Journal of The Korea Society of Computer and Information
Vol. 17 No. 7 , July 2012

2012-17-7-4-2

비마커 증강현실을 위한 색상 및 깊이 정보를 융합한 Mean-Shift 추적 기반 손 자세의 추정

이 선 형, 한 헌 수**, 한 영 준**

The Estimation of Hand Pose Based on Mean-Shift Tracking Using the Fusion of Color and Depth Information for Marker-less Augmented Reality

Sun-Hyoung Lee*, Hernsoo Hahn**, Youngjoon Han**

요 약

본 논문은 비마커 증강현실(Marker-less Augmented Reality)을 위한 색상 및 깊이 정보를 융합한 Mean-Shift 추적 알고리즘 기반 손 자세의 추정 기법을 제안한다. 기존 비마커 증강현실의 연구는 손을 검출하기 위해 단순한 실험 배경에서 피부색상 기반으로 손 영역을 검출한다. 그리고 손가락의 특징점을 검출하여 손의 자세를 추정하므로 카메라에서 검출할 수 있는 손 자세에 많은 제약이 따른다. 하지만, 본 논문은 3D 센서의 색상 및 깊이 정보를 융합한 Mean-Shift 추적 기법을 사용함으로써 복잡한 배경에서 손을 검출할 수 있으며 손 자세를 크게 제약하지 않고 손 영역의 중심점과 임의의 2점의 깊이 값만으로 정확한 손 자세를 추정한다. 제안하는 Mean Shift 추적 기법은 피부 색상정보만 사용하는 방법보다 약 50픽셀 이하의 거리 오차를 보였다. 그리고 증강실험에서 제안하는 손 자세 추정 방법은 복잡한 실험환경에서도 마커 기반 방법과 유사한 성능의 실험결과를 보였다.

▸ Keyword :비마커 증강현실 , Mean Shift 추적 기법, 3D 센서, 손 자세, 손 검출

Abstract

This paper proposes a new method of estimating the hand pose through the Mean-Shift tracking algorithm using the fusion of color and depth information for marker-less augmented reality. On

• 제1저자 : 이선형 • 교신저자 : 한영준
• 투고일 : 2012. 05. 08, 심사일 : 2012. 05. 22, 게재확정일 : 2012. 05. 29.
*숭실대학교 전자공학과(Dept. of Electronic Engineering, Soongsil University)
** 숭실대학교 정보통신전자공학부 (Dept. of Information Communication & Electronic Engineering, Soongsil University)
※ 감사의 글 : "본 논문은 2012년 정부 교육과학기술부(한국학술진흥재단)의 2단계 BK21사업의 재원으로 지원을 받아 수행된 연구임."
※ "본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 육성지원 사업의 연구결과로 수행되었음" (NIPA-2012-H0301-12-2006)

marker-less augmented reality, the most of previous studies detect the hand region using the skin color from simple experimental background. Because finger features should be detected on the hand, the hand pose that can be measured from cameras is restricted considerably. However, the proposed method can easily detect the hand pose from complex background through the new Mean-Shift tracking method using the fusion of the color and depth information from 3D sensor. The proposed method of estimating the hand pose uses the gravity point and two random points on the hand without largely constraints. The proposed Mean-Shift tracking method has about 50 pixels error less than general tracking method just using color value. The augmented reality experiment of the proposed method shows results of its performance being as good as marker based one on the complex background.

▸ Keyword : Marker-less Augmented Reality , Mean Shift Tracking, 3D Sensor, Hand Pose, Hand Detection

I. 서 론

증강현실(Augmented Reality)은 가상과 현실을 혼합한 개념으로서 1990년대 중반부터 사람에게 필요한 정보를 효과적으로 제공하기 위하여 활발히 연구되고 있다. [1, 2]

증강현실은 그래픽스 기술과 디스플레이 장치, 컴퓨터 기술이 발전되면서 컨텐츠 개발에 중요하다. 증강현실 시스템에서는 사용자와 컴퓨터 간의 상호작용을 위하여 실세계 환경의 특징점을 이용하여 객체를 등록하는 과정이 필요하다. 기존 마커 기반 증강현실 시스템은 영상 내에 항상 마커가 존재하여야 하며 기울기, 조명의 영향에서 자유롭지 못하다. [3] 이러한 이유와 사용자의 몰입감 향상을 위해 다양한 비마커 기반 증강현실 시스템에 대한 많은 연구가 수행되고 있다. [3] 그 중 추가적 장비가 필요하지 않고 객체 조작 과정에서 몰입감 향상에 효과적인 손을 이용한 연구들이 활발하게 진행되고 있다. [4, 5, 6]

손을 이용한 증강현실에서 손은 가상객체와 실제물체 사이를 연결하므로 손의 정확한 자세의 추정이 전체 시스템의 몰입감과 그 기능에 중요한 영향을 준다. 손의 자세를 추정하기 위하여 손 영역 검출 및 추적, 좌표계 생성의 과정이 필요하다. 손 영역의 검출단계에서는 피부색상을 검출 및 영상분할이 필요하다. 안정적으로 손 영역을 검출하기 위해 손 영역을 추적하는 방법이 사용되어야 한다. 그리고 손 영역의 검출 후에 자세 정보를 위하여 평면 생성 및 방향을 결정해야 한다.

논문[3]은 Rotating Calipers algorithm을 이용해 손 영역 최적 사각형을 찾고 공간좌표를 생성하여 스크린 좌표로 투영 시키는 방법을 제시하였다. 피부색 기반으로 손을 검출하기 때문에 조명에 의한 깜박임 현상으로 손의 특징 점 변화가 일어나기 때문에 추가적으로 칼만 필터를 사용하여 보완하였다. Shen는 Distance transform 알고리즘으로 손 끝점을 찾고 이 점을 기준으로 객체 좌표계를 생성하였다. [4] 하지만 손 영역을 영상 내 가장 큰 객체로 제한하였다. Chun의 연구에서도 Distance transform을 사용하여 손 끝점을 찾은 뒤 좌표를 생성하였으며 피부색과 구분하기 위해 단순한 배경을 사용하였다. [5] Choi의 연구에서는 입력 손 영상과 데이터 베이스 간의 히스토그램 유사도를 판단하여 손 자세를 정의하는 방법을 제시하였다. [6]

기존의 연구들은 일반적으로 손을 배경으로 구분하기 위해 실험환경을 단순한 배경으로 제한한다. 현실공간을 제한함으로써 증강현실 시스템의 현실-가상공간의 융합 목적 측면에서 몰입감 저하시킨다. 또한 2차원 영상내의 정보로 3차원 객체 자세 정보를 얻기 위해 [4, 5]에서는 복잡한 검출 기법을 사용하여 손 끝의 특징 점들을 추출한다. 본 연구에서는 이런 문제들을 해결하기 위해 3D센서로부터 획득된 색상 및 깊이 정보를 융합한 Mean-Shift 추적 기반 손 자세 추정 기법을 제안한다. 먼저 색상 영상에 infrared laser 센서에서 획득한 깊이 값을 동일한 점으로 대응시키기 위해 카메라 보정을 수행한다. 피부 색상정보를 이용하여 손 영역을 검출한다. 본 논문에서는 검출된 손 영역을 안정적으로 검출하기 위해 3D 센서의 색상 정보와 깊이 정보를 융합한 특징을 갖는 Mean-shift 추적 기법을 사용한다.

검출할 수 있는 손 자세를 크게 제한하지 않으면서 추적된 손 영역의 중심점과 손 바닥의 임의의 2개 점만을 이용하여 손 자세를 추정한다. 제안하는 손 자세의 추정 기법은 3D 센

서가 제공하는 깊이와 색상 정보를 융합하여 특징벡터를 사용하기 때문에 기존 연구들과 달리 복잡한 배경에서도 안정적으로 손의 자세를 추정한다. 색상 및 깊이 정보의 이변량확률밀도함수(Bivariate Probability Density Function)을 사용한 Mean-Shift 추적 기법은 피부 색상의 물체들이 놓인 실험환경에서도 강인하게 손을 추적한다. 또한 추적된 손 영역의 중심점과 손 바닥의 임의의 2개 점만을 이용하여 정확한 손 자세를 추정한다.

본 논문의 구성은 다음과 같다. 2장에서 색상-깊이 영상 사이의 대응 점 추출을 위한 색상 및 깊이 영상의 정합, 3장은 색상-깊이 정보를 융합한 특징벡터 구성 및 Mean-Shift 알고리즘의 추적 방법, 4장에서는 깊이 값을 이용한 손 자세 추정 방법 5장은 실험 및 결과를 보인다. 6장에서 본론 및 향후 연구를 기술하였다.

II. 3D 센서 카메라 보정을 통한 색상 및 깊이 영상의 정합

제안하는 Mean-shift 알고리즘은 색상 및 깊이 정보를 융합하여 손 영역을 추적한다. 두 정보를 융합하기 위해 우선 카메라 보정 기술을 이용하여 깊이 영상의 좌표점을 칼라 영상의 대응점으로 정합하는 과정이 필요하다.

그림 1. Microsoft 사의 KINECT 센서
Fig. 1. KINECT Sensor (microsoft)

본 논문의 실험에서 3D센서로 Microsoft 사의 KINECT 센서를 사용하였다. KINECT 센서는 그림 1와 같이 중앙 칼라 카메라와 infrared laser 센서로 구성된다.

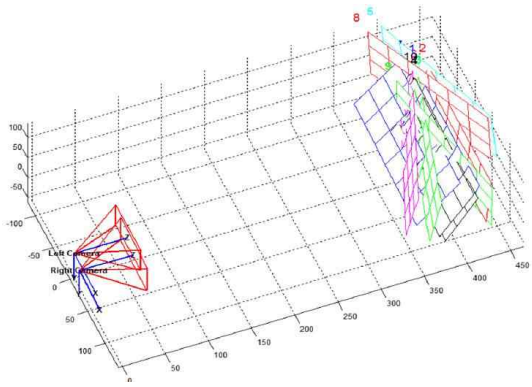

그림 2. 키넥트 센서 보정(Calibration)
Fig. 2. KINECT Sensor calibration

그림 2는 Zhang의 카메라 Calibration 툴을 이용한 두 카메라의 위치 관계를 보여준다(left :색상, Right : 깊이 카메라). 각 카메라의 영상을 얻기 위하여 OpenNI 라이브러리를 사용하였다.[9]

총 두 센서 각각 12장의 이미지를 사용하였으며 카메라 보정 결과 infrared laser 센서는 칼라 카메라를 기준으로 기준선(Base Line)을 따라 2.46 cm 오른쪽에 위치해 있으며 회전은 없는 것으로 나타났다

그림 3는 infrared laser 센서와 칼라 카메라의 관계를 보여준다. 또한 수식 (4)~(6)으로부터 infrared laser센서의 좌표 점을 칼라 영상의 좌표 점으로 대응시키는 변환 관계를 보여준다.

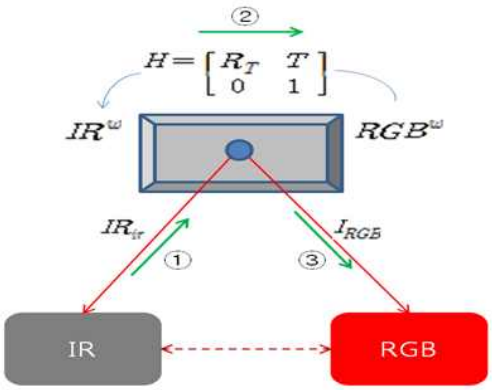

그림 3. 깊이 영상과 칼라 영상의 관계도
Fig. 3. Relation between Depth and color map

그림 3는 infrared laser 센서와 칼라 카메라의 관계를 보여준다. 또한 수식 (4)~(6)으로부터 infrared laser 센서의

좌표 점을 칼라 영상의 좌표 점으로 대응시키는 변환 관계를 보여준다. 수식 (1)는 infrared laser 센서에서 획득된 깊이 영상 좌표계와 infrared laser 센서의 좌표계 관계를 보여준다. 이 식으로부터 2D 깊이 영상 좌표 점을 infrared laser 센서의 3D 좌표 점으로 대응시킬 수 있다.

$$IR^{W} = IR_{ir}^{-1} \times IR \quad (1)$$

여기서, IR은 깊이 영상의 좌표 점, IR_{ir}은 infrared laser 센서의 내부 보정 행렬 (Internal Calibration Matrix), 그리고 IR은 infrared laser 센서의 좌표 점이다. 수식 (2)는 infrared laser 센서의 좌표계와 칼라 카메라의 좌표 관계를 보여준다. 이 식으로부터 칼라 카메라의 좌표 점을 칼라 영상의 좌표 점으로 대응시킬 수 있다.

$$RGB^{W} = H^{-1} \times IR^{W} \quad (2)$$

여기서 H는 infrared laser 센서의 좌표계와 칼라 카메라의 좌표계 사이의 관계를 표현하는 변환 행렬 (Transformation Matrix)이다. 그리고 RGBw는 칼라 카메라의 좌표점이다. 식 (3)은 칼라 카메라의 좌표계와 칼라 영상의 좌표계의 관계를 표현한다.

$$RGB = I_{RGB} \times IR_{ir}^{-1} \times IR \quad (3)$$

위 식으로부터 깊이 영상의 한 점으로부터 칼라 영상의 대응점을 계산할 수 있다. 그림 7은 정합 전의 영상과 식 (3)에 의해 정합 후의 영상을 보여준다. 정합 전 영상인 그림 7(a)에서 볼 수 있듯이 3D 센서로부터 멀어질수록 오차가 커지는 것을 확인할 수 있다. 예를 들면, 그림 4(b)와 달리 천장에 형광등 위치가 상당히 불일치 하는 것을 볼 수 있다.

그림 4. 정합 전(좌) 후(우) 및 정합 결과 영상
Fig. 4. Mapping result images

III. 색상 및 깊이 정보를 융합한 Mean-Shift 추적 알고리즘

본 논문에서 사용한 Mean-Shift 추적 기법은 검출된 손 영역을 초기 탐색 윈도우의 색상 정보와 깊이 정보를 융합하여 특징벡터를 구성한다. Mean-Shift 추적 알고리즘은 추적 영역의 특징 값의 분포와 현재 관심 영역의 특징 값의 분포로부터유사도(Similarity)를 계산하여 그 중심점을 찾는 알고리즘이다.

1. Mean-Shift 추적 알고리즘의 특징벡터의 분포

Mean-shift 알고리즘의 성능은 추적하고자 하는 물체의 측정값으로 구성된 특징벡터에 따라 크게 좌우된다. 일반적으로 사용되는 Mean-Shift 알고리즘은 추적물체에 대해 다양하게 측정된 값들을 표본화(Sampling)하여 다음과 같이 양자화(Quantization)한 1차원 특징벡터를 구성한다.

$$X(N) = [X_1(N_1) | \ldots X_k(N_k) | \ldots X_n(N_n)] \quad (4)$$

여기서, 특징 벡터 X의 양자화 비트(bit)는 N ($N_1 + N_2 + \cdots + N_n$) 이다. $X_k(N_k)$는 추적물체에 관한 k 번째 표본화 측정 값으로 N_k비트로 양자화한 값이며 특징벡터 X(N)의 양자화 N비트 중에서 N_k비트를 구성한다. 예를 들면, 칼라 값을 사용하는 특징벡터를 갖는 Mean-Shift 알고리즘은 256단계 갖는 R 채널(Channel)에 대해 16개 값을 표본화하여 4비트로 양자화하고, 유사하게 G 채널과 B 채널 값을 각각 4비트로 양자화하여 총 12 비트로 특징벡터를 구성한다.

N 비트 양자화한 특징벡터는 X = { x | x ∈ {x0, x1, … x2N} }로 표현할 수 있다. 여기서 x는 표본공간 상의 특징벡터 값이다. 그리고 표본공간 상의 특징 벡터 값의 최대 개수는 2N이다.

Mean-Shift 알고리즘에 사용되는 1차원 특징벡터 값의 분포를 얻기 위해 특징 벡터 값의 확률밀도함수(Probability Density Function)을 구한다. 다음의 식 (5)은 임의의 특징벡터에 대한 확률 밀도 함수를 나타낸다.

$$p_x(x_k) = \frac{n_k}{n} \quad (5)$$
$$0 \le p_x(x_k) \le 1, \sum_{k=0}^{L-1} p_x(x_k) = 1$$

여기서, n은 특징 벡터 X의 전체 측정 개수이고, nk는 k

번째 특징벡터 값 xk를 갖는 측정값의 개수이다. 그림 5는 1차원 칼라 특징벡터 값으로 표현된 확률밀도함수의 예를 보여준다.

기존 Mean-Shift 알고리즘은 수행 시간과 알고리즘의 복잡도를 줄이기 위해 각각의 특징벡터 값을 식(4)와 같이 적은 표본화 개수로 양자화 한다.[12]

이들 양자화한 값들을 직렬화하여 1차원 특징벡터를 구성한다. 하지만 특징벡터 값의 확률분포의 분석 없이 측정값의 상위 값들만으로 표본화하여 양자화한다. 따라서 특징벡터의 분포가 갖는 고유 확률분포의 특성을 잃을 가능성이 높다.

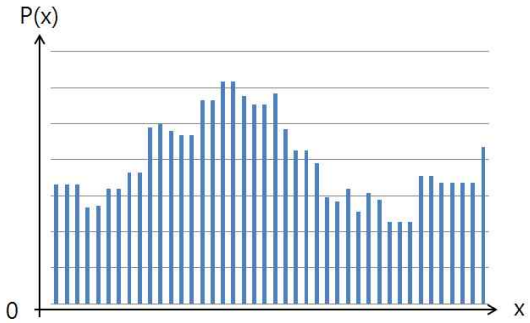

그림 5. 1차원 특징벡터(X)의 확률밀도함수
Fig. 5. PDF of 1 Dimension feature vector X

따라서 본 논문은 특징 벡터를 구성하기 위해 각 측정값의 확률분포를 분석하여 중요한 값으로 표본화한다. 각 특징벡터의 분포의 고유 특성을 최대한 유지하기 위해 다음과 같이 다차원 특징 벡터를 구성한다.

$$X = [X_1(N_1) \cdots X_k(N_k) \ldots X_n(N_n)]^T \quad (6)$$

여기서 특징벡터 X는 n개의 다른 측정값으로 구성된 n차원 벡터이다. 각 원소는 측정값을 표본화하여 다른 크기의 비트로 양자화한 값이다. 특징벡터 X의 Xn(Nn) 원소는 Nn비트로 표현되었으며 표본화 수는 2Nn이다.

N 차원 특징 벡터는 X = { X | X ∈ ($x_1\, x_2 \cdots x_k \cdots x_n$)T , $x_k \in \{x_k_0 , x_k_1, \cdots x_k_2N_k\}$ }로 표현할 수 있다. 여기서 X는 표본 공간상의 n차원 특징 벡터 값을 나타내며 x_k는 특징벡터의 k번째 원소이다. 그리고 xk 값은 2N k_r 표본화 개수를 갖는다. 다차원 특징 벡터의 분포를 얻기 위해 표본화 공간에서 측정된 특징벡터의 다변량 확률 밀도함수(Multi variate Probability Density Function)을 계산한다. 다음의 식 (7)은 임의의 N차원 특징벡터에 대한 다변량 확률밀도함수를 나타낸다.

$$p_x(x_1, x_2, \cdots, x_n) = P(X_1 = x_1, \cdots X_n = x_n) = \frac{n_{x_1 \ldots x_n}}{n}$$
$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} p_x(x_1, x_2, \cdots, x_n) = 1 \quad (7)$$

여기서, n은 표본화 공간에서 측정된 특징벡터들의 전체 개수이고, n$x_1 \cdots x_n$는 이들 중에 특징벡터 값 X=(x_1 x_2 $\cdots x_n$)를 갖는 측정값의 개수이다.

본 논문은 제안하는 Mean-Shift 알고리즘으로 손 영역 추적을 위해 3D 센서에서 측정된 색상 값(X_c)과 깊이 값(X_{ir})을 융합하기 위해 2차원 특징벡터를 구성하였다.

$$X = [\, X_{color}(8)\, X_{ir}(5)]^T \quad (8)$$

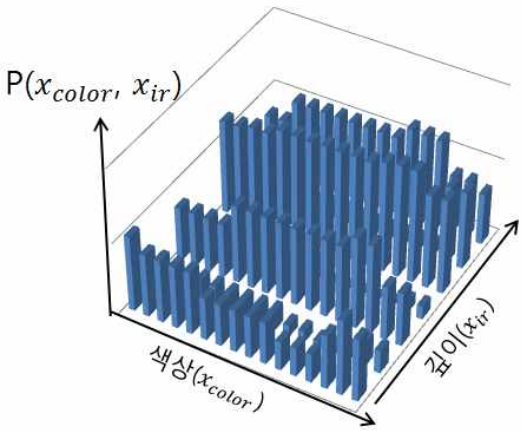

그림 6. 2차원 특징벡터(X)의 확률밀도함수
Fig. 6. PDF of 2 Dimension feature vecor X

여기서 X 특징벡터의 Xcolor(8)는 색상 값을 8비트로 양자화한 값이며 256단계를 갖는다. 그리고 X_{ir}(5)는 깊이 값을 5비트로 양자화한 값이며 32단계 값을 갖는다. 그림 6은 2차원 색상과 깊이 값으로 구성된 특징벡터의 이변량 확률 밀도함수(Bivariate Probability Density Function)를 보여준다.

서로 다른 특성을 갖는 측정 데이터 값을 융합하기 위해 측정값의 상위 값만 표본화하여 양자화를 통해 1차원 특징벡

터의 확률밀도함수를 사용하는 것보다 측정값의 중요한 값들을 표본화하여 다차원 벡터의 구성을 통해 다변량 확률 밀도 함수를 사용하는 것이 더 효과적이다.

2. Mean-Shift 알고리즘을 이용한 물체 추적

Comniciu가 제안한 Mean-Shift 추적 알고리즘[12]을 따라 이전 영상 프레임의 물체 위치 y_0과 추적하고자 하는 물체에 대한 2차원 특징벡터의 이변량 확률 밀도함수 모델 q가 주어졌다고 가정한다.

1) 현재 영상 프레임에서 물체의 초기 위치를 y0로 초기화하고 후보 특징벡터의 이변량 확률 밀도함수 모델 p(y0)를 추정한다. 그리고 Bhattacharyya Coefficient ρ[p(y0),q]를 계산한다.

$$\rho[p(y_0), q] = \sum_i \sqrt{p_i(y_0) q_i} \tag{9}$$

2) 후보 모델의 이변량확률밀도함수의 가중치(wi)를 다음 식 (xx)로 계산한다

$$w_i = \sum_i \sqrt{\frac{q_i}{p_{i(y_0)}}}\, \delta[b(x_i) - i] \tag{10}$$

3) 다음 수식 (xx)로 새로운 위치 y1를 추정한다.

$$y_1 = \frac{\sum_i x_i w_i g\left(\left|\frac{y_0 - x_i}{h}\right|^2\right)}{\sum_i w_i g\left(\left|\frac{y_0 - x_i}{h}\right|^2\right)} \tag{11}$$

4) 새로운 위치 y1에서 후보모델과 객체 모델간의 Bhattacharyya Coefficient ρ[p(y1),q]를 계산한다.

5) ρ[p(y1),q]가 ρ[p(y0),q] 보다 작을 동안 반복하여 y1 새로운 위치 값을 y0과 y1의 평균값을 y1로 대체한다.

6) |y1-y0|이 ε보다 작으면 수렴한 것으로 판단한다. 그렇지 않은 경우 y0을 y1로 대체하여 (가) 단계부터 다시 반복 수행한다.

Mean-Shift 알고리즘은 배경에 추적하는 물체와 유사한 색상을 갖는 물체가 존재하더라도 추적 물체의 색상 분포를 고려하기 때문에 다른 추적 알고리즘에 비해 성능이 뛰어나다. 하지만 추적 시간과 알고리즘의 복잡도를 줄이기 위해 특징벡터를 구성 할 때에 표본 공간을 단순화하고 양자화 수를 줄인다. 이 경우에 Mean-Shift 알고리즘은 특징벡터의 분포가 왜곡되어 관심 물체를 추적하기 어렵다. 이 문제를 해결하기 위해 본 논문은 색상 정보와 특성이 다른 거리 정보를 융합한다.

(a) 기존 방법

(b) 제안하는 방법

그림 7. 제안하는 Mean-Shift 추적 결과
(a) 기존 방법, (b) 제안하는 방법
Fig. 7. Proposed Mean-Shift tracking result
(a) Existing method by color, (b) Proposed method

앞서 설명했듯이 본 논문은 색상과 거리의 중요한 값을 표본화한 후 그 값을 양자화하여 2차원 특징벡터를 구성한다.

그림 7의 (b)는 피부색과 유사한 분포를 갖는 배경이 존재하는 경우에도 색상과 깊이 값을 융합한 2차원 특징벡터의 분포로 인해 물체를 강인하게 추적하는 것을 확인 할 수 있다.

IV. 비마커 증강현실을 위한 손 자세 추정

증강현실에서 객체를 증강시키기 위해서 객체의 자세를 결정하는 기준 좌표가 필요하다. 본 논문에서는 객체를 증강하기 위한 마커 대신에 손의 추적을 통해 검출한 손을 사용한다.

비마커 증강현실의 위한 색상 정보만을 이용하여 손을 검출하는 기존의 방법들은 실험환경의 배경을 단순화시켜 손을 검출하였다. 그리고 2차원 영상내의 정보로부터 3차원 손 자세 정보를 추정하기 위해 손 끝점을 추출한다. 이런 손의 불변 특징들을 검출해야 하기 때문에 검출한 손의 자세에 상당히 제약함으로써 증강현실 시스템의 현실-가상 공간에서 몰입감을 저하시킨다.

본 논문에서 제안하는 손 자세의 추정 기법은 3D 센서가 제공하는 깊이와 색상 정보를 융합한 2차원 특징벡터를 사용한다. Mean-Shift 추적 기법은 특징벡터의 이변량 확률 분포함수를 사용하여 피부 색상과 유사한 물체들이 놓인 실험환

경에서도 강인하게 손을 추적한다. 또한 DB를 사용하지 않고 간단하게 infrared laser 센서에서 제공하는 손 영역의 깊이 값을 이용하기 때문에 단순한 알고리즘으로 정확한 손의 자세를 측정한다.

그림 8는 비마커 증강현실을 위한 손 자세결정 방법에 대한 순서도를 보여준다. 우선 Mean-Shift 추적 기법으로 손을 추적하기 위해 피부 색상값과 거리 값을 사용하여 손 영역을 검출한다. 그림 9(a)은 피부 색상의 차 영상으로부터 검출된 손 후보 영역을 보여준다.

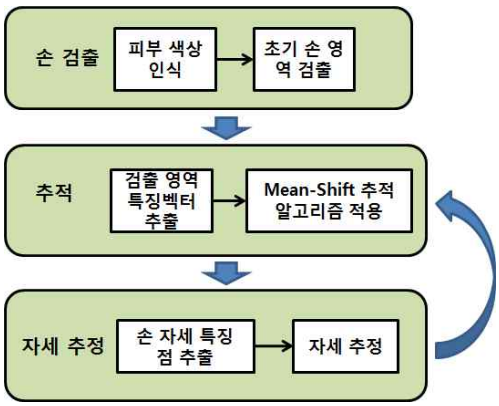

그림 8. 제안하는 손 자세 결정 알고리즘
Fig. 8. Proposed hand posture estimation algorithm

그리고 그림 9(b)는 9(a)의 손 후보 영역으로부터 최 근접 거리 값을 갖는 초기 손 영역을 검출한 결과를 보여준다. 이는 초기 손 검출단계에서 손의 배경에 포함된 피부 영역을 제거하기 위한 과정이다.

(a) 피부색 영역 검출 (b) 손 검출
그림 9. 피부색 영역 검출로부터 거리 정보를 이용한 손 검출
(a)피부색 영역 검출결과 (b) 손 검출 결과
Fig. 9. Hand region recognition by depth information
(a)Detected skin region (b) Detected hand

본 논문은 손 위에 안정적인 객체 좌표계를 설정하기 위해 손 영역의 무게 중심을 선택하였다. 수식 (12)은 Mean-Shift 알고리즘에 의해 추적된 손 영역에 대한 무게중심이다.

$$C(x,y)=\frac{\sum_{i=0}^{N} I_h(x_i,y_i)}{N} \tag{12}$$

여기서, Ih(x,y)는 Mean-Shift 추적 알고리즘에 의해 검출된 손 영역이며 N는 손 영역의 전체 화소 수이다. 손의 중심점과 손 바닥에 2 점을 검출하면 이들 3점을 포함하는 평면의 방정식을 계산할 수 있다. 손 바닥을 평면 방정식에 근사화시켜 손 자세를 추정된다.

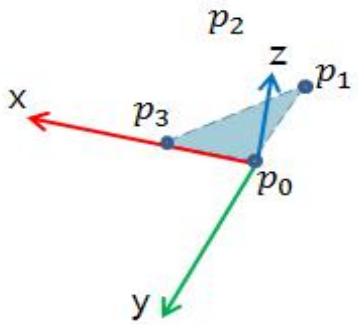

그림 10. 손 좌표계 설정
Fig. 10. Set the hand coordinate

비마커 증강현실에서 손을 마커로 사용하는 기존의 연구들은 손의 불변 특징을 사용하므로 검출하는 손의 자세에 대해 상당한 제약을 두었다. 하지만 본 논문은 infrared laser 센서로부터 손에 대한 깊이 정보를 얻을 수 있기 때문에 복잡한 알고리즘을 사용하지 않고 손의 자세를 추정한다. 그림 10는 근사화한 평면의 방정식을 통해 손 영역의 중심에 손 좌표계를 설정한 예를 보여준다. ((p_0 p_1 p_2)¯ : 삼각 평면)

손의 자세를 추정할 때에 infrared laser 센서로부터 획득된 깊이 값에 임펄스(Impulse)의 랜덤(Random) 잡음이 더해진다.

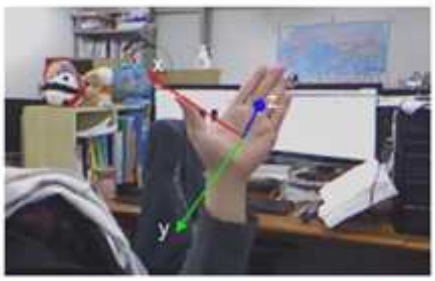

그림 11. 손 영역 위에 설정된 객체 좌표계
Fig. 11. Object coordinate on the hand

임펄스 잡음은 일반 랜덤 잡음과 달리 저주파 필터(Low

Frequency Filter)인 Smoothing 필터를 사용하여 제거하기 어렵기 때문에 주로 Median 필터로 사용한다. 영상 입력 장치나 오 검출에 의해 생기는 임펄스 잡음을 5x5 크기의 Median 필터로 제거하여 증강된 객체의 불안정한 흔들림을 제거한다. 그림 11은 임펄스 잡음을 제거한 후 증강현실을 위해 손 좌표계에 설정한 객체 좌표계를 보여준다.

객체 좌표는 위의 그림 11와 같이 영상 내의 고유한 위치를 갖고 있으며 증강현실에서 객체를 표현하기 위해 객체 좌표계를 카메라 좌표계로 변환하는 과정이 필요하다. 이를 위해 2장의 카메라 보정에서 측정된 내부 변수들의 값을 이용하여 객체 좌표를 카메라 좌표로 변환한다.

V. 실험 및 결과

본 논문은 비마커 증강현실을 위해 깊이와 색상 정보를 융합한 2차원 특징벡터의 분포를 사용하는 Mean-Shift 추적 기반 손 자세의 추정 기법을 제안하였다. 3D 센서의 색상 및 깊이 정보를 융합한 Mean-Shift 추적 기법을 사용함으로써 복잡한 배경에서 손을 검출할 수 있으며 손 자세를 크게 제약하지 않고 손 영역의 중심점과 임의의 2점의 깊이 값만으로 정확한 손 자세를 추정한다.

표 1은 본 논문에서 비마커 증강현실을 위해 손 자세의 측정을 위한 실험 환경을 보여 준다. Open Library에서 각 센서로부터 영상정보 및 깊이 정보를 얻기 위한 OpenNI, 객체 증강을 위하여 OpenGL, ARtoolkit [8]을 사용하였다. 컴파일 환경은 Windows 7 운영체제 아래 Visual Studio 2008을 사용하였고 깊이 값 센서로서 MS KINECT를 사용하였다.

표 1. 시스템 환경
Table 1. System Environment

구성		시스템
S/W	Operating System	Windows 7
	Language	C++
	Open Library	OpenCV, OpenNI OpenGL, ARtoolkit
	Compiler	Microsoft Visual Studio 2008
H/W	CPU	Intel i7 core
	RAM	4GB
	3D Sensor	MS KINECT

본 논문은 제안하는 알고리즘의 성능을 평가하기 위해 Mean-Shift 추적 실험과 증강된 객체로 손 자세의 추정 실험으로 나누었다. 제안하는 Mean-Shift 알고리즘의 손 추적 성능을 검증하기 위하여 색상 정보만을 사용한 추적 실험과 infrared laser 센서에서 획득된 깊이 정보를 융합한 2차원 특징벡터의 이변량 확률 분포함수를 사용한 비교실험을 수행하였다. 제안하는 정확한 손 자세의 추정을 위해 마커의 자세와 손의 자세를 비교 실험을 수행하였다. 그리고 증강현실에서 마커와 손 객체의 검출을 통해 가상의 객체를 증강하는 비교 실험을 수행하였다.

1. Mean-Shift 알고리즘의 손 추적 실험

손 추적 실험은 색상 정보만을 사용할 경우와 제안한 방법을 사용했을 경우를 비교 분석하였다. 입력 영상에서 추적하려는 객체인 손과 유사한 색상을 갖는 얼굴 피부색을 배경으로 가졌을 경우 두 방법의 결과를 관찰하였다.

표 2는 Ground Truth을 정보를 갖는 동일 동영상에 각각 색상정보(a)만을 이용했을 경우와 색상정보와 Depth 정보를 융합했을 경우(b) 추적 결과를 보여준다. 영상에서 프레임 별로 손이 얼굴 앞으로 지나갔을 경우 각각의 추적윈도우의 위치를 확인할 수 있다. 왼쪽 열(a)의 영상에서는 추적윈도우가 계속 얼굴에 수렴하지만 오른쪽 열(b)에서는 배경색의 영향을 받지 않고 손을 추적하는 것을 확인 할 수 있다.

또한 손 영역에 Ground Truth를 지정하고 그 중심 값과 Mean-Shift 추적 윈도우의 중심과의 유클리드 거리(Euclidean Distance)를 계산하여 식 (13)처럼 오차 E를 정의하였다.

$$E = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{13}$$

여기서, (x2,y2)는 Ground Truth에서 주어지는 손의 중심점이고 (x1,y1)는 추정된 손의 중심 값이다.

표 2. 색상정보(a)와 거리정보(b)를 융합한 Mean-Shift 추적 결과
Table 2. Mean-Shift tracking result (a) Color information (b) Merged color & depth information

Frame	(a) 색상정보	(b) 제안 방법
102		
114		
136		

그림 12는 97 프레임으로 구성된 동일 동영상에 대해 색상 정보만 사용했을 경우(파란 색)와 색상정보와 거리정보를 융합해서 사용 했을 경우(빨간 색)에 추적 오차를 보여준다. 제안하는 Mean-Shift 추석 실험은 색상정보만 사용할 경우에 평균 66.67 픽셀 오차, 색상 및 거리 정보를 융합한 경우에 평균 16.77 픽셀 오차를 보였다. 제안하는 Mean Shift 추적 기법은 피부 색상정보만 이용하는 손 추적 방법보다 약 50픽셀 이하의 거리 오차를 보였다.

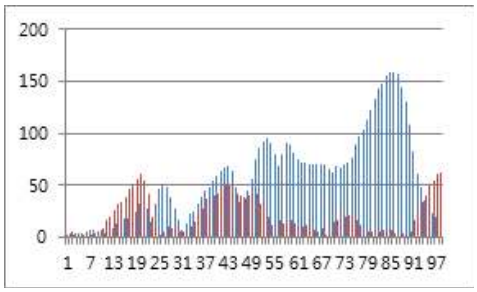

그림 12. 제안하는 Mean-Shift 알고리즘의 손 추적 오차
파란 색 : 색상정보만 이용한 결과
빨간색 : 제안한 방법에 의한 결과
Fig. 12. Proposed Mean-Shift tracking result
Blue : Result using Color information
Red : Result using Proposed method

2. 비마커 증강현실을 위한 손 자세의 추정 실험

비마커 증강현실을 위한 제안하는 손 자세의 측정 실험은 마커의 자세와 손 자세를 추정하여 비교하였다. 그림 13에 보인 것처럼 마커를 손과 동일한 자세 변화를 갖도록 손등에 마커를 부착하였다.

그림 13. 자세 표현 비교를 위한 실험환경
Fig. 13. Environment for pose experiment

그 이유는 손 위의 마커를 고정시켜 Ground Truth 생성하고 제안한 방법과 자세 변화에 대한 비교를 해야 하지만 동일한 동영상에 대하여 비교할 수 없기 때문이다.

제안하는 손 자세의 추정에 대한 정확성을 평가하기 위한 실험은 추정된 손 자세와 마커 자세의 각도 변화에 대한 유사성를 비교하였다. 아래의 그림 14에 나타난 그래프에서 각각의 두 자세에 대한 X축, Y축, 그리고 Z축의 각도변화를 보여주고 있다. (x축 :Frame, y축 : 각도(Degree)

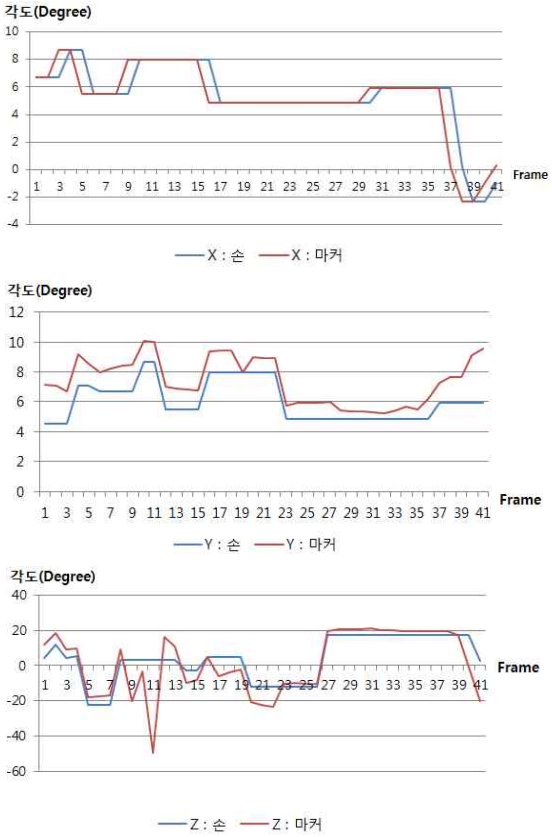

그림 14. 마커와 제안방법(손)의 축별 각도 변화
(상 : x축, 중: y축, 하 : z축)
Fig. 14. Comparing degree changes between Marker and Hand (Top : x axis, middle : y axis, bottom : z axis)

X, Y, Z 축의 평균 측정 각도 차이는 각각 0.15, 4.66, 1.00도 이하였으며 이러한 차이가 발생하는 요인은 두 좌표가 같은 평면 상에 존재하지 않는 이상 정확한 좌표축의 동기화를 할 수 없기 때문이다. 그림 14의 그래프는 손과 마커의 세 축의 각도 변화량을 보여주고 있다. 손과 연결된 마커가 동일하게 움직인다는 가정 하에 각도 변화량을 측정하였다. 세축의 각도 변화량이 모두 유사한 궤적을 그리고 있음을 볼 수 있다.

그림 15. 증강 실험을 위한 마커 환경
Fig. 15. Environment for rendering experiment based marker

그림 17의 (a)마커 기반 실험과 (b)제안하는 방법의 증강 실험에서 알 수 있듯이 제안하는 방법이 마커 기반의 방법만큼 손의 자세를 잘 추정하고 있음을 알 수 있다.

그림 16. 제안 방법 실험을 위한 배경환경
Fig. 16. Environment for rendering experiment based on proposed method

증강 실험을 위하여 그림 15과 같은 손 위에 마커를 부착하고 제안 방법과의 객체 증강 실험을 수행하였다. 또한 본 논문에서 제안한 방법의 증강실험은 그림 16과 같이 얼굴, 책상 등의 피부색을 갖는 다수의 물체를 포함한 실험환경에서 수행하였다. 그림 16은 피부색 추출 후 이진화된 영상이다.

(a)

(b)

그림 17. 피부색을 포함한 복잡한 실험환경에서 증강 실험 결과 (a) 마커에 의해 증강된 객체 (b) 제안한 방법에 의해 손 위에 증강된 객체
Fig. 17. Rendering result in complex environment including skin color (a) rendered by marker (b) rendered by proposed method on hand

VI. 결론 및 향후 연구

본 논문은 비마커 증강현실을 위한 색상 및 깊이 정보를 융합한 Mean-Shift 추적 알고리즘 기반 손 자세의 추정 기법을 제안하였다. 강인한 손 추적을 위해 3D센서에서 측정된 색상정보와 깊이 정보를 융합한 Mean-shift 추적 알고리즘을 사용하였다. 제안하는 Mean-Shift 추적 알고리즘은 복잡한 실험 배경에도 강인하게 손 영역을 추적하는 성능을 색상 정보만 사용하는 추적 방법과 비교실험을 통해 확인하였다. 그리고 infrared laser 센서로부터 획득된 손의 중심점과 손 위에 임의의 2점에 대한 깊이 값만을 사용하기 때문에 손의 자세를 크게 제약하지 않고 손 자세를 정확하게 추정하였다. 증강현실을 위해 마커 자세의 측정과 제안하는 손 자세의 추정 방법에 대한 비교 실험에서 두 방법의 성능의 차이가 없음을 확인하였다. 증강현실의 실험에서도 제안하는 손 자세의 추정 방법은 마커를 사용한 방법만큼 성능이 우수하였다.

앞으로의 연구에서는 손의 위치와 자세변화에 따른 객체의 자연스러운 움직임에 대한 개선이 필요하며 다양한 증강현실에서 다양한 성능 실험이 필요하다.

참고문헌

[1] Ju-Hyun Lee, “A Development Strategy of Augmented Reality Contents in the Contextual Environments,” Human Contents, no.19,

pp.179-218, 2010 November.

[2] Kyung-Hee Noh, Hyung-Keun Jee, Sukhyun Lim, "Effect of Augmented Reality Contents Based Instruction on Academic Achievement, Interest and Flow of Learning," Korea Contents, Vol.10, no.2, pp.1-13, 2010 February.

[3] Keon-Hee Park, Guee-Sang Lee, "Hand Gesture Interface for Manipulating 3D Objects in Augmented Reality," Korea Contents, Vol.10, no.5, pp.20-28, 2010 May.

[4] Y. Shen, "Vision-Based Hand Interaction in Augmented Reality Environment," International Journal of Human-Computer Interaction, vol.27, no.6, pp.523-544, May 2011.

[5] Junchul Chun, Byungsung Lee, "Dynamic Man ipulation of a Virtual Object in Marker-less AR system Based on Both Human Hands," KSII Transactions on Internet and Information Systems, vol.4, no.4, pp.618-632, August 2010.

[6] Junyeong Choi, Hanhoon Park, Jungsik Park, and Jong-Il Park, "Implementation of Hand-Gesture Bas ed Augmented Reality Interface on Mobile Phone," The korea Society of Broadcast Engineers, Vol.16, no.6, pp.941-950, 2011 November.

[7] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations," Proc. Int'l Conf. Computer Vision, pp. 666-673, 1999.

[8] ARTOOLKIT. Human Interface Technology Laboratory, http://www.hitl.Washin gton.edu/artoolkit

[9] OpenNI. (n.d.). OpenNI documentation. http:// www.openni.org/ images/ stories/ pdf / OpenNI_UserGuide_v4.pdf

[10] Rovelo, G. (n.d.). ARTOOLKIT II. from https: //jira.ai2.upv.es/confluence/download/attachments/12222496/WGiM18_ARToolKitII.pdf?version=1&modificationDate=1304095263000

[11] Kwangsoo Kim, Sooyoun Hong, Sooyeong Kwak, Jungho Ahn, Hyeran Byun, "Multiple Human Tracking using Mean Shift and Depth Map with a Moving Stereo Camera," The korean Institute of Ingormation Scientists and Engineers, Vol.34, no.10, pp.937-944, 2007 October.

[12] Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.17, pp.790-799, 1995.

[13] D. Comaniciu, "Kernel-Based Object Tracking," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 25, no. 5, May 2003.

[14] Sang-Geol Lee, Cheol-Ki Kim, Eui-Young Cha, "Retouching Method for Watercolor Painting Effect Using Mean Shift Segmentation," The Korea Society of Computer Information, Vol.15, no.9, pp.25-33, 2010 September.

[15] Heeman Lee, "Implementing Augmented Reality By Using Face Detection, Recognition And Motion Tracking," The Korea Society of Computer Information, Vol.17, no.1, pp.97-104, 2012 January.

저 자 소 개

이 선 형

2011: 숭실대학교 정보통신전자공학부 공학사.

현 재: 숭실대학교 전자공학과 석사과정

관심분야: 영상 처리

Email : showgoon@ssu.ac.kr

한 헌 수

1981: 숭실대학교 전자공학과 공학사.

1983: 연세대학교 전자공학과 공학석사.

1991: University of Southern Califonia 전기공학과 공학박사

현 재: 숭실대학교 정보통신전자공학부 교수

관심분야: 자동화시스템, 센서융합, 물체인식

Email : hahn@ssu.ac.kr

한 영 준

1996: 숭실대학교 전자공학과 공학사.
1998: 숭실대학교 전자공학과 공학석사.
2003: 숭실대학교 전자공학과 공학박사
현 재: 숭실대학교 정보통신전자공학부 부교수
관심분야: 로봇 비전, 컴퓨터 비전, 비주얼 서보잉
Email : young@ssu.ac.kr