

동적 베이지안 네트워크를 이용한 멀티모달센서기반 사용자 행동인식

(Activity Recognition based on Multi-modal Sensors using Dynamic Bayesian Networks)

양 성 익 [†] 홍 진 혁 [†]
 (Sung-Ihk Yang) (Jin-Hyuk Hong)

조 성 배 ^{**}
 (Sung-Bae Cho)

요 약 최근 유비쿼터스 컴퓨팅에 대한 관심이 높아지면서 유비쿼터스 환경에서의 서비스를 위한 인간과 컴퓨터의 상호 작용, 특히 인간의 행동을 인식하는 연구가 활발히 진행되고 있다. 기존의 영상기반 연구와는 달리 모바일 환경에 적합하도록 가속도 센서, 생리신호 센서 등 다양한 센서들을 활용하여 사용자의 행동을 인식하는 기법이 연구되고 있다. 본 논문에서는 멀티모달 센서들을 통합하고 동적 베이지안 네트워크를 계층적으로 구성하여 사용자의 행동을 인식하는 방법을 제안한다. 연산량이 비교적 적은 베이지안 네트워크로 전반적인 사용자 행동을 추론하고 획득된 각 행동의 확률순으로 동적 베이지안 네트워크를 구성한다. 동적 베이지안 네트워크는 OVR(One-Versus-Rest) 전략으로 학습되며, 확률순으로 행동이 검증되어 임계치를 넘는 경우 선택된 행동보다 낮은 확률의 행동에 대한 동적 베이지안 네트워크를 검증하지 않아 추론 연산량을 줄인다. 본 논문에서는 가속도 센서와 생리적 신호 센서를 기반으로

총 8가지의 행동을 인식하는 문제에 제안하는 방법을 적용하여 평균적으로 97.4%의 분류 정확률을 얻었다.

키워드 : 동적 베이지안 네트워크, 행동 인식, 멀티모달 센서

Abstract Recently, as the interest of ubiquitous computing has been increased there has been lots of research about recognizing human activities to provide services in this environment. Especially, in mobile environment, contrary to the conventional vision based recognition researches, lots of researches are sensor based recognition. In this paper we propose to recognize the user's activity with multi-modal sensors using hierarchical dynamic Bayesian networks. Dynamic Bayesian networks are trained by the OVR(One-Versus-Rest) strategy. The inferring part of this network uses less calculation cost by selecting the activity with the higher percentage of the result of a simpler Bayesian network. For the experiment, we used an accelerometer and a physiological sensor recognizing eight kinds of activities, and as a result of the experiment we gain 97.4% of accuracy recognizing the user's activity.

Key words : Dynamic Bayesian networks, activity recognition, multi-modal sensors

1. 서 론

최근 상황인지를 활용하고 있는 다양한 서비스들이 제공되면서, 사용자의 현재 상황을 인식하는 방법이 활발히 연구 되고 있다[1]. 온도 등 사용자 주변 환경, 사용자의 위치, 사용자의 기분, 사용자의 행동 등이 대표적인 컨텍스트이며, 이 중에서 사용자의 행동을 알아내는 것은 복잡하고 어려운 일이다. 사용자의 행동을 인식하는 연구로는 적외선 카메라로 찍은 비디오 영상을 기반으로 인식하는 연구[2]와 GPS를 기반으로 하는 연구들이 있는데[3], 비디오 영상을 기반으로 행동을 인식하는 경우에는 영상 처리 등과 같이 복잡한 전처리 과정이 있어 연산량이 많고, GPS를 기반으로 하는 연구는 위치와 이동 경로에 따른 행동 추론을 하기 때문에 상세한 신체 동작을 인식하는 것은 어렵다. 이에 따라 최근의 많은 연구들은 사용자의 행동을 직접적으로 표현해주는 센서를 기반으로 행동을 인식하고 있으며, 센서의 데이터를 수집 및 전처리를 일괄처리하고 있는 것에 초점이 맞춰져 있다. 사용자에게 서비스를 제공해주기 위해서는 시계열 분석을 통한 정확하고 적은 연산량이나 분산 처리 가능한 행동 인식 방법이 요구된다.

본 논문에서는 가속도 센서와 생리적 신호 센서들을 통하여 사용자의 행동 로그를 수집하고, 실시간 전처리, 그리고 계층적 동적 베이지안 네트워크를 이용하여 실시간으로 행동을 인식한다.

· 본 연구는 (주)삼성전자의 지원에 의해 수행되었습니다.
 · 이 논문은 2008 한국컴퓨터종합학술대회에서 '동적 베이지안 네트워크를 이용한 멀티모달센서기반 사용자 행동인식'의 제목으로 발표된 논문을 확장한 것이다

[†] 학생회원 : 연세대학교 컴퓨터과학과
 unikys@sclab.yonsei.ac.kr
 hjinh@sclab.yonsei.ac.kr

^{**} 종신회원 : 연세대학교 컴퓨터과학과 교수
 sbcho@cs.yonsei.ac.kr

논문접수 : 2008년 8월 28일
 심사완료 : 2008년 11월 16일

Copyright©2009 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

2. 배경

2.1 센서 기반의 행동 인식

최근 센서 자원이 발전하여 행동 인식 분야에서 카메라, 마이크, 가속도 센서와 생리적 신호 센서 그리고 GPS 등을 활용하는 연구들이 진행되고 있다. Ravi는 3축 가속도 센서를 활용하였으며[4], Ben-Arie는 카메라의 영상 처리로 사용자의 위치를 분석하여 행동을 인식하였다[5]. 이렇게 단일 센서가 아닌 여러 센서들을 활용하게 될 경우 단일 센서보다 사용자의 행동을 인식하는데 유리하여, 인간의 컨텍스트를 수집하기 위한, 멀티모달 센서 네트워크를 통해 행동을 인식하는 연구들이 있다. Meijer는 사용자의 가속도 센서와 물건의 모션센서를 통하여 물건과 사용자 행동간의 연관 관계를 통하여 행동을 인식하였고[6], Lieberman의 연구에서는 PDA와 GPS를 이용하여 사용자의 행동 및 컨텍스트를 추론했다[7].

이렇게 멀티모달 센서들을 활용하는 경우, 각기 다른 입력 방식과 입력 포맷을 가지고 있어서 센서 통합 시스템을 통한 통합이 필요하다. 따라서, 본 논문에서는 멀티모달 센서 네트워크를 통해서 실시간으로 센서 데이터를 수집 및 통합할 수 있으며, 실시간 전처리를 지원하며, 전처리 결과를 직접 모니터링할 수 있는 통합 시스템을 활용한다.

2.2 동적 베이지안 네트워크

행동을 인식하기 위해서는 다양한 센서로부터 수집되는 각종 데이터들을 바탕으로 사용자의 행동을 인지하는 모델이 필요로 하다. 베이지안 네트워크는 변수들 사이의 확률적 의존성을 나타내는 방향성 비순환 그래프로, 관측된 증거값을 바탕으로 목적 변수의 확률을 계산하는 모델이다. 동적 베이지안 네트워크는 시계열 정보 분석을 위해 그림 2와 같이 현재 변수의 확률을 계산할 때, 이전 시점의 정보를 함께 고려하는 베이지안 네트워크이다. 시계열 정보 분석을 위해 단순한 베이지안 네트워크보다 많은 수의 노드를 가지며 구조도 복잡하다.

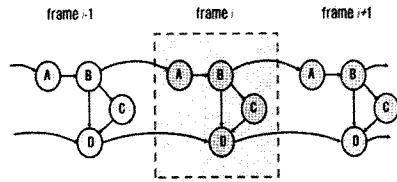


그림 2 동적 베이지안 네트워크의 구조

이렇게 시계열 분석에 효율적인 동적 베이지안을 활용하여 행동을 인식하는 연구들이 진행되어 왔다. Raj는 가속도 센서와 조도 센서 그리고 GPS를 기반으로 동적 베이지안 네트워크를 활용하였고[8], Tapia는 간단한 모션 센서를 기반으로 동적 베이지안 네트워크의 한 종류인 계층적 은닉 마르코프 모델로 사용자 행동을 인식하였다[9]. 이 연구들은 동적 베이지안 네트워크를 그대로 활용하고 있어서 실시간으로 서비스를 제공하기에 연산량이 많다. 따라서, 본 논문에서는 연산량을 줄이며, 분산 시스템에도 적용 가능한 계층적 동적 베이지안 네트워크를 제안한다.

3. 제안하는 방법

제안하는 방법의 전체 흐름은 그림 1과 같다. 먼저 사용자가 행동을 하면, 센서로부터 센서 데이터를 수집하게 된다. 센서 데이터는 각 센서별로 수집 모듈을 통해서 입력받고, 멀티모달의 센서를 동일한 숫자형으로 변환 후 통합한다. 통합한 센서 데이터는 전처리 단계와 컨텍스트 추론 단계에서 사용되며, 전처리 단계의 데이터는 다시 컨텍스트 추론을 하는데 활용된다. 이 두 단계의 결과는 행동 인식기에서 활용하게 되며, 그 후 나온 행동 인식기의 결과는 공유 데이터 메모리에 저장하여 다른 서비스에서 활용할 수 있도록 제공하고 있다.

3.1 실시간 데이터 수집과 전처리

멀티모달 센서의 데이터 수집은 각 센서별 수집 모듈을 통하여 입력이 되고, 센서 통합 시스템에서 데이터를 통합한다. 통합된 센서 데이터를 바로 사용하게 되면 잡

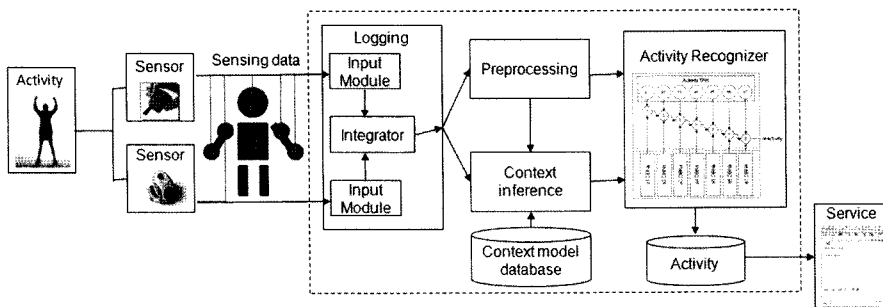


그림 1 제안하는 방법의 전체 흐름도

음이 많고 데이터가 연속적이기 때문에 잘못된 결과를 유도할 수 있으므로, 행동 인식에 바로 사용하기에 적합하지 않다. 따라서, 통합한 데이터의 잡음을 없애주고 연속적인 데이터를 이산화 시켜주는 전처리가 필요하다.

시간 t 에 해당하는 데이터의 평활화한 데이터를 구하는 함수 $SqrFlat(t)$ 는 데이터 x 가 시간 t 에서 $t-n$ 까지 소속되어있는 데이터 윈도우의 크기 n 에 대하여 식 (1)과 같은 방법으로 구한다. 해당하는 수식의 파라미터 값인 데이터 윈도우의 크기(n)이 커지게 되면 잡음 처리는 더 효율적으로 되지만, 센서의 동작에 대한 데이터의 민감도는 떨어진다. 따라서, 이 두 가지를 고려하여 파라미터를 설정해야한다. 본 논문에서는 윈도우 크기를 4로 설정하여 데이터 평활화를 하고 있다.

$$SqrFlat(t) = \frac{\sum_{i=t-n}^t x_i}{n} \quad (1)$$

데이터 이산화를 위한 방법으로는 K개의 군으로 데이터를 이산화 시켜주는 K-means 군집화 기법을 활용했다. K-means 군집화 기법은 특정 공간상에 특징을 표현할 때 유사한 특징은 유사한 위치에 있다는 것을 이용해서 서로 가까운 점들끼리 묶어서 K개의 군으로 나누는 분류 기법이다. 이는 각 군집에 속해 있는 데이터들의 분산을 최소화하여 식 (2)의 V값을 최소화하게 되는 기법이다.

$$V = \sum_{i=1}^k \sum_{j \in S_i} |x_j - \mu_i|^2 \quad (2)$$

3.2 동적 베이지안 네트워크를 통한 행동 인식

제한하는 방법은 다양한 행동 인식을 위한 확장성을 위해 다수의 확률 추론모델로 구성된 분산형 추론모델을 설계하여, 사용자의 행동을 인식한다. 분산형 추론모델은 그림 3과 같이 전역 추론모델과 지역 추론모델로 나누게 된다. 이 모델은 전역 추론모델을 통하여 행동을 선분류하고 지역 추론모델을 통해 후분류를 하게 된다. 전역 추론모델은 연산량이 적은 베이지안 네트워크로 추론하게 된다. 이는 저급의 센서 데이터를 논리적으로 해석한 고급 컨텍스트 정보를 활용하기 때문에, 그 결과를 빠르게 추론할 수 있다. 하지만 베이지안 네트워크를 통해서 행동을 인식하는 것은 한 순간의 센서 데이터에 의존적이기 때문에 연속적인 행동의 인식은 어렵다. 따라서, 지역 추론모델은 연속적인 행동의 인식에 용이한 동적 베이지안 네트워크를 활용하여 저급의 센서 데이터를 시계열 분석하게 된다.

베이지안 네트워크를 이용하는 이유는 변수들의 결합 확률 분포를 나타내는 것이 비용이 크기 때문이다. 즉 n 개의 변수가 있을 때 n 이 증가함에 따라 파라미터의 개수도 기하급수적으로 증가하기 때문에 정확하게 결합

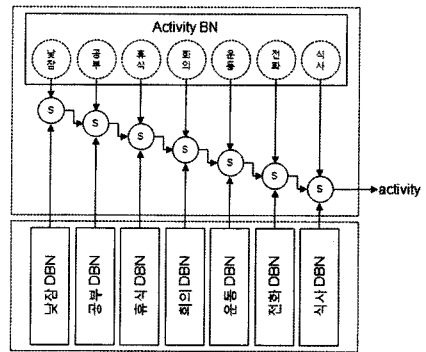
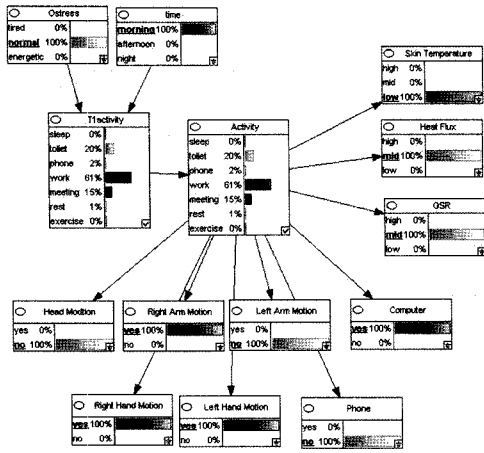


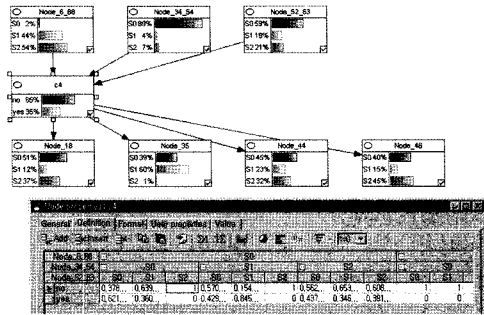
그림 3 행동 컨텍스트 통합추론 구조

확률 분포를 나타내는 것을 불가능하며 근사적인 방법으로 가급적 가까운 확률 분포를 이용한다. 베이지안 네트워크는 변수들 사이의 확률적인 의존관계를 이용하여 특정 변수들이 서로 확률적 의존이라는 가정을 한다. 이렇게 하면 고려되어지는 변수의 개수가 크게 줄어든다. 다양한 센서를 이용하면 증거 변수의 수가 많아지므로 그를 효율적으로 결합하고, 센서 데이터가 가지고 있는 불확실성을 처리하기에 적합한 추론 모델로 베이지안 네트워크를 이용한다.

행동을 추론하기 위한 전역 추론모델의 베이지안 네트워크는 그림 4(a)와 같이 이루어져 있으며, 각 부위별로 부착된 센서를 기반으로 특징 추출한 결과와 생리적 신호를 통하여 들어온 센서 데이터를 증거 값으로 사용한다. 각 행동은 동적 베이지안 네트워크를 OVR전략으로 학습하여 그림 3과 같이 각각의 지역 추론모델이 존재한다. 그림 4(b)는 낮잠에 대한 동적 베이지안 네트워크이며, 구조와 조건부 확률은 행동별로 학습된다. OVR전략으로 각 행동별로 추론모델을 설계하게 될 경우, 동시에 다수의 행동에 대하여 동일한 추론 결과가 나올 경우 한 가지 행동을 선택하는 방법이 필요하다. 지역 추론모델들의 결과를 선택하는 방법은 전역 추론모델을 통한 추론결과의 행동들을 확률순으로 정렬하여 우선순위를 주고, 우선순위가 높은 행동부터 해당하는 지역 추론모델로 검증한다. 특정 행동의 추론모델에 대한 검증 결과가 특정 임계치 이상이면, 그 행동이 행동 인식의 결과가 된다. 지역 추론모델의 검증 결과가 임계치 이하로 나오면, 다음으로 우선순위가 높은 행동의 지역 추론모델로 넘어가 검증을 하게 된다. 이러한 단계를 거쳐서 임계치를 넘는 지역 추론모델이 나오게 되면, 나머지 우선순위가 낮은 행동들의 지역 추론모델에 대해서 검증을 하지 않고 결과를 구할 수 있다. 이는 전체 OVR모델에서 전체 행동인식 모델에서 필요한 연산량을 줄여주고, 각 베이지안 네트워크는 분산 시스템 환경에서 분산이 가능하여 연산량의 균형 조절에도 유용하다.



(a) 행동 추론을 위한 베이지안 네트워크



(b) 낮잠의 OVR 동적 베이지안 네트워크

그림 4 행동 인식을 위한 베이지안 네트워크의 예

3.3 동적 베이지안 네트워크의 학습

본 논문에서는 동적 베이지안의 구조를 설계하는데, 유전자 알고리즘을 이용하였다. 기존의 개체 단위의 학습 기법과는 달리, 유전자 알고리즘은 개체집합을 단위로 학습시키기 때문에 지역해 문제 등에 대해 보다 강인하다. 유전자 알고리즘은 다음과 같이 진행된다.

먼저, 진화를 위한 집단을 초기화 한다. 집단은 다수의 개체로 구성되어 있으며, 각 개체는 하나의 동적 베이지안 네트워크를 나타낸다. 행동 인식을 위한 추론 모델이기 때문에, 사전에 관측 노드와 결과 노드를 설정하고, 링크는 관측 노드와 결과 노드 사이에만 존재하도록 제한한다. 그 후, 적합도 평가, 만족, 선택, 교차, 돌연변이의 진화 과정을 원하는 해를 획득할 때까지 반복한다.

진화를 통해 획득된 구조에 대한 현재 노드 간의 조건부 확률 테이블은 데이터를 기반으로 구하였고, 이전 증거 변수와의 조건부 확률 테이블은 EM 알고리즘을 이용했다. OVR(One-Versus-Rest) 전략으로 모델을 구축하면 데이터에서 하나의 클래스가 참, 나머지 클래스가

거짓이 되기 때문에 각 클래스의 학습 데이터 수가 동일하지 않다. 따라서 적합도 함수는 대상으로 하는 행동 T의 샘플 수와 나머지 행동의 샘플 수가 동일하도록 비율을 조절하고 식 (3)의 f와 같이 표현한다. 다음과 같은 적합도 함수를 사용함으로써 전체 테스트 데이터에 대한 실제 분류율을 판별하고 그 유용성을 검증할 수 있다.

R: 샘플 수		실제 데이터	
		행동 T	나머지
분류결과	행동 T	R ₁₁	R ₁₂
	나머지	R ₂₁	R ₂₂

$$f = \frac{R_{11} \times \frac{R_{12} + R_{22}}{R_{11} + R_{21}} + R_{22}}{2 \times (R_{12} + R_{22})} \quad (3)$$

4. 실험 및 결과

4.1 실험 환경

실험에서 인식하는 행동으로는 사무실 환경에서 일어날 수 있는 식사, 낮잠, 휴식, 전화, 공부, 화장실, 운동, 회의 등의 8가지 행동을 정의하여 데이터를 수집했다. 센서 데이터의 수집 주기는 20Hz로, 시퀀스 별로 약 300여개의 데이터가 들어가며, 각 행동별로 시퀀스 데이터 10개씩 수집하였다.

실험에서 사용한 멀티모달 센서들은 생리적 신호 센서인 Bodymedia의 암밴드 스트리밍 모듈과 3축 가속도, 각속도, 자력 센서 모듈이 5개 들어있는 XSens의 XBus-Master를 사용했다. 암밴드는 콘솔 입력을 통하여 입력되기 때문에 입력 파이프라인을 이용하여 입력을 받고, XSens는 제품과 함께 제공되는 SDK를 활용하여 수집 모듈을 만들었다. 암밴드 센서에서는 생리적 신호인 피부온도와 열유량 그리고 피부 전기 반응도를 수집하였고, Xsens의 가속도 센서는 각 머리, 오른팔, 왼팔, 오른 손목과 왼 손목에 착용하여 상반신의 동작에 대한 데이터를 수집하였다. 각 센서의 데이터 값은 3단계로 이산화하여 그 결과 값으로 부여된다.

실험은 5-폴드 교차 검증을 통하여, 로그 데이터를 5개의 폴드로 나누어 4개를 학습 데이터로 사용하고 나머지 세트에 대하여 테스트를 하는 방식으로 이루어졌다. 동적 베이지안 네트워크가 OVR 전략으로 학습되어 있으므로, 행동별로 실제 데이터와 추론 결과를 비교하여, 행동을 맞게 추론한 경우, 다른 행동을 행동이 다르다고 추론한 경우, 다른 행동을 추론하는 행동이 맞다고 추론한 경우와 맞는 행동을 행동이 다르다고 추론하는 4가지 경우에 대하여 결과를 얻게 된다.

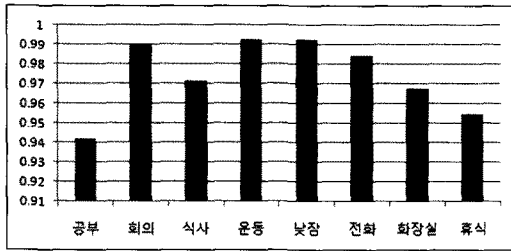


그림 5 행동별 분류 정확률

표 1 행동인식 혼동행렬(%)

%	공부	회의	식사	운동	낮잠	전화	화장실	휴식
공부	94.08	4.52	0	0	0	0	0	1.40
회의	0	99.94	0.06	0	0	0	0	0
식사	0.08	0.32	99.3	0.23	0	0	0	0.07
운동	0	2.88	0	96.84	0.29	0	0	0
낮잠	0.36	1.78	0	0	96.73	1.12	0	0
전화	0.16	0.15	0	0	0	97.72	1.97	0
화장실	0	0.98	0	0	0	0	98.78	0.24
휴식	0.67	0.81	0	0	0	0	0	98.52

4.2 결과분석

실험의 결과는 각 행동별로 인식 확률이 그림 5과 같이 나왔다. 전체 분류 확률이 97.4%로 나왔고, 행동별 혼동 행렬이 표 1과 같이 나왔다. 모든 행동이 대체로 94%이상의 높은 정확률을 보여주고 있다. 각 행동은 틀리게 추론한 경우는 다른 행동을 맞다고 인식한 경우와 행동을 인식하지 못하는 경우도 있었다. 행동을 인식하지 못하는 경우는 모든 지역 추론모델에서 해당하는 센서 데이터 입력 값에 대하여 만족하는 결과를 보이지 않은 경우이다. 해당하는 공부, 화장실, 휴식의 행동에서 맞는 행동에 대하여 인식하지 못하는 결과가 나왔고, 이 행동들은 분류율도 다른 행동들보다 낮게 나왔는데, 이들은 동일한 행동 패턴으로 나타내기 어렵고 반복적이지 않으며 매번 움직임이 달라지는 행동들이다.

다른 행동이 맞다고 인식하는 경우를 개선하기 위해서는 전역 추론모델의 정확률 개선을 통해서 알맞은 지역 추론모델을 선택할 수 있도록 해야 하고, 행동을 인식을 하지 못하는 경우는 개선하기 위해서는 지역 추론모델의 성능을 개선해야 한다.

5. 결론

본 논문에서는 멀티모달 센서를 기반으로 계층적 동적 베이지안 네트워크를 활용하여 실시간으로 행동을 인식하는 방법을 제안하였다. 실시간으로 행동을 인식하기 위하여 멀티모달 센서를 통합하는 프레임워크를 활용하여, 데이터 수집과 데이터 전처리 과정이 실시간으로 이루어졌고, 이 데이터를 연산량이 적고 분산 시스템에 적용 가능한 계층적 동적 베이지안 네트워크에서 변수 값으로 사용하여 행동을 인식하였다. 실험 결과로 동적 베이지안 네트워크를 통한 행동 인식의 결과가 97.4%로 행동 인식을 위한 베이지안 네트워크의 성능을 입증하였다.

향후에는 개인별 행동의 크기와 세기의 차이를 극복하고 자연스럽게나 모호한 행동을 인식하는 방법과, 실제로 분산 시스템에 적용할 경우 추론의 로드를 고려하는 방법에 대한 연구가 필요할 것이다.

참고 문헌

- [1] J. Kolari, T. Laakko, T. Hiltunen, V. Ikonen, M. Kulju, R. Suihkonen, S. Toivonen and T. Virtanen, "Context-aware services for mobile users," *VTT Information Technology*, 2004.
- [2] J. Han and B. Bhanu, "Human activity recognition in thermal infrared imagery," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol.3, pp. 17-25, 2005.
- [3] L. Liao, D. Fox and H. Kautz, "Hierarchical conditional random fields for GPS-based activity recognition," *Robotics Research*, Vol.28, No.3, pp. 487-506, 2007.
- [4] N. Ravi, N. Dandekar, P. Mysore and M. Littman, "Activity recognition from accelerometer," *Proc. Seventeenth Innovative Applications of Artificial Intelligence Conf.*, pp. 11-18, 2005.
- [5] J. Ben-Arie and Z. Wang, "Human activity recognition using multidimensional indexing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.24, No.8, pp. 1091-2002, 2002.
- [6] G. Meijer, K. Westerer, F. Verhoeven, H. Koper and F. Hoor, "Methods to assess physical activity with special reference to motion sensors and accelerometers," *IEEE Trans. Biomedical Engineering*, Vol.38, No.3, pp. 221-229, 1991.
- [7] H. Lieberman and T. Selker, "Out of context: Computer systems that adapt to, and learn from, context," *IBM Systems Journal*, Vol.39, No.3&4, pp. 617-632, 2000.
- [8] A. Raj, A. Subramanya, D. Fox and J. Bilmes, "Rao-blackwellized particle filters for recognizing Activities and spatial context from wearable sensors," *Experimental Robotics*, Vol.39, pp. 211-221, 2008.
- [9] E. Tapia, S. Intille and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," *Lecture Notes in Computer Science, Int. Conf. on Pervasive Computing*, Vol.3001, pp. 158-175, 2004.