

로봇 인터페이스 활용을 위한 가속도 센서 기반 제스처 인식

장민수

한국전자통신연구원
로봇/인지시스템연구부
(minsu@etri.re.kr)

조용석

건양대학교 전자정보공학과
(yscho@konyang.ac.kr)

김재홍

한국전자통신연구원
로봇/인지시스템연구부
(jkhim504@etri.re.kr)

손주찬

한국전자통신연구원
로봇/인지시스템연구부
(jcsohn@etri.re.kr)

로봇 자체 또는 로봇에 탑재된 콘텐츠와의 상호작용을 위해 일반적으로 영상 또는 음성 인식 기술이 사용된다. 그러나 영상·음성인식 기술은 아직까지 기술 및 환경 측면에서 해결해야 할 어려움이 존재하며, 실적용을 위해서는 사용자의 협조가 필요한 경우가 많다. 이로 인해 로봇과의 상호작용은 터치스크린 인터페이스를 중심으로 개발되고 있다. 향후 로봇 서비스의 확대 및 다양화를 위해서는 이들 영상·음성 중심의 기존 기술 외에 상호보완적으로 활용이 가능한 인터페이스 기술의 개발이 필요하다. 본 논문에서는 로봇 인터페이스 활용을 위한 가속도 센서 기반의 제스처 인식 기술의 개발에 대해 소개한다. 본 논문에서는 비교적 어려운 문제인 26개의 영문 알파벳 인식을 기준으로 성능을 평가하고 개발된 기술이 로봇에 적용된 사례를 제시하였다. 향후 가속도 센서가 포함된 다양한 장치들이 개발되고 이들이 로봇의 인터페이스로 사용될 때 현재 터치스크린 중심으로 된 로봇의 인터페이스 및 콘텐츠가 다양한 형태로 확장이 가능할 것으로 기대한다.

논문접수일 : 2011년 02월 09일 논문수정일 : 2011년 02월 20일 게재확정일 : 2011년 02월 26일
투고유형 : 국문급행 교신저자 : 조용석

1. 서 론

다양한 의사소통 채널을 통해 인간과 로봇이 자연스럽게 상호작용할 수 있도록 해주는 인간-로봇 상호작용(Human-Robot Interaction) 기술은 로봇이 가지는 자율성, 상호작용의 양방향성, 상호작용 또는 제어 수준의 다양성 등이 존재하는 지능형 서비스 로봇의 핵심 기술이다(김계경, 2005; 곡근창, 2007). 일반적으로 인간-로봇 상호작용 기술은 비디오 기반 상호작용 기술, 오디오 기반 상호작용

기술 및 이들과 타 기술들을 융합한 멀티모달 정보 기반 상호작용 기술로 분류할 수 있다.

비디오 기반 인간-로봇 상호작용 기술은 로봇에 부착된 카메라로부터 얻어진 영상정보를 근거로 하여 얼굴 검출, 얼굴 인식, 얼굴 검증, 표정인식, 사람 추출, 사용자 인식, 사람 추적, 호출자 식별, 제스처 인식, 포스처 인식, 행동인식 등의 수행을 통해 로봇과 상호작용하는 기술을 포함한다.

오디오 기반 인간-로봇 상호작용 기술은 로봇에 부착된 마이크로폰으로부터 얻어진 음성정보

* 본 연구는 지식경제부 및 한국산업기술평가원의 산업원천기술개발사업의 일환으로 수행하였음(KI001836, 인간-로봇 상호작용 매개기술 개발).

에 근거해서 음성 인식, 화자 인식, 음원 추적, 음원 분리 등의 수행을 통해 로봇과 상호작용하는 기술을 포함한다. 멀티모달 정보 기반 상호작용 기술은 비디오, 오디오 및 다른 센서 기반의 인식 기술을 융합하여 사용자 인식 등을 포함한 로봇과의 상호작용 기술을 포함한다.

이들 기술 중 제스처 인식은 먼 거리나 잡음 환경에서 유용한 정보전달 수단이 될 수 있으며 로봇의 동작 제어, 사용자에게 대한 서비스 제공 및 게임 등에 활용될 수 있다. 일반적으로 카메라 기반 제스처 인식은 카메라 영상 기반 제스처 인식과 3D 기반 제스처 인식으로 나눌 수 있다. 전자의 카메라 영상 기반 제스처 인식은 하나 혹은 그 이상의 카메라를 이용하여 제스처 영상을 획득한 다음 영상 및 모션 등과 같은 특징 정보로부터 의미있는 제스처를 추출하여 인식하는 방법이다. 후자의 3D 기반 제스처 인식 방법은 카메라로부터 획득한 입력 영상으로부터 3차원 데이터를 추출하여 제스처를 인식하는 것이다.

카메라 기반 제스처 인식을 위하여 여러 가지 방법들이 제안되었으나 대부분의 연구가 극히 제한된 실험실 환경에서 이루어졌거나 소수의 실험 대상자들에 대해서 제스처 데이터를 얻고 인식 결과를 도출하여 왔다. 또한, 다양한 조명 환경 및 배경을 가지는 실세계 환경에서 제스처를 검출하여 인식하는 것은 상당히 어렵다(김계경, 2005). 이러한 이유로 카메라 영상 기반 제스처 인식 기술의 경우 인식 대상 제스처의 종류를 10개 이내로 제한하는 경우가 많다.

최근에는 센서 기술의 발전으로 저가의 센서를 활용한 모바일 폰, 웨어러블 컴퓨터, 정보기기 및 로봇 등에 대한 자연스러운 인터페이스에 대한 대안 연구가 많이 행해지고 있으며, 제스처가 그 중 가장 자연스러운 인터페이스 중의 하나로 인식되고 있다.

앞에서 설명한 바와 같이 일반적으로 제스처 인식은 컴퓨터 비전 기술에 의해 행해지고 있으나, MEMS 타입의 관성 센서의 도입이후 관성센서, 특히 가속도 센서는 제스처 인식에 많이 활용되고 있다(Milner, 1999; Sawada and Hashimoto, 1997; Farella et al., 2007; Benbasat et al., 2001).

가속도 센서 기반의 제스처 인식은 카메라 기반의 인식과 비교할 때, 조명 환경, 배경 및 카메라 시야각 등의 영향을 받지 않고 처리속도가 빠르다는 장점을 가진다. 현재로서는 로봇에 부착된 카메라가 아닌 별도의 장치가 필요하다는 단점을 가지나 미래의 유비쿼터스 환경에서는 이들 센서들이 시계, 모바일 폰 등의 생활 필수품에 내재될 수 있으므로 미래 환경에서의 인터페이스로서 활용 가능성이 높다고 할 수 있다.

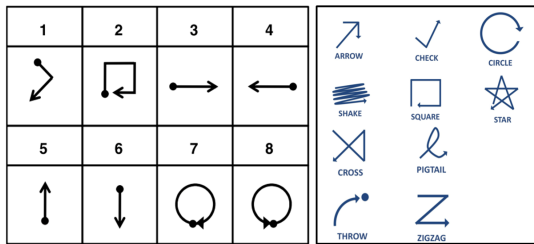
본 논문에서는 이러한 맥락에서 로봇 인터페이스 활용을 위한 가속도 센서 기반의 제스처 인식 및 이를 적용한 사례에 대해 기술한다. 본 논문에서 제안하는 기술을 활용하여 기존에 터치스크린 활용 중심으로 개발된 로봇 콘텐츠들이 향후 여러 형태의 장치 등장 및 유비쿼터스 컴퓨팅 환경의 정착과 함께 로봇의 인터페이스 및 콘텐츠의 다양한 확장이 가능할 것으로 기대한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 가속도 센서를 이용한 제스처 인식 기술의 동향 및 문제점과 문제 해결을 위한 본 논문에서의 접근방법에 대해 기술한다. 제 3장에서는 로봇 인터페이스 활용을 위한 가속도 센서 기반 제스처 인식 기술 알고리즘에 대해 기술하고, 제 4장에서 제스처 인식 실험 결과를 제시한다. 제 5장에서는 가속도 센서 기반 제스처 인식 기술의 로봇 콘텐츠에의 적용 사례를 통해 향후 활용가능한 분야를 예를 제시하고, 마지막으로 제 6장에서 결론을 맺는다.

2. 가속도 센서기반 제스처 인식 기술

2.1 관련 기술 동향 및 문제점

가속도 센서를 포함한 관성 센서 기반 제스처 인식에 관한 기존 연구의 대부분은 휴대장치를 기울이는(Tilt) 것과 같은 단순 동작을 감지하거나, 장치 제어를 위한 소규모 제스처 집합을 인식하는 연구에 집중되어 있다. 이러한 부류의 연구에서 다루는 제스처 패턴의 전형적인 예는 <그림 1>에서 보는 바와 같다.



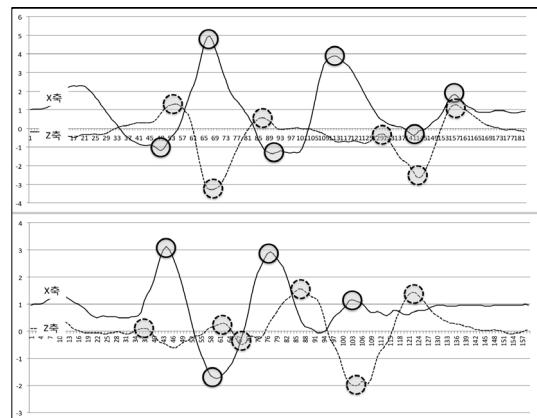
<그림 1> 통상적인 제스처 패턴 사례(Liu et al., 2008; Leong et al., 2009)

이와 같은 제스처들은 패턴 구조가 단순할 뿐 아니라 패턴 간 변별력이 높아 특징 선정이 용이하고 패턴 인식 성능이 높다. 기존의 연구 사례들을 보면 가속도 신호를 그대로 특징으로 사용하거나 가속도 신호 값의 크기에 따라 간단하게 양자화한 결과를 특징으로 활용하여 95% 이상의 높은 제스처 인식률을 획득하였다(Liu et al., 2008; Leong et al., 2009; Kallio et al., 2006).

또 다른 연구 사례에서는 휴대 전화를 이용한 단축번호 다이얼링 등의 응용을 위해 숫자 인식을 다루었다(Oh et al., 2004; Lim et al., 2007). 숫자 패턴은 <그림 1>의 패턴들에 비해 복잡도가 상대적으로 높을 뿐 아니라 '0'과 '6', '1'과 '7', '2'와 '3' 등 상호 변별이 어려운 패턴들이 포함되어 있어

인식이 어렵다. 숫자 패턴을 대상으로 한 인식을 실험에서 Oh and Cho(2004)는 가속도 센서와 자이로 센서 융합을 통해 궤적을 추출하여 인식 실험을 수행한 결과 93% 내외의 인식률을 획득하였고, Lim et al.(2007)은 가속도 신호를 직접 특징 벡터로 활용하여 91% 수준의 인식률을 획득하였다.

기존 연구들은 대체로 사용자 종속적 패턴 인식을 다루었다(Milner, 1999; Sawada and Hashimoto, 1997; Farella et al., 2007; Bebasat et al., 2001; Oh et al., 2004; Lim et al., 2007; Liu et al., 2008). 일반적으로 관성센서기반 제스처 인식 실험 결과 사용자 독립 인식 성능은 사용자 종속 인식 성능에 비해 매우 낮은 것으로 보고되었다(Leong et al., 2009; Kratz et al., 2007). 예를 들어, Leong et al.(2009)은 <그림 1>의 우측에 보인 10개의 제스처를 대상으로 사용자 종속 및 독립 인식 실험을 수행하였는데, 사용자 종속 인식률은 97%에 이르는 데 반해 사용자 독립 인식률은 72%에 그치는 것으로 나타났다.



<그림 2> 제스처 입력 가속도 신호의 개인별 차이

사용자 독립 인식 성능이 저조한 이유는 동일한 동작을 수행하더라도 힘의 세기 및 시간에 따른 분

배, 동작 속도 변화 등이 사람에 따라 많은 편차를 나타내기 때문이다. <그림 2>는 두 사람이 영문 알파벳 'A'를 입력한 동작의 가속도 신호를 각각 보여준다. 위와 아래의 파형을 비교해 보면 주요 최고/최저점의 개수 및 상대적 위치, X축과 Z축의 교차 위치 및 위상 등이 상이함을 관찰할 수 있다.

Kallio et al.(2006)은 <그림 1>의 좌측에 보인 8개의 제스처를 대상으로 가속도 센서 기반 사용자 독립 제스처 인식을 연구하여 평균 96% 내외의 우수한 성능을 이끌어내기도 하였다. 그러나 이 연구에서는 인식 대상 패턴 데이터가 단순할 뿐 아니라 실험 대상 인원이 7명으로 제한되어 있어 일반화된 결과로 보기 어렵다. 뿐만 아니라 Kallio 등과 유사한 제스처를 대상으로 하고 동일한 인식 알고리즘을 적용한 다른 연구의 경우 사용자 독립 인식률은 매우 낮았다(Leong et al., 2009; Kratz et al., 2007).

이상에서 살펴본 바와 같이 가속도 센서를 이용한 제스처 인식은 다양한 장점을 제공함에도 불구하고, 인식 성능이 패턴의 복잡도에 민감하고 사용자 독립적인 인식기 개발이 어렵다는 문제점을 안고 있다.

본 연구는 로봇 인터페이스 활용을 위한 3축 가속도 센서 기반 제스처 인식 기술 개발을 목적으로 한다. 로봇 기반 콘텐츠 및 서비스의 다양한 시나리오 하에서 사용자는 다양한 제스처를 통해 로봇과 상호작용할 수 있으나, 본 논문에서는 영어 교육에 활용할 수 있는 26개의 영어 필기체 대문자 알파벳 인식을 중심으로 다룬다.

2.2 본 논문에서의 접근 방법

이미 언급한 바와 같이 가속도 센서를 이용하여 필기 제스처를 인식하기 위한 노력들이 있었다. 그러나 이러한 연구들은 10개의 숫자와 2~3개의 부

가적인 제스처를 포함하는 총 10~13개 정도의 소규모 제스처 인식에 제한되어 있다. 이에 비해 영문 알파벳은 패턴의 복잡도가 높을 뿐 아니라 클래스의 수도 많기 때문에 앞서 살펴본 가속도 기반 제스처 인식 기술의 문제점을 포함하는 도전적인 문제이다. 이에 본 연구에서는 상기 문제를 극복하기 위하여 아래와 같은 접근 방법을 취하였다.

2.2.1 궤적 특징 활용

가속도 센서 기반 제스처 인식의 경우 크게 두 가지 종류의 특징이 사용된다. 하나는 가속도 신호 자체의 특성을 나타내는 원시신호 특징(Raw Signal Features)이다. 전처리를 거친 가속도 신호, 최고점과 최저점의 개수, 가속도 평균값, 영교차 등이 원시신호 특징에 속하며, 이것들은 가속도 신호에서 통계적으로 추출될 수 있다. 다른 하나는 제스처의 궤적(Trajectory)이다. 제스처의 궤적은 가속도 신호를 두 번 적분함으로써 얻을 수 있다.

기존의 많은 연구는 원시신호 특징을 기반으로 한다. 원시신호 특징은 계산 비용이 적어 제한된 계산 자원을 보유한 장치에 적용하기가 용이하여 폭넓게 활용되고 있으나, 인식 대상 패턴의 복잡도가 낮아서 원시신호만으로도 충분한 인식 성능을 얻을 수 있다는 점도 작용하였다. 단순 동작 제스처보다 상대적으로 복잡한 숫자 패턴 인식의 경우 원시신호 특징을 사용한 경우보다 궤적을 사용한 경우 인식률이 5% 내외 개선되었다(Oh et al., 2004; Lim et al., 2007).

이와 같이 인식 대상 패턴의 복잡도가 높고 패턴 간 모호성이 커지면 원시신호 특징으로 충분한 인식 성능을 획득하기 어렵다. 숫자보다 더 복잡한 패턴을 가지는 영문 알파벳 제스처를 대상으로 이를 검증하기 위해 본 연구에서는 원시신호 특징과 궤적을 이용하여 상대적인 성능을 비교하기 위한

사전 실험을 수행하였다.

원시신호 특징으로 제스처 한 개 당 3축 가속도 신호에 대해 각 축 별로 최대 가속도 값, 최소 가속도 값, 피크(Peak) 개수, 첫 피크(Peak)의 방향, 최종 피크(Peak)의 방향, 최대 가속도 값과 최소 가속도 값의 차를 추출하였다. 원시신호 특징을 활용한 인식기로는 다층 퍼셉트론(MLP) 신경망 모델을 적용하여 학습 및 테스트하였다. 궤적을 이용한 인식 실험은 제 3장에서 소개할 인식 방법을 기반으로 수행하였다. 실험 결과 원시신호 특징 기반 사용자 종속 인식 성능은 평균 87%를 기록하여, 궤적을 기반으로 한 95% 내외의 인식 성능보다 낮았다.

2.2.2 사용자 적응형 온라인 학습 기법 활용

본 논문에서는 사용자 종속성 문제에 대한 해결책으로 강력한 사용자 독립적인 제스처 인식기를 만드는 대신 사용자 그룹에 최적화된 인식기를 손쉽게 제작할 수 있는 적응 메커니즘(Adaptation Mechanism)을 제공하는 방법을 채택하였으며, 크게 2가지 단계의 전략으로 구성된다. 1단계에서는 사용자 독립적이지만 사용자 별로 최적의 성능을 보장하지 않는 인식기를 제작한다. 2단계에서 사용자 그룹 내의 개별 사용자는 1단계에서 제작된 인식기를 사용하게 되며 인식기는 단순하고 직관적인 적응 알고리즘을 통해 사용자 그룹에 최적화된 인식기로 진화해간다. 2단계를 통해 인식기가 충분히 최적화되면 인식기는 해당 사용자 그룹의 제스처에 대해 최상의 인식 성능을 제공할 수 있다.

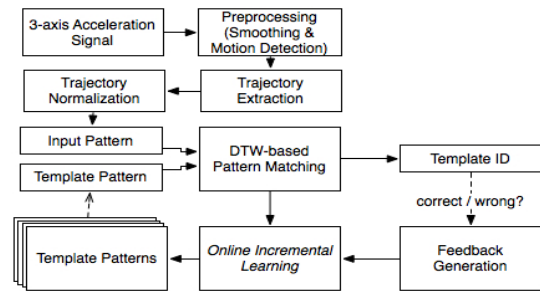
2단계의 인식기 적응 기능을 실현하기 위해 사용자의 제스처를 점증적으로 학습하여 인식기에 반영할 수 있어야 한다. 즉 인식기는 기존에 학습된 모든 데이터의 망각(Catastrophic Forgetting) 없이 점진적으로 새로운 학습 내용을 축적(Online Incremental Learning) 할 수 있어야 한다. 이러한

특성을 만족시키는 대표적인 학습 알고리즘은 인스턴스 기반 학습(Instance-based Learning)이다.

인스턴스 기반 학습은 학습용 패턴 데이터의 인스턴스를 기억하고 이를 차후 입력 대상 패턴과 비교함으로써 인식을 수행한다. 인스턴스 기반 학습은 매우 단순하고 일반화 기능이 결여되어 있으나 점증적인 온라인 학습을 통한 적응형 인식기 구현에 적합하다. 본 연구에서는 DTW(Dynamic Time Warping)와 1-NN(1-Nearest Neighbor)을 기반으로 한 인스턴스 기반 학습 알고리즘을 사용하여 사용자 적응형 제스처 인식기를 구현하였다.

3. 제스처 인식 알고리즘

본 장에서는 본 연구를 통해 개발한 가속도 기반 제스처 인식기인 GIT(Gesture Interface Technology)의 구현 알고리즘을 설명한다. GIT의 전체적인 구조는 <그림 3>에서 보는 바와 같다.

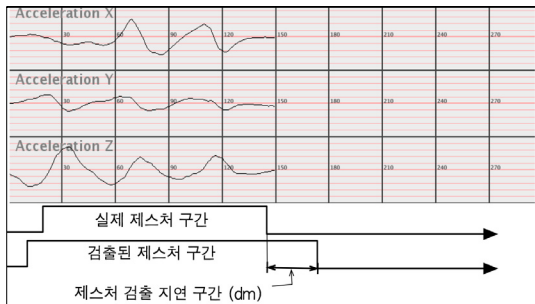


<그림 3> GIT의 구조

3.1 제스처 구간 검출

제스처 구간 검출은 제스처가 시작되는 시점과 종료되는 시점을 구별하는 기능이다. 기존의 많은 연구에서는 자동화된 제스처 구간 검출을 고려하지 않았으나, 자연스러운 제스처 기반 상호작용을 위해 제스처 구간 검출 기능은 필수적이다.

기존의 제스처 구간 검출 방법들은 제스처 구간 내의 가속도 신호 에너지가 제스처가 없는 구간에 비해 높다는 사실을 기반으로 동작한다(Liu et al., 2008). 제스처 에너지는 일정 길이의 구간을 대상으로 계산한 가속도 값의 표준편차 또는 분산이다. 제스처의 시작과 종료 시점은 제스처 에너지 값을 정해진 문턱값과 비교함으로써 검출한다.



<그림 4> 제스처 검출 지연 구간 정의

<그림 4>는 제스처 구간 검출 알고리즘의 동작 특성을 보여준다. 그림에서 보듯이 제스처 구간 검출 알고리즘은 실제 제스처 종료 시점보다 조금 늦게 제스처 종료 시점을 검출한다. 이러한 검출 지연은 일정 길이의 구간을 대상으로 제스처 에너지를 검출하기 때문에 발생한다.

사용자는 입력 조작 후 시스템이 150ms~200ms 사이에 응답을 줄 때 지연시간을 거의 느끼지 못한다는 연구 결과가 있다(Nittono, 2007). 따라서 본 연구에서는 GIT의 응답 시간을 200ms 이내로 유지하기 위해 <표 1>과 같이 다양한 제스처 검출 알고리즘을 적용·시험하였다. 제스처 검출 지연 시간 측정을 위해 사용자가 제스처를 입력할 때 입력 장치에 부착된 버튼을 눌러 직접 시작 및 종료 시점을 표시한 패턴 데이터를 활용하였다. 제스처 검출 지연 시간과 제스처 검출 정확도를 측정 한 실험 결과는 <표 1>과 같으며, 실험에 사용된

데이터는 <표 2>의 1차 수집 데이터이다.

<표 1> 제스처 검출 알고리즘 성능 평가

알고리즘	검출 지연시간	실제 제스처 개수	검출 제스처 개수	검출 오차
AMD1	279ms	1,294	1,263	31
AMD2	193ms	1,294	1,252	42
AMD3	176ms	1,294	1,281	13

최적의 성능을 기록한 AMD3 알고리즘은 다음과 같이 함수 $M(\cdot)$ 으로 표현할 수 있다. M 의 출력 값이 1이면 제스처 구간이며, 0이면 제스처 구간이 아니다.

$$M(V_t) = 1 \text{ if } V_t > th_b, \\ 0 \text{ if } V_t \leq th_b.$$

$$V_t = LP(\text{Norm}(\text{VAR}_{x,y,z}(A_{t-W}, A_{t-W+1}, \dots, A_t)))$$

위 식에서 A_t 는 t 시점의 3축 가속도 벡터를 나타낸다. V_t 는 t 시점의 제스처 에너지로서 길이가 W 인 제스처 구간인($t-W \sim t$)를 대상으로 계산된다. $\text{VAR}_{x,y,z}$ 는 제스처 구간 내에 속한 가속도 값 ($A_{t-W}, A_{t-W+1}, \dots, A_t$)에 대해 x, y, z 각 축 별로 분산을 계산하는 함수이며, Norm 은 계산된 3축 분산벡터의 L2-Norm을 계산하는 함수이다. LP 는 저역통과필터(LPF)를 나타낸다. th_b 는 제스처 구간을 구별하기 위한 문턱값이다.

대개의 경우 문턱값을 기준으로 한 제스처 구간 검출은 제스처의 세기나 속도에 민감하게 반응하여 높은 검출 오류를 발생시키게 되는데 저역통과필터(LP)를 적용함으로써 제스처 검출 알고리즘을 안정적으로 동작시킬 수 있었다.

AMD1과 AMD2는 각각 다음과 같이 분산을 두 번에 걸쳐 적용하는 방식으로 제스처 에너지를 계

산한다. 분산을 두 번 계산함으로써 제스처 구간 검출의 안정성을 확보할 수 있으나 대신 제스처 검출 지연 구간이 증가하는 문제점을 보였다.

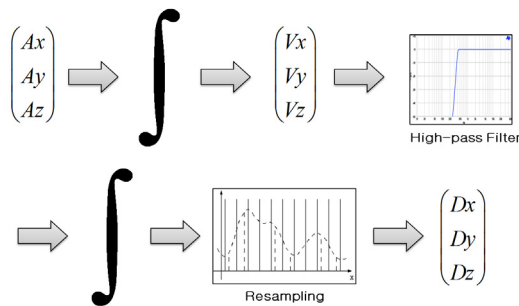
$$AMD1 : V_t = VAR(Norm(VAR_{x,y,z}(G(A_{t-w}, A_{t-w+1}, \dots, A_t))))$$

$$AMD2 : V_t = VAR(Norm(VAR_{x,y,z}(A_{t-w}, A_{t-w+1}, \dots, A_t)))$$

위 식에서 G는 1차원 가우시안 필터를 나타낸다.

3.2 궤적 추출

GIT는 가속도 신호로부터 궤적을 추출하여 패턴 특징으로 활용한다. 궤적 추출 과정은 <그림 5>와 같다. 가속도 벡터 A는 전처리 과정을 통해 잡음을 감쇄시킨 신호이다. 본 연구에서는 1차원 가우시안 필터를 적용하여 잡음을 감쇄하였다. 궤적 추출 알고리즘은 제스처 구간 추출 함수 M(·)의 값이 1인 경우에만 작동한다.



<그림 5> 궤적 추출 과정

가속도 벡터를 입력으로 1회 적분을 수행하면 속도 벡터 V를 얻는데 V의 값은 가속도 값의 축적으로 빠르게 증가해 버리므로 그대로는 사용할 수 없다. 따라서 V에 고역통과필터(HPF)를 적용하여 값의 지속적인 증가 파형을 제거한 결과를 2

차 적분에 적용한다. 2차 적분이 완료되면 제스처의 궤적이 출력된다. 궤적은 제스처의 수행 시간에 따라 가변적인 길이와 크기를 갖게 되는데, 이러한 패턴 간 변이를 감쇄시키기 위해 정규화를 거친다. 정규화를 위해 궤적을 구성하는 점의 개수를 통일하는 리샘플링(Resampling)과 궤적의 크기를 통일하는 스케일링(Scaling)을 수행한다(Wobbrock et al., 2007). D는 정규화를 거친 3차원 궤적 벡터이다.

3.3 패턴 학습 및 인식

앞서 언급한 바와 같이 GIT는 적응형 제스처 인식기로 제스처 패턴을 온라인 학습하여 점증적으로 인식 기능을 사용자의 동작 특성에 최적화시킬 수 있는 기능을 포함한다. 이러한 적응형 학습 시스템을 구현하기 위해 템플릿 정합을 기반으로 한 인스턴스 기반 패턴인식 알고리즘을 적용하였다. 인스턴스 기반 패턴인식은 패턴 샘플을 템플릿 저장소에 저장하는 것만으로 학습이 가능하므로 점증적인 온라인 학습에 적합하다.

본 연구의 인식 대상인 제스처는 시간 축 상에 위상 변이를 포함하므로 탄력적인 패턴 매칭이 가능한 DTW(Dynamic Time Warping)를 템플릿 정합에 적용하였다. 패턴 인식은 최상의 정합 결과 한 개를 인식결과로 결정하는 1-NN(1 Nearest Neighbor) 알고리즘을 적용하였다.

3.4 템플릿 집합 최적화

GIT는 온라인 학습을 통해 사용자의 제스처 패턴을 지속적으로 학습한다. 본 연구에서는 온라인 학습의 입력으로 제스처 패턴과 함께 패턴이 속한 클래스 이름이 함께 입력되는 상황을 가정한다. 즉, 온라인 상황에서 지도학습(Supervised Learning)이 가능해야 한다.

온라인 학습이 진행되면 GIT의 템플릿 집합은 지속적으로 확장된다. GIT의 초기 템플릿의 집합을 TS0라 하고 N번의 온라인 학습이 이루어진 후의 템플릿의 집합을 TSN이라 하면, TSN의 크기는 TS0보다 N만큼 증가된다. 템플릿 집합의 크기가 커지면 정합 대상 템플릿의 개수가 많아지므로 인식 속도가 느려지고, 템플릿 간 상호 간섭으로 인해 거짓 양성(False Positive)의 빈도수가 증가하여 인식률이 하락한다.

템플릿 집합 최적화는 이와 같은 문제점을 해결하기 위해 주기적으로 활용성이 낮은 템플릿을 제거하는 기능이다. 활용성이 낮은 템플릿은 패턴 인식에 활용되지 않는 템플릿들과, 거짓 양성(False Positive : FP)에 많이 기여하는 템플릿들이다.

GIT는 활용성이 낮은 템플릿을 구별하기 위해 TCM을 활용한다. TCM은 템플릿 혼동 행렬(Template Confusion Matrix)을 나타내는데, TCM의 각 열(Column)은 클래스 이름으로 구성되고 행(Row)은 템플릿 아이디로 구성된다. TCM(i, j)는 클래스 j에 속하는 제스처 패턴과 템플릿 i가 정합된 횟수이다. 만약 템플릿 i가 클래스 j를 대표하는 템플릿 집합에 속한다면 해당 정합은 진양성(True Positive : TP) 사례에 해당한다. 템플릿 i가 클래스 j를 대표하지 않는다면 거짓양성(FP) 사례가 된다. 따라서 TCM을 이용하면 각 템플릿이 TP와 FP에 기여하는 정도를 계산할 수 있다. TCM을 기반으로 템플릿의 품질을 계산하는 식은 다음과 같다.

$$PC(t) = \frac{TCM(t, L(t))}{\sum_{i \in T(L(t))} TCM(i, L(t))},$$

$$NC(t) = 1 - \frac{TCM(t, L(t))}{\sum_{i \in L} TCM(i, c)}$$

PC(t)는 템플릿 t의 긍정적 기여 정도(Positive

Contribution)를 나타내며, NC(t)는 t의 부정적 기여 정도(Negative Contribution)를 나타낸다. L(t)는 템플릿 t가 대표하는 클래스 이름을 나타내고, T(l)은 클래스 l을 대표하는 템플릿들의 집합을 나타낸다. 따라서 T(L(t))는 템플릿 t가 속한 클래스를 대표하는 모든 템플릿들의 집합을 나타낸다. L은 모든 클래스 이름의 집합이다.

위 식에 의하면, PC(t)는 t가 속한 클래스를 대표하는 모든 템플릿들의 TP 정합 횟수에 대한 t의 TP 정합 비율이다. 즉, PC(t)는 t가 TP 정합에 기여한 비중을 대표하는 값이다. NC(t)는 템플릿 t가 TP와 FP를 모두 포함하여 정합에 기여한 총 횟수 중 NP에 기여한 비중을 나타낸다. 즉, NC(t)는 t가 FP 정합에 기여한 비중을 대표하는 값이다.

만약 템플릿 t에 대하여 PC(t)가 0이면 t는 TP에 해당하는 패턴 인식에 전혀 기여하지 않은 것으로서 제거 대상이다. 또한 NC(t)가 0.5보다 크다면, t는 FP를 유발한 빈도수가 높은 것으로서 역시 제거 대상이다.

4. 실험

본 장에서는 제 3장에서 제시한 알고리즘을 기반으로 한 영문 알파벳 인식 실험 결과 및 분석 내용을 설명한다.

4.1 패턴 데이터 수집

제스처 데이터 수집을 위한 입력 장치로 Nintendo®사의 WiiTM Remote(Wiimote)를 사용하였다. Wiimote는 하나의 3축 가속도 센서를 포함하고 있으며, 100Hz의 주기로 샘플링된다. 제스처 인식기가 동작하는 호스트 컴퓨터와 Wiimote는 블루투스(Bluetooth) 무선 인터페이스를 통해 연결된다.

제스처 인식기의 인식 성능 실험을 위해 <표 2>와 같이 2차에 걸쳐 약 8,000여 개의 알파벳 제스처 데이터를 수집하였다. 1차 패턴 데이터는 성인 5명을 대상으로 각 알파벳을 10번씩 입력하도록 하였다. 또한 제스처를 입력할 때 사용자에게 제스처의 시작 시점과 종료 시점에 Wiimote의 버튼을 누르도록 지시하였다. 버튼을 누른 시점을 패턴 데이터에 제스처 구간을 표시하는 구분자를 기록하여 자동 제스처 구간 검출 알고리즘의 성능을 평가하기 위한 근거 데이터로 활용하였다. 제 3.1절의 제스처 구간 검출 알고리즘 평가 데이터는 1차 수집 패턴 데이터를 대상으로 실험한 결과이다.

2차 패턴 데이터는 총 50명의 성인에게 각 알파벳을 5회씩 쓰도록 지시하여 수집한 패턴 데이터이다. 제스처 패턴 데이터는 제스처 구간 구분자를 포함하지 않으며 제스처를 연속 수행한 패턴들로 구성되어 있다. 알파벳 필기 제스처 인식 성능 측정을 위해 2차 패턴 데이터를 사용하였다.

<표 2> 영문 알파벳 제스처 패턴 데이터베이스

구분	연령대	인원	규모	동작구간 표시
1차	30~40	5	1,294개	O
2차	20~50	50	6,500개	X

4.2 실험 결과 및 분석

<표 2>의 2차 패턴 데이터를 대상으로 제스처 인식 실험을 수행하였다. 인식 실험은 사용자 종속 인식 방법으로 수행하였으며, 인식률은 10차 교차 검증 방법을 이용하여 산출하였다. 테스트 단계에서 제 3.4절에 설명한 템플릿 최적화 알고리즘이 수행되는데, 템플릿 최적화 수행 결과 제스처 인식기의 최종 템플릿 개수는 초기 템플릿 개수의 25% 수준으로 축소된다. 각 클래스 별 인식률은 <표 3>

에 보는 바와 같으며, 평균 인식률은 95.48%이다.

<표 3> 알파벳 제스처 인식 성능

(단위 : %)

클래스	인식률	클래스	인식률	클래스	인식률
A	100.00	J	100.00	S	100.00
B	100.00	K	100.00	T	100.00
C	95.65	L	96.15	U	75.00
D	88.00	M	100.00	V	100.00
E	96.65	N	88.46	W	100.00
F	95.45	O	95.65	X	100.00
G	100.00	P	74.07	Y	100.00
H	95.65	Q	100.00	Z	100.00
I	81.82	R	100.00		

<표 4> 타 기술과 성능 비교

출처	Kela et al., 2006	Leong et al., 2009	Oh et al., 2004	본 연구
제스처 종류	가전제어	액션게임	숫자 및 심볼	영문 알파벳
제스처 개수	8	10	11	26
제스처 DB 규모	7인×20회 총 1,120개	50인×10회 총 5,000개	100인×3회 총 3,300개	50인×5회 총 6,500개
인식 알고리즘	HMM	DTW	BN+SVM	DTW
인식률	96.99%	97.00%	96.9%	95.48%
응답시간	8.3ms (동작구간 검출 시간 미포함)	NA	NA	190ms (동작구간 검출 시간 포함)
제스처 구간 자동검출	X	X	O	O
제스처 인식 개인화	X	O	X	O

<표 4>는 본 논문에서의 제스처 인식기의 성능 및 기능을 타 기술과 비교한 것이다. 제스처의 종류 측면에서 타 연구는 가전제품 제어 제스처, 액

션 게임 제스처, 숫자 및 간단한 심볼 등 제스처의 종류가 8~11개로 소규모이고, 개별 제스처의 패턴 복잡도도 본 연구의 영문 알파벳에 비해 단순하므로 본 연구의 인식 성능은 매우 우수하다고 판단할 수 있다.

실험 대상 패턴의 규모는 Kela et al.(2006)을 제외하면 모든 연구가 비교적 대규모 패턴 데이터를 대상으로 하므로 실험 결과를 일반화하는데 무리가 없다. 다만, 타 연구에서는 자동 제스처 구간 검출을 위한 패턴 데이터베이스를 별도로 구축한 사례가 없을 뿐 아니라 객관적인 제스처 구간 검출 성능 지표가 제시되어 있지 않다.

인식 알고리즘은 Leong et al.(2009)과 본 연구의 경우 DTW를 사용하였으며 Kela et al.(2006)의 경우 HMM을, Oh et al.(2004)의 경우 베이지안 네트워크와 SVM(Support Vector Machine)을 사용하였다. 인식률의 경우 대체적으로 비슷하나 본 연구의 경우 클래스의 수가 26개로 가장 많고 개별 패턴도 복잡한 것을 감안하면 가장 좋은 성능을 나타낸다고 할 수 있다.

응답시간은 제 3.1절의 <그림 4>를 통해 보인 동작 검출 지연 시간에 패턴 인식 시간을 합한 시간이다. Kela et al.(2006)의 경우 제스처 구간 자동 검출 기능을 제공하지 않기 때문에 제스처 검출 시간을 포함하지 않는 순수 패턴 인식 시간이므로, 실제 온라인 상황 하에서 시스템의 응답 성능을 정확히 알 수 없다. 본 연구는 상대적으로 우수한 인식 성능 뿐 아니라 200ms 이내의 시스템 응답 성능을 보장하므로 사용성 측면에서 비교 시스템 중 우수하다고 판단된다.

기능적인 측면에서 본 연구의 경우 제스처 구간을 자동으로 검출하는데 반해 Kela et al.(2006) 및 Leong et al.(2009)의 경우 그러한 기능을 제공하지 않는다. 제스처 구간 자동 검출의 정확도에 따

라 최종 인식 성능이 달라질 수 있다는 점을 감안하면, 이들 두 연구의 경우 제스처 구간을 수동 검출하였기 때문에 이것이 인식률의 향상에 기여했을 것으로 유추해볼 수 있다.

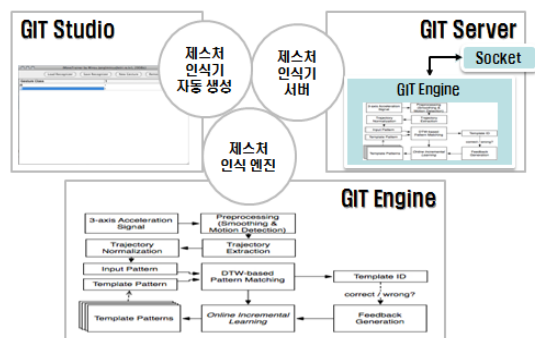
온라인 학습을 통한 인식기의 사용자 적응 기능은 CMU와 본 연구가 모두 제공하는 기능이나 CMU의 경우 템플릿 집합을 최적화하는 등의 방법을 통해 템플릿 집합의 규모와 인식 성능을 적정 수준으로 유지하는 방안에 대하여는 방안을 제시하고 있지 않다.

5. 응용 사례 연구

본 장에서는 제 3장에서 제시한 알고리즘을 기반으로 구현한 소프트웨어 시스템의 구성과 기능을 설명하고, 로봇기반 서비스 응용 사례를 소개한다.

5.1 시스템 패키지화 및 기업 적용

제 3장에서 설명한 알고리즘을 기반으로 개발한 제스처 인식기 패키지는 <그림 6>과 같이 GIT Engine, GIT Studio, GIT Server로 구성된다. GIT Engine은 제스처 인식기의 핵심 구성요소로서 제 3장에서 설명한 제스처 인식 알고리즘을 구현한



<그림 6> 제스처 인식기 패키지 구성

다. GIT Studio는 인식 대상 제스처 데이터를 수집하고 수집한 데이터를 기반으로 제스처 인식기를 생성하고 시험하는 도구이다. 마지막으로 GIT Server는 GIT Studio에 의해 자동 생성된 제스처 인식기를 클라이언트-서버 구조로 손쉽게 다양한 응용에 연동할 수 있도록 해 준다.

<그림 7>은 제스처 인식기 패키지의 활용 순서를 도시한 것이다. 그림에 보인 바와 같이 제스처 인식기 패키지를 활용하여 빠르게 원하는 제스처 인식기를 제작하여 응용 콘텐츠와 연동하는 과정은 다음과 같다.

<그림 7>(a)는 인식하고자 하는 제스처의 패턴 클래스 목록을 입력하는 과정이다. 영문 대문자 알파벳 26자를 인식대상으로 할 경우 A~Z까지를 입력하면 된다.

<그림 7>(b)는 각 제스처별로 제스처 데이터 샘플을 입력하는 단계이다. 그림에는 나타나 있지 않으나 제스처를 입력할 때 3축 가속도 센서를 내장한 휴대 장치를 이용한다. 입력 가능한 제스처 샘플의 개수에는 특별한 제한이 없으며 GIT Studio는 입력된 제스처 샘플을 사용하여 전처리 및 제스처 구간 검출 알고리즘을 수행하여 그 결과 패턴을 템플릿으로 보관한다.

<그림 7>(c)는 제스처 인식기를 생성하여 저장하는 단계를 보여준다. <그림 7>(b) 단계에서 제스처 샘플 입력이 완료된 후 “Save Recognizer” 버튼을 누르면 저장할 인식기의 이름을 입력하는 창이 뜨고 인식기의 이름을 입력하면 궤적 추출, 궤적 정규화 및 DTW 기반 패턴 매칭을 수행하고 학습된 인식기를 파일로 저장한다.

<그림 7>(d)는 <그림 7>(c) 단계에서 저장한 제스처 인식기에 대한 기능시험을 위한 준비 과정을 보여준다. 그림에 보이는 “Load Recognizer” 버튼을 누르면 저장된 인식기들의 리스트를 보여주

며, 그 중에서 사용할 인식기 파일을 선택하면 인식기가 메모리에 적재된다.

<그림 7>(e)는 불러온 제스처 인식기를 시험하는 과정을 보여준다. 시스템은 시험할 제스처 인식기의 인식 대상 패턴을 임의의 순서대로 사용자에게 제시하고 사용자는 제시된 제스처를 입력 장치를 이용하여 입력하게 된다. 입력한 제스처 패턴에 대해 인식 과정을 수행하여 그 결과를 기록함으로써 인식 성능을 측정한다.

그림에서 보듯이 화면의 왼쪽에 사용자가 입력해야 할 패턴 이름이 표시되고 우측에 사용자가 입력한 패턴에 대한 인식 성공·실패 여부가 표시된다. 이 단계에서 온라인 점층 학습 모드를 선택할 경우 학습이 안 되어 있는 새로운 패턴이 나타난 경우 해당 패턴을 온라인 학습 기능을 통해 템플릿 집합에 추가함으로써 인식기의 기능을 확장시킬 수 있다. 따라서 <그림 7>(e)의 시험 과정은 패턴 데이터 수집 및 인식기 학습 용도로 활용할 수 있다.

<그림 7>(f)는 <그림 7>(a)~<그림 7>(e)의 과정을 통해 생성한 제스처 인식기를 콘텐츠와 연동하여 활용하는 장면을 나타낸 것이다. 화면상에 표시된 콘텐츠는 동물 그림과 동물에 대한 영어 이름이 멀리서 가까이 다가올 때 동물의 영문이름 첫 글자를 입력 장치로 허공에 써서 입력하는 영어 학습용 에듀테인먼트 콘텐츠이다.

콘텐츠와의 연동을 위해 GIT Server가 활용되는데, 콘텐츠는 클라이언트 역할을 수행하여 입력 장치로부터 가속도 신호를 받아 GIT 서버에 전달하고 그 응답으로 GIT 서버가 제공하는 인식 결과를 수신하여 게임 진행에 활용한다. 클라이언트인 콘텐츠 입장에서는 게임의 진행에 따라 인식 대상 제스처 집합이 변경되더라도 해당 제스처 인식기를 확보하여 GIT 서버에 탑재하기만 하면 다양한

제스처 인식 기능을 활용할 수 있다.

본 제스처 인식기 패키지는 총 5개 기업을 통해 실제 응용 콘텐츠 연동 및 실증 실험을 진행하였다. 5개 기업 중 3개 기업과는 제스처 인식기 패키지를 제공한 후 지속적으로 의견 교환이 있었으며 이 과정에서 관찰한 결과 각 기업들에서 만든 콘텐츠와의 연동을 위해 필요한 시간은 샘플 데이터 수집을 포함하여 수 시간 정도로 매우 짧았다.

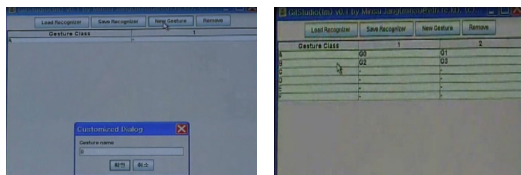
이를 통해 본 논문에서 개발한 제스처 인식기 패키지가 사용자의 관점에서 볼 때 자신들이 원하는 제스처 인식기를 만들어 자신들의 콘텐츠와 즉시 연동하여 활용해 볼 수 있는 수단으로 매우 유용함을 확인할 수 있었다.

서 동작하는 체감형 콘텐츠를 위한 제스처 기반 상호작용 인터페이스로서 GIT의 활용 가능성을 살펴보기 위해 다양한 형태의 장치 및 콘텐츠에 GIT를 적용한 사례를 그림과 설명을 통해 소개한다.

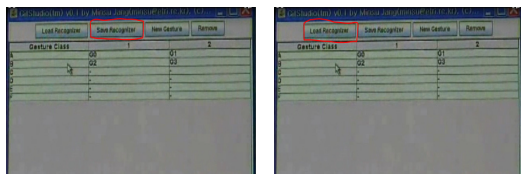
<그림 8>(a)는 제 4.1절에서 설명된 동물 이름 맞추기 게임이다. 이 콘텐츠는 두 번의 전시회(2008 로보월드, 2009 신성장동력박람회)에 출품하여 총 7일간 100여명의 어린이들이 실제 체험해 본 것으로 알파벳을 어느 정도 알고 있는 어린이의 경우 쉽게 사용법에 적응하였다. <그림 8>(b)는 알파벳 인식기능을 이용한 퍼즐 게임이다. 게임의 방법은 일반적인 퍼즐 맞추기와 동일하며 전시회(2009 로보월드)에 출품하여 어린이 및 일반인들이 체험하였다.

<그림 8>(c)는 산술학습 게임 콘텐츠로 화면 상단에서 떨어지는 우주선에 표시된 문제의 답을 맞히면 화면 하단의 대포에서 포가 발사되어 우주선을 폭파시키는 게임이다. 예를 들어, 우주선에 “10-4=”과 같은 문제가 출제되면 정답 “6”을 쓸 경우 그 문제를 제시한 우주선을 대포를 이용해서 폭파시킬 수 있다. <그림 8>(d)는 <그림 8>(b)의 퍼즐 게임과 유사한 게임을 아이폰의 앱으로 개발한 것이다. 제 3장에서 설명한 제스처 인식기는 아이폰 용으로도 개발되어 아이폰 상에서의 제스처 인식기 개발도 쉽게 할 수 있다.

<그림 8>(e)는 기존에 개발된 콘텐츠와의 연동 사례이다. 콘텐츠는 국내 모 업체에서 인터넷 상에서 상용으로 서비스를 제공한 영어 교육용 파닉스 프로그램이다. 기존 시스템의 입력 수단은 문자 입력을 위한 키보드와 화면 클릭 및 아이콘 조작을 위한 마우스였으나, 본 기술 개발 결과물을 활용하며 가속도 센서 기반 영어 알파벳 인식기, 숫자 인식기를 입력 수단으로 대체하였다. 드래그의 경우는 Wiimote에서 제공하는 IR 센서/카메라의 출력 정보를 활용하여 구현하였다.



(a) 대상 제스처 이름 입력 (b) 제스처 샘플 입력



(c) 제스처 인식기 저장 (d) 제스처 인식기 로드



(e) 제스처 인식기 테스트 (f) 콘텐츠와 연동

<그림 7> 제스처 인식기 패키지 활용 순서

4.2 적용 사례

본 절에서는 로봇 또는 PC 및 모바일 기기 상에



<그림 8> 제스처기반 체감형 콘텐츠 인터페이스에 적용



<그림 9> 로봇 콘텐츠 인터페이스에 적용

<그림 9>는 로봇 플랫폼을 대상으로 한 콘텐츠와의 상호작용을 나타낸 것이다. <그림 9>(a)는 유아용 로봇 플랫폼 상의 영어 단어 퀴즈 콘텐츠와의 상호작용 사례이며, <그림 9>(b)는 초등학생

을 대상으로 기획된 총 10종의 교육용 콘텐츠와의 상호작용에 활용된 사례이다.

상기 분야 외에도 로봇을 제어하기 위한 명령 제스처, 유아용 로봇 플랫폼을 활용한 참참참 게임, 댄스 게임 등의 콘텐츠에 본 논문의 기술을 적용하였으며, 새로운 인터페이스로의 가능성을 확인하였다.

6. 결 론

본 논문에서는 기존의 로봇과의 상호작용을 위해 사용되는 영상·음성 인식 기술을 보완할 수 있는 가속도 센서 기반의 제스처 인식 기술 및 상기 기술의 로봇 콘텐츠에의 적용에 대해 기술하였다. 본 논문에서 제시하는 기술의 기술적 검증을 위해 제스처 인식 문제로서는 복잡한 문제이면서 콘텐츠에의 활용 가능성이 높은 영문 알파벳 인식을 대상으로 성능을 검증하였다. 성능 평가 결과 타 연구에서 비교적 단순한 제스처를 대상으로 한 인식 결과와 비슷한 성능을 알파벳을 대상으로 한 실험에서 달성하였다.

실제 로봇에의 적용 가능성을 살펴보기 위해 로봇 상에서 동작하는 콘텐츠와 연동하여 미취학 아동 및 초등학생을 대상으로 실험을 하였으며 영상이나 음성인식 기술의 대안 또는 보완적인 기술로 활용이 가능함을 확인하였다. 본 논문에서 제안하는 기술을 활용하여 기존에 터치스크린 활용 중심으로 개발된 콘텐츠들이 앞으로 다양한 형태의 장치의 등장과 함께 로봇의 인터페이스 및 콘텐츠가 다양한 형태로 확장이 가능할 것으로 기대한다.

참고문헌

김계경, 김혜진, 조수현, 이재연, “인간-로봇 상호

- 작용을 위한 제스처 인식 기술”, *전자통신동향분석*, 20권 2호(2005).
- 곽근창, 김혜진, 배경숙, 윤호섭, “오디오 기반 인간 로봇 상호작용 기술”, *전자통신동향분석*, 22권 2호(2007).
- Milner, B., “Handwriting recognition using acceleration-based motion detection”, *IEE Colloquium on Document Image Processing and Multimedia*(Ref. No. 1999/041), 1999.
- Sawada, H. and S. Hashimoto, “Gesture Recognition Using an Acceleration Sensor and Its Application to Musical Performance Control”, *Electronics and Communications in Japan, Part III*, Vol.80, No.5(1997), 9~17.
- Farella E., L. Benini, B. Riccò and A. Acquaviva, “MOCA : A Low-Power, Low-Cost Motion Capture System Based on Integrated Accelerometers”, *Advances in Multimedia*, 2007.
- Benbasat, A. Y. and A. Paradiso, “An Inertial Measurement Framework for Gesture Recognition and Applications”, *International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, London, 2001.
- Wilson, D. H. and A. Wilson, “Gesture Recognition Using The XWand”, Technical Report CMU-RI-TR-04-57, Robotics Institute, Carnegie Mellon University, 2004.
- Kela, J., P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo and D. Marca, “Accelerometer-based gesture control for a design environment”, *Personal and Ubiquitous Computing*, Vol.10, No.5(2006), 285~299.
- Bailador, G., D. Roggen, G. Tröster and G. Triviño, “Real time gesture recognition using Continuous Time Recurrent Neural Networks”, In Proceedings of the ICST 2nd international conference on Body area networks, 2007.
- Oh, J. K., S. J. Cho, W. C. Bang, W. Chang, E. S. Choi, J. Yang, J. K. Cho and D. Y. Kim, “Inertial Sensor Based Recognition of 3-D Character Gestures with an Ensemble of Classifiers”, In Proceedings of Ninth International Workshop on Frontiers in Handwriting Recognition, 2004.
- Lim, J. G., Y. I. Sohn and D. S. Kwon, “Real-time Accelerometer Signal Processing of End Point Detection and Feature Extraction for Motion Detection”, *International Federation of Automatic Control-Human Machine System*, Seoul, Korea, 2007.
- Liu, J., Z. Wang, L. Zhon, J. Wickramasuriya and V. Vasudevan, “uWave : Accelerometer-based Personalized Gesture Recognition”, TR0630-08, Rice University and Motorola Labs, 2008.
- Leong, T. S., J. Lai, J. Panza, P. Pong and J. Hong, “Wii Want to Write : An Accelerometer Based Gesture Recognition System”, 2009.
- Kratz, L., M. Smith and F. J. Lee, “Wiizards : 3D gesture recognition for game play input”, *Proceedings of the 2007 conference on Future Play*, Toronto, Canada, 2007.
- Kallio, S., J. Kela, P. Korpupaa and J. Mäntyjärvi, “User Independent Gesture Interaction For Small Handheld Devices”, *International Journal of Pattern Recognition*, Vol.20, No.4 (2006), 505~524.
- Nittono, H., “Event-Related Brain Potentials Corroborate Subjectively Optimal Delay in Computer Response to a User’s Action”, *HCI 2007, LNAI 4562*(2007), 575~581.
- Wobbrock, J. O., A. D. Wilson and Y. Li, “Gestures without libraries, toolkits, or training: a \$1 recognizer for user interface prototypes”, In Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology(2007), 159~168.

Abstract

Accelerometer-based Gesture Recognition for Robot Interface

Minsu Jang^{*} · Yong-Suk Cho^{**} · Jaehong Kim^{*} · Joo-Chan Sohn^{*}

Vision and voice-based technologies are commonly utilized for human-robot interaction. But it is widely recognized that the performance of vision and voice-based interaction systems is deteriorated by a large margin in the real-world situations due to environmental and user variances. Human users need to be very cooperative to get reasonable performance, which significantly limits the usability of the vision and voice-based human-robot interaction technologies. As a result, touch screens are still the major medium of human-robot interaction for the real-world applications. To empower the usability of robots for various services, alternative interaction technologies should be developed to complement the problems of vision and voice-based technologies.

In this paper, we propose the use of accelerometer-based gesture interface as one of the alternative technologies, because accelerometers are effective in detecting the movements of human body, while their performance is not limited by environmental contexts such as lighting conditions or camera's field-of-view. Moreover, accelerometers are widely available nowadays in many mobile devices. We tackle the problem of classifying acceleration signal patterns of 26 English alphabets, which is one of the essential repertoires for the realization of education services based on robots. Recognizing 26 English handwriting patterns based on accelerometers is a very difficult task to take over because of its large scale of pattern classes and the complexity of each pattern. The most difficult problem that has been undertaken which is similar to our problem was recognizing acceleration signal patterns of 10 handwritten digits. Most previous studies dealt with pattern sets of 8~10 simple and easily distinguishable gestures that are useful for controlling home appliances, computer applications, robots etc.

Good features are essential for the success of pattern recognition. To promote the discriminative power upon complex English alphabet patterns, we extracted 'motion trajectories' out of input acceleration signal and used them as the main feature. Investigative experiments showed that classifiers based on trajectory performed 3%~5% better than those with raw features e.g. acceleration signal itself or statistical figures. To minimize the distortion of trajectories, we applied a simple but effective set of smoothing filters and band-pass filters.

* Electronics and Telecommunications Research Institute

** Konyang University

It is well known that acceleration patterns for the same gesture is very different among different performers. To tackle the problem, online incremental learning is applied for our system to make it adaptive to the users' distinctive motion properties. Our system is based on instance-based learning (IBL) where each training sample is memorized as a reference pattern. Brute-force incremental learning in IBL continuously accumulates reference patterns, which is a problem because it not only slows down the classification but also downgrades the recall performance. Regarding the latter phenomenon, we observed a tendency that as the number of reference patterns grows, some reference patterns contribute more to the false positive classification. Thus, we devised an algorithm for optimizing the reference pattern set based on the positive and negative contribution of each reference pattern. The algorithm is performed periodically to remove reference patterns that have a very low positive contribution or a high negative contribution.

Experiments were performed on 6500 gesture patterns collected from 50 adults of 30~50 years old. Each alphabet was performed 5 times per participant using Nintendo® Wii™ remote. Acceleration signal was sampled in 100hz on 3 axes. Mean recall rate for all the alphabets was 95.48%. Some alphabets recorded very low recall rate and exhibited very high pairwise confusion rate. Major confusion pairs are D(88%) and P(74%), I(81%) and U(75%), N(88%) and W(100%). Though W was recalled perfectly, it contributed much to the false positive classification of N. By comparison with major previous results from VTT (96% for 8 control gestures), CMU (97% for 10 control gestures) and Samsung Electronics(97% for 10 digits and a control gesture), we could find that the performance of our system is superior regarding the number of pattern classes and the complexity of patterns.

Using our gesture interaction system, we conducted 2 case studies of robot-based edutainment services. The services were implemented on various robot platforms and mobile devices including iPhone™. The participating children exhibited improved concentration and active reaction on the service with our gesture interface. To prove the effectiveness of our gesture interface, a test was taken by the children after experiencing an English teaching service. The test result showed that those who played with the gesture interface-based robot content marked 10% better score than those with conventional teaching.

We conclude that the accelerometer-based gesture interface is a promising technology for flourishing real-world robot-based services and content by complementing the limits of today's conventional interfaces e.g. touch screen, vision and voice.

Key Words : Human-Robot Interaction, Gesture Pattern Recognition, Accelerometer Signal Processing

저자 소개



장민수

현재 한국전자통신연구원 로봇/인지시스템연구부에서 선임연구원으로 재직 중이다. 서강대학교 전자계산학과에서 인공지능 전공으로 석사학위를 취득하였으며, 현재 한국과학기술원 전산학과에서 박사과정 중이다. 시맨틱웹, 매개인터페이스 등 분야에서 국책과제에 참여하였다. 주요 관심분야는 추론과 온톨로지, 패턴인식 등이다.



조용석

성균관대학교 전기공학과에서 공학사(1989), 공학석사(1992년), 공학박사(1998) 학위를 취득하였으며, 현재 건양대학교 전자정보공학과에서 부교수로 재직 중이다. 주요 관심 분야는 다변수 제어시스템, 로봇제어, 센서 및 생체계측, 재활보조기기제어 및 USN 네트워크시스템이다.



김재홍

현재 한국전자통신연구원 로봇/인지시스템연구부에서 팀장으로 재직 중이다. 경북대학교 컴퓨터공학과에서 자연어처리 전공으로 석사 및 박사학위를 취득하였다. 시맨틱웹, 감성로봇, 매개인터페이스, 실버케어로봇 등 분야에서 국책과제에 참여연구원 및 과제책임자로 활동하고 있다. 주요 관심분야는 인간로봇 상호작용, 실버케어 로봇 등이다.



손주찬

현재 한국전자통신연구원 로봇/인지시스템연구부에서 부장으로 재직 중이다. 생산관리시스템, 전자적자원관리시스템, 지능형웹서비스, 감성로봇 등의 분야에서 국책과제의 책임자로 활동하였다. 지능형 로봇을 기반으로 한 사회 복지 증진과 로봇의 산업화를 위해 많은 노력을 기울이고 있다.