

지능적인 홈을 위한 사용자 모션 트래킹 시스템에 관한 연구

최순용, 최종화, 신동규, 신동일
세종대학교 컴퓨터공학과

e-mail: artjian@gce.sejong.ac.kr, com97@sejong.ac.kr,
shindk@sejong.ac.kr, dshin@sejong.ac.kr

Research on Human Motion Tracking System for Intelligent Home

Soon Yong Choi, Jong Hwa Choi, Dongkyo Shin, Dongil Shin
*Dept of Computer Science, Sejong University

요 약

유비쿼터스 환경에서의 사용자의 위치인식 및 행동인식은 매우 중요하다. 인식을 하기 위해 카메라를 쓰는 것은 센서를 이용하는 것에 비하여 여러 가지 장점들이 존재한다. 본 논문은 여러 대의 네트워크 카메라를 이용한 실내에서의 사용자의 위치인식 및 행동인식을 위한 시스템을 제안한다. 사용자의 위치인식, 행동인식을 위하여 시스템에서는 영상처리기법들이 사용된다. 또한 행동인식에서는 추가적으로 SVM을 이용한 학습 및 예측 방법이 사용된다.

1. 서론

최근 들어, 카메라를 이용하여 여러 대상을 인식하는 연구에 대한 관심이 높아지고 있다. 이러한 관심은 유비쿼터스(Ubiquitous)[1] 환경에 관한 연구가 활발히 진행되고 있는 것과 무관하지 않다. 유비쿼터스 환경에서의 서비스는 단순히 사용자의 제어로 움직이는 단순한 것이 아니다. 유비쿼터스 환경이 스스로 사용자의 상태 및 주변의 상황을 파악하여 어떠한 서비스를 해주어야 하는지를 결정하는 능동적인 것이다. 능동적인 서비스를 제공하기 위해서는 사용자의 대략적인 위치와 어떠한 상태인지 혹은 어떠한 행동을 취하고 있는지를 알아야 한다. 인식을 하기 위해서는 사용자의 몸에 센서를 부착하는 방법, 건물에 센서를 설치하는 방법, 카메라를 이용하는 방법 등 다양한 방법들이 존재한다. 그러나 센서를 이용한 방법은 사용자가 거부감을 느낄 수도 있고, 센서를 설치해야 하는 등 다소 부담이 있다. 반면에 카메라를 이용하면 이를 해소할 수 있고 보다 자연스러운 서비스를 제공할 수 있다는 장점이 있다.

본 연구는 여러 대의 네트워크 카메라를 이용한

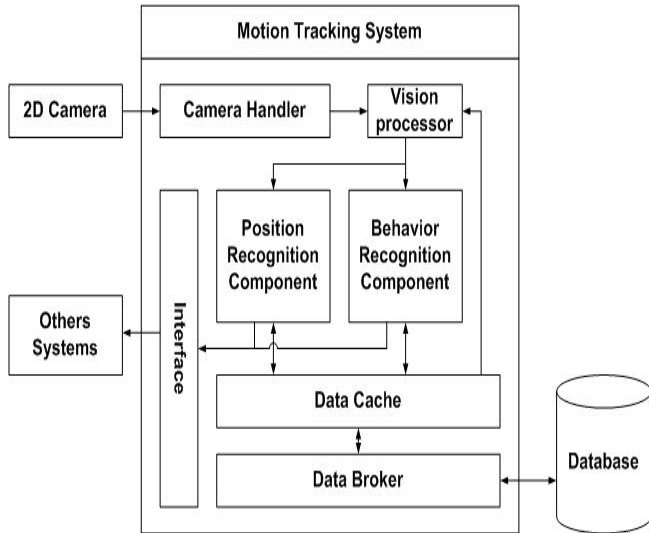
실내에서의 사용자의 위치인식 및 행동인식을 하는 시스템에 관한 연구로서 필요한 몇 가지 사항들이 있다. 본 연구에서의 사용자의 위치인식 및 행동인식을 위해서는 입력영상과 배경영상의 차이를 이용하는 방법[2][3], 영상의 에러를 줄이는 방법, 영상에서 에지(Edge)를 추출하는 방법 등의 여러 가지 영상처리기법을 이용해야 한다. 또한 사용자의 행동인식을 하기 위해서는 영상 데이터를 입력으로 하고 인공지능을 이용한 학습을 사용하여야 한다. 본 연구에서는 사용자의 행동인식을 위해서 인공지능 기법 중에 하나인 SVM(Support Vector Machine)[4]을 채택하고 있다.

3장에서는 본 논문에서 제시하는 사용자의 위치 및 행동을 인식하기 위한 시스템의 구조를 보여준다. 4장에서는 인식에 사용하는 SVM 알고리즘에 대해서 간단하게 살펴볼 것이다. 5장에서는 사용자의 위치인식과 행동인식을 위한 알고리즘에 대해서 설명한다. 마지막으로 6장에서는 결론과 앞으로의 연구의 진행방향에 대해서 기술한다.

3. 모션 트래킹 시스템

3.1 전체 구조

(그림 1)은 사용자의 위치 및 행동을 인식하기 위한 시스템의 전체 구성을 보여준다. 이번 장에서는 에이전트의 전체 구성에 관하여 설명한다.



(그림 1) 시스템의 구성

사용자의 위치 및 행동의 인식은 2개의 컴포넌트에서 담당한다. 각 컴포넌트에서는 오직 한 가지의 인식만을 한다. 만약 영상을 이용하여 할 수 있는 또 다른 인식분야가 있다면 새로운 컴포넌트로 만들어 추가해주면 된다.

에이전트는 네 대의 카메라로부터 얻어진 영상을 바탕으로 사람의 위치를 인식한다. 하나의 영상으로 실내에서의 사용자의 위치를 인식하기란 사실상 불가능하다. 따라서 본 논문에서는 실내 네 곳의 구석에 각각 한 대의 카메라를 설치하여 총 네 대의 카메라를 설치하는 방법을 사용한다. 한편, 행동인식에서는 사람이 존재한다고 판단된 카메라의 영상만을 사용한다.

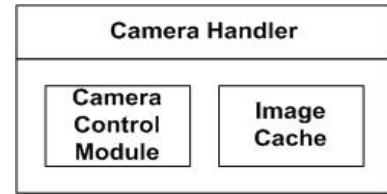
3.2 구성 컴포넌트

에이전트를 구성하는 주요 구성요소는 카메라 핸들러(Camera Handler), 영상처리기(Vision Processor), 위치인식 컴포넌트(Position Recognition Component), 행동인식 컴포넌트(Behavior Recognition Component), 인터페이스(Interface) 등이다.

3.2.1 Camera Handler

카메라 핸들러는 카메라 조작 및 영상을 얻어내는 역할을 하는 컴포넌트이다. 카메라 핸들러는 Camera Control Module, Image Cache로 구성된다. Camera Control Module은 카메라의 각도조절, 카메

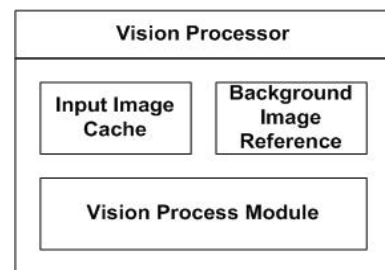
라로부터의 영상획득 등을 담당한다. Image Cache는 카메라에서 획득한 영상을 임시 보관하는 메모리이다.



(그림 2) 카메라 핸들러의 구조

3.2.2 Vision Processor

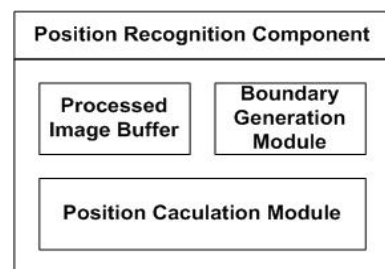
영상처리기(Vision Processor)는 카메라 핸들러로부터 제공되는 영상을 조작하는 기능을 한다. 영상처리기는 (그림 3)과 같이 구성된다. 배경영상(Background Image Reference)은 입력영상(Input Image Cache)에서 사용자만을 추출하기 위해서 필요하다. Vision Process Module은 영상의 압축 및 압축해제, 에지검출, 두 영상간의 차이(XOR)를 구하는 등 영상에 관한 처리를 전담한다.



(그림 3) 영상처리기의 구조

3.2.4 Position Recognition Component

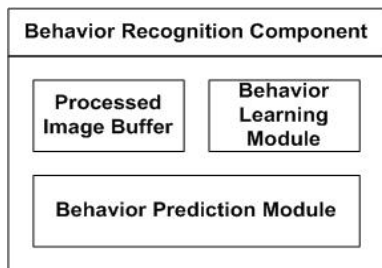
위치인식 컴포넌트(Position Recognition Component)에서는 실내에서의 사용자의 위치를 인식하는 역할을 한다. 사용자 외곽선 추출모듈(Boundary Generation Module)은 영상처리기에서 처리한 영상으로부터 사용자의 외곽범위를 계산하는 일을 한다. 위치계산모듈(Position Calculation Module)은 계산된 외곽정보와 전체 영상을 가지고 실내에서의 사용자의 위치를 판단한다.



(그림 4) 위치인식 컴포넌트의 구조

3.2.5 Behavior Recognition Component

행동인식 컴포넌트(Behavior Recognition Component)는 행동학습모듈(Behavior Learning Module)과 행동예측모듈(Behavior Prediction Module)의 두 부분으로 구성이 된다. 우선 각 부분을 설명하기에 앞서 행동에 대해 설명한다. 행동은 서있는 모습, 앉아있는 모습, 누워있는 모습, 팔을 들고 있는 모습 등 여러 가지 패턴으로 구분을 한다. 행동학습부분에서는 구분하길 원하는 행동의 표본 영상과 행동의 라벨(예: 서있는 모습)을 학습 데이터로 하여 SVM알고리즘을 이용하여 학습을 한다. 한편, 행동예측부분에서는 학습된 SVM 분류기를 이용하여 실시간으로 입력되는 영상에서 사용자가 어떠한 행동을 취하고 있는지 예측한다.



(그림 5) 행동인식 컴포넌트의 구조

3.2.5 Interface

인터페이스는 외부 시스템과의 통신을 위해서 존재한다. 시스템에서는 실내에서의 사람의 위치 및 행동을 인식한 결과를 외부로 알려줘야 할 필요가 있는데, 인터페이스가 그 창구역할을 하게 된다.

4. SVM 알고리즘

SVM은 생물정보학(Bioinformatics), 텍스트 분류, 얼굴 이미지 인식 등의 패턴인식 분야에 이용되어 왔다. 기본적인 SVM은 변형된 입력 공간에 놓여지는 최대극한평면(Maximum Margine Hyperplane)을 생성한다. 이진 분류 SVM의 경우 두 집단 사이의 공간을 최대한 멀리 떨어지도록 분류함으로써 가장 근접하는 이질 집단 데이터들 간의 거리가 최대가 되는 최대극한평면을 중심으로 두 집단을 나누게 된다.

데이터가 선형으로 분류될 수 있는 경우 SVM은 오류없이 평면과 가장 가까운 데이터 지점들 간의 거리를 최대로 하는 최적의 평면을 찾기 위해 Linear Machine을 학습시킨다. 서로 다른 집단을 분류하는 평면과 가장 가까운 거리에 있는 훈련용 데이터를 Support Vector라고 하며, 그 이외의 데이터들은 집단을 분류하는 경계를 결정짓는데 아무런 영

향을 끼치지 않는다. 선형으로 분류되지 않는 대부분의 현실 문제의 경우, SVM은 훈련용 데이터 셋의 오류 수를 최소화시키는 평면을 찾아내기 위해 Nonlinear Machine을 사용한다. 이러한 비선형 문제의 경우 커널함수를 사용함으로써 분류를 가능하게 한다.

학습용 데이터를 $[x_i, y_i]$ 라고 할 때, 입력벡터를 $x_i \in \mathbb{R}^n$, 결과값을 이진 분류의 경우 $y_i \in \{-1, 1\}$, 다수의 집단으로 분류될 경우를 $y_i \in \{1, 2, \dots, m\}, i=1, \dots, n$ 이라고 정의한다. 선형 분류가 가능한 경우에는 두 집단을 분류하는 최적평면이 Support Vector에 대해 정의하는 결정 규칙은 다음의 등식과 같다.

$$Y = \text{sign}\left(\sum_{i=1}^n y_i \alpha_i (x \cdot x_i) + b\right) \quad (1)$$

비선형 분류의 경우에는 등식(2)와 같이 정의된다.

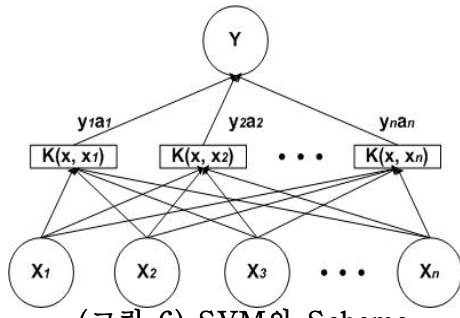
$$Y = \text{sign}\left(\sum_{i=1}^n y_i \alpha_i K(x, x_i) + b\right) \quad (2)$$

$K(x, x_i)$ 는 커널(Kernel)함수로서, 입력공간에서 다양한 형태로 입력벡터들을 분리하기 위해서 쓰인다. 공통적으로 많이 사용되는 SVM의 커널함수로는 다항(Polynomial), RBF(Radial Basis Function), 이중신경망(Two-layer Neural Network)함수가 있다. 이 중 RBF가 현실 문제에 가장 많이 사용되며 함수식은 등식(3)과 같다. 이때, λ^2 는 RBF함수의 대역폭이 된다.

$$K(x, x_i) = \exp(-(x-x_i)^2/\lambda^2) \quad (3)$$

SVM 분류 방법은 Support Vector와 매개변수 b 와 α_i 를 찾아내기 위해 선형으로 제한된 QP(Quadratic Programming)을 풀어나가면서 구현된다. 완전한 분류가 가능한 경우에는 등식(1)에서 계수 α_i 의 하위경계가 0이 된다. 그렇지 않을 경우에는 SVM은 하위 경계뿐만 아니라 계수 α_i 의 상위 경계인 C값을 줌으로써 일반화될 수 있다[5].

SVM의 결정함수를 만들어내기 위한 학습과정은 인공신경망과 유사한 두 개의 층 구조로 나타낼 수 있다. 하지만 학습방법에 있어서 SVM은 통계적 학습이론에 기반하여 잘못된 분류를 최소화하는 최적화 이론을 적용하여 학습한다. 첫 번째 층에서는 상응하는 특징공간에서 최적의 평면을 만들어내게 된다. (그림 6)은 SVM의 구조를 나타낸다.



(그림 6) SVM의 Scheme

5. 인식 알고리즘

본 연구에서의 위치인식은 카메라가 비추는 영역이 겹치지 않게 조절되어 있다. 따라서 입력영상과 배경영상과의 차이에서 사용자의 에지정보가 검출된다면 이를 사용자의 위치로 판단한다. 즉, 네 영역으로 위치를 판단한다. 한편, 사용자의 행동인식은 에지정보에 라벨을 붙여 학습을 한다. 학습이 완료되면 입력 영상에서 에지정보를 추출하여 그 정보를 가지고 사용자가 어떠한 행동을 취하고 있는지 예측하게 된다. (그림 7)은 사용자의 행동인식을 위한 알고리즘을 나타낸다.



(그림 7) SVM 알고리즘

6. 결론

본 연구에서는 실내에서의 사용자의 위치와 행동을 인식하기 위한 시스템을 제안하였다. 제안한 시스템은 네트워크 카메라를 이용함으로써 센서를 이용하는 방법에 비하여 사용자에게 비교적 거부감을 덜 주게 되고, 위치인식 및 행동인식을 동시에 함으로 효율을 극대화한다. 또한 각각의 인식 분야를 컴포넌트로 제작하는 방식으로 설계함으로써, 카메라

영상을 이용하여 할 수 있는 또 다른 인식 분야가 있다면 컴포넌트로 제작하여 포함할 수 있게 하였다.

앞으로의 연구에서는 사용자의 위치인식을 함에 있어서 영역을 더 세분화할 수 있는 알고리즘을 연구할 것이다. 또한 사용자의 행동인식에서는 더 다양한 알고리즘들을 사용하여 최적의 알고리즘을 찾아낼 것이다.

참고문헌

[1] Yih-Farn Robin Chen and Charles Petrie, "Ubiquitous Mobile Computing," IEEE INTERNET COMPUTING, Mar. 2003, pp.16-17
 [2] C. Wrem, A. Azarbajejani, T.Drrell, A. pentland, "Pfinder: Real-Time Traking of the Human Body," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.19, no.7, July 1997.
 [3] I.Haritaoglu, D.Harwood, "Who? When? Where? What? A Real Time System for Detecting and Tracking People," International Conference on Face and Gesture Recognition, Arpil, 1998
 [4] K.M. Lin, C.J. Lin, "A study on reduced support vector machines", Neural Networks, IEEE Transaction on, Vol.14, Issue:6, pp.1449-1459, 2003
 [5] Y.S. Kwon, I.G. Han, K.C. Lee, "Ordinal pair-wise partitioning OPP approach to neural networks train bond rating", Intelligent Systems in Accounting, Finance and Management, Vol.6, 1997, pp.23-40.