

관측 시점에 강인한 손 모양 인식을 위한 손 모양과 손 구조 사이의 학습 기반 유사도 결정 방법

Learning Similarity between Hand-posture and Structure for View-invariant Hand-posture Recognition

장효영¹, 정진우², 변증남³

¹ 대전시 유성구 구성동 한국과학기술원 전자전산학과
E-mail: hyjang@gmail.com

² 대전시 유성구 구성동 한국과학기술원 인간친화 복지 로봇시스템 연구센터
E-mail: jinwoo@ctrsys.kaist.ac.kr

³ 대전시 유성구 구성동 한국과학기술원 전자전산학과
E-mail: zbien@ee.kaist.ac.kr

요 약

본 논문에서는 비전 기술에 기반을 둔 손 모양 인식 시스템의 성능 향상을 위해 학습을 통해 손 모양과 손 구조 간 유사도를 결정하는 방법을 제안한다. 비전 센서에 기반을 둔 손 모양 인식은 손의 높은 자유도로 인한 자체 가림 현상과 관찰 방향 변화에 따른 입력 영상의 다양함으로 인해 인식에 어려움이 따른다. 따라서 비전 기반 손 모양 인식의 경우, 카메라와 손 간의 상대적인 각도에 제한을 두거나 여러 대의 카메라를 배치하는 것이 일반적이다. 그러나 카메라와 손 간의 상대적 각도에 제한을 두는 경우에는 사용자의 움직임에 제약이 따르게 되며, 여러 대의 카메라를 사용할 경우에는 각 입력된 영상에 대한 인식 결과를 최종 인식 결과에 반영하는 방식에 대해서 추가적으로 고려해야 한다. 본 논문에서는 비전 기반 손 모양 인식의 이러한 문제점을 개선하기 위하여 인식의 과정에서 사용되는 손 모양 특징을 손 구조적인 각도 정보와 손 영상 특징으로 나누고, 학습을 통해 각 특징 간 연관성을 정의한다.

Key Words : View-invariant, Similarity Decision, Hand-posture Recognition, Hand-posture, Hand-Structure

1. 서 론

컴퓨터 환경의 발전에는 그에 적합한 사용자 인터페이스(user interface; UI)의 변혁이 수반된다. 최근의 유비쿼터스(Ubiquitous) 패러다임이 추구하는 사용자를 중심으로 하는 기술 개발의 경향은 지능을 가진 자동 인터페이스로의 전환을 촉구하고 있으며, 손 제스처는 이에 대한 하나의 안으로 부각되고 있다.

손 제스처는 일상생활에서 사람 간 언어적인 특질을 지닌 독립적인 대화 수단으로 친숙하다. 또한 소음이 심한 공사 현장이나 수중, 운동 경기 등에서는 음성을 대체하는 수단이 되

며, 특히 위치나 방향 등 공간 정보를 전달하는 데에는 음성보다 용이하다. 이와 같은 손 제스처의 장점으로 인하여, 1980년대로부터 가상현실(virtual reality)과 수화 인식 분야를 포함한 여러 분야에서 손 제스처 인식에 기반을 둔 사용자 인터페이스에 대한 연구가 진행되어 왔다.^{[2][3]} 또한 최근에는 저가의 USB 카메라 사용이 보편화되면서 비전 기반의 손 제스처 인식에 대한 연구가 활발하다.

그러나 여전히 카메라를 이용한 손 제스처 인식 기술은 상용화된 예를 찾아보기가 쉽지 않다. 이에 대한 가장 큰 이유는 사람의 손이 높은 자유도를 가지고 있고 관측 방향에 따라

다양한 형태로 관측될 수 있음으로 인해 인식 과정에 어려움이 존재한다는 것이다. 또한 대개의 경우, 손의 형태적인 변형과 관찰 방향 변화는 복합적으로 발생한다. 따라서 기존 연구들의 경우, 사전에 손바닥과 카메라의 상대적인 방향을 정하고 손의 움직임에 제한을 두어 손 모양을 인식한다. 그러나 이러한 가정은 사용자 편의성 관점에서 바람직하지 못하다.

본 논문은 비전 기반 손 제스처 인식에서 발생하는 손 모양 인식의 이와 같은 문제를 해결하기 위하여 인식의 과정에서 사용되는 손 모양 특징을 손의 구조적인 각도 정보와 영상 특징으로 나누고, 학습을 통해 각 특징 간 연관성을 정의하는 방식을 제안한다.

2. 손 특징 데이터베이스

사용한 손 특징은 크게 손의 구조적인 각도 정보와 영상 특징으로 나뉜다. 대상으로 하는 손 모양은 그림 1과 같다.



그림 1. 대상 손 모양

각 손 모양에 대한 각도 정보와 영상 특징은 세 대의 카메라와 데이터글러브(DataGlove™) 장치를 사용하여 동시에 취득하였다. 이 과정에서 세 대의 카메라를 사용한 것은, 하나의 손 모양에 대하여 한 번에 여러 방향에서 영상을 취득함으로써 데이터 취득 시의 반복 시행 횟수를 줄이기 위한 것이다.

3.1 손 각도 정보

데이터글러브를 통해 그림 2에 나타낸 14개 위치의 각도 정보를 얻을 수 있다.

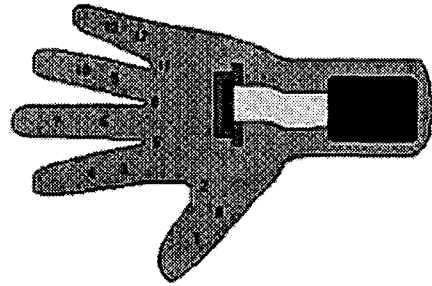


그림 2. 데이터 글러브
(0~13: 각도 취득을 위한 센서 위치)

데이터글러브에서 얻은 14개의 값은 식 (1)과 같이 정규화하여 사용한다.

$$normalized\ data = \frac{val - min}{max - min} \times max \quad (1)$$

여기서 *val*은 하나의 센서에 대해 입력된 값이며, *max*와 *min*은 각각 사용자가 데이터글러브를 착용한 후 해당 센서에 줄 수 있는 최대값과 최소값을 의미한다.

3.2 손 영상 특징 정보

손 모양 영상 특징을 얻기 위해 기준 각도로 부터의 각도 차에 대해 손 영역 중심으로부터 경계까지의 거리를 구하는 함수를 정의하고, 동일한 간격으로 각도를 분할하여 측정된 중심과 경계간의 거리로 이루어진 벡터를 손 영상 특징으로 사용하였다.

이 함수를 손 중심으로 손 영역 경계 추출영상에 적용하기 위해서는 각도 분할의 중심축과 기준 0°의 위치 및 각도 증가 방향을 정해야 한다. 이에 대해 추출된 손 영역의 중심을 회전의 중심축으로 하였으며, 손 영역의 장축을 기준 0°로 하였다. 각도 증가는 반시계방향을 정방향으로 정의하였다. 측정된 벡터의 값은 다시 해당 벡터 내 최대값을 기준으로 하여 0과 1 사이의 값으로 정규화한다. 그림 3에 추출된 손 영상 특징 정보의 예를 보였다.

3.3 손 특징 데이터베이스의 구축

데이터베이스 구축의 과정에서는 특정 손 모양의 영상 특징 정보와 각도 정보 및 손 모양 식별자가 입력으로 주어진다.

이에 대하여 기존에 데이터베이스 내에 존재

하는 영상 특징정보 중 유클리드 거리 측면에서 새롭게 입력된 영상 특징 정보와 가장 유사한 것을 찾는다. 만약 이 둘 사이의 유클리드 거리가 임계치보다 작을 경우, 입력된 영상 특징 벡터 x_p , x_i 와 가장 유사한 데이터베이스 내의 기존 영상 특징 x_k 그리고 그 다음으로 유사한 기존 영상 특징 x_{k-1} 의 평균 x_k' 로 x_k 를 대체한다. 기존 데이터베이스 내의 영상 특징 중 새롭게 입력된 영상 특징 x_i 와 유사한 것이 없을 경우에는 x_i 가 데이터베이스에 등록된다.

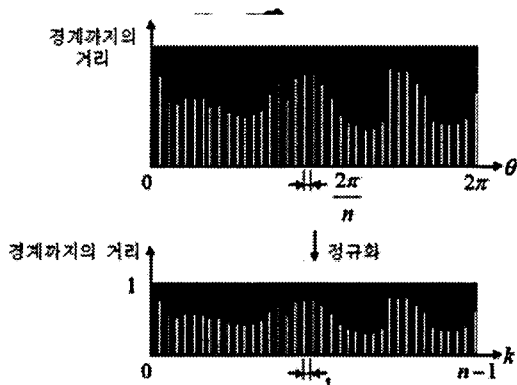


그림 3. 손 영상 특징 추출의 예

x_i 가 데이터베이스에 등록되거나 x_k 가 x_k' 으로 대체될 경우, 이와 함께 입력된 각도 정보 y_i 역시 기존에 데이터베이스에 존재하는 각도 정보와의 비교를 통해서 동일한 과정을 통해 새롭게 등록되거나 기존 데이터베이스 내의 가장 유사한 각도 정보를 갱신하는 데에 쓰인다.

이와 같은 과정에서 동시에 입력된 영상 특징과 각도 정보는 연관성 1로 초기화되고, 기존 데이터가 수정된 경우, 수정된 기존 데이터는 새롭게 입력된 <영상 특징 정보, 각도 정보, 식별자>의 영상 특징 정보 및 각도 정보 간 연관성 및 식별자를 상속하게 된다.

그 결과로서 데이터베이스 내 영상 특징 정보와 각도 정보 간에는 다대다 관계가 형성되게 되며, 각 영상 특징 및 각도 정보에는 그에 연관된 복수개의 손 모양 식별자가 동시에 기록되게 된다. 이와 같은 형태로 손 특징 데이터베이스를 구축한 것은, 동일한 손 영상 정보라 하더라도 관측 방향을 달리하여 관측된 서로 다른 손 모양일 수 있음을 반영한 것이다.

3. 손 특징 간 연관성 학습

손 모양 간 템플릿 정합 결과가 높은 것이

	ㄱ	ㄴ	ㄷ
추출된 손 영역			
가능한 실제 손 모양			

반드시 왼 손 모양에 가까운 손모양인 것은 아닙니다. 그림 4에 그 예를 보았다. 마찬가지로 유클리드 거리 측면에서 가장 가까운 손 각도 정보를 갖는 손 모양이 실제로 가장 유사한 손 모양을 의미하지는 않는다.

그림 4. 템플릿 정합에 유사도 정의의 오류

다시 말하자면 손모양의 특징을 기술하는 것으로 '정한' 수치가 실제 대상의 의미와 차이를 가지게 마련이라는 것인데, 실제 인식 시스템의 설계 과정에서는 얻을 수 있는 정보량의 한계 및 계산 편의 등을 이유로, 손 모양 간 유사도를 영상 특징 정보 또는 각도 정보 중 어느 한 쪽의 측면에서만 기술하는 것이 일반적이다.

그러나 다양한 방향 변화에 대한 영상을 통해 방향 변화에 대하여 영향을 받지 않는 인식 시스템을 개발하고자 할 경우, 이 두 가지 정보를 아우르는 유사도가 필요하게 된다.

본 논문에서는 주어진 손 모양 영상 특징 정보와 각도 정보 간 유사도를 정의하는 과정을 장기적인 관점에서 최적인 해를 찾아가는 경로 탐색 과정으로 보고 Q-학습을 적용하였다. 이를 위해, 하나의 특징에서 다른 특징 노드로의

이동에 대해 유클리드 거리가 먼 것이 작은 것보다 선택될 가능성이 크다고 가정한다. 그러나 유클리드 거리에 의해 유사하더라도 그것이 반드시 실제 손 모양에 연관되는 특징 노드라는 보장은 없으므로, 해당 손 모양에 대해 주어진 손 모양 식별자를 이용하여 잘못된 이동을 판단하고 이에 대해 벌점(penalty)을 준다. 아래에 Q-학습을 적용하기 위한 설정을 요약하였다.

■Q-테이블의 초기 값:

즉시 보상 (immediate reward) 값

■행동(action): 특징 간 이동

■상태(state): 각 특징 노드 간 유클리드 거리와 에이전트들의 이동에 의해 특징 간 연결에 할당된 값

■즉시 보상 (immediate reward)

- 잘못된 선택일 경우: -5
- 옳은 선택일 경우:

$$Normaize\{1/Euclidean\ Dist\} \times 10$$

4. 시뮬레이션 및 결과 고찰

실험을 위하여 5개의 대상 손 모양에 대하여 총 15개의 손 방향을 조합하여 1회당 75개의 데이터를 얻었다. 총 3인에 대해 데이터를 얻었으며, 한 사람 당 5회의 반복을 통해 데이터를 얻었다. 이렇게 취득한 총 3375장(=3인 x 3 카메라 x 5(손 모양) x 15(손 방향) x 5회)의 영상 중 임의로 추출된 1500 장을 사용하였다.

실험을 통해 Q-학습 과정에서의 탐색(Exploration)과 이용(Exploitation) 간의 균형을 확인하였다 (그림 5). 본 실험에서 탐색은 유클리드 거리 측면에서 보았을 때 가장 근사하나, 실제로는 서로 다른 손모양일 경우에 대한 고려를 의미한다.

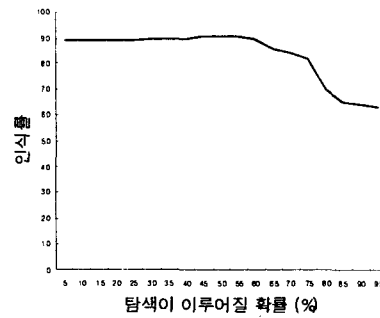


그림 5. 탐색 확률 변화에 의한 인식률 변화

탐색 확률이 커짐에 따라 인식률은 개선되고 수렴까지의 속도 또한 빨라짐을 보였다(그림 6). 그러나 탐색 확률이 지나치게 커질 경우, 오히려 인식률이 급격히 떨어지고 수렴까지의 반복 횟수도 증가한다. 이와 같은 결과에 근거하여 0.55가 적절한 탐색 확률인 것으로 결론 지을 수 있다.

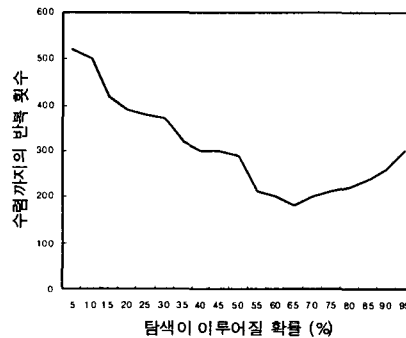


그림 6. 탐색 확률에 따른 수렴 반복 횟수 변화

감사의 글

본 연구는 과학기술부 / 한국과학재단 우수 연구센터육성사업의 지원으로 수행되었음 (R11-1999-008).

참고 문헌

[1] Mark Weiser, The Computer for 21st century, Sci. Amer., 1991.

- [2] Jung-Bae Kim, Kwang-Hyun Park, Won-Chul Bang and Z. Zenn Bien, "Continuous gesture recognition system for Korean sign language based on fuzzy logic and hidden markov model," Proc. of FUZZ-IEEE, 2000.
- [3] Chan-Su Lee, Sang-Won Ghyme, Chan-Jong Park, and Kwang-Yuñ Wohn, "Virtual reality software and technology archive," Proc. of the ACM symposium on virtual reality and technology 1998, pp.59-65, 1998.