x1080(1080p)

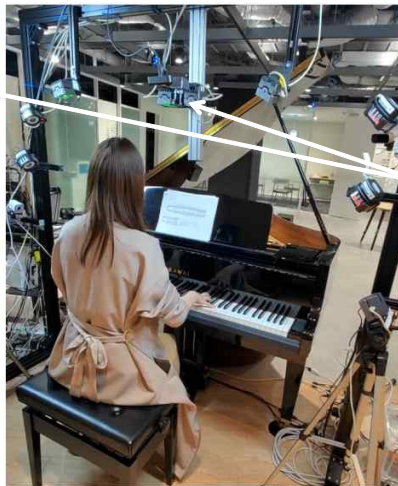


Ultra-wide FOV for optical tracking in CAVEs and other compact spaces.

[제품 링크 바로가기](#)



Capture Studio A (EP)



Capture Studio B (GP)



Prime Color FS

Part number: PCC0001-FS

[제품 링크 바로가기](#)

RGB 카메라

- 피아노 상단 중앙에 고정

# 1. Build Environment(Data Capturing – PianoHand2.5M)

## 3. Makers Placement – Optitrack hemisphere 4mm facial reflective markers



### Facial Markers- 4 mm

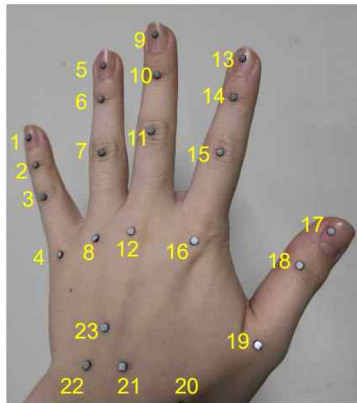
Part number: MCP1130

\$110

Set of 4 mm **hemisphere** facial markers, adhesive backed. Includes 50 count **hemispheres**.

제품 링크 바로가기

한 손당 20개의 마커를 각 관절, 손가락 끝에 부착하고  
추가로 손목 근처에 3개를 삼각형 모양으로 부착해서 총  
양손에 46개의 마커를 부착



마커 정보, 각 플레이의 MIDI도 기록, 피아노 장면도 녹화

# 1. Build Environment(Data Capturing – PianoHand2.5M)

## 4. 기록하는 정보

- 마커 정보
- 각 연주의 MIDI
- 연주 중 피아노 장면
- 피아노 건반의 실시간 깊이(최대 10mm) ~ 피아노 건반 뒤에 내장된 IR(적외선) 센서

## 5. 추가 사항

- 데이터 수집 절차를 생명윤리위원회(IRB)에 승인을 받았다고 함

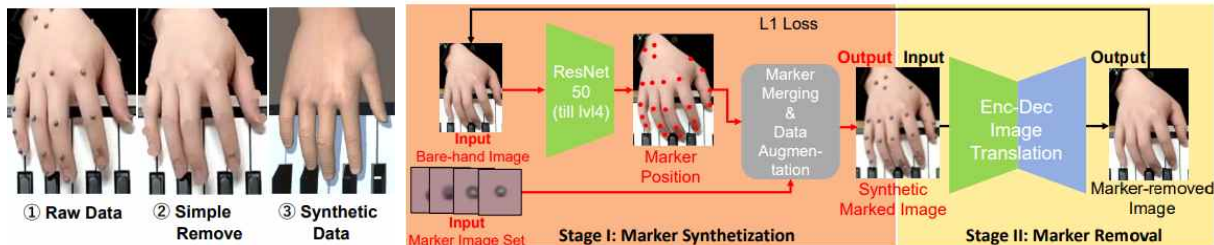


## 2. Generation of Data(PianoHand2.5M)

MR-Net(마커 제거 네트워크) & PiaNET

## 2. Generation of Data(PianoHand2.5M)

- 합성 마커 생성을 이용한 마커 제거 네트워크(Marker-removal Network)



⑤ Proposed Method: *Marker-removal Network* using synthetic marker generation

- MR-Net은 2단계로 구성
  - 1단계: 마커 합성
    - 맨손 이미지에서 ResNet50 백본을 통해 전달하여 marker estimation을 위한 시각적 특징 추출
    - 시각적 특징을 기반으로 2d 마커를 생성(size:2x21)
      - 백본은 직접 고른 CycleGAN 기반의 몇 가지 좋은 결과를 사용하여 사전 훈련되었음(is pretrained)
  - 2단계: 마커 제거

## 2. Generation of Data(PianoHand2.5M)

- GAN 기반 마커 제거와 마커 제거 네트워크의 몇몇 실패 사례

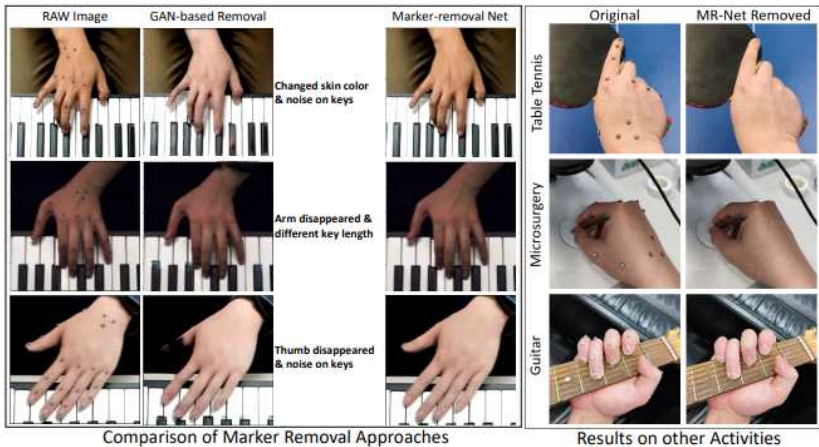


Figure 4. (Left) Some failures cases of GAN-based marker removal compared with the proposed marker-removal net. (Right) Examples of MR-Net on the other hand activities.

## 2. Generation of Data(PianoHand2.5M)

- 데이터셋 세부정보

| Data Split  | Task                 | Subjects | Keystroke | No. of Frame |      |       |              |
|-------------|----------------------|----------|-----------|--------------|------|-------|--------------|
|             |                      |          |           | R            | L    | B     | Sum          |
| Train (EP)  | R,L,B1-B6            | 6        | no        | 150K         | 110K | 400K  | 660K         |
| Train (GP)  | R,L,B1-B11           | 10       | yes       | 263K         | 217K | 834K  | 1314K        |
| Val (EP+GP) | B1-B6(EP),B1-B11(GP) | 1+1      | no        | 0            | 0    | 155K  | 155K         |
| Test (EP)   | Free Play (R,L,B)    | 1        | no        | 23K          | 17K  | 65K   | 105K         |
| Test (GP)   | Free Play (R,L,B)    | 2        | yes       | 50K          | 36K  | 166K  | 252K         |
| Total       | /                    | 21       | /         | 513K         | 403K | 1696K | <b>2486K</b> |

Table 1. The details of the PianoHand2.5M Dataset. EP: data taken in the electric piano studio. GP: data taken in the grand piano studio.

- 13명의 여성 피아니스트와 8명의 남성 피아니스트
- 평균 연령: 27.75
- 대부분 4살 때 시작했으며 평균 경력 22.7년
- 그 중 2명은 국제적인 콘서트에서 연주한 전문 피아니스트, 4명은 피아노 강사 또는 관련 전문가, 나머지는 피아노 전공 학생
- PiaNet에 사용될 훈련 세트: 2.0M(6EP, 10GP), 검증 세트 및 테스트 데이터 세트 표 참고

## 2. Generation of Data(PianoHand2.5M)

- 진행 과정

- 그룹 A

- 전문가 2명과 학생 6명이 EP(전자 피아노) 스튜디오에서 연주

- 그룹 B

- 나머지 13명의 학생은 GP(그랜드 피아노) 스튜디오에서 연주

- 각 단위별 과제는 두 실험집단 모두 동일

- 2가지 유형의 task 수행

- 1) 오른쪽(R) 또는 왼쪽(L) 손 중 한 손만 사용해서 연주

- 오른손으로 10개, 왼손으로 8개의 phrase를 연주하며 각 피실험자는 각 구절을 5번 반복해야 함
      - 그 결과, 오른손 phrases는 50개, 왼손 phrase는 40개를 얻음

- 2) 양손(L+R)을 모두 사용해서 연주

- 11개의 음악 중 특정 발체곡이 선택되고 모든 피실험자는 약 1분동안 연주 ~ 약 11분 가량의 데이터가 기록

## 2. Generation of Data(PianoHand2.5M)

- 데이터 캡처 및 PiaNET

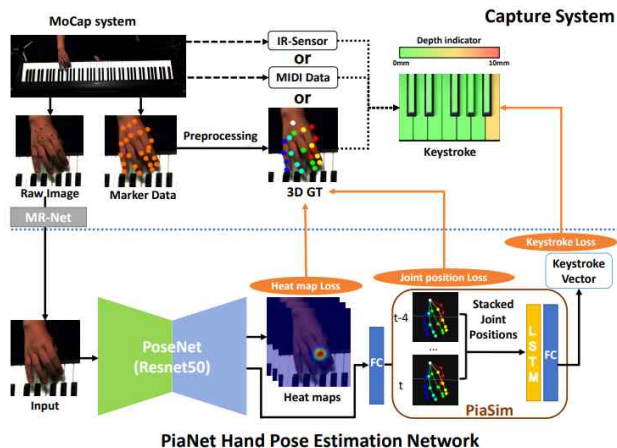
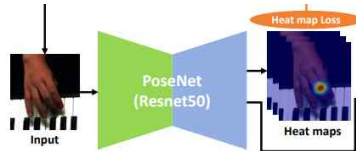


Figure 5. An overview of the data capture and the proposed PiaNET.

### 1. 3D Hand Key Point estimation

- 3D hand key point estimation을 위해 인코더-디코더 PoseNet 아키텍처와 ResNet50 백본을 사용



### 2. PiaSim 모듈

- 피아노 연주에서는 건반 누르는 순간(소리의 타이밍)을 가장 중요한 요소로 간주
- 따라서 훈련을 강화하기 위해 시계열 동작을 추출하는 LSTM 계층과 키 입력을 재현하는 완전 계층으로 구성된 PiaSim 네트워크를 개발

## 2. Generation of Data(PianoHand2.5M)

- 실험별 결과 비교

### 1. 데이터셋별 비교

| Training Set  | Result on Validation Set (Same task) |            |             |
|---------------|--------------------------------------|------------|-------------|
|               | MPJPE(mm)                            | MPJAE(°)   | PCK(%)      |
| Raw           | 14.44                                | 9.9        | 66.1        |
| Simple R.     | 12.04                                | 8.9        | 74.1        |
| Synthetic     | 11.12                                | 7.9        | 75.2        |
| GAN-based     | 9.98                                 | 7.8        | 78.8        |
| <b>MR-Net</b> | <b>9.22</b>                          | <b>7.3</b> | <b>81.6</b> |
|               | Result on Test Set (Different task)  |            |             |
|               | MPJPE(mm)                            | MPJAE(°)   | PCK(%)      |
| Raw           | 22.40                                | 13.5       | 54.3        |
| Simple R.     | 21.38                                | 13.0       | 57.2        |
| Synthetic     | 19.11                                | 12.1       | 60.2        |
| GAN-based     | 10.97                                | 8.5        | 76.7        |
| <b>MR-Net</b> | <b>9.95</b>                          | <b>7.9</b> | <b>80.3</b> |

Table 2. Result of the between-datasets comparison.

### 2. Ablation Study

| Ablation               | MPJPE       | MPJAE      | PCK         |
|------------------------|-------------|------------|-------------|
| w/o PiaSim (Direct R.) | 10.72       | 8.6        | 76.1        |
| PiaSim + pose          | <b>9.95</b> | 7.6        | 80.2        |
| PiaSim + MIDI          | 10.09       | 7.9        | 79.7        |
| PiaSim + Keystroke     | <b>9.95</b> | <b>7.5</b> | <b>80.3</b> |

Table 3. Result of the ablation study.

### 3. 방법간 비교

| Method                 | 2D Position |             | 3D Position  |             | 3D Angle Error |             |             |             |
|------------------------|-------------|-------------|--------------|-------------|----------------|-------------|-------------|-------------|
|                        | MAE         | PCK         | MPJPE        | PCK         | MCP            | PIP         | DIP         | MPJAE       |
| FreiHand [36]          | 8.4         | 89.1        | 24.47        | 61.6        | 8.09           | 12.34       | 13.33       | 11.58       |
| InterHand [15]         | 7.3         | 94.1        | 21.32        | 64.4        | 7.41           | 11.34       | 12.00       | 10.25       |
| Zhou et al. [33]       | 6.8         | 95.0        | 16.11        | 67.3        | 7.02           | 8.12        | 12.56       | 9.22        |
| PiaNet (Synthetic)     | 7.2         | 94.2        | 14.72        | 70.5        | <b>6.95</b>    | 8.39        | 11.82       | 9.05        |
| <b>PiaNet (MR-Net)</b> | <b>5.6</b>  | <b>98.1</b> | <b>12.49</b> | <b>73.7</b> | 7.04           | <b>7.12</b> | <b>7.66</b> | <b>7.30</b> |

Table 4. Quantitative results of the compared models tested on the piano sequences from the CMU Panoptic Hands dataset.