

유비쿼터스 스마트 홈을 위한 위치와 모션인식 기반의 실시간 휴먼 트랙커

박세영⁰ 신동규 신동일 Nguyen Quoc Cuong
세종대학교 컴퓨터공학과

spark@gce.sejong.ac.kr, shindk@sejong.ac.kr, dshin@sejong.ac.kr, cuongnqc@gce.sejong.ac.kr

Real-Time Human Tracker Based Location and Motion Recognition for the Ubiquitous Smart Home

Seyoung Park⁰ Dongkyoo Shin Dongil Shin Nguyen Quoc Cuong
Dept. of Computer Engineering, Sejong University

요 약

유비쿼터스 스마트 홈(ubiquitous smart home)은 인간과 홈의 컨텍스트(context) 정보를 이용하여 인간에게 자동적인 홈 서비스(Home service)를 제공해줄 수 있는 미래의 환경이다. 인간의 위치와 모션은 유비쿼터스 스마트 홈에서 굉장히 중요한 컨텍스트이다. 본 논문은 유비쿼터스 스마트 홈에서 인간의 위치와 모션을 예측할 수 있는 실시간 휴먼 트랙커(tracker)를 연구하였다. 실시간 휴먼 트랙커를 위해 우리는 4개의 네트워크 카메라를 사용하였다. 본 논문에서는 실시간 휴먼 트랙커의 구조를 설명하고, 인간의 위치와 모션을 자동적으로 예측 및 판단하는 알고리즘을 제안하였다. 인간 위치를 위해서 3개의 배경이미지를 이용하였다(이미지1: 빈 방, 이미지2: 가구 및 가전, 이미지3: 이미지 2와 거주자를 포함). 실시간 휴먼 트랙커는 3개의 이미지를 비교하여 각 이미지로부터 추출되는 특징 값을 결정하고, 이들 특징 값을 SVM(Support Vector Machine)을 이용하여 각각의 모션을 예측하였다. 3개의 배경 이미지를 이용한 인간 위치 인식실험은 평균 0.037 초가 소요 되었다. SVM을 이용한 모션 인식 요소에서, 우리는 각 동작에 대하여 1000번씩 측정했고, 모든 모션의 정확도 평균은 86.5%의 정확도를 보였다.¹⁾

Abstract

The ubiquitous smart home is the home of the future that takes advantage of context information from the human and the home environment and provides an automatic home service for the human. Human location and motion are the most important contexts in the ubiquitous smart home. We present a real-time human tracker that predicts human location and motion for the ubiquitous smart home. We used four network cameras for real-time human tracking. This paper explains the real-time human tracker's architecture, and presents an algorithm with the details of two functions (prediction of human location and motion) in the real-time human tracker. The human location uses three kinds of background images (IMAGE1: empty room image, IMAGE2:image with furniture and home appliances in the home, IMAGE3: image with IMAGE2 and the human). The real-time human tracker decides whether the human is included with which furniture (or home appliance) through an analysis of three images, and predicts human motion using a support vector machine. A performance experiment of the human's location, which uses three images, took an average of 0.037 seconds. The SVM's feature of human's motion recognition is decided from pixel number by array line of the moving object. We evaluated each motion 1000 times. The average accuracy of all the motions was found to be 86.5%.

1. 서 론

유비쿼터스 스마트 홈은 인간과 홈의 컨텍스트를 비교하여 자동 홈 서비스를 제공한다 [1]. 사람의 행동 패턴을 분석하기 위해서 '누가', '언제', '어디서', '무엇을', '어떻게', '왜'의 정보들이 필요하다 [10]. 이러한 정보들을 위해서 인간의 위치와 동작은 유비쿼터스 스마트 홈에서 아주 중요한 컨텍스트이다. 예를 들어, TV 프로그램을

보기 위해서 쇼파에 앉을 때 유비쿼터스 스마트 홈은 인간의 선호하는 채널을 조사하여 자동적으로 홈 서비스를 제공한다. 유비쿼터스 스마트 홈 서비스는 인간과 홈의 환경으로부터 많은 컨텍스트 정보를 얻게 되지만 이보다 더 중요한 정보는 인간의 위치와 동작에 관한 것이다 [2]. 위치인식에 관한 연구는 어떤 센서장비를 사용하였는가에 따라서 시스템을 분류할 수 있다. 크게 카메라의 사용여부로 나누어 볼 수 있다. 카메라를 사용하지 않고 압력센서, 적외선 등을 사용하는 방법과 카메라를 사용하는 방법이 있다. 현재의 위치 인식 연구에는 카메라를 이용한 방법이 많이 사용되고 있다. 하지만, 카메라를 이

1) 본 논문은 서울시 지원 연구과제 (과제번호: 11098) 에 의하여 지원 받은 연구결과임

용하여 위치 인식을 할 때에는 여러 가지 문제점이 발생할 수 있다. 첫 째, 빛의 세기의 변화에 따라 이미지가 서로 다르게 인식될 수 있다. 둘째, 물체의 이동과 사람의 이동을 서로 구분해야 한다. 셋 째, 가구 및 가전용품의 재배치 및 이사를 통하여 배경이미지가 변할 수 있다. 이러한 문제점을 해결하기 위해서 간단한 조작으로 원활한 위치인식을 할 수 있는 방법이 필요하다.

인간의 위치를 추적하고 동작을 예측하기 위해서는 사물 또는 사물과 겹쳐져 있는 사람을 구분해 내는 일은 중요한 일이다. 이를 구분해 내기 위해서 본 논문에서는 실시간으로 3개의 이미지(이미지1: 빈 방, 이미지2: 가구 및 가전, 이미지3: 이미지 2와 거주자를 포함)를 비교하고 인간의 위치를 결정하였다. 인간의 동작을 예측하기 위해서 SVM을 이용하였고, 분류를 위한 요소들은 실시간 휴먼 트래커에 의해 포함된 사람의 이미지 픽셀 분류를 이용해서 사람의 위치를 추적하고 사람의 동작을 예측할 수 있는 실시간 트래커를 본 논문에서는 제안하였다.

2. 관련 연구

스마트 스페이스(Smart space)를 위한 휴먼 트래커는 많은 방법으로 연구되었다. Pfinder 는 사람을 추적하기 위한 실시간 시스템이다. 이는 배경으로부터 사람을 분류해내기 위해서 색과 모양의 다중 클래스 통계 모델을 하였다 [3]. 이는 시각 범위 아래의 사람의 머리와 손을 찾고 추적한다. Tominaga 와 Hongo는 Percept-Room에서 캡처 된 다중 채널 동작 이미지로부터 사람의 움직임과 손 동작을 찾기 위한 방법을 제시하였다 [4]. 인간의 위치와 손을 드는 동작들은 배경이미지와 프레임의 제거를 통하여 다중 카메라의 윤곽들의 통합에 의해서 추출된다. KidRooms 은 이미 알고 있는 제한된 공간에서의 트래킹 시스템이다 [5]. Guohui Li 와 Jun Zhang은 어둠과 소음에 의한 방해와 조명의 변화에 대한 다중 기술의 통합과 배경 템플릿에 근거한 비디오 스트림으로부터 동적 객체를 발견하기 위한 효과적인 접근법을 제안하였다 [6]. Microsoft의 EasyLiving에서는 거실에서 다수의 사람들의 트래킹을 위해서 2대의 컬러 스테레오 카메라를 사용하였다 [7]. 스테레오 이미지들은 사람들의 위치를 위해 사용되고 컬러 이미지들은 사용자 간의 식별을 위해서 사용 되었다. 하지만, EasyLiving에서는 물체의 이동을 고려하지 않고, 소파의 앉았는지의 여부를 위해서 압력센서를 한다는 한계를 보여준다. MIT의 House_n에서는 천장에 카메라를 설치하여 작은 방에서 다수의 사용자 위치를 추적하는 방법을 제시하였다 [11]. 카메라를 천장에 설치하였기 때문에 넓은 지역에서 더 많은 카메라를 사용해야 하고 사용자와 다른 물체를 구분하기가 어렵다는 단점이 있다. Percept Room[12] 프로젝트에서는 벽의 상층부에 8대 벽의 중간 높이에 8대 총 16대의 카메라를 설치하였다. 각 카메라에 비춰진 사용자가 존재할 수 있는 영역들의 교집합을 구해서 사용자의 위치를 알아내는 방법을 사용하였다 [13]. 하지만 실내의 가구 및 가전용품과 사람의 위치와의 관계를 고려하지 않아서 상황에 맞는 서비스를 하기 힘든 점이 있다. 그림 1은 Percept Room의 카메라 배치

도를 보여준다. 이와 같은 기존의 연구들과 달리 본 논문에서 제안하는 실시간 휴먼 트래커는 4대의 카메라를 이용하여 가구 또는 가전 등과 함께 있는 인간을 구별해내고 동작을 예측 할 수 있다. 그림 2는 실시간 휴먼 트래커의 카메라 배치도를 보여준다. 인간의 실시간 이미지를 얻기 위해서 4개의 디지털 네트워크 카메라를 사용하였고 실험에서의 카메라 시스템은 10 base-T LAN을 사용하였다. 1개의 서버 카메라(32bit RISC CPU)와 3개의 클라이언트 카메라로 구성된다. 4대 카메라의 시각은 위, 아래, 오른쪽, 왼쪽 90도로 한정되도록 설치 하였다.

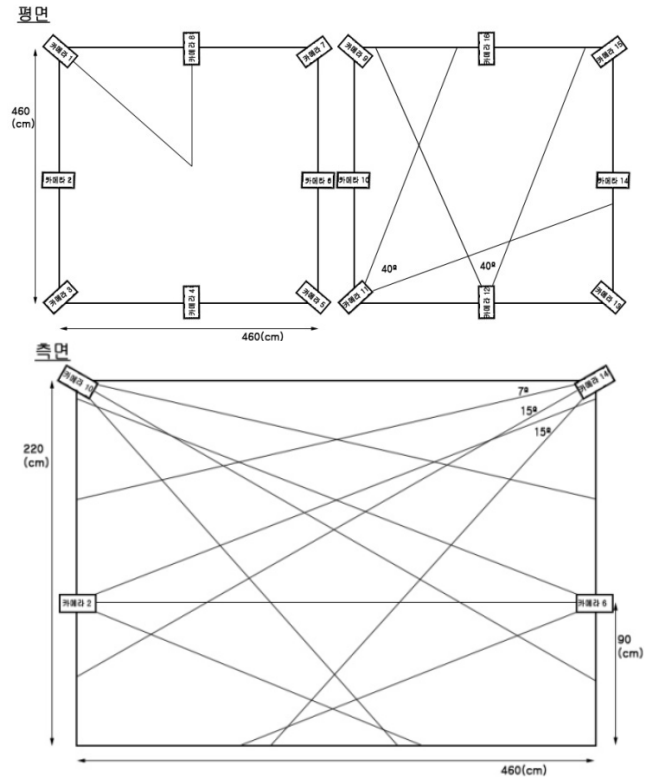


그림 1. Percept room 구조도

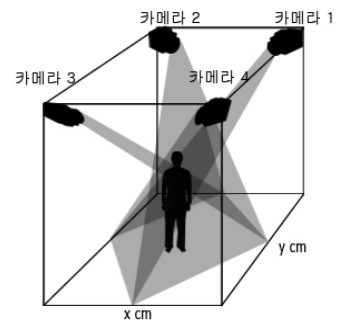


그림2. 실시간 휴먼 트래커 카메라 배치도

3. 실시간 휴먼 트래커

본 논문에서 제안된 실시간 휴먼 트래커는 유비쿼터스 스마트 홈을 위한 사용자의 위치와 동작 예측을 수행한다. 그림 3은 실시간 휴먼 트래커 구조를 보여준다. 인간의 위치를 결정하기 위해서, 카메라 핸들러는 2초마다 디지털 네트워크 카메라로부터 컬러이미지(720x486)를 가지고 온다. 또한 카메라의 동작여부를 체크해서 네트워크 전송오류의 검출 및 복구, 이미지의 해상도 및 선

명도의 조정 등 카메라에 관련된 모든 정보를 관리한다. 동적 객체 탐지기는 움직이는 객체의 공간을 배경이미지와 획득한 이미지 사이의 차이를 이용해서 이동하는 물체의 영역을 검출하여 위치 인식기로 검출한 영역과 함께 배경 처리된 전체이미지를 전송하고, 객체 분류기는 검출된 영역의 이미지만을 전송한다. 객체 분류기는 포착된 이동하는 객체가 사람인지 아닌지의 판별을 담당한다. 그리고 위치인식기로 동적 객체로 검출된 영역이 사람인지 아닌지 판별한 정보를 전송한다. 위치 인식기는 유비쿼터스 스마트 홈에서 가구와 가전용품 정보를 이용해서 사람의 위치를 추정한다. 4대의 카메라로부터 전송된 이미지를 바탕으로 사람으로 구분된 영역의 절대 위치 좌표를 계산하고, 정적 객체 매니저에서 설정된 주요 가구 및 가전의 위치 정보를 바탕으로 사용자의 권역 정보를 계산한다. 정적 객체 매니저는 권역정보를 계산하기 위해서 필요한 가구 및 가전의 위치를 설정하기 위한 도구를 제공하며 설정된 가구 및 가전의 위치 데이터를 관리하고 갱신하는 역할을 담당한다.

휴먼 트래커는 인간이미지의 픽셀 좌표를 얻고 가구와 가전 등의 픽셀 좌표와 비교 한다. 이렇게 함으로써 인간의 위치를 분석한다.

3.1.2 사람의 동작인식

사람의 동작인식은 위치인식과 달리 객체 분류기를 이용하여 분류한다. 사람의 여러 가지 동작을 예측하기 위해서는 미리 학습 시켜놓은 알고리즘에 각 이미지의 특징 값을 넣음으로써 동작을 예측해야 한다. 본 논문에서는 사람의 동작 인식을 위해서 교사학습 방법 중에 하나인 SVM(Support Vector Machine)을 사용하였다 [9]. SVM중에서 실험에서 사용하는 세부 알고리즘은 선형(linear) SVM이다. 동적 객체 탐지에서 검출한 전체 이미지 중 움직이는 부분인 부분 이미지와 카메라 핸들러에서 전송한 이미지와 배경이미지와의 차이정보인 흑백 이미지를 통하여 선형 SVM을 학습 시켰다. 실제로 인간의 동작을 예측하기 위해서 선형 SVM에 입력되는 요소는 흑백 부분 이미지에서 추출한 에지(Edge)정보이다. 이 에지 정보를 가지고 인간의 동작을 분류하고 예측 하게 된다. 본 논문에서는 인간의 동작 중 4개 동작(누워있는 동작, 앉아있는 동작, 서있는 동작과 걷는 동작)을 학습 시키고 인식하였다. 이러한 각 동작을 분별하기 위해서 4개의 선형 SVM을 각각 따로 구성하고, 이미지에서 추출한 요소들을 가지고 각 동작을 판단하였다. 각 선형 SVM에서 판단한 결과를 바탕으로 입력된 이미지가 사람인지 아닌지를 구별한다. 그림5는 객체 분류기의 구조이다.

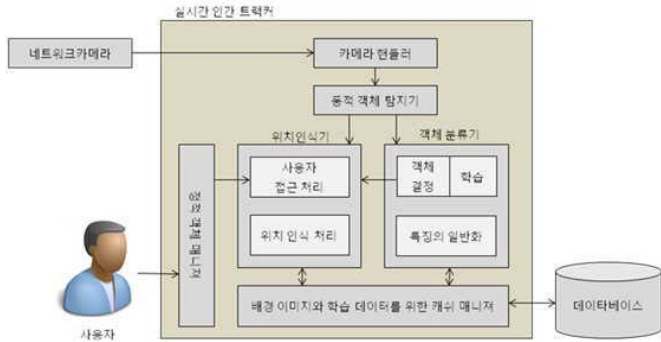


그림 3. 실시간 휴먼 트래커의 구조

3.1. 실시간 휴먼 트래커의 내부 알고리즘

실시간 휴먼 트래커는 인간의 실시간 이미지를 얻기 위해서 4개의 디지털 네트워크 카메라를 사용한다. 실험에서의 카메라 시스템은 10 base-T LAN을 사용하였고 1개의 서버 카메라(32bit RISC CPU)와 3개의 클라이언트 카메라로 구성된다. 이미지에 사람과 함께 가구 또는 가전이 포함되어 있는지를 결정하기 위해서 입력된 이미지에서 미리 저장되어있는 이미지를 제거하는 방법을 사용하였고, 이를 위해 3개의 이미지(이미지1: 빈 방, 이미지2: 가구 및 가전, 이미지3: 이미지 2와 거주자를 포함)를 사용했다. 그림 4는 사용된 이미지의 예제이다. 실시간 휴먼 트래커는 이미지1과 이미지2사이의 이미지가 다른 부분을 찾아내고 이를 통해 유비쿼터스 스마트 홈에서의 가구 및 가전의 위치를 알아낼 수 있다. 그리고 알아낸 가구의 위치 좌표는 실루엣 메소드를 사용하여 결정한다 [8].

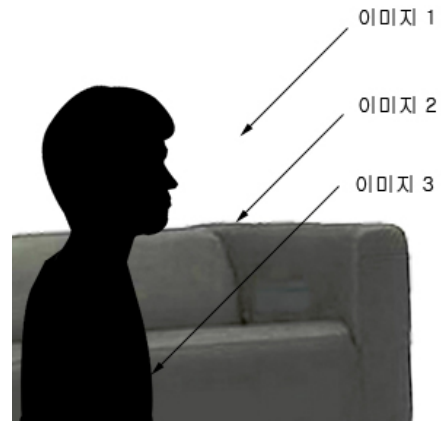


그림 4. 인간의 구분을 위한 이미지 예제

3.1.1. 사람의 위치인식

사람의 위치 인식은 위치인식기를 이용하여 결정한다. 이미지1과 이미지2의 차이를 통하여, 가구와 가전용품들의 위치를 얻고, 위치 정보를 통해 픽셀 좌표(x, y)를 얻어서 가구와 가전의 절대좌표를 구한다. 인간이 집안으로 들어온다면, 이미지 2와 이미지 3의 차이를 통해서



그림 5. 객체 분류기의 구조

4. 실험과 평가

실시간 휴먼 트래커는 인간의 위치와 동작을 결정한다. 만약에 인간의 위치가 유비쿼터스 스마트 홈에 배치된 가구(가전) 등과 함께 포함되어있다면, 실시간 휴먼 트래커는 학습된 모션 판단에 의해서 사람의 동작을 인식한다. 그림 6은 사람 동작 인식을 위해 선택될 수 있는 요소들과 오브젝트 분류기에 의해서 분석된 인식의 결과를 보여주고 표 1은 사람의 위치와 동작을 인식하는 실험의 성능 결과를 보여준다.

표 1과 같이 3개의 이미지를 이용한 인간의 위치 실험 성능은 평균 0.037 초가 걸렸다. 인간의 모션 인식은 픽셀 SVM의 요소처럼 동적 객체의 어레이 라인에 의해 픽셀 번호로부터 결정된다. 우리는 각 모션을 1000번씩 측정했다. 모든 모션의 정확도 평균은 86.5%의 정확도를 보였다.

	1	2	3	4	5	6	7	8	9	10
눅다 1	0.013	0.019	0.021	0.032	0.037	0.040	0.044	0.053	0.056	0.066
눅다 2	0.006	0.009	0.010	0.012	0.013	0.014	0.016	0.017	0.020	0.019
눅다 3	0.008	0.014	0.016	0.016	0.018	0.021	0.016	0.023	0.030	0.031
앉다 1	0.001	0.001	0.001	0.004	0.002	0.005	0.003	0.004	0.008	0.004
앉다 2	0.013	0.01	0.016	0.017	0.027	0.035	0.038	0.051	0.041	0.065
앉다 3	0.003	0.019	0.036	0.045	0.047	0.049	0.051	0.046	0.048	0.038
서있다 1	0.022	0.021	0.014	0.018	0.021	0.019	0.025	0.023	0.031	0.026
서있다 2	0.043	0.049	0.063	0.045	0.046	0.042	0.034	0.032	0.035	0.033
서있다 3	0.028	0.026	0.031	0.032	0.037	0.037	0.030	0.040	0.041	0.038
걸다 1	0.032	0.045	0.061	0.058	0.067	0.068	0.063	0.055	0.050	0.045
걸다 2	0.012	0.018	0.014	0.017	0.025	0.018	0.024	0.025	0.022	0.020
걸다 3	0.042	0.046	0.048	0.059	0.059	0.086	0.093	0.097	0.088	0.089

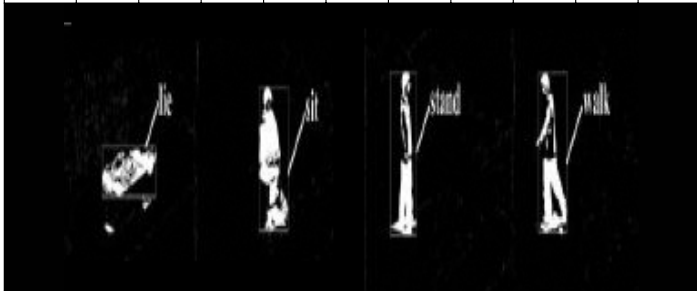


그림 6. SVM을 위한 요소 맵

표 1. 인간의 위치와 동작 인식에 대한 성능 실험 결과

구분	객체 위치	객체 동작			전체 정확도		
	소요시간 (초)	SV의 수	LV의 평균	KE의 수	전체	성공	정확성
눅다	0.03699	171	1.75761	14204	1000	930	93.0%
앉다	0.03719	125	2.02598	13085	1000	911	91.9%

5. 결 론

본 논문은 유비쿼터스 스마트 홈을 위해서 인간의 위치와 동작을 예측할 수 있는 실시간 휴먼 트래커를 제안하였다. 이를 위해 4개의 네트워크 카메라를 사용하였다. 본 논문은 실시간 인간 위치 트래커의 구조를 설명하고, 실시간 휴먼 트래커에서 2가지 기능(인간의 위치인식과 동작을 인식)을 가지는 알고리즘을 제안하였다. 실시간 휴먼 트래커는 인간이 가구 또는 가전용품등과 함께 포함되어 있는지 아니든지 3개의 이미지의 분석을 통해서 판단을 하고 SVM을 이용해서 인간의 동작을 예측하였다. 인간의 위치의 성능 실험은 3개의 이미지를 사용했고 평균 0.037초의 시간이 소요되었다. 인간의 동작 인식의 SVM의 요소는 동적 객체의 어레이 라인을 통해 결정된다. 우리는 각 모션을 1000번씩 측정했다. 모든 동작의 평균정확도는 86.5%이다. 향후에는 실시간 휴먼 트래커에서 어두울 때와 밝기가 바뀔 때의 알고리즘에 대해서 연구할 것이다.

6. 참고문헌

- [1] Das. S.K, Cook. D.J, "Guest Editorial - Smart Homes. Wireless Communications", IEEE. vol 9, Issue 6, (2002) 62 - 62
- [2] Jonghwa Choi, Dongkyoo Shin and Dongil Shin, "Research and implementation of the contextaware middleware for controlling home appliances", Consumer Electronics, IEEE Transactions on. vol 51, Issue 1, (2005) 301 - 306
- [3] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder : Real-time Tracking of the human Body". Pattern Analysis and Machine Intelligence, IEEE Transactions on. vol 19. Issue 7. (1997) 780 - 785
- [4] Tominaga M, Hongo H, Koshimizu H, Niwa Y, Yamamoto K, " Estimation of human motion from multiple cameras for gesture recognition" Pattern Recognition. (2002) 401-404
- [6] S.Intille, J.Davis, A.Bobick, "Real-time recognition of activity using Temporal Templates", In Proc. Third IEEE Workshop on Application of Computer Vision. (1996) 1233-1251
- [6] Guohui Li, Jun Zhang, Hongwen Lin, Tu D, Maojun Zhang, " A moving object detection approach using integrated background template for smart video sensor", Instrumentation and Measurement Technology Conference, IMTC 04. Proceedings of the 21st IEEE. vol 1.(2004) 462 - 466
- [7] Krumm J, Harris S, Meyers B, Brumitt B, Hale M, Shafer S, " Multi-camera multi-person tracking for EasyLiving" Visual Surveillance. Proceedings, Third IEEE International Workshop on. (2000) 3-10
- [8] T. Matsuyama, "Cooperative Distributed Vision: Research Achievements and Future Directions", The 7th Symposium on Sensing via Image Information (SSII2001). (2001) 187-198
- [9] C.J.C.Burges, "A tutorial on support vector machines for patten recognition", Data Mining Knowl. Disc. (1998) 1-47
- [10] S.Jang, W.Woo, "A unified context-aware application model", Lecture Notes in Computer

- Science, Volume 2680/2003, August 2003, pp178-189
- [11] R. Khalaf and S. S. Intille, "Improving multiple people tracking using temporal consistency", Massachusetts Institute of Technology, Cambridge, MA, MIT Dept, of Architecture House_n Project Technical Report, 2001.
- [12] H. Hongo, H. Watanabe, M. Yasumoto, Y. Niwa, K. Yamamoto, "Eye Regions Extraction for Omni-directional Gaze Detection Using Multiple Cameras", Proc, IASTED Conference on Signal Processing, Pattern Recognition and Applications(SPPRA2001), Jul2001, pp.241-246
- [13] M. Tomminage, H. Hong, H. Koshimizu, Y. Niwa, K. Yamamoto, "Estimation of human motion from multiple cameras for gesture recognition", Pattern Recognition, 2002. Proceedings. 16th International Conference on Volume 1, 11-14 Aug. 2002, pp401-404 vol.1