

레터논문 (Letter Paper)

방송공학회논문지 제25권 제5호, 2020년 9월 (JBE Vol. 25, No. 5, September 2020)

<https://doi.org/10.5909/JBE.2020.25.5.808>

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

## 랜덤 무향 칼만 필터를 이용한 RGB+D 카메라 센서의 잡음 보정 기법

권오설<sup>a)‡</sup>

### Noise Reduction Method Using Randomized Unscented Kalman Filter for RGB+D Camera Sensors

Oh-Seol Kwon<sup>a)‡</sup>

#### 요 약

본 논문은 랜덤 무향 칼만 필터를 이용하여 키넥트 카메라 센서의 오차를 최소화하는 방법을 제안한다. RGB 값과 깊이(Depth) 정보를 제공하는 RGB+D 카메라는 센서의 오차로 인해 뼈대 검출 과정에서 성능 저하의 원인을 제공한다. 기존에는 다양한 필터링 기법을 이용하여 오차를 제거하였으나 비선형 잡음을 효과적으로 제거하는데 한계가 있었다. 이에 본 논문에서는 비선형 잡음 특성을 예측하고 업데이트하기 위해 랜덤 무향 칼만 필터를 적용하였으며 이를 바탕으로 뼈대 검출 성능을 높이고자 하였다. 실험 결과 제안한 방법은 기존의 방법에 비해 정량적 오차를 줄였으며 뼈대의 3D 검출 시 우수한 결과를 확인할 수 있었다.

#### Abstract

This paper proposes a method to minimize the error of the Kinect camera sensor by using a random undirected Kalman filter. Kinect cameras, which provide RGB values and depth information, cause nonlinear errors in the sensor, causing problems in various applications such as skeleton detection. Conventional methods have tried to remove errors by using various filtering techniques. However, there is a limit to removing nonlinear noise effectively. Therefore, in this paper, a randomized unscented Kalman filter was applied to predict and update the nonlinear noise characteristics, we next tried to enhance a performance of skeleton detection. The experimental results confirmed that the proposed method is superior to the conventional method in quantitative results and reconstructed images on 3D space.

Keyword : Kinetic camera, RUKF, and skeleton detection

a) 창원대학교 전기전자제어공학부(School of Electronical Electronics and Control Engineering, Changwon National University)

‡ Corresponding Author : 권오설(Oh-Seol Kwon)

E-mail: [oskl@changwon.ac.kr](mailto:oskl@changwon.ac.kr)

Tel: +82-55-213-3669

ORCID: <http://orcid.org/0000-0002-1077-9615>

\* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning No.2019R1F1A1058489).

· Manuscript received August 31, 2020; Revised September 17, 2020; Accepted September 17, 2020.

## I. Introduction

일반적으로 의료영상 분야에서 인체 내 정보를 획득하기 위해서 X-ray, CT, MRI 등의 장비를 사용하고 있다. 이러한 장비들은 정확성은 우수하나 고가이면서 특히 병원과 같은 특정한 장소에서만 사용할 수 있는 단점이 있다. 최근 마이크로소프트사에서는 RGB 값과 깊이(Depth) 정보를 바탕으로 저가의 키넥트 카메라를 이용하여 뼈대 및 관절 정보를 검출하는 인터페이스를 개발하였다<sup>[1]</sup>. 이 시스템을 이용하면 사용자의 32개의 관절을 실시간으로 추적할 수 있기 때문에 의료분야뿐만 아니라 증강현실 및 HCI 분야<sup>[2]</sup>에서 다양한 응용이 가능하다. 뿐만 아니라 제공된 인터페이스 기반의 뼈대 검출을 통해 자세 인식을 포함한 체형 유지와 건강 증진 등으로 활용할 수 있는 장점이 있다. 그러나 마이크로소프트사에서 제공된 RGB+D 기반 키넥트 카메라는 저가형으로써 센서의 특성이 기존 의료제품에 비해 우수하지 않는 단점이 있다. 특히 뼈와 관련된 데이터는 잡음이 많이 포함되어 있기 때문에 보정 단계가 필요하다. 특히 관찰자의 움직임이 없는 경우에도 센서의 변동성이 심한 경우에는 오차가 많이 발생하여 자세 인식 등의 응용에서 성능이 떨어지는 원인이 된다. 이를 해결하기 위해서 기존의 방법은 다양한 필터 등을 이용하여 센서에서 발생하는 잡음을 제거하고자 하였다. 가장 대표적인 기법이 바로 칼만 필터(Kalman Filter)를 이용하는 것이다. 이 방법은 측정값에 대한 확률적인 오차가 특정 시점에서의 상태가 이전 시점의 상태와 선형적인 관계를 가지고 있는 특성을 이용하여 오차를 최소화하는 방법이다. 그러나 키넥트 센서와 같이 잡음이 비선형적인 특성을 가질 경우 오차를 줄이는데 한계가 있었다. 이를 보완하는 방법이 무향 칼만 필터(Unscented Kalman Filter, UKF)이다<sup>[3]</sup>. 이 방법은 기존의 선형모델 가정을 기반으로 하는 칼만 필터와 달리 비선형 모델을 위해 개발된 방법으로 평균 주변에 시그마 점(sigma point)의 최소 집합을 얻기 위하여 무향 변환 기법을 사용한다. 최근에는 Amir et al.<sup>[4]</sup>은 발목에 부착된 센서로부터 UKF 방법을 적용하여 엉덩이나 몸통의 자세를 추정하는 방법을 제안하였다. 센서 입력에 대한 UKF 적용에서 시그마 점들은 비선형 함수를 통해 전달되고, 변환된 점들에 대해 평균과 공분산을 계산함으로써 보다 정확한 결과를 얻

을 수 있게 된다. 본 논문에서는 애저 키넥트 카메라(Asure Kinetic Camera)에서 제공된 데이터의 잡음을 제거하고 3D 뼈대 검출을 실시간으로 제공하는 방법을 제안한다. 제안한 시스템에서는 RGB+D 카메라 센서 오차를 최소화하기 위하여 랜덤 무향 칼만 필터(Randomized Unscented Kalman Filter, RUKF)를 기반으로 비선형적 잡음을 제거하고 이를 통해 3D 자세를 보정하였으며 RMSE 계산을 통해 정량적으로 분석하였다.

## II. 제안한 RGB+D 카메라로부터 실시간 뼈대 검출 잡음 제거 시스템

마이크로소프트사의 키넥트 카메라는 RGB 영상과 함께 깊이(Depth) 정보를 제공한다. RGB+D 카메라로 입력된 정보를 바탕으로 SDK 인터페이스에서는 초당 8프레임씩 관절에 관한 정보를 3D 정보로 제공한다. 이것은 기존의 저가형 장치들이 손을 이용하거나 혹은 옷을 입는(Wearable) 형태 방법들에 비해서 매우 간단하고 특별한 조작이 없이 카메라 앞에 서는 것만으로도 신체 정보를 획득할 수 장점이 있다. 그러나 RGB+D 카메라의 센서는 비선형적 잡음을 매우 많이 포함하고 있기 때문에 고정된 자세에서도 관절 포인트의 위치가 시간에 따라 변하게 되는 한계가 있다. 이를 해결하기 위해서 본 논문에서는 랜덤 무향 칼만 필터를 이용한 잡음 제거 기법을 적용하여 관절 정보에 대한 잡음을 최소화하고 3D 뼈대 검출의 정확도를 높이고자 하였다. 그림 1은 RGB+D 카메라 및 SDK에서 제공된 관절 포인트의 예시와 실제 시뮬레이션으로 구성된 3D 공간에서의 결과이다. 먼저 카메라의 출력인 3차원 관절 포인트 정보  $x$ 는 각각의 형태가 카메라 관찰함수(Observation function)  $g$ 를 통해  $y$ 로 변환된다. 랜덤 변수  $x \in R^{n_x}$ 와  $y \in R^{n_y}$ 는 다음과 같은 수식으로 표현된다.

$$y = g(x) = [g_1(x), \dots, g_{n_y}(x)]^T \quad (1)$$

랜덤 변수  $x$ 는 평균  $\hat{x}$ 와 공분산 행렬  $P_x$ 로 계산할 수 있으며, 또한  $y$ 의 평균  $\hat{y}$ , 공분산  $P_y$ ,  $x$ 와  $y$ 의 교차 공분산 행렬  $P_{xy}$  등을 계산할 수 있다.

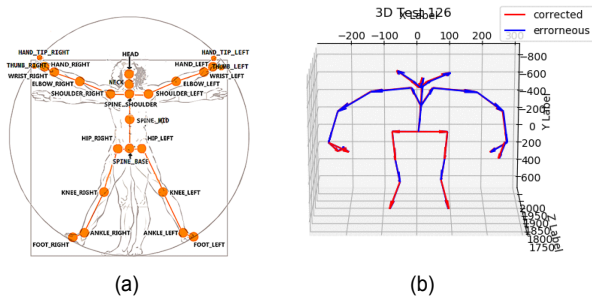


그림 1. RGB+D 카메라에서 제공되는 관절 포인트의 개념[1]과 실제 카메라 입력, (a) 관절 포인트, (b) 3D 시뮬이션 결과  
 Fig. 1. Concept of skeleton joints and simulated camera responses by the RGB+D cameras, (a) Joint point, (b) 3D simulation

다음으로 무향 변환(Unscented Transform)의 개념으로의 확장은 평균 주변에 시그마 점( $\sigma$ -point)으로 불리는 샘플 포인트의 최소 집합  $\{\chi_i\}$ 를 얻기 위해 결정론적인 샘플링 기술을 사용한다. 이 시그마 점들은 비선형 함수를 통해 전달되고, 변환된 점들에 대해 새로운 평균과 공분산을 구하는 형태를 가진다. 이때  $\{\chi_i\}$ 에 대응되는 가중치를  $\{\omega_i\}$ 로 정의한다[5].

$$\chi_0 = \hat{x}, \omega_0 = \frac{k}{n_x + k}, \quad (2)$$

$$\chi_i = \hat{x} + (\sqrt{(n_x + k)P_x})_i, \omega_i = \frac{k}{2(n_x + k)}, \quad (3)$$

$$\chi_{n_x+i} = \hat{x} - (\sqrt{(n_x + k)P_x})_i, \omega_{n_x+i} = \omega_i, \quad (4)$$

여기에서  $i = 1, \dots, n_x$ 이고, 이때 각 시그마 점( $\sigma$ -point)은 비선형 함수로써 다음과 같이 변형된다.

$$y_i = g(\chi_i), \forall i, \quad (5)$$

그 결과 무향 변환(Unscented Transform, UT)의 결과는 다음과 같이 표현할 수 있다.

$$\hat{y}^{UT} = \sum_{i=0}^{2n_x} \omega_i y_i, \quad (6)$$

$$\hat{P}_y^{UT} = \sum_{i=0}^{2n_x} \omega_i (y_i - \hat{y}^{UKF})(y_i - \hat{y}^{UKF})^T, \quad (7)$$

$$\hat{P}_{xy}^{UT} = \sum_{i=0}^{2n_x} \omega_i (\chi_i - \hat{x}^{UKF})(y_i - \hat{y}^{UKF})^T, \quad (8)$$

랜덤 무향 변환은 기존의 무향 변환의 개선된 특징을 가지고 있다. 비록 무향 변환이 랜덤 변수  $x$ 에 관한 계산으로 설계되어 있으나 통계적 적분 기법[5]을 적용하여 다음과 같이 근사화가 가능하다.

$$\mu = \int_{R^n} \varphi(x) (2\pi)^{-n_x/2} e^{-\frac{1}{2}(x-\hat{x})^T P^{-1}(x-\hat{x})} dx, \quad (9)$$

여기에서  $\varphi(x)$ 는 임의의 비선형 함수이며, 수식 (9)는 랜덤 변수  $x$ 로 이루어진 함수  $\varphi(x)$  기댓값 (Expected values), 즉  $\mu = E[\varphi(x)]$ 을 계산하는 과정이다.

### III. 실험 및 결과

제안한 알고리즘의 성능을 평가하기 위하여 16개의 동작에 대해 실험하였다. 하나의 동작은 약 30초간의 동영상으로 구성되며 하나의 프레임은 또한 32개의 관절 정보를 제공하기 때문에 최종적으로는 약 10만개의 관절에 대한 잡음 데이터를 분석하였다. 인체를 구성하기 위해서는 3D 좌표를 두 개의 모델(상체와 하체)로 나누어서 구성되며, 각 모델은 기본(root)좌표에 대한 위치변환 매트릭스를 이용하여 상대 오일러각과 뼈의 길이로 표현된다. 그림 2는 다양한 관절 포인트 중에서 Chest, Ankle, Hand, Neck에 대한 프레임별 데이터를 기존 UKF 기법과 제안한 방법에 대하여 적용한 결과이다. 모션이 없는 입력이기 때문에 모든 관절의 위치 변화가 없어야 하는 것이 이상적이거나 센서의 값은 시간에 따라 변동성이 심하게 나타나는 것을 확인할 수 있다. 각 4 가지 경우에 대하여 위쪽 행은 UKF 방법과 RUKF 방법의 결과이며 아래쪽 행은 원본(original)과의 오차이다. 대부분은 푸른(Blue)선에 의한 피크가 나타나는 지점에서 오차가 많이 발생하는 것을 확인할 수 있다. 표 1은 4개의 관절 포인트에 대한 그에 대한 정량적 결과이다. 카메라 응답을 Kinect로 나타내었으며, UKF 방법과 제안한 방법의 결과를 원본에 대하여 RMSE와 표준편차로 비교하였다. 카메라 응답의 오차가 UKF방법을 이용하여 줄었으나, 제안한 방법은 기존의 방법보다 모든 관절에서 오차 및 편차에서 줄어듦을 확인하였다.

그림 3은 3D 좌표상에서 자세에 따른 알고리즘별 성능을 비교한 결과이다. 그림 3은 16개의 자세 중에서 4개를 선택

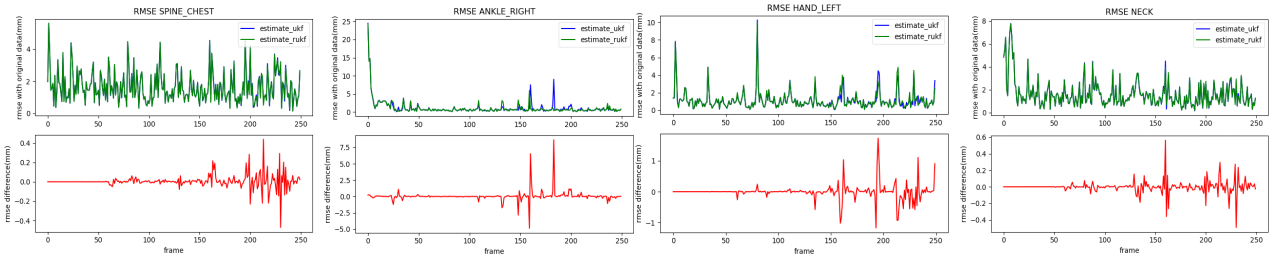


그림 2. 프레임 분석에 의한 UKF와 RUKF 방법의 시간 오차(temporal error) 성능 비교  
 Fig. 2. Comparison of UKF and RUKF methods

표 1. 알고리즘에 따른 RMSE 및 표준편차 오차 비교  
 Table 1. Quantitative evaluation of RMSE and its standard deviation

no. of joints	Kinect	UKF	Proposed
Spine_Chest	12.14±8.27	7.42±4.16	1.63±0.96
Knee_Right	10.23±4.53	6.52±4.51	1.68±1.69
Pelvis	10.21±6.93	4.76±2.63	2.51±1.46
Neck	9.94±6.65	4.73±2.72	2.30±1.31

두 알고리즘 모두 원본과 유사하게 도출되었으나 움직임이 있는 아래 행에서는 제안한 방법이 원본과의 오차가 줄어들음을 확인할 수 있다. 이것은 랜덤 무향 칼만 필터가 기존의 방법에 비해 상대적으로 움직임이 있는 경우 발생하는 잡음을 잘 예측하고 제거할 수 있다는 결과이다.

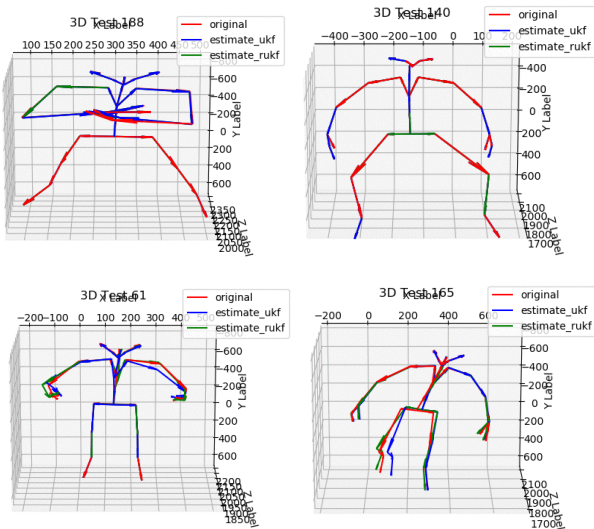


그림 3. 3D XYZ 좌표상에서 자세에 따른 알고리즘별 성능 비교  
 Fig. 3. Comparison of algorithm's performance for each pose on 3D XYZ coordinates

하고 자세별로 하나의 프레임을 캡처한 것이다. 기존의 원본 좌표를 붉은색(red), UKF 결과를 푸른색(blue), 제안한 방법을 녹색(green)으로 표현하였다. 자세의 움직임에 따른 결과를 분석하기 위해 위쪽 행은 비교적 움직임이 없는 데이터이고 아래 행은 상대적으로 자세 변화가 많은 데이터로 분류하였다. 실험 결과 움직임이 없는 쪽에서는 결과가

#### IV. 결론

본 논문은 RGB+D 카메라 센서 입력에 대해 랜덤 무향 칼만 필터를 이용하여 잡음을 제거하는 기법을 제안하였다. 랜덤 무향 칼만 필터는 예측 단계와 업데이트 과정을 거치면서 잡음 제거의 성능을 높일 수 있었으며, 이를 통해 획득된 관절 정보를 바탕으로 3D 뼈대 검출 단계에서 성능을 향상시킬 수 있었다. 실험한 결과 제안한 방법은 기존의 방법에 비해 잡음의 제거 성능이 우수함을 확인하였다.

#### 참고 문헌 (References)

- [1] Microsoft Robotics. Kinect Sensor. from <http://msdn.microsoft.com/en-us/library/hh438998.aspx>. (accessed Oct. 06, 2017)
- [2] G. Han, I. Lee, H. and Choi, "MPEG-U based advanced user interaction interface system using hand posture recognition," *The Korean Institute of Broadcast and Media Engineers*, Vol. 19, No. 1, pp. 83-95, 2014.
- [3] M. Goulao, "Kinematic data filtering with unscented kalman filter," Thesis of Tecnico Lisboa, 2017.
- [4] A. Baghdadi, L. Cavuoto, and J. Crassidis, "Hip and trunk kinematics estimation in gait through kalman filter using IMU data at the ankle," *IEEE Sensors Journal*, Vol. 18, No. 10, pp. 4253-4259, 2018.
- [5] O. Straka, J. Dunik, and M. Simandl, "Randomized unscented kalman filter in target tracking," *IEEE International Conference on Information Fusion*, Singapore, pp. 503-510, July 2012.