

사람의 움직임 추적에 근거한 다중 카메라의 시공간 위상 학습

(Learning Spatio-Temporal Topology of a Multiple Cameras Network by Tracking Human Movement)

남 윤 영 † 류 정 훈 †† 최 유 주 ††† 조 위 덕 ††††
(Yunyoung Nam) (Junghun Ryu) (Yoo-Joo Choi) (We-Duke Cho)

요 약 본 논문은 유비쿼터스 스마트 공간에서 중첩 FOV와 비중첩 FOV에 대한 카메라 네트워크의 시공간 위상을 표현하는 새로운 방법을 제안한다. 제안된 방법을 이용하여 다중 카메라들간의 움직이는 객체들을 인식 및 추적하였으며 이를 통해 카메라 네트워크의 위상을 결정하였다. 다중 카메라의 영상으로부터 여러 객체들을 추적하기 위해 여러 가지 방법들을 사용하였다. 우선, 단일 카메라에서 객체들의 겹침 문제를 해결하기 위해서 병합-분리(Merge-Split) 방법을 사용하였으며, 보다 정확한 객체 특성을 추출하기 위해 그리드 기반의 부분 추출 방법을 사용하였다. 또한, 비중첩 FOV를 포함하는 다중 카메라의 보이지 않는 지역에 대한 객체 추적을 위해 등장과 퇴장 영역간의 전이시간과 사람들의 외형 정보를 고려하였다. 본 논문에서는 다양한 등장과 퇴장 영역간의 전이시간을 추정하고 전이확률을 이용하여 무방향 가중치 그래프로써 카메라 위상을 가시적으로 표현하였다.

키워드 : 유비쿼터스, 지능공간, 영상감시, 다중 카메라, 사람 추적, 위상, 추론

Abstract This paper presents a novel approach for representing the spatio-temporal topology of the camera network with overlapping and non-overlapping fields of view (FOVs) in Ubiquitous Smart Space (USS). The topology is determined by tracking moving objects and establishing object correspondence across multiple cameras. To track people successfully in multiple camera views, we used the Merge-Split (MS) approach for object occlusion in a single camera and the grid-based approach for extracting the accurate object feature. In addition, we considered the appearance of people and the transition time between entry and exit zones for tracking objects across blind regions of multiple cameras with non-overlapping FOVs. The main contribution of this paper is to estimate transition times between various entry and exit zones, and to graphically represent the camera topology as an undirected weighted graph using the transition probabilities.

Key words : ubiquitous, smart space, surveillance, multiple cameras, people tracking, topology, inference

· 본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스 컴퓨팅 및 네트워크원천기술개발사업의 지원에 의한 것임

† 학생회원 : 아주대학교 유비쿼터스시스템연구소 선임연구원
youngman@ajou.ac.kr

†† 학생회원 : 아주대학교 전자공학과
ryujunghun@gmail.com

††† 정 회 원 : 서울벤처정보대학원대학교 컴퓨터공학과 교수
yjchoi@suv.ac.kr

†††† 종신회원 : 아주대학교 유비쿼터스시스템연구소 전자공학부 교수
chowd@ajou.ac.kr

논문접수 : 2007년 9월 6일

심사완료 : 2007년 12월 6일

: 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.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제13권 제7호(2007.12)
Copyright©2007 한국정보과학회

1. 서론

카메라와 네트워크 기술이 진보되고, 감시 시스템의 설치비용이 두드러지게 저렴해짐으로써, 감시 카메라의 사용이 급증하게 되었다. 전형적인 영상 감시 시스템은 은행, 카지노, 공항, 군사시설과 같은 보안을 필요로 하는 넓은 지역에서 사용된다. 이러한 영상 감시 시스템은 실시간 영상 데이터 저장 및 특정 이벤트를 모니터링하는데 사용될 수 있다. 예를 들어, 교통 모니터링 시스템은 교통량 측정과 혼잡이나 사고를 감지하여 즉각적인 지원을 수행하였으며, 공공지역에 대한 감시 시스템은 주거지 또는 주차장과 같은 지역에서의 범죄들을 줄이는 역할을 하였다.

일반적으로 영상 감시 시스템은 넓은 지역에 분포된 카메라의 광대한 네트워크를 포함한다. 이러한 카메라 네트워크에서, 카메라들간의 관계를 알아내는 것은 지능적인 감시 시스템을 구축하기 위한 중요한 이슈중의 하나이다. 본 논문에서는 중첩 FOV(overlapping fields of view) 뿐만 아니라 비중첩 FOV(non-overlapping fields of view)를 포함하는 카메라 네트워크의 시공간 위상을 표현하기 위한 방법을 제안한다. 시공간 위상을 결정하기 위해서 중첩 FOV에서의 동일한 객체 인식, 겹쳐진 객체 인식, 비중첩 FOV에서의 다중 객체 추적을 고려하였다.

우선, 객체 인식을 위해서 카메라로부터 다양한 거리에 있는 움직이는 사람의 머리와 손 영역을 추출하기 위해 그리드 기반의 추출 방법[1]을 사용하였다. 이 방법은 다음과 같은 단계로 이루어진다. 첫째로 배경영상 제거(background subtraction) 기법을 색상과 채도 정보 기반의 이미지 시퀀스에 적용하여 전경영상과 배경영상을 분리하며, 배경 영상이 제거된 이미지는 그리드 패치로 나누어진다. 다음 단계에서는 그리드 패치의 특성 값 히스토그램 분석에 기반하여 배경영상, 피부가 아닌 전경영상, 그리고 피부 전경 영상 클래스로 나누어 추출한다.

객체의 겹침 문제는 단일 카메라에서 처리하는 것은 매우 어려울 수 있으며, 다중 객체 추적에서 다른 객체로 인해 가려지는 문제로 추적하는 객체를 잃어버리는 경우가 있다. 이러한 가려짐은 객체들의 영상에서 전체적으로 발생하거나 부분적으로 발생할 수 있으며, 다른 전경 객체나 배경 객체에 의해서 발생할 수 있다. 본 논문에서는 등장, 퇴장, 병합, 분리, 연속과 같은 이벤트들과 블랍의 속성을 이용하는 병합-분리(Merge-Split)방법을 사용하였다.

멀티 카메라를 이용한 감시 시스템에서 카메라의 FOV가 중첩되는 경우가 많지 않다. 즉, 넓은 영역을 모니터

링 해야 하기 때문에 항상 중첩된 카메라 뷰를 포함하는 것은 현실 상황에서는 어려운 경우가 많다. 이러한 한계점을 극복하기 위해서 카메라 핸드오프(handoff) 방법을 이용하였으며, 이 방법을 통해 다중 카메라의 FOV로부터 객체들을 추적할 수 있도록 하였다. 카메라 핸드오프에서 중심적인 부분은 비중첩 FOV에서 다중 카메라 트래킹을 통해 보이지 않는 지역에서 객체를 추적하고 목표물을 일치시키는 데 있다. 본 논문에서는 이러한 카메라 핸드오프처리를 위해 카메라간에 이동하는 사람의 외형정보와 등장과 퇴장 영역간에 전이시간을 고려하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 소개하고 3장에서는 객체인식과 식별방법을 서술한다. 4장에서는 카메라 네트워크 위상을 학습하는 방법에 대해서 설명하며, 5장에서는 실험결과를 보이고 6장에서 결론을 맺는다.

2. 관련연구

다중 카메라에 대한 과거의 연구들은 움직이는 객체를 검출(detection), 인지(recognition), 추적(tracking)하는 방법을 다루고 있다. 이러한 연구 중에 추적은 인식과 인지가 전처리 되어 가능하며, 추적이 단일 카메라에서 다중 카메라로 확장되려면 카메라 보정(camera calibration)을 거쳐야 한다. 최근, 많은 연구자들은 다중 뷰 포인트로부터 객체를 인지하고 추적을 하기 위해 카메라 네트워크 보정을 제안하였다. 카메라 영상으로부터 객체를 인식하기 위하여 색상 정보는 매칭 단계에서 자주 사용된다. Black[2]은 색상 특성은 보존하면서 조도 변화에 대해 대처할 수 있도록 HSI 색상 공간의 비균등 양자화(non-uniform quantization)를 사용하였다. KaewTraKulPong과 Bowden[3]은 Sturges[4]에 의해 제안된 CCCM(Consensus-Colour Conversion of Munsell colour space)을 사용하였다. 이것은 색상 카메라를 보정할 필요 없이 일치된 색상 표현을 제공하였다.

다중 카메라들간의 객체 추적에 관한 연구는 일반적으로 공통적인 영상 특징이나 움직이는 객체들을 추적을 다루고 있다. 중첩된 카메라의 경우, 추적 알고리즘[5,6]은 카메라 보정과 카메라들간의 추적된 객체들의 핸드오프의 계산이 필요하다. 이를 위해서는 많은 공통적인 FOV를 공유하는 것이 필요하다. 그러나 중첩된 카메라들의 환경을 위해서는 많은 카메라들이 필요하고 이들의 위치를 물리적으로 연결하는 것은 상당히 비현실적이다. 따라서, 이러한 보이지 않는 비중첩 지역을 처리할 수 있어야 한다. 대부분의 단일 카메라 추적 알고리즘은 칼만(Kalman) 필터[7]와 같은 기법을 사용하여 이전의 관찰된 속도와 부드러운 움직임으로 이후의

위치를 예측한다.

그러나 카메라간의 움직임은 부드럽지 못하기 때문에, 다중 카메라를 이용한 여러 가지 기법들은 보이지 않는 지역을 처리하고 객체들의 핸드오프를 개선할 수 있도록 개발되어 왔다. Haritaoglu[8]는 사람의 위치뿐만 아니라 머리, 손, 발, 등, 몸통 등 신체 부위와 외형모델을 이용하여 추적 시스템을 개발하였다. 또한 Haritaoglu [9]은 스테레오 정보를 시스템에 적용하였다. Kettner와 Zabih[10]는 중첩되지 않는 지역에서의 다중 카메라들간의 객체를 추적하도록 베이지안을 이용한 방법을 제안하였다. 이 방법을 통해 위치, 객체 속도, 이동시간과 함께 카메라간의 객체의 움직임을 연계하여 이용하였다. 여기에서 베이지안식은 다중 카메라간의 객체들의 이동경로를 재구성하는 데 이용되었으며, 움직임이 가능한 경로의 위상과 전이 확률에 대한 입력을 필요로 한다. Huang과 Russell[11]은 고속도로에서의 자동차들의 추적을 위해 확률적인 방법을 이용하였으며, 비중첩 카메라에서 자동차들의 전이 시간과 외형 매칭을 종합적으로 사용하였다. 자동차들의 외형은 색상의 평균값과 가우시안 분포로써 모델링한 전이 시간으로 계산하였다. Cai와 Aggarwal[12]은 카메라간의 전환 방법을 사용하여 단일 카메라 추적 시스템을 확장시켰다. 중첩된 FOV에서의 보정된 카메라를 사용하였으며, 객체간의 일치는 기하학적이고 외형적 특징을 매칭함으로써 이루어졌다.

Javed[13]는 Parzen 윈도우를 이용하여 객체들의 카메라 위상과 이동경로 확률을 학습함으로써 더욱더 일반적인 시스템을 제공하였다. 개별적인 트랙은 시공간 그리고 색상 외형의 이후 확률을 최대화함으로써 일치시키고 카메라간의 변화를 고려하여 개선시켰다. 전이 확률은 많은 레이블된 궤적들을 이용하여 지도 학습(supervised learning) 되었다. Dick[14]은 시야각 간에 관찰된 사람들의 움직임 패턴을 기술하기 위해 확률적인 전이 행렬을 사용하였으며, 트레이닝 단계(training phase)가 필요하다. Ellis[15]는 일치나 트레이닝 단계는 필요하지 않지만 오랜 기간의 움직임을 관찰하고 등장과 퇴장에 대한 정보를 축적하는 것이 필요하다. 입장과 퇴장간에 이동시간에 대한 시간적 분포도에서 최고점을 찾기 위해서 임계값 기법을 이용하였으며, 최고점을 기반으로 카메라가 연결되어 있다고 가정한다. 이후에 이 방법은 통계를 기반으로 추출된 전이정보를 이용하여, Stauffer[16]와 Tieu[17]에 의해서 확장되었다. 중첩과 비중첩 FOV에서 다중 카메라들간에 추적하여 전체 카메라 구성을 위한 일치 모델을 구축하였으며, 2차원 평면으로 가정하고 카메라들간의 일치 관계를 찾아내었다. 최근 Gilbert[18]은 점층 학습을 이용하여 카메라의 주 등장과 퇴장 영역을 확률에 기반하여 자동적으로 찾기

위한 방법을 보였다.

본 논문에서는 다중 카메라가 설치된 환경에서 움직이는 사람들을 추적하기 위해 중첩 또는 비중첩 카메라에서 움직이는 객체들에 대한 등장-퇴장과 출현-전이 시간 모델을 이용하여, 카메라간의 관계를 자율적으로 학습하고 다중 카메라 네트워크의 시공간 위상을 자동적으로 구성되도록 하였다.

3. 객체 인식

3.1 배경 제거(Background Subtraction)

배경 제거는 움직이는 사람을 인식하고 추적하기 위해서 중요하고 필요한 단계이다. 배경 제거를 위해 색상 이미지 처리와 분석에 널리 사용되고 있는 색상(Hue), 채도(Saturation), 명도(Intensity)의 칼라 모델(HSI colour model)을 사용할 수 있다[2-4]. 본 논문에서는 HSI 칼라 모델을 기반으로 노이즈 특징을 분석하였으며, 노이즈를 줄이기 위해 적응적인 공간 필터링 기법과 HSI 칼라 공간에서의 노이즈 변화의 비일률성(non-uniformity)을 개발하였다.

색상은 일반적으로 조도 변화를 최소화하기 위해 색상의 자체 유일 값으로 표현되며, 채도는 색상의 순도를 나타낸다. 명도는 색상의 밝은 정도이며, 장시간 동안에 여러 영상처리를 하는 배경 제거에는 어려움이 있다.

본 논문에서는 색상과 채도에 대하여 정적인 배경 이미지의 모든 픽셀을 가우시안 분포로 모델링 하였으며, 서로 다른 조명에서의 50개 프레임들에 대한 HSI 색상 공간에서의 참조 배경 영상을 통계적으로 분석한다.

배경 이미지를 분석한 후에, 이미지들의 시퀀스를 HSI 색상 이미지로 변환한다. 만약 차이값이 배경 이미지의 변수 값을 기반한 유도된 임계 값보다 크다면, 이러한 픽셀들은 전경 픽셀로 결정된다. 즉, 아래의 식에 기반하여 픽셀을 전경 또는 배경 클래스로 분류시킨다.

$$R(P_{ij}) \begin{cases} 1(\text{foreground}), & \text{if } |H_{ij} - H_{b_j}| > \omega_1 \sigma(H_{b_j}) \\ & \text{and } |S_{ij} - S_{b_j}| > \omega_2 \sigma(S_{b_j}), \\ 0(\text{background}), & \text{otherwise} \end{cases} \quad (1)$$

H_{ij} : Hue, S_{ij} : saturation, P_{ij} : current pixel

H_{b_j} : average Hue, S_{b_j} : average Saturation

$\sigma(H_{b_j})$: Hue variance, $\sigma(S_{b_j})$: Saturation variance

3.2 그리드 기반의 ROI 추출

객체 인식을 개선하기 위해서, 본 논문에서는 카메라로부터 떨어진 다양한 거리에서의 움직이는 사람의 머리와 손 영역을 강건하게 추출하기 위한 그리드 기반의 추출 방법[1]을 사용하였다. 제안된 방법은 원래의 입력 이미지보다 저해상도에서의 전경과 배경 정보를 계속적으로 유지하는 그리드 이미지를 정의하여 사용한다. 그

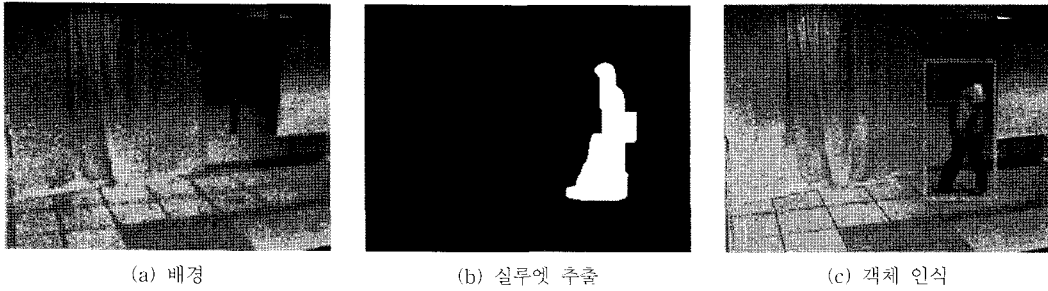


그림 1 객체 인식의 예

리드 이미지는 전체 입력 이미지 픽셀들에 관하여 전경과 피부 픽셀 부분의 분석에 기반하여 적응적으로 배경, 피부가 아닌 전경영상, 피부전경영상으로 구분한다. 즉, 저해상도에서의 그리드 이미지를 이용하여 피부영역을 효과적으로 추출하고 노이즈를 줄였다.

배경 제거 후 결과 이미지를 8*8 픽셀의 그리드 패치(patch)로 분할시키고 각 그리드의 전경 픽셀과 피부 픽셀의 개수를 계산하였다. 그리드(F_{ij})에서의 64 픽셀에 대한 전경 픽셀 개수의 비율과 그리드(S_{ij})에서의 전경 픽셀에 대한 피부 픽셀 개수의 비율이 패치의 특징 값으로 사용된다.

$$F_{ij} = \frac{\# \text{ of foreground pixels in Patch}(i, j)}{64}$$

$$S_{ij} = \frac{\# \text{ of skin pixels in Patch}(i, j)}{\# \text{ of foreground pixels in Patch}(i, j)} \quad (2)$$

그리드 패치는 두 가지의 패치 특징 값 F_{ij} 와 S_{ij} 를 기반으로 배경 클래스, 피부가 아닌 전경 클래스, 피부 전경 클래스로 구분하고 이러한 세 가지의 클래스로 구성된 그리드 이미지를 구성한다.

$$G_{ij} = \begin{cases} \text{foreground patch, if } F_{ij} \geq X_{fg} \text{ and } S_{ij} < X_{sk} \\ \text{skin patch, if } F_{ij} \geq X_{fg} \text{ and } S_{ij} \geq X_{sk} \\ \text{background patch, otherwise} \end{cases} \quad (3)$$

$$X_{fg} = \arg \min(f_{fg}(x)), \quad f_{fg}(x) > (1 - P_{fg})$$

$$X_{sk} = \arg \min(f_{sk}(x)), \quad f_{sk}(x) > (1 - P_{sk})$$

where

$$f_{fg}(x) = \sum_{i=0}^x H_{fg}(i), \quad f_{sk}(x) = \sum_{i=0}^x H_{sk}(i)$$

$$P_{fg} = \frac{\text{width}(ROI) \times \text{height}(ROI)}{\text{width}(I) \times \text{height}(I)}$$

$$P_{sk} = \frac{\# \text{ of skin pixels}}{\text{width}(ROI) \times \text{height}(ROI)}$$

실험에서, 고정된 임계값을 이용하여 정적인 패치 분류와 패치 히스토그램 분석기반으로 제안된 적응적인 패치 분류를 수행하였다. 그림 2에서 보는 바와 같이, 두 가지의 방법에 대한 ROI 추출 결과를 비교하였다. 정적인 패치 분류의 경우, 먼 거리의 이미지에서 작은 피부 영역에 대한 피부 ROI 인식이 실패하였으며, 제안

	Using static threshold	Using adaptive threshold
1		
2		
3		

그림 2 648*468 이미지 크기에서 전경 영상 추출한 결과 비교

된 적응적인 패치 분류에서는 성공적으로 추출하였다.

3.3 겹쳐진 물체 인지(Occluded Object Recognition)

여러 가지 상황에서 객체를 추적하기 위해서는 영상에서 객체와 구조물 또는 객체간의 겹침을 고려해야 한다. 객체간의 겹침 문제를 해결하기 위해 두 가지 방법이 제안되었다[20]. 한 가지 방법은 Merge-Split(병합 분리) 방법은 겹침이 발생한 경우 새로운 블랍(blob)을 생성하는 것이다. 블랍은 하나 혹은 그 이상의 객체들을 뜻하며 내부에 있는 객체들의 특성들을 포함하고 있으며, 이러한 블랍을 기준으로 겹쳐진 객체 인식이 이루어진다. 다른 방법은 Straight-Through 방법이며, 객체의 외형을 모델링하여 겹침 현상이 발생했을 때 객체를 병합하지 않고 기존의 외형을 유지하며 계속적으로 인지 및 식별하는 방법이다. 이때 색상 정보를 이용하여 겹쳐졌을 경우에도 객체간의 공간적 관계를 파악할 수 있다.

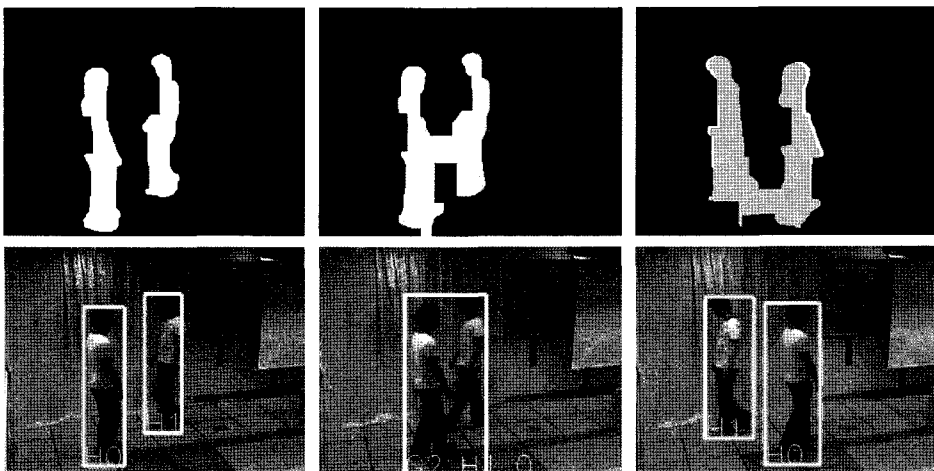
본 논문에서는 겹쳐지는 객체를 인지하기 위해서 등장, 퇴장, 연속, 병합, 분리와 같은 5 가지 이벤트를 이용하여 처리하였다. 등장과 퇴장은 새로운 블랍이 FOV에 나타나고 사라질 때이다. 병합은 전 프레임의 여러 개의 블랍이 현재 프레임에서 합쳐졌을 경우이다. 이 때 병합되기 이전의 블랍의 속성들이 병합된 블랍의 속성에 상속된다. 병합을 판단하는 방법은 현재 프레임의 블랍 수가 이전 프레임의 블랍 수보다 줄어들고 이전 프레임에서 여러 개의 블랍들과 현재 프레임에서 하나의 블랍이 차지하는 공간 영역이 공통적으로 포함되는 경우이다. 분리는 이전 프레임에서 하나의 블랍이 현재 프레임에서 여러 개의 블랍들로 분리된 경우이다. 분리를 판단하는 방법은 현재 프레임의 블랍 수가 이전 프레임의 블랍 수보다 늘어나고 이전 프레임에서 하나의 블랍

과 현재 프레임에서 여러 개의 블랍들이 차지하는 공간 영역이 공통적으로 포함되는 경우이다. 분리된 블랍들은 병합되기 이전의 블랍들과 매칭되어야 한다. 이때 블랍이 가지고 있는 속성 중의 하나인 블랍들의 일반화된 색상 히스토그램을 이용하여 이전 블랍과의 일치성을 파악한다. 연속은 등장, 퇴장, 병합, 분리 이벤트를 처리한 후, 기존 블랍들의 일치성을 파악하는 과정이다. 이 과정에서 이전 프레임과 현재 프레임 블랍의 위치에 대한 유클리드 거리뿐만 아니라 블랍의 이동 방향, 크기, 색상도 고려한다. 이를 통해, 시간에 따른 동일한 객체의 이동을 파악할 수 있다.

그림 3은 겹침이 일어날 때 객체를 인지 및 추적한 결과이다. 프레임 662에서는 두 명의 사람이 H0과 H1으로 인식되었다. 프레임 669에서는 두 명의 사람이 겹치는 장면으로 H0과 H1이 병합되어 G2라는 새로운 블랍이 생성되었다. 프레임 705에서는 블랍 G2인 두 명의 사람이 분리되는데 병합되기 이전의 식별자인 H0과 H1으로 인식되어 계속적으로 객체 인지 및 추적이 되었다.

4. 카메라 네트워크 위상의 학습 기법

임의의 카메라 네트워크 위상을 학습하기 위해서 카메라간의 시공간 관계를 고려하고, 이것은 카메라 네트워크간에 객체들의 이동경로를 예측하는데 이용될 수 있다[13,15]. 과거의 연구는 단순히 객체의 등장과 퇴장의 시간을 기록하여 이동경로를 예측하였으나, 동일한 객체가 동일한 카메라에서 재 출현하거나 등장과 퇴장의 지역이 동일한 경우에 다르게 처리해줘야 한다. 또한, 출현과 이동시간을 표현할 수 있는 모델을 정립하고 이 모델을 기반으로 시공간 관계를 구조화하여 가시화



(a) Frame 662

(b) Frame 669

(c) Frame 705

그림 3 단일 카메라에서의 겹치는 다중 객체를 처리하여 추적하는 화면

해야 한다. 본 논문에서는, 공간 관계와 시간 관계를 위해 객체들의 등장-퇴장(entry-exit)과 출현-전이시간(travel-transition) 모델을 사용하였으며 이러한 모델을 기반으로 카메라의 위치를 노드로 표시하는 가시적인 그래프로 표현하였다. 링크는 연결된 카메라 위치들간의 이동 가능한 경로라고 표현하고 이동 시간과 객체가 첫 번째 나타난 위치에서 두 번째 카메라에서 보일 확률을 이용하여 표현하였다.

그림 4와 그림 5는 객체가 등장하고 퇴장할 때 처리하는 프로세스 순서도를 보이고 있다. 우선, 등장-퇴장 모델은 객체가 나타나고 퇴장하는 등장 영역과 퇴장 영역을 정의한다. 등장 영역은 객체 식별에 따라 새로운 등장 영역일 수도 있고 재등장 영역일 수도 있다. 새로운 등장 영역은 다중 카메라에서 객체가 처음으로 인식 되었을 때 결정된다. 이에 반해서, 재등장 영역은 한 카메라에서 나타난 객체가 사라진 이후에 또 다른 카메라에서 재등장할 때 결정된다. 객체가 새로운 등장 영역에서 퇴장 영역으로 이동할 때, 이러한 이동 시간을 출현 시간이라고 하였다. 또한 객체가 퇴장 영역에서 재등장 영역으로 이동할 때, 전이 시간이라고 하였다. 따라서 출현 시간은 FOV의 크기로 나타낼 수 있으며, 전이 시간은 카메라들간의 거리를 나타낼 수 있다.

본 논문에서는 카메라 네트워크 위상을 구축하기 위해

중첩 노드(overlapping node), 비중첩 노드(non-overlapping node), 가상 노드(virtual node)를 정의하였다. 중첩 FOV에서, 객체가 단일 카메라에 출현하고 사라지기 전 그 객체가 다른 카메라에 다시 나타날 수 있다. 이런 경우, 중첩 노드는 중첩 에지(overlapping edge)를 표현하기 위해 필요하다. 비중첩 FOV에서 비중첩 노드는 객체가 출현하는지 퇴장하는지에 따라 등장 노드 또는 퇴장 노드일 수 있다. 특히, 객체가 한 카메라에서 사라진 후, 같은 카메라에서 재등장할 수도 있다. 이런 경우, 비가시 에지(invisible edge)를 표현하기 위해 가상 노드(virtual node)를 정의하였다.

결론적으로, 카메라 네트워크를 무방향 가중치 그래프를 표현할 수 있다. 그래프 G 는 두 개의 요소로 구성이 된다. 두 개의 요소란 정점(vertex) $V = \{v_1, v_2, \dots, v_n\}$ 의 집합과, 정점 V 를 연결하는 에지(edge) $E = \{e_1, e_2, \dots, e_n\}$ 의 집합이다. 정점은 카메라 노드, 중첩 노드, 비중첩 노드, 가상 노드를 표현하며, 에지는 일반 에지, 중첩 에지, 비중첩 에지로 표현한다. 에지 $e_i \in E$ 는 튜플(tuple) $\langle v_i, v_j, w_{ij} \rangle$ 에 의해 표현된다. 이것은 카메라 네트워크에서의 두 노드들 간의 연결을 뜻한다. 또한, 가중치 w_{ij} 는 v_i 에서 v_j 까지 거리에 대한 소요시간으로 정한

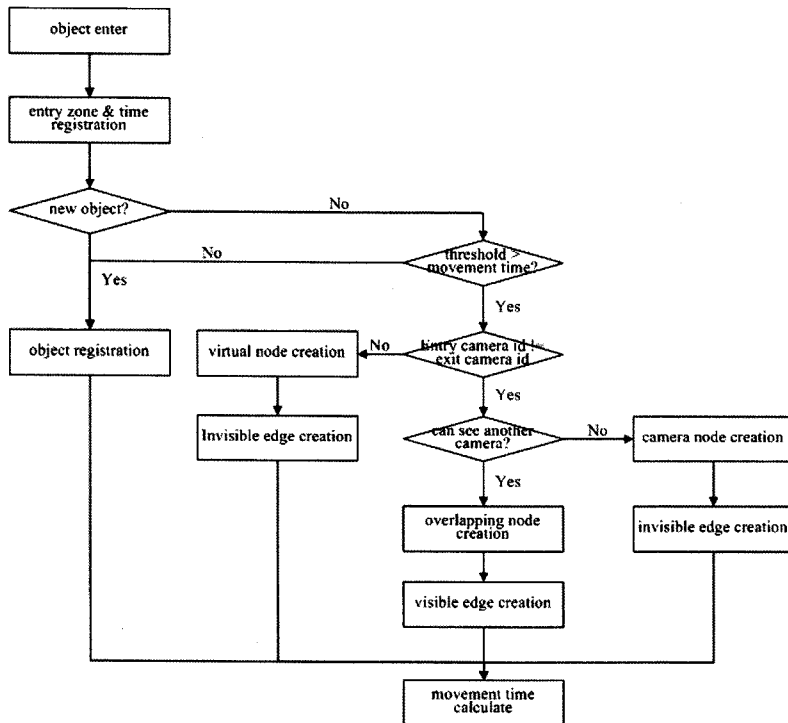


그림 4 객체가 등장할 때 처리하는 순서도

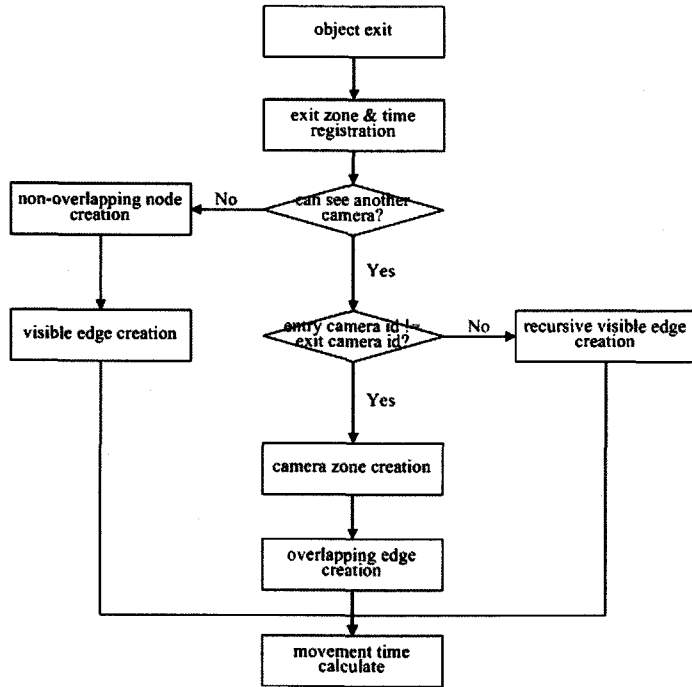


그림 5 객체가 퇴장할 때 처리하는 순서도

다. 여기에서 거리는 카메라들간의 출현 시간과 전이 시간에 의해 계산이 된다.

$$G = \langle V, E, W \rangle \tag{4}$$

$$w_{ij} = |T(C_i) - T(C_j)| \tag{5}$$

일반적으로, 그래프는 인접 행렬(adjacency matrix)을 이용하여 표현할 수 있다. 무방향 그래프 G 에 대한 인접 행렬은 $|V| \times |V|$ 배열이다. 이 배열을 A 라고 하고 만약 그래프 G 의 i 와 j 사이의 에지가 존재한다면 $A(i,j) = A(j,i) = \text{edge}(i, j)$ 의 거리이다. 만약 에지가 존재하지 않는다면, $A(i,j) = 0$ 이다.

$$A(i, j) = \begin{cases} w_{ij}, & \text{if } (i, j) \in E \\ 0, & \text{if } (i, j) \notin E \end{cases} \tag{6}$$

그림 9에서 만약 그래프가 무방향이라면 모든 엔트리(entry)는 그 에지의 양 두 끝점을 포함하는 두 개의 노드의 집합이다. 인접 매트릭스는 많은 0 값을 포함하고 있기 때문에 불필요한 메모리 영역이 발생하게 된다. 또한 무방향 그래프이기 때문에 행렬은 대칭적이기 때문에 이러한 공간적 효율성을 증대시키기 위하여, 그래프에 대한 인접 리스트를 사용하였다. 그래프 G 는 리스트 $Adj[1...|V|]$ 로 표현하였고 아래와 같은 데이터 구조에 의해 구현되었다.

5. 실험

표 1 연결 리스트를 위한 데이터 구조

index of adjacent vertex	weight	reference to next edge node
--------------------------	--------	-----------------------------

시스템 성능을 평가하기 위해서 다섯 개의 CCD 카메라(320×240)를 이용하였으며 펜티엄 4 프로세서와 1기가 바이트 메모리인 하드웨어 플랫폼에서 수행하였다. DBMS는 마이크로소프트 SQL Server 2000을 사용하였다. 또한 영상 처리를 위해 C++와 OpenCV 라이브러리[19]를 이용하여 구현하였다.

실험은 그림 6에서 보는 바와 같이 3개의 방과 1개의 복도에 설치된 5대의 카메라를 이용하여 테스트 하였다. 트레이닝 단계는 5분 동안 이루어졌으며 테스트는 10분 동안 실행하였다. 표 2는 이러한 실험환경에서 입장-퇴장 영역과 3명의 움직이는 객체들(p1, p2, p3)의 이동시간에 대한 결과의 일부분을 보이고 있다. 표 2에서 보는 바와 같이, p1은 09:45:22:12에 카메라1(C1)에 처음으로 나타났음을 알 수 있다. 따라서 이 경우, p1이 등록되고, 카메라 노드(c1)가 생성된다.

한편 p1은 09:45:30:34에 카메라2(C2)에 다시 나타나는데, 카메라1(C1)에서도 계속해서 나타나고 있다. 따라서 이 경우, 중첩 노드(o1)가 생성되고, 카메라 노드(c1)로부터 중첩 노드(o1)로의 가시 에지가 생성된다. 한편 p1이 카메라1(C1)에 나타난 후 다시 카메라2(C2)에 나타날 때까지 약 8초가 소요되었으므로, 이 때 중첩 노드

표 2 카메라에 따른 등장과 퇴장 및 출현과 전이 시간

Time	Camera	New entry	Re-entry	Exit	Travel time	Transition time
09:45:22:12	c1	p1				
09:45:26:87	c4	p2				
09:45:30:34	c2		p1		8	
09:45:31:68	c4			p2	5	
09:45:32:56	c1			p1	10	
09:45:32:81	c3		p2			1
09:45:37:41	c2			p1	7	
09:45:37:94	c3			p2	5	
09:45:38:23	c3		p1			1
09:45:39:18	c3			p1	1	
09:45:40:76	c4		p1			1
09:45:40:56	c5		p2			3
09:45:45:24	c4			p1	5	
09:45:46:74	c3		p1			1
09:45:48:21	c5			p2	8	
09:45:51:81	c3			p1	5	
09:45:54:81	c5		p1			3
09:46:02:64	c5			p1	8	
09:46:03:37	c5	p3				
09:46:05:15	c5		p1			-3
09:46:11:37	c5			p3	8	

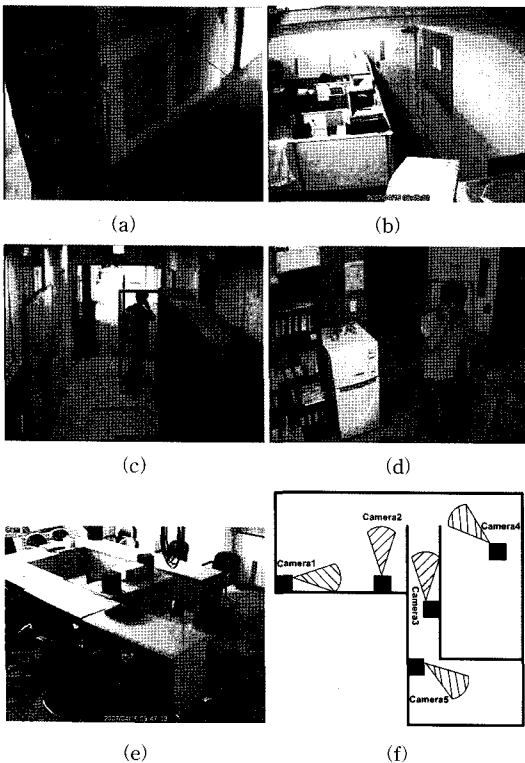


그림 6 (a)-(e) 실험 환경. (f) 카메라 시스템의 조감도

(o1)로의 가시 예지의 가중치로 8이 등록된다.

p1은 09:45:32:56에 카메라1(C1)에서 사라진다. 그러나 아직 카메라2(C2)에서는 보이고 있기 때문에 카메라

노드(c2)가 생성되고 중첩 노드(o1)로부터 카메라 노드(c2)로의 가시 예지가 생성되고 가중치로 2가 등록된다. 이후에, p1은 09:45:37:41에 카메라2(C2)에서 사라지고 다른 카메라에서도 보이지 않기 때문에, 비중첩 노드(no1)를 생성하고 카메라 노드(c2)로부터 비중첩 노드(no1)로의 가시 예지가 생성되고 가중치 5가 등록된다.

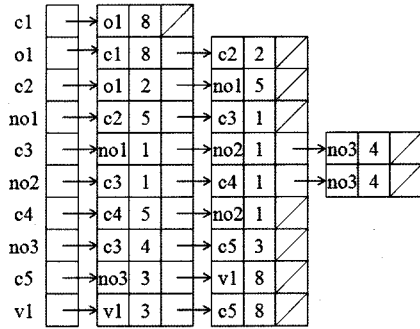
p1은 09:45:38:23에 카메라3(C3)에서 재등장하여 카메라 노드(c3)가 생성되고 비가시 예지가 생성되고 가중치 1이 등록된다. p1은 09:45:39:18에 카메라3(C3)에서 퇴장하여 비중첩 노드(no2)가 생성되고 카메라 노드(c3)로부터 비중첩 노드(no3)로의 가시 예지가 생성되고 가중치 1이 등록된다.

다시 p1은 09:45:40:76에 카메라4(C4)에서 재등장하여 카메라 노드(c4)가 생성되고 비가시 예지가 생성되고 가중치 1이 등록된다. p1은 09:45:45:24에 카메라4(C4)에서 퇴장하는데 이때 퇴장하는 카메라가 재등장 카메라이므로 회귀 가시 예지를 생성하고 가중치 5가 등록된다.

p1은 09:45:46:74에 카메라3(C3)에서 재등장하는데 이미 노드와 예지가 생성되어 있으므로 전이시간만 확인한다. p1은 09:45:51:81에 카메라3(C3)에서 퇴장하여 카메라5(C5)에서 재등장하며 이때, 비중첩 노드(no3)와 카메라 노드(c5)가 생성되고 각각 가시 예지와 비가시 예지와 가중치 4와 3이 생성된다.

p1은 09:46:02:64에 카메라5(C5)에서 사라진 후, 다시 09:46:05:15에 카메라5(C5)에서 재등장하는데 이때 가상 노드(v1)가 생성되고 각각 가시 예지와 회귀 비가시 예지를 생성하고 가중치 8과 3이 생성된다.

	c1	o1	c2	no1	c3	no2	c4	no3	c5	v1
c1	0	8	0	0	0	0	0	0	0	0
o1	8	0	2	0	0	0	0	0	0	0
c2	0	2	0	5	0	0	0	0	0	0
no1	0	0	5	0	1	0	0	0	0	0
c3	0	0	0	1	0	1	0	4	0	0
no2	0	0	0	0	1	0	1	0	0	0
c4	0	0	0	0	0	1	5	0	0	0
no3	0	0	0	0	4	0	0	0	0	0
n5	0	0	0	0	0	0	0	0	0	8
v1	0	0	0	0	0	0	0	0	8	3



(a) (b)

그림 7 (a) p1에 대한 인접 행렬 (b) 인접 리스트

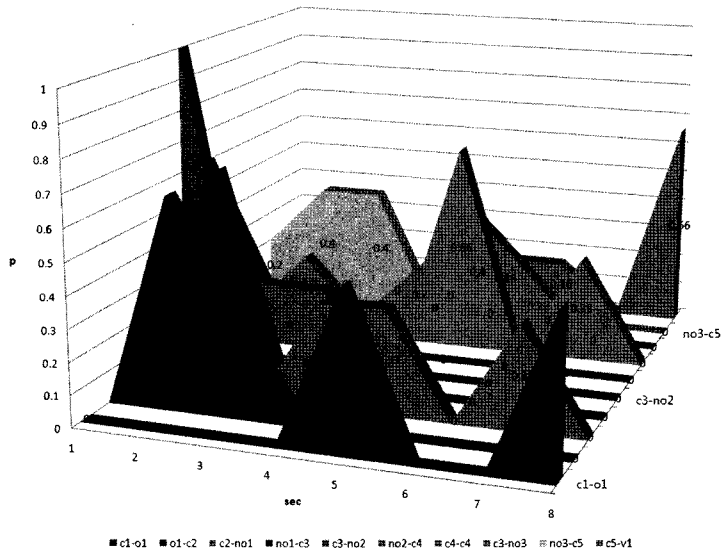


그림 8 10개 노드에 대한 출현과 전이 시간을 보이는 확률 분포도

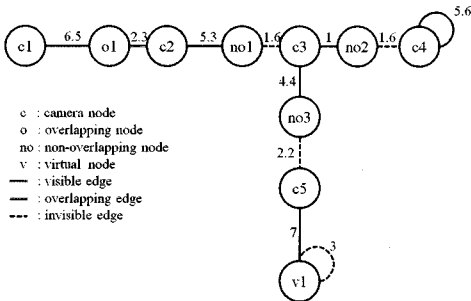


그림 9 무방향 가중치 그래프에 의해 표현된 카메라 네트워크의 위상

그림 7은 표 2를 기반으로 p1에 대한 그래프를 인접 행렬과 인접 리스트를 보이고 있다. 표에서 보는 바와 같이 행렬은 대칭적이며 많은 0 값을 포함하고 있다. 그

림 8은 5분 동안 10개의 에지에 대한 시간적인 확률 분포를 보이고 있으며, 임계값 범위 내에서의 각 에지에 대한 이동시간의 평균값으로 가중치를 결정한다. 결과적으로 그림 9는 무방향 가중치 그래프에 의해 표현된 카메라 네트워크 위상을 보이고 있다.

6. 결론

본 논문에서는 사람들의 움직임 추적을 근거한 등장-퇴장과 출현-전이 시간 모델을 이용하여 다중 카메라 네트워크의 시공간 위상을 자동적으로 구성하는 기법을 보였다. 이러한 기법은 일정 시간의 분석된 데이터를 이용하여 자율적으로 학습하는 방식을 사용하였으며, 중첩 또는 비중첩 카메라에서 입력받은 영상으로부터 이미지의 객체를 매칭하여 카메라간의 관계를 결정하였다. 이러한 카메라 네트워크의 시공간적인 위상은 무방향

가중치 그래프로 표현하였다.

강건하고 효과적으로 객체 추출 및 인식을 위해, 카메라와 움직이는 사람간의 거리에 따라 피부 영역을 그리드 기반의 추출 방법을 사용하였다. 또한, 겹침 블랍과 HSI 색상 공간을 고려하여 겹치는 객체들에 대한 인식을 효과적으로 개선하였다.

향후 연구로는 시공간 위상을 계단과 엘리베이터를 포함하는 건물에서 3 차원적인 위상 기법을 사용하여 확장하고, 객체들의 복합된 행동과 복잡한 이벤트를 인지하도록 고려하여 개발하는 것이다.

참 고 문 헌

[1] Y. Choi, K. Kim, W. Cho, "Grid-Based Approach for Detecting Head and Hand Regions," International Conference on Intelligent Computing, Qingdao China, August 21-24, 2007.

[2] J. Black, T. Ellis, and D. Makris, "Wide Area Surveillance with a Multi-Camera Network," Proc. IDSS-04 Intelligent Distributed Surveillance Systems, 2003, pp. 21-25.

[3] P. KaewTrakulPong and R. Bowden, "A Real-time Adaptive Visual Surveillance System for Tracking Low Resolution Colour Targets in Dynamically Changing Scenes," Journal of Image and Vision Computing. Vol.21, Issue 10, Elsevier Science Ltd, 2003, pp. 913-929.

[4] J. Sturges and T. Whitfield, "Locating Basic Colour in the Munsell Space," Colour Research and Application, 1995, pp. 364-376.

[5] Q. Cai and J. Agrarian, "Tracking Human Motion using Multiple Cameras," Proc. International Conference on Pattern Recognition, 1996, pp. 67-72.

[6] P. Kelly, A. Katkere, D. Kuramura, S. Moezzi, and S. Chatterjee, "An Architecture for Multiple Perspective Interactive Video," Proc. of the 3rd ACE International Conference on Multimedia, 1995, pp. 201-212.

[7] G. Welch and G. Bishop, "An Introduction to the Kalman Kilter," Technical Report 95-041, University of North Carolina at Chapel Hill, 1995.

[8] I. Haritaoglu, D. Harwood, L. Davis, "W4:Who, When, Where, What: A Real Time System for Detecting and Tracking People," Third International Conference on Automatic Face and Gesture, 1998.

[9] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4S: A realtime system for detecting and tracking people in 2 1/2D," 5th European Conference on Computer Vision, Freiburg, Germany, 1998.

[10] V. Kettner, R. Zabih, "Counting People from multiple cameras," in IEEE ICMCS, Florence, Italy, 1999, pp. 267-271.

[11] T. Huang and S. Russell, "Object Identification in a Bayesian Context, Proc. International Joint Con-

ference on Artificial Intelligence (IJCAI-97), Nagoya, Japan, 1997, pp. 1276-1283.

[12] Q. Cai and J.K. Aggarwal, "Automatic Tracking of Human Motion in Indoor Scenes Across Multiple Synchronized video Streams," 6th International conference on Computer Vision, Bombay, India, 1998, pp. 356-362.

[13] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. "Tracking Across Multiple Cameras with Disjoint Views," Proc. IEEE International Conference on Computer Vision, 2003, pp. 952-957.

[14] A. Dick and M. Brooks, "A Stochastic Approach to Tracking Objects Across Multiple Cameras," Australian Conference on Artificial Intelligence, 2004, pp. 160-170.

[15] T. J. Ellis, D. Makris, and J. Black, "Learning a multi-camera topology," In Joint IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), 2003, pp. 165-171.

[16] C. Stauffer, "Learning to track objects through unobserved regions," In IEEE Computer Society Workshop on Motion and Video Computing, 2005, pp. 96-102.

[17] K. Tieu, G. Dalley, and W. Grimson, "Inference of nonoverlapping camera network topology by measuring statistical dependence," In Proc. IEEE International Conference on Computer Vision, 2005, pp. 1842-1849.

[18] A. Gilbert, R. Bowden, "Tracking Objects Across Cameras by Incrementally Learning Inter-camera Colour Calibration and Patterns of Activity," In Proc European Conference Computer Vision, 2006, pp. 125-136.

[19] Intel. Open computer vision library. <http://sourceforge.net/projects/opencvlibrary/>

[20] P. Gabriel, J. Verly, J. Piater, A. Genon, "The State of the Art in Multiple Object Tracking Under Occlusion in Video Sequences," Advanced Concepts for Intelligent Vision Systems, pp. 166-173, 2003.



남 윤 영

2001년 아주대학교 정보 및 컴퓨터공학과(학사). 2003년 아주대학교 정보통신전문대학원 정보통신공학과(석사). 2007년 아주대학교 정보통신전문대학원 정보통신공학과(박사). 2007년~현재 아주대학교 유비쿼터스시스템연구센터 선임연구원. 관심분야는 데이터베이스, 멀티미디어 시스템, 영상처리, 패턴인식, 컴퓨터 비전, 유비쿼터스 컴퓨팅



류 정 훈

2007년 아주대학교 전자공학과(학사). 2007년~현재 아주대학교 일반대학원 전자공학과(석사). 관심분야는 컴퓨터비전, 패턴 인식, 유비쿼터스 컴퓨팅



최 유 주

1989년 이화여자대학교 전자계산학과(학사). 1991년 이화여자대학교 전자계산학과(석사). 2005년 이화여자대학교 컴퓨터학과(박사). 1991년~1993년 한국컴퓨터주식회사 기술연구소 주임연구원. 1994년~1999년 포스데이터주식회사 기술연구소 주임연구원. 2005년~현재 서울벤처정보대학원대학교 컴퓨터응용기술학과 전임강사. 관심분야는 컴퓨터 그래픽스, 가상현실, HCI, 컴퓨터비전, 의료영상처리 등



조 위 덕

1981년 서강대학교 전자공학과(학사). 1983년 한국과학기술원 전기 및 전자공학과(석사). 1987년 한국과학기술원 전기 및 전자공학과(박사). 1990년 금성전기(현 LG전자) 기술연구소 DSP 연구실장. 1990년~1991년 한국생산기술연구원 전자정보시스템연구부 팀장/조교수. 1991년~2003년 전자부품연구원 시스템연구본부 본부장. 2003년~현재 유비쿼터스컴퓨팅사업단 단장, 아주대학교 전자공학부 교수. 관심분야는 유비쿼터스 컴퓨팅/네트워크, 센서 네트워크, Post-PC(차세대 Smart PDA), Interactive DTV 방송기술, 고품질 홈서버/게이트웨이기술, 디지털방송/이동통신 연계 융합플랫폼기술, 무선인터넷응용기술