
3차원 손 특징을 이용한 손 동작 인식에 관한 연구

배철수*

A study on hand gesture recognition using 3D hand feature

Cheol-soo Bae*

요 약

본 논문에서는 3차원 손 특징 데이터를 이용한 동작 인식 시스템을 제안하고자 한다. 제안된 시스템은 3차원 센서에 의해 조밀한 범위의 영상을 생성하여 손 동작에 대한 3차원 특징을 추출하여 손 동작을 분류한다. 또한 다양한 조명과 배경하에서의 손을 견실하게 분할하고 색상 정보와 상관없이 수화와 같은 복잡한 손 동작에 대해서도 견실한 인식능력을 나타낼 수가 있다. 제안된 방법의 전체적인 순서는 3차원 영상 획득, 팔 분할, 손과 팔목 분할, 손 자세 추정, 3차원 특징 추출, 그리고 동작 분류로 구성되어 있고, 수화 자세에 대한 인식 실험으로 제안된 시스템의 효율성을 입증하였다.

ABSTRACT

In this paper a gesture recognition system using 3D feature data is described. The system relies on a novel 3D sensor that generates a dense range image of the scene. The main novelty of the proposed system, with respect to other 3D gesture recognition techniques, is the capability for robust recognition of complex hand postures such as those encountered in sign language alphabets. This is achieved by explicitly employing 3D hand features. Moreover, the proposed approach does not rely on colour information, and guarantees robust segmentation of the hand under various illumination conditions, and content of the scene. Several novel 3D image analysis algorithms are presented covering the complete processing chain: 3D image acquisition, arm segmentation, hand -forearm segmentation, hand pose estimation, 3D feature extraction, and gesture classification. The proposed system is tested in an application scenario involving the recognition of sign-language postures.

키워드

gesture recognition, 3d depth image, hand segmentation, feature extraction

I. 서 론

최근 막대한 양의 컴퓨터의 보급과 정보유입으로 인해 사람과 컴퓨터 상호간의 정보 교환 수단으로 기존의 인터페이스 방법인 마우스나 키보드가 가지는 의사소통의 제약성을 극복하고 사람과 사람간의 의사소통과 같은 보다 자연스럽고 지적인 인터페이스를 구현하기 위한 연구가

활발히 진행되고 있다.

그 중 손 동작인식은 빠른 의사 전달의 특징과 함축적 의미의 동작을 통해 많은 정보를 전달할 수 있으므로 활발한 연구가 진행되고 있다. 손 동작 인식은 크게 글로브 기반 방법(glove-based method)[1]과 비전에 기반한 방법(vision-based method)[2]으로 나눌 수 있다. 전자는 실시간으로 손의 모양과 손가락의 움직임을 검출할 수 있으나,

장치 착용에 따른 불편함과 손의 운동 범위와 같은 여러 가지 제약 조건을 수반한다. 반면 후자의 방법은 카메라를 통한 입력 영상을 사용하는 방법으로 전자의 방법보다 장비가 간단하며, 행동반경이 자유롭고, 사용자가 불편함을 느끼지 않으면서 자연스럽게 손 동작의 인식이 가능하다는 장점이 있다. 그러나 손의 회전 시에 발생하는 그림자 처리 문제와 주위 환경에 따른 입력 영상 변화에 의한 인식률의 변동이 크다는 단점[3]이 있고, 이와 같은 문제점을 해결하고자 다중 카메라와 수동적 입체 센서에 의한 3차원 정보를 이용[4,5,6,7]한다. 하지만 수화와 같은 복잡한 손 동작 인식에서는 만족할 만한 결과를 보이지 못하고 있다.

본 논문에서는 3차원 손 특징 데이터를 이용한 동작 인식 시스템을 제안하고자 한다. 제안된 시스템은 3차원 센서에 의해 조밀한 범위의 영상을 생성하여 손 동작에 대한 3차원 특징을 추출하여 동작을 분류한다. 또한 다양한 조명과 배경하에서의 손을 견실하게 분할하고 색상 정보와 상관없이 수화와 같은 복잡한 손 동작에 대해서도 견실한 인식능력을 나타낼 수가 있다. 제안된 방법의 전체적인 순서는 3차원 영상 획득, 팔 분할, 손과 팔목 분할, 손 자세 추정, 3차원 특징 추출, 그리고 동작 분류로 구성된다. 그림 1에 제안된 손 동작 인식 알고리즘의 순서도를 나타내었다.

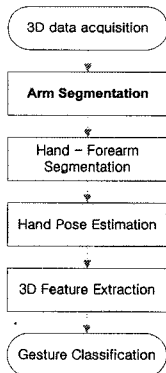


그림 1. 손 동작 인식 순서도
Fig 1. A flowchart of hand gesture recognition

II. 3차원 특징 추출

2.1 3차원 데이터 획득

3차원 데이터 획득을 위하여 구조화된 빛(Structured Light)을 사용하여 그림 2와 같이 능동적 3각 측량 원리[8]를 이용하였다. 객체가 카메라로부터 대략 1 미터에 위치하도

록 하였기 때문에 획득된 영상의 3차원 깊이의 평균 정확성은 1mm 미만이다. 실험에서, 투영된 빛 패턴의 깜빡임을 줄이기 위하여 적합한 범위의 영상만 획득하도록 하였다. 그림 3은 본 논문에서 획득한 3차원 영상 데이터를 나타내었다.

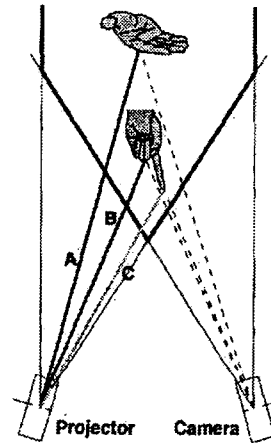


그림 2. 3차원 영상 획득 장치의 원리.
Fig 2. Principle of 3D image acquisition system.



(a)



(b)

그림 3. 3차원 영상 데이터.
(a) 획득한 칼라 영상. (b) 3차원 깊이 영상.
Fig 3. 3D Image Data.
(a) acquired Color image (b) 3D depth image.

2.2 손과 팔목 분할

비전에 근거한 동작 인식에 가장 중요한 과제는 다른 신체 부위나 사물과 같은 배경으로부터 손을 분할 해내는 것이다. 그러나 기존의 피부색을 이용한 분할 방법은 조명, 움직임, 카메라 설정 및 작업 공간에 매우 민감하게 반응하는 문제점을 가지고 있고, 3차원 깊이 정보의 문턱치 값을 이용한 영상 분할은 사용자의 몸이 카메라 면에 대하여 기울어지게 되면 견실하게 동작하지 못한다. 그러나 인간과 컴퓨터의 자연스러운 상호 작용을 이루기 위해서는 아무런 제약 없이 영상분할을 해야 한다.

그러므로 본 논문에서는 계층적 군집화 절차에 의한 팔의 분할 방법을 제안하고 있다. 제안된 방법은 팔, 몸통, 머리 몸체 같은 다양한 부분이 3차원 군집 형태를 이룬다는 관측에 근거를 두고 있다. 제안된 분할 알고리즘은 다음과 같은 단계로 구성되어진다.

1. 초기 군집화는 순차적으로 깊이 정보를 검색하여 얻는다. 그리고 Z 공간에서 이전 화소와의 거리를 유클리디안 거리법에 의해 구하여 각 화소를 분류한다. 이 절차에서 작은 영역 들이 큰 값을 이룬다.
2. 작은 영역을 큰 영역으로 병합하면서 좀 더 큰 영역을 이루어 나간다. 가장 작은 군집이 선택되어지고 군집 분산 SB 사이를 최소화하는 군집으로 병합된다. 이러한 반복작업은 군집의 수가 일정한 수에 이르면 종료된다.
3. 최종적으로, 인접한 군집의 계층 합병이 수행된다. 만약 합성된 군집 ST의 전체적인 확산 범위가 최소이면 2개의 군집은 병합된다. 이 과정은 클러스터가 특정수에 이르면 종료된다.

사전 지식에 의하여 카메라와 근접한 3~4 개의 군집을 선택하고, 그 중 상대적으로 팔에 상응하는 기다란 형태를 선택한다.

팔목으로부터 손바닥과 손가락을 분할해내는 것은 손의 자세와 연속적인 특징을 정확히 추출하는데 중요하다. 손-팔목 분할을 위하여 3차원 공간에서 팔의 지점을 통계학적 모델링하였다. 식 (1)의 가우시안을 혼합하여 3차원 지점의 확률적 분포 x 를 모델링한다.

$$P(x) = P(hand)P(x|hand) + P(forearm)P(x|hand) \quad (1)$$

$$= \pi_1 N(x; \mu_1, \Sigma_1) + \pi_2 N(x; \mu_2, \Sigma_2)$$

여기서 π_1, π_2 는 손과 손목 각각의 이전 확률이고, $N(x, \mu, \Sigma)$ 는 3차원 가우시안 분포이다.



그림 4. 손과 팔목 분할 결과.
Fig 4. Hand-Forearm Segmentation results

3차원 데이터로부터 미지의 매개변수인 $\pi_k, \mu_k, \Sigma_k (k=1, 2)$ 의 최대우도 추정 은 기대치최대화 알고리즘의 평균에 의해 얻는다. 반복되는 절차의 수렴은 적절한 초기 매개변수에 달려있는데 여기서는 팔의 기하학적 사전 지식을 이용하였다. k 에 대한 3차원 지점 x_n 의 분류는 최대우도 표준에 의하여 얻을 수 있었다. 그림 4에 손과 팔목에 대한 분할 결과를 나타내었다.

2.3 자세 추정 및 보정

분류 알고리즘은 입력 패턴에 의해 강체 변환하는 경향이 있다. 3차원 정보는 손의 방향에 대하여 추정 및 보정을 가능하게 하여 효율적인 추정을 이루게 한다. 3차원 자세 추정에 있어 오류를 최소화 하고자, 동작 분류를 위하여 사용한 학습 집단은 정규화된 위치 주위 패턴에 의해 증대된다.

손의 3차원 방향 추정은 확산 행렬 S_T 의 고유벡터 u_1 에 해주어지는 3차원 데이터의 방향에 대한 연산에 의하여 얻을 수 있다. 손가락의 외형적 제한으로 손바닥 중앙으로부터의 손가락까지의 거리 $\|d\|$ 에 의하여 각각의 3차의 지점은 가중화 되고, 방향 벡터를 추정할 수 있다. 3차원 센서의 잡음에 대한 영향을 줄이기 위하여 최소중위제곱 (LMS : Least Median of Squares)을 사용하였다.



그림 5. 3차원 자세 추정 및 보정 결과.
Fig 5. 3D pose estimation and compensation results.

3차원 자세를 보정하기 위하여 집단의 중심 m 과 u_1, u_2 와 u_3 로 구한 x, y 그리고 z 축으로 정규 좌표 프레임을 규정하였다. 카메라의 중앙으로부터의 집단의 중심 m 까지의 거리 \tilde{z} 에서의 카메라 프레임과 정규 좌표 프레임을 정렬한다. 그림 5에 3차원 자세 추정 및 보정 결과를 나타내었다.

III. 특징 추출 및 동작 분류

각각의 정규화된 깊이는 다음 절차에 의하여 추출된 특징 벡터에 의해 표현된다. $s(x, y)$ 를 정규화된 깊이 영상에서 손에 대한 화소를 나타낸 영역의 이진 분할이라 하면, 거리 변환 d_i 는 s 로부터 근사 챔퍼링 알고리즘을 사용하여 계산한다. 거리 변환은 형태의 경계로부터 각각의 화소에 최소 거리를 주고 2차원 형태를 나타낸다. 정규화된 깊이 영상의 깊이 값과 d_i 값 사이의 초과치는 손의 원형구조를 이용하여 제거해 나간다. 손바닥의 중심 c 는 $d(x, y)$ 에서 최대값을 검색하여 정할 수 있다. 이것은 s 에서 2차원 형태에서 최대 원형 위치를 선정하는 것과 동등한 역할을 한다. 그리고, 거리 함수는 깊이 영상과 마찬가지로 균등하게 중심 c 와 k_r 에서 샘플링되고, 원형 샘플링율인 l_c 에서 벡터 $f_i (i = 1, \dots, d = k_r l_c)$ 를 순차적으로 저장한 값에 대응하는 반지름 $r_i (i = 1, \dots, k_r)$ 를 증가시킨다. 고차원 d 와 특징 벡터는 $k_r = 6-8$ 과 $l_c = 100-200$ 에서 기본 값을 유도한다. 그리고 저주파 계수를 보존하고 각각의 원과 일치하는 값을 푸리에 변환을 수행하여 감소해 나간다.

입력된 깊이 영상의 분류는 기억기반 학습을 사용한 학습집단에 의해 이미 정의된 손의 자세 분류 중 하나로 표현된다. 입력되는 영상은 구분되고 근접 학습 샘플 k 가운데서 가장 빈번히 나타난 것을 특징 벡터로 추출한다. 입력 벡터의 근사치는 학습 샘플 벡터에서 정규화된 상호 상관 지수를 사용하여 계산한다

IV. 실험결과 및 고찰

실험을 위하여 수화의 10개의 손 모양을 선택하여 실험에 사용하였다. 10개의 손 모양은 수화에서 0부터 9까지를 나타내는 손 모양을 사용하였다. 그 중 학습에 사용

된 영상은 각각 20개의 샘플 중 6개를 이용하였으며, 3차원 손 모델에 의해 생성하였다. 손가락 구조와 형상이 일치하는 매개변수들의 조합으로 구성하였다.

주어진 손 모양에 대한 강직하고 연식 변화를 생성하고 손가락 길이와 같은 특별한 속성을 실험하기 위해 모델 매개변수들을 선택하였다. 그리고 생성된 모델들 중 부분 집합을 선택하고 카메라로부터 3차원 모델 표면의 거리를 측정하여 각각에 대하여 깊이 영상을 생성하였다. 제안된 학습 절차의 장점은 변이성 매개변수를 직접 조정함으로써 새로운 모양을 추가할 수 있다는 것이다. 또한, 학습된 샘플들은 증첩된 영역과 같은 3차원 센서로 생길 수 있는 어떤 오류에도 영향을 받지 않는다는 것이다.

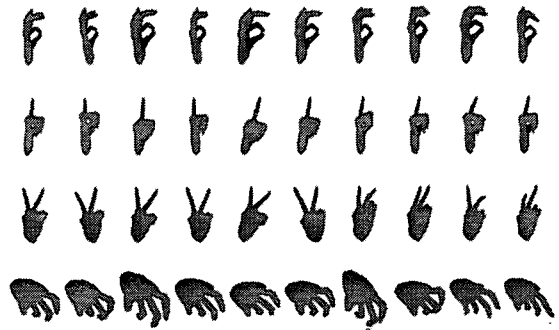


그림 6. 학습 집단의 부분집합
Fig 6. Subset of the training set

제안된 동작 인식시스템은 펜티엄4 1.4GHz의 PC 플랫폼에서 구현하였다. 화소 수 180×144 샘플 영상을 동작 인식하는데 있어 약 1초의 처리시간이 소요되었는데, 초기 분할 과정이 전체 처리시간의 80%를 차지하였다. 몸체에서 손 분할을 수월하게 할 수 있는 제약 사항을 두게 되면 처리시간은 상당부분 감소 될 수 있다는 것을 의미한다.

분류 알고리즘의 성능을 실험하고자 2가지 실험을 수행하였다. 첫 번째 실험은 샘플 중 학습에 참여한 샘플을 제외한 나머지 영상을 가지고 동작 분류를 한 결과 97%의 인식률을 얻을 수 있었다. 그 결과를 표 1에 나타내었고, 두 번째 실험은 자기 10개의 모양에 대하여 각각 100개의 영상을 사용하였다. 그 결과 평균 95%의 인식률을 얻을 수 있었고 그 결과를 표 2에 나타내었다.

표 1. 학습에 참여하지 않은 영상에 대한 실험 결과.
Table 1. Result of recognition experiment without training set.

recognition posture	Success	Fail	Rate (%)
pose 0	14	0	100.0
pose 1	14	0	100.0
pose 2	14	0	100.0
pose 3	13	1	92.9
pose 4	14	0	100.0
pose 5	13	1	92.9
pose 6	13	1	92.9
pose 7	14	0	100.0
pose 8	13	1	92.9
pose 9	14	0	100.0
Average	136	4	97.1

표 2. 각 10개 모양에 대한 100개 영상에 대한 실험 결과.
Table 2. Result of recognition experiment with each 100 image for pose

recognition posture	Success	Fail	Rate (%)
pose 0	97	3	97.0
pose 1	97	3	97.0
pose 2	100	0	100.0
pose 3	94	6	94.0
pose 4	95	5	95.0
pose 5	93	7	93.0
pose 6	91	9	91.0
pose 7	95	5	95.0
pose 8	92	8	92.0
pose 9	99	1	99.0
Average	953	47	95.3

V. 결 론

본 논문에서는 3차원 특징 데이터를 이용한 정적인 손 동작 인식 시스템을 제안하였다. 제안된 시스템은 범위에 대한 데이터만으로 이루어지므로 주위배경이나 조명 변화에 제한 없이 견실하게 동작하였다. 3차원 형태 정보를 합병한 특징 벡터 식별에 의해 범위 영상을 표현하고 구

분된 손 자세를 얻을 수 있었다. 실험 결과 3차원 정보에 의한 동작 식별의 효율성과 견실성을 증명하였다.

향후, 분할 과정을 간략화 하여 현재 약 1초의 처리 시간을 단축한다면 인간과 컴퓨터 자연스러운 상호작용을 위한 실시간 인식이 가능할 것이라고 사료된다.

참고문헌

- [1] V. I. Pavlovic, R. Sharma, T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review.", *IEEE transactions on pattern analysis and machine intelligence*, v.19 no.7, pp.677-695, 1997.
- [2] D. M. Gavrila, "The Visual Analysis of Human Movement: A Survey.", *CVIU*, v.73 no.1, pp.82-98, 1999.
- [3] C. C. Lien, C. L. Huang, "Model-based articulated hand motion tracking for gesture recognition.", *Image and vision computing*, v.16 no.2, pp.121-134, 1998.
- [4] A. Utsumi and J. Ohya. "Multiple hand gesture tracking using multiple cameras.", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.
- [5] C. Jennings. "Robust finger tracking with multiple cameras.", in *Proc. Intl. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, 1999.
- [6] R. Grzeszczuk, G. Bradski, M. H. Chu, and J. Y. Bouguet. "Stereo based gesture recognition invariant to 3d pose and lighting.", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2000.
- [7] N. Jovic, B. Brumitt, B. Meyers, S. Harris, and T. Huang. "Detection and estimation of pointing gestures in dense disparity maps.", in *Proc. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.
- [8] F. Forster, M. Lang, and B. Radic. "Real-time 3d and color camera.", in *Proc. ICAV3D 2001*, Mykonos, Greece, May 2001.
- [9] R. O. Duda, E. Hart, and D. G. Stock. "Pattern Classification.", Wiley, 2001.

저자소개



배철수(Cheol-Soo Bae)

1979년 2월 명지대학교 전자공학과 졸업(공학사)

1981년 2월 명지대학교 대학원 전자공학과 졸업(공학석사)

1988년 8월 명지대학교 대학원 전자공학과 졸업(공학박사)

1999년 3월~2001년 5월 관동대학교 공과대학 학장

2000년 3월~2002년 2월 관동대학교 양양캠퍼스 창업
보육센터 소장

2001년 6월~2003년 8월 관동대학교 평생교육원장

2001년 3월~현재 해양정보통신학회 강원지부장

2003년 1월~현재 한국통신학회 국내저널 편집부위원장

2003년 1월~현재 대한전자공학회 이사

1981년~현재 관동대학교 정보통신공학부 교수

※ 관심분야: 영상처리, 신호처리시스템, 영상압축