

# **다중 영상기반 신체추적 신호의 칼리브 레이션과 칼만필터 잡음 제거 기법**

**Multiple image-based body tracking  
signal calibration and noise reduction  
technique using Kalman Filter**

**한국사원기술대학교 지식기반기술·에너지대학원**

디지털엔터테인먼트학과전공

**차 종 위**

# **다중 영상기반 신체추적 신호의 칼리브 레이션과 칼만필터 잡음 제거 기법**

**Multiple image-based body tracking  
signal calibration and noise reduction  
technique using Kalman Filter**

**한국사원기술대학교 지식기반기술·에너지대학원**

디지털에터테이먼트학과전공

**차 조 위**

# 다중 영상기반 신체추적 신호의 칼리브 레이션과 칼만필터 잡음 제거 기법

지도 교수 이 태 히

이 논문을 공학석사학위 청구논문으로 제출함.

2021년 12월

한국사원기술대학교 지식기반기술·에너지대학원

디지털에터테이머트학과전공

차 조 위

**차족원이 고향선사한위 노무은 이주한**

**심사위원장** **인**

**심 사 위 원** **인**

**심 사 위 원** **인**

**한국산업기술대학교 지식기반기술·에너지대학원**

2021년 12월

# 목 차

그림목차 .....	i
국문요약 .....	ii
제1장. 서론 .....	1
제2장. 문제점 및 해결 방안.....	3
제1절 Joint 차폐 문제와 해결 방안.....	3
제2절 간헐적 잡음 발생 문제와 해결 방안 .....	9
제3장. 실험 및 실험 결과.....	13
제1절 다중 카메라를 이용한 Joint 차폐문제 해결 실험.....	13
제2절 칼만 필터 기법 사용을 이용한 간헐적 잡음 발생 문제 해결 실험 .....	15
참고문헌 .....	18
Abstract .....	20

## 그림 목차

[그림 2-1] Joint-hierarchy of Azure Kinect

[그림 2-2] Multiple Azure Kinect

[그림 2-3] Transform of skeleton

[그림 2-4] table of confidence level

[그림 2-5] Example of Confidence Level

[그림 2-6] Visualization of Confidence Level Noise

[그림 2-7] Example of Confidence Level Status Determination

[그림 2-8] Update process of Unified Skeleton data

[그림 2-9] Intermittent Noise of Body Tracking

[그림 2-10] Kalman filter Sequence

[그림 3-1] Occlusion issue of Right Hand Joint

[그림 3-2] Right hand Joint's Position Y in single and multiple cameras

[그림 3-3] Comparison of Raw data, Low Pass Filter data and Kalman Filter data

[그림 3-4] Comparison of Low Pass Filter Error, Kalman Filter Error

## 국문요약

본 논문에서는 영상 기반 신체 추적 센서인 Azure Kinect를 사용할 때의 문제점인 Joint 차폐문제와 간헐적 잡음 발생 문제를 다중 카메라 사용과 칼만 필터 기법으로 해결하는 방안을 제안한다.

우선 Joint 차폐 문제는 장애물로 인하여 Joint들이 차폐됨에 따라서 측정할 수 없는 문제를 말한다. 본 논문에서는 해결 방안으로 다중 카메라 사용을 제안한다. 다중 카메라로 연결되어 있다면 카메라마다 시점이 달라서 특정 카메라에서는 보이지 않던 Joint가 또 다른 카메라에서는 보일 수 있기 때문이다.

두 번째로 간헐적 잡음 발생 문제가 있다. Joint들이 장애물에 의하여 차폐되지 않아도 간헐적으로 잡음이 포함된 데이터가 측정되는 문제이다. 본 논문에서는 칼만 필터 기법을 해결 방안으로 제시한다. Azure Kinect의 데이터에 잡음이 포함되고 이전시점과 현재 시점이 선형적인 관계를 가지고 있기 때문에 최적화된 잡음 제거 기법이다.

Joint 차폐문제는 단일 카메라와 다중 카메라로 신체 추적 기능을 사용했을 때 실험을 통하여 해결됨을 확인하였다.

또한, 간헐적 잡음 발생 문제는 오른손 Joint의 Position을 추적하여 나온 결

과를 Raw data, Low Pass Filter data, Kalman Filter data순으로 비교하여 Filtering 한 Data와 Raw Data와의 오차를 계산하여 비교하는 실험을 진행하였다.

실험결과 1프레임부터 102프레임까지 측정하였을 때, Low Pass Filter의 오차 평균은 80.802(mm), Kalman Filter의 오차 평균은 21.723(mm)으로 나타났다.



## 제 1 장. 서론

최근 AR/VR 기술이 발전함에 따라서 RGB 와 depth 센서가 결합된 RGB-D 센서 카메라가 보편화 되었고 다양한 분야에서 널리 사용되고 있다.

마이크로소프트(Microsoft)사도 이와 같은 흐름에 발맞춰 VR 과 AR 을 경험할 수 있게 해주는 Azure kinect 를 만들었다. Azure kinect 는 물리적인 장치를 신체에 부착하지 않아도 RGB 카메라와 심도 카메라를 이용하여 사용자의 신체를 인식하여 게임과 엔터테인먼트를 경험할 수 있게 해주는 주변 기기이다. 키넥트는 심도 카메라(Depth Camera)와 RGB 카메라를 이용하여 사람의 움직임에 대한 3 차원 깊이 정보를 추적하여 신체 관절의 위치와 각도를 추적한다. 이를 신체 추적(Body Tracking)기능이라고 한다. 이 기능을 이용하여 신체 내부 안에 있는 관절(Joint)들을 감지하여 관절의 위치와 각도 값을 측정할 수 있다. 2011 년 공개된 키넥트의 소프트웨어 개발키트(Software Development Kit)로 인하여 일반 개발자들도 키넥트를 이용하여 자신만의 프로그램을 만들 수 있게 되었다. 실제로 키넥트는 연구 및 산업분야에서 다양한 형태로 활용되고 있다.

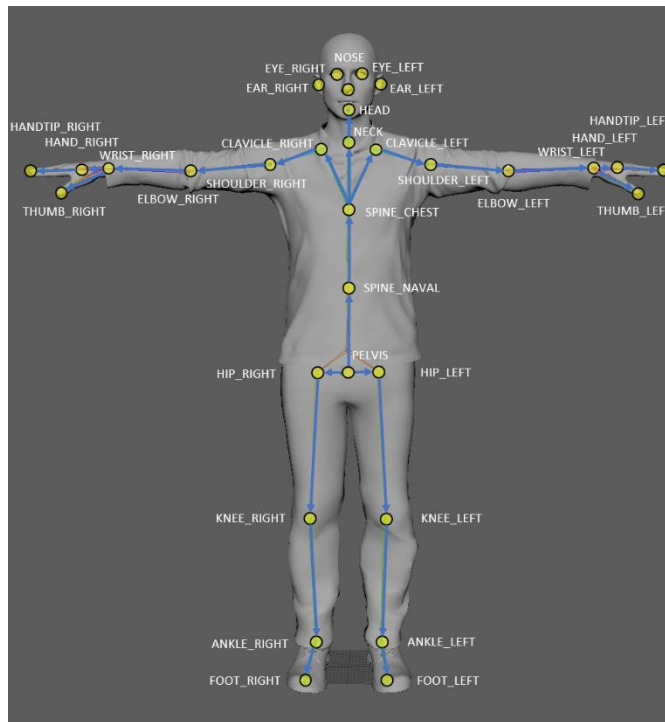
Azure Kinect 의 depth 카메라는 적외선 센서를 이용하는데, 신호가 너무 강하거나 약할 때, 다중 경로 간섭 등의 이유로 일부 픽셀에 대해 올바른 값을 제공하지 않을 수 있다. 이 때문에 Azure kinect 의 신체 추적 기능에서는

추적한 신체 스켈레톤의 관절에 오차가 포함되는 현상이 나타나는 것이다. 이 오차가 포함된 값 즉, 로데이터(raw data)를 사용하여 증강현실을 제작한다면 실제 움직임과 다르게 보이게 되어 사용자들은 몰입을 제대로 할 수 없을 것이다. 본 논문에서는 다중 카메라 칼리브레이션과 칼만 필터 알고리즘을 사용하여 키넥트의 신체추적 기능으로 얻어진 관절 값의 오차를 보정하는 방법을 제시한다.

## 제 2 장. 문제점 및 해결방안

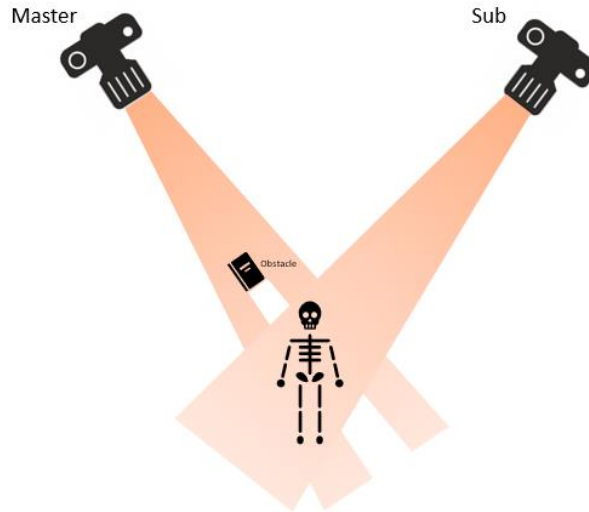
### 제 1 절 Joint 차폐문제와 해결 방안

Joint 란 신체의 관절 중 Azure Kinect 가 감지할 수 있는 관절을 뜻한다.



[그림 2-1] Joint-hierarchy of Azure Kinect

단 한 대의 카메라를 사용할 경우에는 장애물 또는 본인의 신체에 의하여 가려진 Joint 는 관측되지 않기 때문에 다중 카메라를 사용하면 가려진 부분을 다른 카메라가 대신 관측하여 오차를 줄일 수 있는 장점이 있다.



[그림 2-2] Multiple Azure Kinect

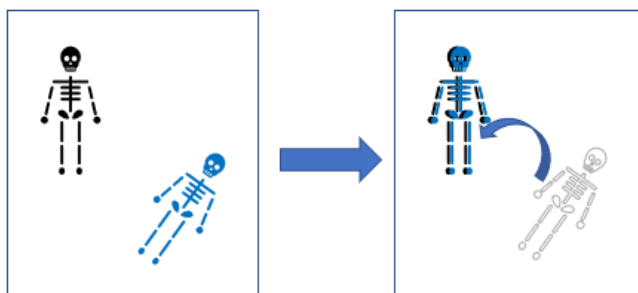
하지만 카메라마다 위치와 각도가 다르기 때문에 모든 카메라를 칼리브레이션 하는 단계가 필요하다. 칼리브레이션의 기준점은 스켈레톤의 root Joint 인 Pelvis Joint 로 설정했다.

칼리브레이션은 다음과 같은 순서를 거친다. 먼저, 연결되어 있는 카메라 중 하나를 Master 카메라로 설정하고 나머지 카메라는 Sub 카메라로 설정한다. 그 다음 모든 카메라들이 Pelvis Joint 의 Position 과 Orientation 을 10 초간 추적하고 그 값들의 평균 값을 저장한다. 10 초가 지나고, Master 카메라를 제외한 모든 카메라는 Master 카메라 내의 pelvis joint 의 Position, Orientation 의 평균 값에서 각 카메라의 pelvis joint 의 좌표, 회전 값을 뺀 값을 저장한다. 이를 변환 파라미터라고 한다. 이 값은 Master 카메라를 제외한 모든 카메라의 스켈레톤을 Master 카메라의 스켈레톤의 Position, Orientation 을 통합시키기 위한 작업에 필요하다. 이 과정 중에는 높은 정확도의

칼리브레이션 결과를 얻기 위해 일관성 있는 기준점이 필요하기 때문에 측정시에는 최대한 신체를 움직이지 않아야 한다.

칼리브레이션 작업이 끝난 이후의 모든 Sub 카메라에서 측정된 Joint 들은 본래 측정된 Position, Orientation 값에 변환 파라미터를 뺀 값을 사용하여 Sub 카메라의 좌표계를 Master 카메라의 좌표계로 변환한다.

이 때, 모든 카메라의 스켈레톤을 변환 파라미터를 통하여 Master 카메라의 스켈레톤으로 통합한 것을 '통합 스켈레톤'(Unified Skeleton)이라고 정의한다.

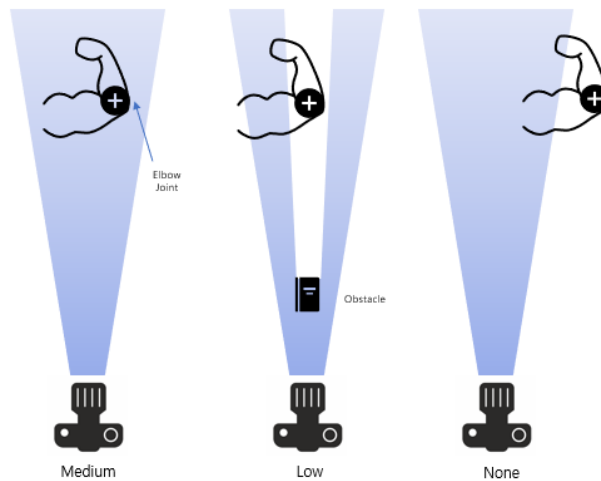


[그림 2-3] Transform of skeleton

Confidence Level 이란 현재 카메라에서 관측되고 있는 Joint 데이터가 얼마나 믿을 수 있는 데이터인지를 나타내는 값이다.

None	The joint is out of range (too far from depth camera)
Low	The joint is not observed (likely due to occlusion), predicted joint pose
Medium	Medium confidence in joint pose. Current SDK will only provide joints up to this confidence level
High	High confidence in joint pose. Placeholder for future SDK
Count	The total number of confidence levels.

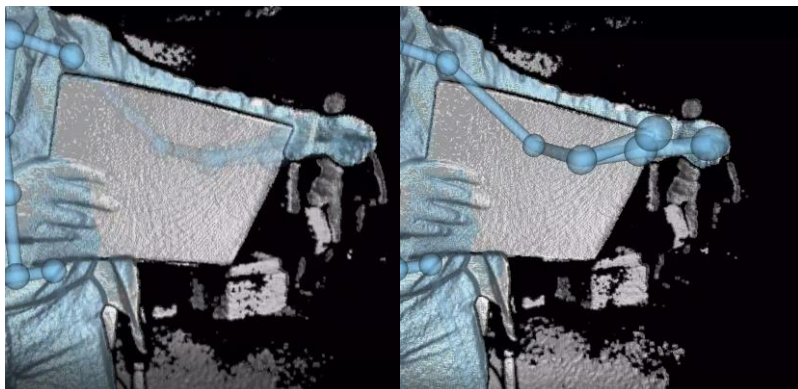
[그림 2-4] table of confidence level



[그림 2-5] Example of Confidence Level

[그림 2-4]의 왼쪽 예시처럼 Joint 가 카메라 영역 안에 존재하고 장애물에 의하여 차폐되지 않았다면 Medium 레벨, 카메라 영역 안에 있지만 장애물에 의해서 차폐된 Joint 는 Low 레벨, 카메라 영역 밖에 있다면 None 레벨로 측정된다.

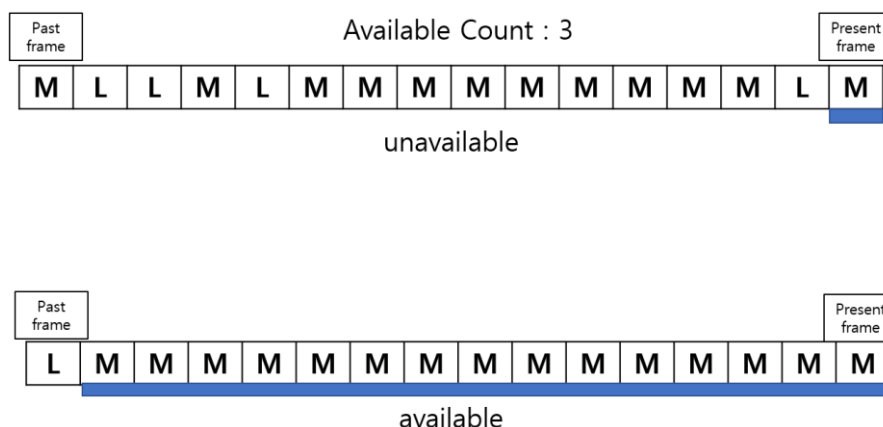
하지만 Confidence Level 에는 잡음이 존재한다.



[그림 2-6] Visualization of Confidence Level Noise

[그림 2-6]은 왼팔 Joint 가 카메라 영역 안에 있고 장애물에 의하여 차폐되어 있지만 Low 레벨 또는 Medium 레벨로 측정되는 것을 보여주는 사진이다.

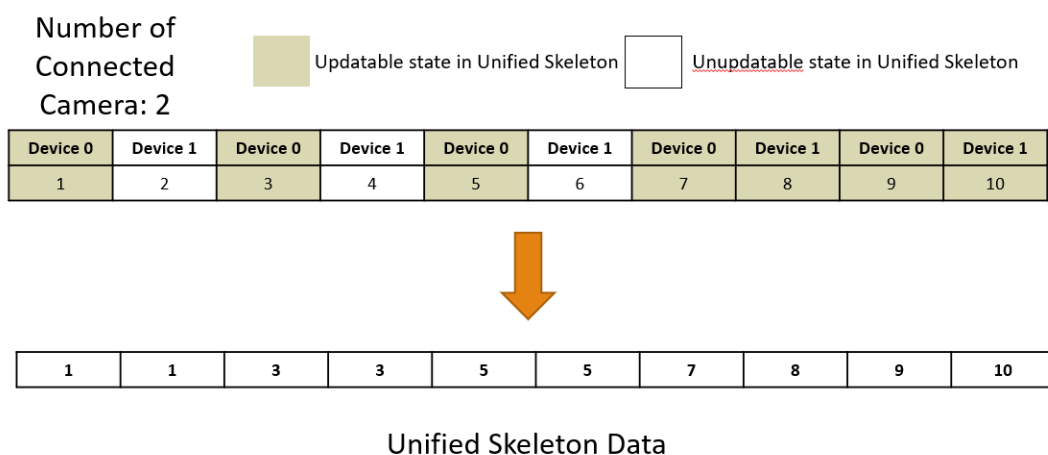
Confidence level 을 이용하여 현재 측정된 Joint 의 data 가 통합 스켈레톤에 업데이트가 가능한 상태인지에 대한 여부를 결정한다.



[그림 2-7] Example of Confidence Level Status Determination

[그림 2-6]을 예시로 들어 스켈레톤 업데이트 상태 결정 알고리즘을 설명하겠다. Available Count 라는 값이 3 이라고 가정하고 [그림 2-6]과 같이 Joint 의 Confidence level 이 측정되었다고 하자. Unavailable 카메라는 현재 프레임 기준으로 하여 연속 1 프레임이 Medium 으로 측정되었다. 반면 Available 카메라는 현재 프레임 기준으로 하여 연속 15 프레임이 Medium 으로 측정되었다. 이럴 때 연속으로 측정된 Medium 프레임의 수가 Available Count 의 값보다 크면 통합 스켈레톤 업데이트 가능 상태라고 정의한다. 따라서 Available 카메라는 현재

프레임 기준으로 통합 스켈레톤에 업데이트가 가능한 상태이며 Unavailable 카메라는 현재 프레임 기준으로 업데이트가 불가능한 상태이다.

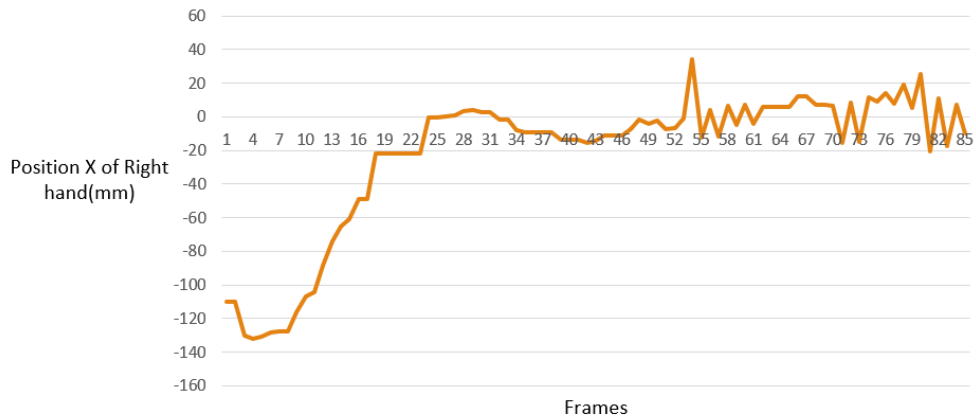


**[그림 2-8] Update process of Unified Skeleton data**

[그림 2-8]과 같이 각 카메라의 스켈레톤 데이터는 통합 스켈레톤 업데이트 가능 상태일 때만 통합 스켈레톤의 데이터를 업데이트하고 그렇지 않다면 업데이트 하지 않는다.



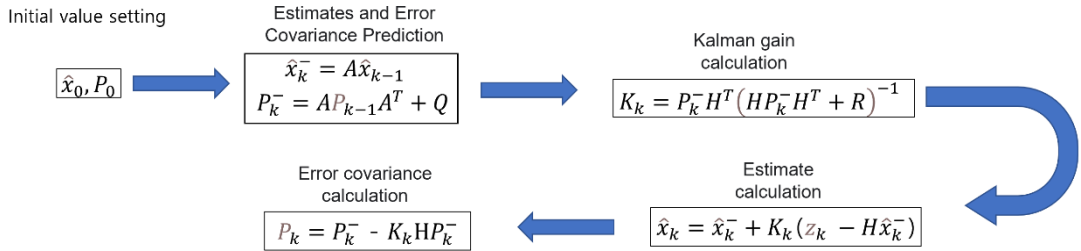
## 제 2 절 간헐적 잡음 발생 문제와 해결 방안



[그림 2-9] Intermittent Noise of Body Tracking

Azure kinect 의 신체 추적 기능을 사용하여 Joint 의 Position, Orientation 값을 측정할 때, 간헐적으로 잡음이 발생하는 문제가 있다. 이 문제는 Joint 가 장애물에 의하여 차폐되지 않을 때도 발생한다. [그림 2-9]는 시간이 흘러감에 따라서 오른손 Joint 의 Position 중 X 값이 변하는 것을 보여준다. 측정 시에는 오른손의 Joint 가 카메라 영역 안에 있고 차폐되지 않았지만 50 프레임 부근부터 잡음이 포함되어 있는 모습을 볼 수 있다. 이 문제를 칼만 필터(Kalman Filter) 기법으로 해결할 수 있다.

칼만 필터는 오차가 포함되어 있는 측정치를 바탕으로 오차를 없애는 필터링 기법 중 하나이다. 칼만 필터 알고리즘은 5 단계들로 이루어져 있다.



[그림 2-10] Kalman filter Sequence

칼만 필터는 [그림 2-10]과 같은 단계로 이루어져 있다. 알파벳 위에 -표시는 추정 값이라는 의미이고 k 는 현재 단계를 의미한다. 첫 번째 단계는 초기값을 설정하는 단계이다. 두 번째 단계는 추정 값과 오차 공분산을 예측하는 단계이다.

$$P_k^- = A P_{k-1} A^T + Q \quad (\text{eq.1})$$

이전 값과 현재 값이 선형 관계일 경우에 현재 값은 상수와 이전 단계 측정 값의 곱으로 나타낼 수 있고 이 때 상수를 A 라고 한다. 하지만 A 를 곱하는 식을 쓴다고 하여 실제 값과는 차이가 있을 것이다. Q는 수학식으로 묘사한 전 측정 값과 현재 측정 값의 차이를 보정하는 상수이다. P 는 오차 공분산(Error Covariance) 즉, 추정 값과 실제 값의 오차를 Gaussian distribution 을 그릴 때의 분산 값이다.

세 번째 단계는 칼만 이득 계산 단계이다.

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (\text{eq.2})$$

칼만 이득  $K$ 는 1 에 가까울수록 센서로부터 측정된 값과 실제 값이 같다는 것을 의미한다. 반대로 0 에 가까울수록 센서 값을 믿지 못하고 이전 단계의 결과 값을 사용하여 예측한 값을 사용하겠다는 의미이다.  $H$ 란 만약  $\hat{x}_k^-$ 가 [위치; 속도]라고 할 때, 위치를 기준으로 잡아 칼만필터를 연산하고 싶다면  $H = [1, 0]$ 이 될 것이고 속도를 기준으로 하고 싶다면  $H = [0, 1]$ 이 된다.  $H$ 로 원하는 값을 필터링 할 수 있는 것이다. 그리고 어느 센서라도 실제 값과 차이가 0 일 순 없다.  $R$ 은 그 차이를 보정하는 상수이다.  $R$ 이 증가하게 되면  $K$ 는 감소하는 구조. 즉, 센서의 오차가 클수록 다음 단계의 측정 값을 그대로 사용하지 않고 계산을 통한 예측 값을 더 신뢰하겠다는 뜻이다. (eq.2)의 식을 계산하면 칼만 이득을 계산할 수 있다.

네 번째 단계는 추정 값 계산이다.

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad (\text{eq.3})$$

$z_k$  는  $k$  단계에서의 측정 값을 의미한다. 이전 세 단계에 걸쳐 구해진 칼만 이득 값을 이용하여 현재 상태의 값을 추정하는 단계이다.

다섯 번째 단계는 오차 공분산 계산이다.

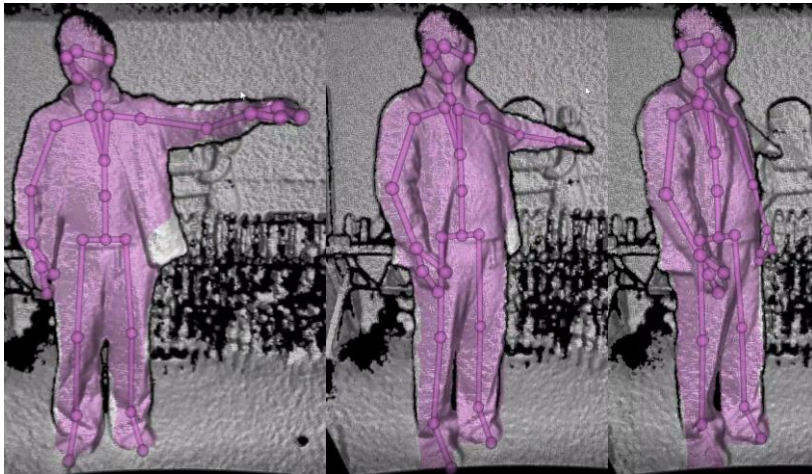
$$P_k = P_k^- - K_k H P_k^- \quad (\text{eq.4})$$

다음 단계를 위해 새롭게 오차 공분산을 계산하는 단계이다. 이 다섯 단계를 거치면 칼만 필터를 이용하여 값을 예측할 수 있다.

이 기법을 통합 스켈레톤에 적용하여 문제점이었던 간헐적으로 발생하는 잡음 문제를 해결할 수 있다.

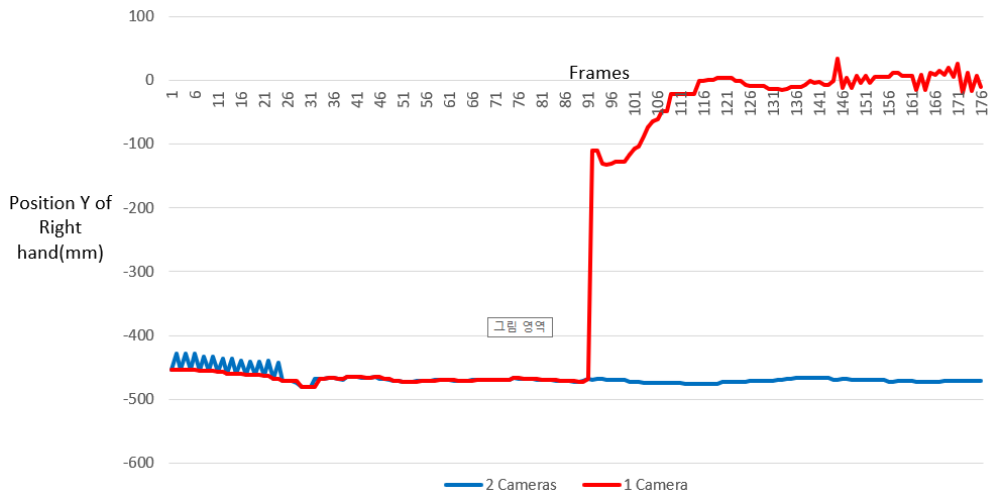
### 제 3 장. 실험 및 실험 결과

#### 제 1 절 다중 카메라 사용을 이용한 Joint 차폐 문제 해결 실험



[그림3-1] Occlusion issue of Right Hand Joint

[그림 3-1]과 같이 오른손을 옆으로 뻗은 후 점차 팔을 카메라의 반대 방향으로 향하게 하는 동작을 수행 하였을 때, 단일 카메라를 사용한 경우와 다중 카메라를 사용한 경우의 오른손 Joint 의 Y Position 을 비교하여 그래프로 나타내는 실험을 진행 하였다.



**[그림 3-2] Right hand Joint's Position Y in single and multiple cameras**

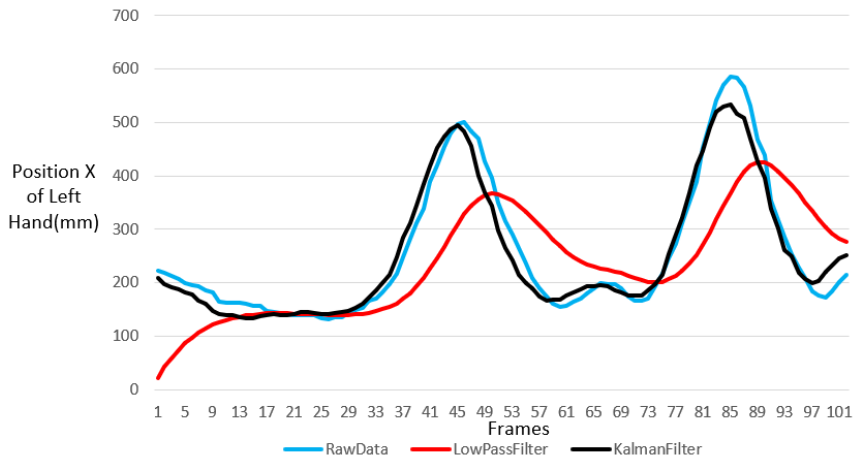
[그림 3-2]의 그래프는 0 프레임부터 176 프레임 사이의 오른손 Joint의 Position 중 높이(Y) 값을 나타낸 것이다. 높이가 낮을수록 Y의 값은 커진다. [그림 3-1]의 동작을 취하게 된다면 오른손 Joint의 높이 값은 거의 일정해야 하지만 카메라를 하나만 사용하였을 경우, 90 프레임 부근에서 값이 비정상적으로 높아지는 현상이 나타난다. 이는 90 프레임 부근부터 오른손이 몸에 의하여 차폐되어 나타나는 현상이다. 반면 2개의 카메라를 사용하였을 경우, 몸이 오른손을 차폐하여도 또 다른 카메라가 오른손을 관측할 수 있기 때문에 값이 일정하게 측정된 것을 알 수 있다.

## 제 2 절 칼만 필터 기법을 이용한 간헐적 잡음 발생 문제 해결 실험

이 실험은 신체 추적 기능으로 얻은 Joint Data 에 간헐적으로 잡음이 발생하는 현상을 칼만 필터 기법으로 해결할 수 있는지 알아보는 실험이다.

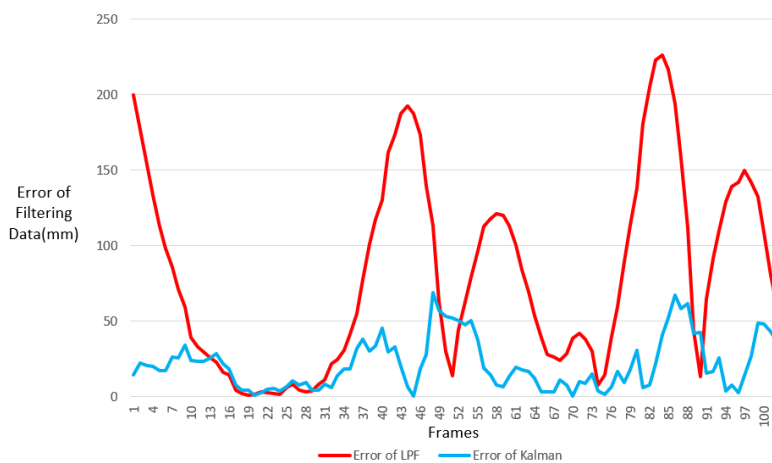
실험에서는 Raw data 와 Raw data 에 Low Pass Filter 기법을 사용한 data, 그리고 Raw data 에 칼만 필터 기법을 사용한 data 를 비교하고 잡음 제거 성능 개선이 이루어졌는가를 알아본다.

Low Pass Filter 는 잡음을 제거할 수 있는 유명한 기법이지만 Latency 가 낮아 최근에 측정된 data 를 빠르게 따라가지 못한다는 단점이 있는데, 칼만 필터의 Latency 가 Low Pass Filter 와 비교하였을 때 성능 향상이 있는지 알아본다. 단, Raw data 는 간헐적으로 잡음이 발생하는 현상이 발생하기 때문에 진리 값은 Raw data 를 인위적으로 오차를 하나하나 제거한 data 를 사용하였고 Low Pass Filter 와 Kalman Filter 는 잡음이 포함되어 있는 Raw data 를 이용하여 Filtering 했다.



**[그림 3-3] Comparison of Raw data, Low Pass Filter data and Kalman Filter data**

[그림 3-3]에 나와있는 그래프를 보면 Low Pass Filter Data 는 잡음이 제거된 모습이지만 Latency 가 높아 최근 측정된 데이터를 빠르게 따라가지 못하는 모습을 보였다. 반면 칼만 필터 Data 는 잡음이 제거되었음은 물론 Latency 가 개선되어 최근 데이터를 Low Pass Filter 보다 빠르게 따라가고 있는 모습을 보였다.



Error of LPF =  $\text{abs}(\text{Raw data} - \text{LPF})$   
 Error of Kalman =  $\text{abs}(\text{Raw data} - \text{Kalman})$



### **[그림 3-4] Comparison of Low Pass Filter Error, Kalman Filter**

#### **Error**

[그림 3-4]의 그래프는 1 프레임부터 102 프레임까지 Raw data 와 Low Pass Filter 와의 오차, Raw data 와 Kalman Filter 와의 오차를 시각화한 것이다. 여기서 오차라는 의미는 Raw data 값에서 각 필터링 데이터를 빼고 절대값으로 변환한 값을 뜻한다. 전체 평균은 Low Pass Filter 가 80.802, Kalman Filter 가 21.723 으로 상당한 개선이 이루어진 것을 확인할 수 있다.

## 참고문헌

- [1] Azure Kinect DK documentation, <https://docs.microsoft.com/ko-kr/azure/kinect-dk/>
- [2] 대한전자공학회 2009 년도 정보 및 제어 심포지움 논문집 2009 May 07 , 2009 년, pp.207 – 209, Object tracking using Kalman filter
- [3] Dan Simon, "Kalman Filtering", Embedded Systems Programming, pp 72-79, June, 2001
- [4] 제어·로봇·시스템학회지 = iCROS v.15 no.4 , 2009 년, pp.38 - 42, 이형근, 벡터, 좌표계, 좌표값 그리고 좌표변환행렬
- [5] 한국정보통신학회지 = The magazine of KIICE v.18 no.1 , 2017 년, pp.36 - 42 ,홍진우

## **Abstract**

Multiple image-based body tracking signal calibration and noise reduction  
technique using Kalman Filter

by Cha Jong-won

Advisor: Prof. Lee Teakhee, Ph.D

Course for Department of Digital  
Entertainment

Graduate School of Knowledge-

In this paper, we propose a method to solve the joint Occlusion issue  
intermittent noise issue, which are issues when using Azure Kinect, and image-  
based body tracking sensor, using multiple cameras and the Kalman filter  
technique.

Joint Occlusion issue refers to an issue that cannot be measured as the joints  
are shielded due to obstacles. In this paper, we propose the use of multiple  
cameras as a solution. This is because, if multiple cameras are connected, the  
viewpoints are different for each camera, so a joint that is not visible in one  
camera can be seen in another camera.

Second, there is the issue of intermittent noise generation. It is an issue that  
data containing noise is intermittently measured even if the joints are not

occluded by obstacles. In this paper, the Kalman filter method is presented as a solution. This is an optimized noise removal technique because the data of Azure Kinect contains noise and the previous time point, and the current time point have a linear relationship.

It was confirmed that the joint occluding issue was solved through experiments when the body tracking function was used with a single camera and multiple cameras.

For the issue of intermittent noise generation, an experiment was conducted to calculate and compare the error between the filtered data and the raw data by comparing the results of tracking the position of the right-hand joint in the order of raw data, Low Pass Filter data, and Kalman filter data.

As a result of the experiment, when measuring from frame 1 to frame 102, the average error of the Low Pass Filter was 80.802(mm) and that of the Kalman Filter was 21.723(mm)