

사용자 의존적 인터페이스를 위한 제스처 명령어 습득 (Learning Gesture Command for User-Dependent Interface)

양선옥, 최형일

(Seon-Ok Yang, Hyung-Il Choi)

송실대학교 컴퓨터학부

(School of Computing Soongsil University)

요 약

명령어로 사용되는 핸드 제스처가 사용자에게 더 많은 친밀감을 주기 위해서는 사용자가 자신의 원하는 형태로 제스처를 정의할 수 있어야 한다. 본 논문에서는 카메라를 통해 입력되는 사용자의 핸드 제스처를 명령어로 이용하는 지능적 사용자 인터페이스에 대하여 소개한다. 지능적 사용자 인터페이스는 제스처 명령어로 이용되는 핸드 제스처의 종류를 사용자가 임의대로 정의할 수 있도록 제스처 명령어 습득 모듈을 포함한다.

1. 서론

인간은 제스처를 사용함으로써 자신의 의도를 간단 명료하게 상대방에게 전달한다. 이와 같이 컴퓨터와의 상호 작용의 과정에서 제스처를 명령어로 이용하려는 시도가 점차 많아지고 있다 [1][2][3][4]. 사람이 눈을 통하여 상대방의 의사를 이해하는 것처럼, 컴퓨터가 컴퓨터 시각을 통하여 사용자의 의사를 알 수 있도록 하는 것은 컴퓨터의 사용이 대중화됨에 따라 나타나는 입력 장치에 대한 사용자의 부담감에 대한 해결하기 위한 방안일 뿐만 아니라, 컴퓨터와의 자연스러운 상호 작용을 위한 환경을 구축하기 위해 요구되는 사항이다. 그러나 컴퓨터가 컴퓨터 시각을 통하여 사용자의 제스처를 이해하기 위해서는 세가지 작업을 거쳐야 한다. 카메라로부터 입력 영상을 획득하는 작업, 획득된 영상으로부터 사용자의 제스

처 부분을 검출하는 작업, 검출 영역의 의미를 판단하는 작업이다.

기존의 컴퓨터 시각에 기반한 사용자 인터페이스의 응용들은 대부분 펜이나 마우스를 대신하는 입력 장치를 개발하기 위한 연구에 주력하였다 [1][2][3]. 여기서 사용되는 제스처 명령어로는 동기화 명령어, draw 명령어, click 명령어 등의 제스처 만이 요구되기 때문에 비교적 단순하고 간단하게 정의될 수 있다. 대부분의 응용 시스템들은 미리 사용 가능한 명령어들을 정의하고, 사용자에게 제스처의 형태를 알려 주는 형식을 취하였다. 이것은 제스처의 형태를 몇 가지로 고정시킨다면 제스처의 인식 방법이 간단하게 되고, 사용자가 제스처 명령어도 쉽게 익히기 때문이다. 또한 다양한 형태의 제스처들을 미리 정의하여 사용한다면, 그 중에는 어떤 사용자가 따라 하기

어려운 형태의 제스처도 있을 수 있기 때문이다.

제스처를 이용하여 설계된 사용자 인터페이스가 사용자에게 더 많은 친밀감을 주기 위해서는 사용자가 자신의 원하는 형태로 제스처를 정의할 수 있어야 한다. 이것은 사용자로부터 제스처를 습득하여 습득된 새로운 지식에 의해 기존의 지식 베이스를 수정, 유지 하는 문제로 생각할 수 있다. 이 문제는 학습하는 시스템(learning apprentice systems)으로 고려할 수 있으며, Mitchell, Mahadevan, Steinberg 는 학습하는 시스템을 사용자에 의해 제공되는 문제 해결 과정을 시스템이 관찰, 분석함으로써 얻어진 새로운 지식을 기존의 지식들에 결합시키는 상호작용에 의한 지식 베이스의 구조로 보고 있다[5].

지식 베이스를 개발, 유지하고자 할 때, 고려해야 할 몇 가지 사항이 있다. 지식 베이스에 존재하는 문제 영역을 어떻게 정의할 것인가, 사용자에게 친숙한 지식베이스를 어떻게 구성할 것인가, 초기에 주어진 지식베이스가 시간에 따라 변화하는 문제를 어떻게 수용할 것인가 하는 것이다. 이 문제를 Jourdan, Dent, McDermott, Mitchell, Zabowski 은 다음의 두 가지를 고려하는 문제로 제시하고 있다[6]. 사용자의 결정을 반영하는 유용한 훈련 예들을 획득하기 위한 방법을 어떻게 구성할 것인가? 훈련 예들을 일반화하여 유용한 결정을 만드는 학습 메커니즘은 어떻게 정의할 것인가?

본 논문에서는 제스처를 이용하는 지능적 인터페이스를 설계시, 명령어로 사용되는 제스처를 사용자가 직접 정의할 수 있도록 고려하여 지식 베이스의 확장 및 수정이 가능하도록 새로운 제스처를 습득하는 모듈을 지능적 인터페이스에 포함되도록 제안한다.

2 절에서는 제스처를 이용하는 지적 인터페

이스 시스템의 구조에 대하여 소개하고, 3 절에서는 새로운 제스처 명령어를 습득하기 위한 방법을 설명한다. 마지막으로 4 절에서는 간단한 예와 함께 결론을 기술한다.

2. 제스처를 이용하는 지적 인터페이스

응용 시스템의 명령어로 사용되는 제스처를 사용자가 자유롭게 추가, 삭제할 수 있다면, 그 종류가 제한되고, 그 형태가 고정된 제스처 명령어를 이용할 때보다는 사용자가 시스템에 대한 친화감은 향상될 수 있다. 이러한 사용자 의존적인 제스처 인터페이스를 만들기 위해서는 인공지능의 기계 학습 기법이 적용되어야 한다. 즉, 시스템은 사용자가 정의하는 제스처들을 인식하기 위한 제스처 명령어 습득 기능을 포함해야 한다.

그림 1은 사용자 의존적 제스처 인터페이스의 구조를 보여준다.

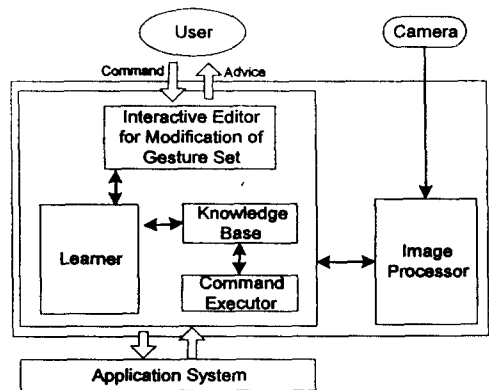


그림 1 사용자 의존적 제스처 인터페이스 구조

그림 1에서 보여준 것처럼 사용자 의존적인 제스처 명령어를 이용하는 인터페이스를 설계하기 위해서 시스템은 사용자가 새로운 제스처의 형태 정보를 입력할 수 있도록 상호작용 편집기

를 제공해야 한다. 또한 카메라로부터 입력되는 사용자의 제스처를 획득하여 처리할 수 있는 영상 처리기와 카메라로부터 획득된 영상이 최적으로 처리하고 표현할 수 있도록 지원하는 지식 베이스, 사용자의 의도에 따라 일반화된 제스처 인식 규칙을 습득하기 위한 학습기가 포함되어야 한다.

상호 작용 편집기는 사용자가 제스처 명령어를 정의하고자 할 때, 그것을 어떤 형태로 표현할지에 대한 정보를 제공하는 도움말에 따라, 사용자는 키보드를 통해 제스처 명령어의 형태를 입력하는 상호작용의 기능을 제공한다. 편집기를 통해 얻게 되는 정보들은 카메라로부터 입력된 영상을 처리하기 위해 필요한 함수들을 호출하는 지식으로 사용된다. 그리고 상호작용 편집기의 기능은 제스처를 습득하는 과정에서만 필요로 한다. 사용자로부터 입력된 정보와 카메라로부터 입력된 제스처 영상이 처리되는 과정은 학습기에 의해 관찰되며, 그 결과는 지식베이스에 통합되어 차후에 제스처를 인식하기 위한 규칙으로 사용된다.

3. 제스처 명령어 습득

새로운 제스처 명령어를 습득한다는 것은 카메라로부터 입력되는 제스처 영상으로부터 그 특징을 추출하여 기존의 제스처 명령어들과 구분할 수 있도록 입력된 제스처 명령어를 묘사하는 것이다.

컴퓨터 시각 시스템에 의해 제스처가 인식되기 위해서는 우선 입력된 영상을 식(1)과 같이 영상의 특징 값이 양자화된 특징 벡터로 표기되어야 한다.

$$F_j = \bigwedge_{i=1}^{n_j} (f_j, v_j^{n_j}) \quad \text{식 (1)}$$

n_j : 특징의 갯수

v_j : j번째 특징 f_j 의 domain, $v_j^{n_j} \in v_j$

이 때, 특징 f_j 는 nil 값을 가질 수 있는데, 그것은 특징이 계층적인 구조를 가질 경우에 발생한다. 예를 들어, 제스처의 손가락 사이의 각도에 대한 특징은 제스처를 표현하는데 2개 이상의 손가락을 사용할 때에만 실제적인 값을 갖게 되며, 아닐 경우에는 nil이 된다.

n개의 서로 다른 제스처가 존재할 때, 각 제스처는 하나의 명령어가 되며, 룰 베이스는 $\{R_1, R_2, \dots, R_n\}$ 으로 구성되며, 각 제스처 G_i 에 대한 인식 룰은 집합 R_i 로 표기된다. 각 제스처의 인식 룰은 특징의 집합 $\{f_1, f_2, \dots, f_m\}$ 으로 구성되는 n_r 차 특징 공간을 제스처 고유의 격자 형태로 분할된다.

명령어로 입력된 제스처를 묘사하기 위해서는 그 제스처만을 구별할 수 있는 고유의 특징들로 표현되어야 한다. 즉, 제스처 명령어로 입력된 예제 제스처들에 공통적으로 존재하는 특성이 있어야 하며, 그 특성은 다른 제스처 명령어의 예제에는 존재하지 않는 유일성을 나타내야 한다. 점차적으로 제스처 명령어를 습득하기 위해서는 제스처 명령어가 한번에 주어지지 않고, 여러 단계로 나뉘어 주어지며, 새로운 제스처 명령어를 인식하는 규칙을 만들기 위해서는 기존에 습득된 제스처 명령어들의 인식 규칙을 고려해야 한다.

제스처 명령어를 묘사하는 특징 공간으로부터 다른 제스처 명령어와 구분되는 특징들을 발견하는 문제는 특징 공간을 탐색하는 과정으로 볼 수 있다. 이것은 어떤 특징이 새로 입력된 제스처를 가장 적절하게 묘사할 수 있는 것인가를

선택하여 Working Set W , $W=\{W_1, W_2, \dots, W_n\}$, 을 형성하는 것으로 시작된다.

본 논문에서는 입력된 제스처 명령어에 대한 Working Set 을 형성하기 위해서, 사용자로부터 입력된 제스처의 형태 정보를 이용하며, 카메라로부터 입력된 제스처 명령어 예제들에만 존재하는 유일성을 갖는 특징을 추론하기 위해서는 Working Set 에 형성된 규칙과 지식 베이스에 정의된 사전 정보들을 이용한다. 특징 추론을 위해 사용되는 지식 베이스에 포함된 정보들은 다음과 같이 분류된다.

1) 명확한 관계 : 불변 사실을 나타낸다.

예) 제스처에 이용되는 손가락은 다섯개이다

$The_Number_Of_Fingers (Hand)= 5$

사용하는 손가락 번호는 1 과 5 사이의 정수 값을 갖는다.

$Use_Finger(X) \quad X = 1..5$

2) 인과/함수 관계 : 영상 처리 함수의 적용 관계를 나타낸다.

예) Body area 를 계산하기 위해서는 Body region 을 구해야 한다.

$Compute_Body_Area(X) \leq Get_Body_Region(Y)$

3) 상관관계 : 경험적 지식을 나타낸다.

예) InsideFingerType 일 경우에는 손가락이 벌어지는 정도에 따라서 Body Area 에 대한 Finger Area 의 비가 다르게 계산된다.

4. 결론

본 논문은 새로운 제스처 명령어를 습득하는 지능적 인터페이스가 갖추어야 할 구성 요소와 새로운 제스처 명령어의 쉬운 습득을 위하여 사

용자와 상호 협력하는 방법을 제안하였다. 사용자가 임의의 제스처를 명령어로 사용하고자 할 때, 사용자가 정의하는 제스처는 카메라로부터 입력되며, 카메라에 의해 입력된 제스처가 기존에 정의된 제스처와 구별되는 특징들을 추출하기 위해서는 시스템은 키보드 또는 마이크(본 논문에서는 키보드를 입력 장치로 이용한다.)를 통하여 입력된 제스처의 형태 정보와 지식베이스에 저장된 사전 정보들을 이용한다.

그림 2는 새로운 명령어를 획득하기 위해 사용자와 시스템의 상호작용 과정을 보여준다. 이것은 "TWO"라는 제스처 명령어에 대한 형태 정보를 사용자로부터 얻기 위하여 과정을 보여준다.

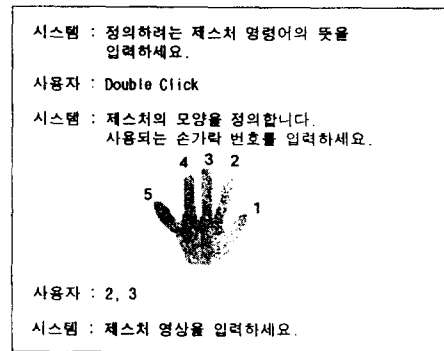


그림 2 사용자-시스템 상호작용 과정

그림 3은 사용자가 제스처 명령어를 입력하는 화면을 보여준다. 그림 4는 본 논문에서 "Double Click" 제스처 명령어를 표현하기 위해 사용된 특징의 의존 관계를 보여준다.

본 논문에서는 특정 제스처를 습득하여 명령어로 이용하는 컴퓨터 시각에 기반한 지능적 인터페이스를 제안하였다. 그러나 이것을 실용적으로 사용하기 위해서는 동적인 제스처를 제스처 명령어에 포함시키기 위한 방법과, 서로 다른 형

태의 제스처가 동일한 명령어로 정의될 때의 제스처 습득 방법도 고려해야 할 것이다.

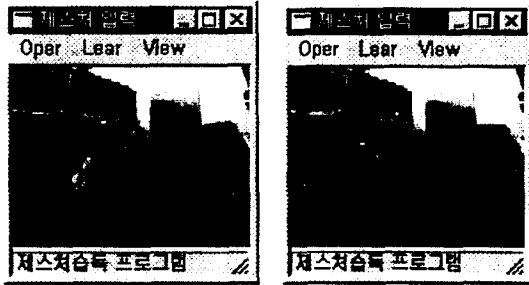


그림 3 제스처 명령어 입력 화면

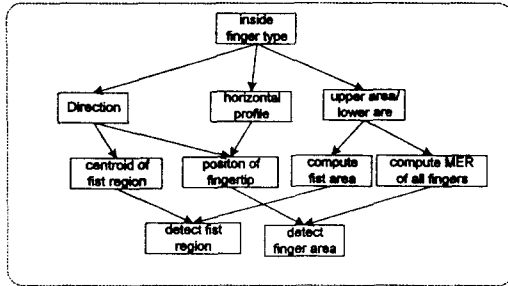


그림 4 제스처 표현 특징의 의존 관계

참고문헌

[1] J.L. Crowley, J. Coutaz, "Vision for Man Machine Interaction", EHCI'95, Grand Targhee, August'95.
 [2] T.Baudel, M.Beaudouin-Lafon, "Charade : Remote Control of Objects Using Free-Hand Gestures", Communications of the ACM, Vol.36, No.7, July 1993, pp.28-35.
 [3] Francis K.H. Quek, Thomas A. Mysliwicz, Meide Zhao, "FingerMouse : A Freehand Computer Pointing Interface", Int. Workshop on Automatic Face-and Gesture-Recognition, Zurich, pp.372-377,1995.
 [4] William T. Freeman, Craig D. Weissman, "Television Control by Hand Gestures", Int.

Workshop on Automatic Face-and Gesture-Recognition, Zurich, pp.179-181,1995.

[5] T. Mitchell, S. Mahadevan, L. Steinberg, "Leap: A Learning Apprentice for VLSI design," Proc. Of the Ninth Int. Joint Conf. On Artificial Intelligence pp.573-580.
 [6] Steven Minton, *Machine Learning Methods for Planning*, Morgan Kaufmann Publishers, pp. 31 - 65, 1993.
 [7] R. S. Michalski, J.G. Carbonell, T.M. Mitchell, *Machine Learning*, Tioga, 1983.
 [8] V. Cantoni, L. Lombardi, "Learning from Examples in ComputerVision:Preliminary Statements", *Image Analysis and Processing*, pp.157-163.
 [9] R. Bareiss, *Exemplar-Based Knowledge Acquisition*, Academic Press, Inc. 1989.