

조명변화와 겹침에 강건한 적응적 모델 기반 다중객체 추적

(Adaptive Model-based Multi-object Tracking Robust to Illumination Changes and Overlapping)

이 경 미 [†] 이 윤 미 ^{**}
(Kyoung-Mi Lee) (Youn-Mi Lee)

요약 본 논문에서는 고정된 카메라로부터 획득된 색상 비디오 프레임에서 조명변화와 겹침으로 인한 왜곡에 강건하게 다수의 사람을 추적하는 방법을 제안한다. 조명변화에 따른 외형변화의 문제점을 해결하기 위하여 시간 비종속적인 본래(intrinsic) 영상을 이용하여 프레임에 존재하는 조명을 제거하며, 매 프레임마다 조명, 영상을 적응적으로 갱신한다. 카메라 내에서 사람을 추적하기 위해 색상정보를 포함하는 계층적 사람모델을 사용함으로써 겹침의 문제를 해결한다. 추적된 사람모델은 사람모델 리스트에 저장되어 해당되는 사람이 카메라에서 사라진 후에도 일정 기간 보존됨으로써, 재등장한 사람의 정보를 복원할 수 있다. 본 논문에서 제안하는 적응적 모델기반 방법은 실내·외 영상을 대상으로 여러 시나리오로 실험되어, 조명변화로 왜곡된 사람의 색상정보를 옳게 보정하였을 뿐만 아니라 사람들이 겹치거나 헤어진 후에도 성공적으로 추적하였음을 확인하였다.

키워드 : 계층적 사람모델, 보안시스템, 조명보정, 사람추적

Abstract This paper proposes a method to track persons robustly in illumination changes and partial occlusions in color video frames acquired from a fixed camera. To solve a problem of changing appearance by illumination change, a time-independent intrinsic image is used to remove noises in an frame and is adaptively updated frame-by-frame. We use a hierarchical human model including body color information in order to track persons in occlusion. The tracked human model is recorded into a persons' list for some duration after the corresponding person's exit and is recovered from the list after her reentering. The proposed method was experimented in several indoor and outdoor scenario. This demonstrated the potential effectiveness of an adaptive model-base method that corrected distorted person's color information by lighting changes, and succeeded tracking of persons which was overlapped in a frame.

Key words : human tracking, surveillance system, hierarchical human model, illumination compensation

1. 서론

사람과 컴퓨터의 상호작용(HCI)의 표현에서 시작된 사람추적은 사용자 인터페이스 설계, 로봇학습, 감시 시스템 등의 많은 응용분야에서 사용되고 있으며, 특히 보안시스템분야에서 카메라를 이용하여 사람을 인식하고 추적하여 위험행위를 자동적으로 찾아낼 수 있는 감시 시스템 개발의 중요성이 날이 커져가고 있다. 현재 많

이 사용되고 있는 보안시스템으로는 백화점이나 마켓 등에서 사용되고 있는 무인감시 시스템, 신분증명 시스템, 범인검색 시스템 등이 있다. 그 외에 보행자를 감시하여 추적하는 지하철 안전사고 예방시스템이나 출입국 통제관리 시스템, 위험예방관리 시스템(예를 들어 사람이 위험지역에 접근하면 경보가 울리거나 특정한 행동을 했을 경우 경고하는 시스템) 등과 같은 보안시스템은 사람의 동작을 실시간으로 인식하고 그 의미를 처리할 수 있기 때문에 사람동작 추적이 주요하게 요구된다.

최근 수십 년 동안 비디오 카메라를 이용하여 실내·외 환경에서 움직이는 사람을 획득하고, 추적하는 방법에 대하여 여러 가지 환경과 방법들이 제안되어 왔다. 이들 방법은 크게 “장소(실내 환경, 실외 환경)”, “범위

· 본 연구는 2004학년도 덕성여자대학교 교내 연구비지원으로 이루어졌음

† 정 회 원 : 덕성여자대학교 컴퓨터과학부 교수

kmlee@duksung.ac.kr

** 비 회 원 : 덕성여자대학교 전산및정보통신대학원 대학원생

blanchia@duksung.ac.kr

논문접수 : 2004년 11월 10일

심사완료 : 2005년 3월 31일

(넓은 지역, 좁은 지역)”, “카메라 수(단일카메라, 다수의 카메라)”, 그 외에 획득된 영상의 “색상정보(색상정보, 그레이정보)”, “추적된 사람 수(한 사람, 다수의 사람들)”에 따라 분류될 수 있다(표 1). 본 논문에서는 실내·외 모두를 실험대상으로 하고 있으며 추적 범위는 비교적 좁은 지역으로 한정되었다. 또한, 비겹침 배치를 구성하는 단일카메라를 이용한다. 비겹침 배치를 이루는 카메라간의 정보전달을 위해 서로 다른 조명 환경을 보정하고 사람추적에 필요한 정보를 계층적 모델로 표현하였다. 본 논문에서는 피부색 및 사람이 착용한 옷 색 등의 색상정보를 이용한다.

사람추적은 감시시스템에서 사용되는 가장 기본적인 처리단계로 추적될 사람의 정보를 비디오를 통해 얻어와 그 사람의 동작이 어떤 의미와 의도를 가지는지 파악하기 위한 과정이다. 일반적으로 사람추적은 비디오나 영상의 기록으로부터 사람을 추출한 후, 추적할 사람 모델을 설계하여 동작이나 행동의 패턴을 분석하고 추적하는 방법을 쓴다. 즉, 위치, 속도, 형태, 색상 등의 특징을 이용하여 연속된 프레임사이의 프레임구조와의 대응을 만드는 것이다. 따라서 추적 과정은 프레임사이에서의 동작이나 형태와 같은 시각적인 정보를 이용하는 매칭을 포함하게 된다. 사람을 추적하기 위해 여러 방법들이 제안되었는데[1,2], “모델기반[3,6]”, “외형기반 [7,8]”, “블랍(blob)의 통계적 특징을 이용한 방법[3,4]” 등이 대표적으로 이용되는 방법들이다(표 1). 본 논문에서는 주변의 잡음이나 빛의 변화에 강건하고 추적대상의 행동 변화에 큰 영향을 끼치지 않는 색상정보를 이용하여, 각 추적대상의 피부 영역, 머리, 상체, 하체 등으로 몸의 블랍을 나누고 인식된 대상을 모델링 시키는 방법을 이용

한다.

본 논문에서는 고정된 카메라를 이용하여 서로 다른 실내·외 환경을 보정하여 다수의 사람을 검출하고, 검출된 다수의 사람을 추적하는 알고리즘을 제안하고자 한다. 조명 보정된 영상을 여러 개의 색으로 구별된 색상 집합들을 블랍으로 분리하고, 분리된 블랍의 위치정보와 색상정보를 이용하여 각 사람의 인체를 머리, 몸통, 다리, 피부색으로 나누어 사람을 추적한다. 그림 1은 본 논문에서 제안하는 적응적 모델 기반 다수사람 추적

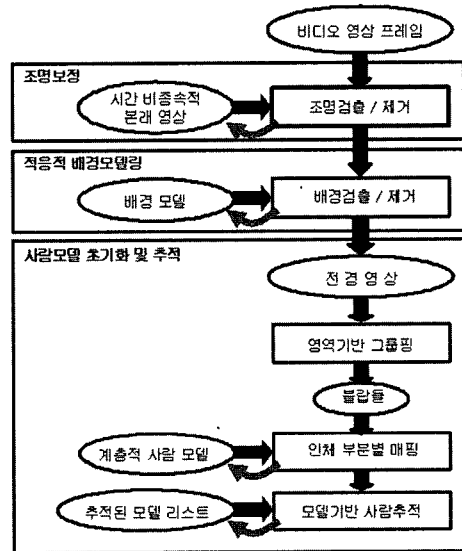


그림 1 제안된 방법에 대한 순서도: 타원은 데이터, 사각형은 영상처리 단계, 곡선 화살표는 갱신(update) 과정이다.

표 1 동영상에 사람 추적 방법에 대한 분류

분류 방법		논문										제안하는 방법
		1	2	3	4	5	6	7	8	9	10	
장소에 따라	실내	o	o	o	o	o			o			o
	실외			o			o	o	o	o	o	o
범위에 따라	넓은 지역						o	o	o	o		
	좁은 지역	o	o	o	o	o				o		o
카메라 수에 따라	단일 카메라	o	o	o	o	o	o			o	o	o
	다수의 카메라							o	o			
색상에 따라	흑백 영상			o			o		o	o		
	컬러 영상	o	o			o		o			o	o
사람 수에 따라	한 사람	o		o	o	o	o					
	다수의 사람들					o			o			
추적 방법에 따라	다수 사람 겹침		o					o		o	o	o
	모델기반	o			o			o			o	o
	외형기반			o		o	o		o			
	블랍의 통계적 특징	o	o									o
	기타									외곽선		

알고리즘에 대한 순서도를 보여주고 있다. 제안된 방법은 1) 시간 비중속 본래영상을 이용한 조명보정, 2) 적응적 배경모델링, 3) 블랍기반의 사람 모델 초기화, 4) 모델기반의 추적, 5) 적응적 사람모델링 및 갱신 등으로 이루어진다. 2장에서 본래영상을 이용한 적응적 조명 보정방법과 배경 모델링을 설명하고, 3장에서는 추적대상의 색상정보와 위치 정보를 이용한 블랍 기반의 적응적 사람 모델링과 추적방법을 자세히 설명한다. 실험 결과 및 결론은 각각 4장과 5장에서 설명한다.

2. 적응적 조명 모델링을 이용한 배경 제거

비디오 카메라에서 획득된 실내·외 영상은 같은 환경에서의 영상이라 하더라도 시간의 경과에 따라, 빛과 조명에 의해 끊임없이 변화된다. 비디오 카메라를 이용한 사람추적은 카메라의 상태, 입력영상의 조명과 잡음에 따라, 추적대상의 정보가 왜곡, 분실되기 쉬우며, 따라서 이러한 상태에서의 사람추적은 매우 어렵게 된다. 일반적으로 비디오영상에서 움직이는 사람을 탐지하기 위하여 전경영상과 배경영상과의 차이를 이용하여 배경을 분리시키는 방법으로 움직이는 물체를 추출한다 [3,6,13,14]. 그러나 불행하게도 이러한 방법으로는 움직이는 물체를 추출할 수는 있지만, 조명이나 여러 가지 잡음 등에 따라 배경이 끊임없이 변화하기 때문에 완전히 전경(사람)만을 분리하기가 쉽지 않고, 움직이는 물체와 함께 있는 조명과 그림자 같은 잡음 등을 감쇄시키는 것은 매우 어렵다. 따라서, 조명에 따라 끊임없이 변화하는 영상에서 배경을 완벽히 분리하기 위해서는 다양한 조명의 변화와 잡음에 강건한 방법이 필요하다.

그림 2는 도서관 외부를 내려다본 환경에서 각각 다른 조명과 빛의 상태를 보여준다. 그림 2(a)는 조명의 영향이 적은 영상, 그림 2(b)는 조명이 거의 없는 어두운 영상, 그림 2(c)는 밝은 조명이 많이 포함된 영상을 보여준다. 조명의 영향이 적은 영상과 비교하여 어두운 영상(그림 2(b))은 주변 환경과 추적대상과의 경계가 모호할 뿐 아니라, 추적대상이 가지고 있는 정보가 거의

유실되었음을 볼 수 있다. 밝은 조명이 많이 포함된 영상(그림 2(c))의 경우 나무와 사람의 여러 가지 그림자와 빛으로 인한 부분적 색상유실을 볼 수 있다.

2.1 적응적 조명 모델링

Y. Matsushita et. al은 정지된 물체의 조명 효과를 제거시키기 위해서 시간에 따라 변하는 음영영상 B' 과 조명영상 d' 을 이용하는 시간 종속적 본래(intrinsic)영상을 제안하였다[14]. 짧은 시간 동안에는 조명의 변화가 없다고 가정하여, 시간 t 에서 입력 영상 C' 는 아래와 같이 조명 환경에 따라 변하는 본래영상을 구할 수 있다.

$$d' = (C' - B'), B' = \text{median}(C') \quad (1)$$

시간 t 동안 촬영된 동영상에서의 반사영상 B' 은 입력 영상 C' 들의 중간값으로 간단하게 구해진다. 식 (1)을 이용하여 시간 t 에서의 조명영상 d' 은 현재영상과 반사영상의 차이로 구해지고, 본래영상은 반사영상과 조명영상의 조합으로 구해질 수 있다.

위에서 제시하는 본래영상을 이용하여 보정한 방식만으로도 효과적인 조명 처리가 이루어졌지만, 각 시간대마다 다른 본래영상을 이용하여 조명 보정을 하는 것은 비효율적이다. 본 논문에서는 하나의 본래영상을 이용하여 다양한 시간대의 현재영상을 변화시킬 뿐 아니라, 적응적 조명 모델링을 이용하여 현재영상의 조명은 이전 프레임에서 생성되었던 조명의 위치로부터 현재영상 내의 조명의 위치를 예측하고, 예측 가능한 조명을 모델링함으로써 예측된 조명을 영상에서 제거 할 수 있다. 따라서, 본 논문에서는 적응적 조명모델링을 이용하여 다음 프레임에서 나타날 조명과 음영을 현재의 프레임에서 예측하고, 예측된 조명을 제거하도록 제안한다. 본래영상은 고정된 카메라에서 획득된 영상은 배경에 대한 기본 프레임으로 적용할 수 있다는 점을 가정하여, 빛의 영향을 가장받지 않은 영상을 본래영상으로 두고, 입력되는 영상을 전경과 배경으로 분리한 후 사람일 가능성이 높은 전경영상에 대하여 처리가 이루어지도록 하였다.



(a) 조명의 영향이 적은 영상



(b) 어두운 조명



(c) 밝은 조명이 많이 포함된 영상

그림 2 조명환경에 따라 변화되는 영상

적용적 조명 모델링의 초기 단계로는 빛에 가장 영향을 받지 않은 영상을 본래영상으로 정의하고, 본래영상(R)과 매 프레임의 현재영상 C'을 조명에 강건한 YCbCr의 색상공간으로 변화시킨다. YCbCr 색상공간은 비디오파워 표준에서 널리 사용되고 있으며, 광도를 분리하여 조명의 영향을 고려할 수 있다. 또한, YCbCr 색상공간은 피부색에 대해 RGB나 HVS공간에 비하여 피부색 영역이 조밀하게 검출된다[17]. 본 논문에서는 피부색 영역을 이용하여 사람의 정보를 구하고 추적을 하기 위해, YCbCr 색상공간을 이용하였다.

비디오 카메라에 입력된 영상은 매 프레임마다 본래영상(R)과 현재영상(C)의 차이의 중간값을 이용하여 정규화하지 않고, 첫 번째 영상 C⁰만 본래영상을 이용하여 보정(C^{0*})시킨다. 현재영상은 이전 프레임에서부터 예측 가능하며, 따라서 이전 프레임에서 존재하는 조명은 현재영상에서 존재 가능한 위치를 예상할 수 있으므로, 두 번째 프레임부터는 이전 프레임에서 계산된 조명(N)의 값으로 현재영상을 보정하게 된다. 첫 번째 프레임을 제외한 두 번째 프레임부터는 아래의 식에 의해 보정된다.

$$C^{0*} = C^0 + \text{average}(R - C^0) \quad (2)$$

$$C^{0*} = C^0 + N^0 \quad (N^0 = \text{average}(R - C^0))$$

$$C^{1*} = C^1 + N^1$$

$$C^{2*} = C^2 + N^2$$

$$C^{(t-1)*} = C^{(t-1)} + N^{(t-1)} \quad (t \geq 2) \quad (3)$$

C^{0*}은 본래영상(R)을 이용하여 정규화 시켜준 영상이며, N₀은 본래영상과 C⁰과의 전체적인 차이의 중간값을 나타낸다. C⁰이 초기 영상이면, C¹~C^(t-1)은 식(3)을 이용하여 정규화 시켜주고, 보정된다. 매 프레임마다 본래영상을 이용하여 영상을 정규화 시키는 방법은 간단하고 어느 정도 효과적인 결과를 보여준다. 하지만, 현재영상은 이전프레임을 통해 예측이 가능하므로, 매 프레임마다 이전 프레임에서 계산된 조명으로 현재영상을 보정함으로써 적용적 조명 모델링이 가능하다. 식(3)은 현재영상의 영상은 이전프레임에서의 계산된 조명(N^(t-1)) 값으로 보정됨을 보여준다. C^(t-1)은 조명이 포함된 조명, C^{(t-1)*}는 이전 프레임을 이용하여 조명보정 후 조명이 최소화된 영상이므로, 두 영상의 차이로 조명을 쉽게 예측할 수 있다.

$$\delta'_N = \sqrt{\frac{(N^t - \mu'_N)^2}{(\sigma'_N)^2}} \quad (4)$$

$$\mu'_N = a^{t-1} \mu'_N^{t-1} + (1 - a^{t-1}) N^t$$

$$\sigma'_N = \sqrt{a^{t-1} W + (1 - a^{t-1})(\mu'_N - N^t)^2} \quad (5)$$

여기서 $W = (\sigma_N^{t-1})^2 + (\mu'_N - \mu_N^{t-1})^2$, $a^{t-1} = \frac{t-1}{t}$ 이다.

시간 t에서 보정 전 조명영상 N'의 각 화소마다 조명 정도가 다르므로, 유클리디안 거리와 같은 구형(spherical) 수식을 이용하여 각 화소의 조명도를 판단하는 것은 적당하지 않다. 따라서 통계학적 자료를 이용할 때, 조명과 영상처럼 변수들 사이의 상관도를 계산하는 타원형(elliptical) 수식인 마할라노비스 거리(δ'_N)를 이용하여 조명도를 판단한다. 마할라노비스 거리 δ'_N 는 시간 t에서의 조명영상과 시간 t-1까지의 평균(μ_N^{t-1})과 표준편차(σ_N^{t-1})로 구해질 수 있다(식(4)). 만일 각 화소가 많은 조명이나 음영을 포함하였다면, 즉, 각 화소의 조명도에 대한 거리 δ'_N 가 임계치보다 크다면, 시간 t-1까지의 각 화소에 대한 평균(μ_N^{t-1})과 표준편차(σ_N^{t-1})는 시간 t에서의 조명영상 N'을 이용하여 매 프레임마다 갱신되어 새로운 평균(μ'_N)과 표준편차(σ'_N)를 구한다(식(5))[15].

그림 3(b),(c),(d)는 그림 2(a),(b),(c)에 식(1)(왼쪽 줄)과 적용적 조명보정(오른쪽 줄)을 이용하여 보정시켜 준 결과이다. 그림 3(b)에서는 빛의 변화가 거의 없기 때문에 그림 2(a)와의 차이가 거의 없음을 볼 수 있지만, 그림 3(d)에서는 밝은 영상을 전체적으로 어둡게 만들어주어 비교적 색을 선명하게 나타내어 줌을 볼 수 있다. 그림 3(c)영상은 조명이 거의 없는 환경인 그림 2(b)에서 보여지는 주변과 사람의 경계가 모호해지고, 사람이 지니고 있는 색상정보가 많이 왜곡된 상태를 정규화 시켜줌으로써 사람의 모습을 분명하고 뚜렷하게 나타내고 있으며, 색상은 원래색상에 가까운 선명한 색상을 보여주는 결과를 볼 수 있다. 적용적 조명보정을 적용한 결과(그림 3(b),(c),(d)의 오른쪽 줄)는 그림 2(a),(b),(c)에 비해 영상이 뚜렷하고 원래의 색상에 가깝게 표현되고 있음을 보여주었다. 또한, 식(1)을 이용한 그림 3(b),(c),(d)의 왼쪽 줄 결과에 비해 색감이 균일함을 볼 수 있다.

2.2 배경제거

본래영상을 이용하여 조명 보정된 영상에서 배경을 분리한다. 배경영상을 분리하기 위해서 식(4)와 식(5)를 이용하여 배경의 표준편차(σ'_B)와 평균값(μ'_B)을 계산한다. 각 프레임들은 YCbCr공간의 각 Y,Cb,Cr 색상공간에서 시간 t 동안의 배경과 현재 영상(C)과의 차이를 이차 판별함수인 마할라노비스 거리(δ')를 이용하여 구한다. 첫 번째 프레임을 배경으로 사용시키므로, 통계적 정보를 0으로 초기화 한다($\mu_B^0 = \sigma_B^0 = 0$). 새로운 프레임이 들어올 때마다, 식(4)에 의해서 배경과 전경

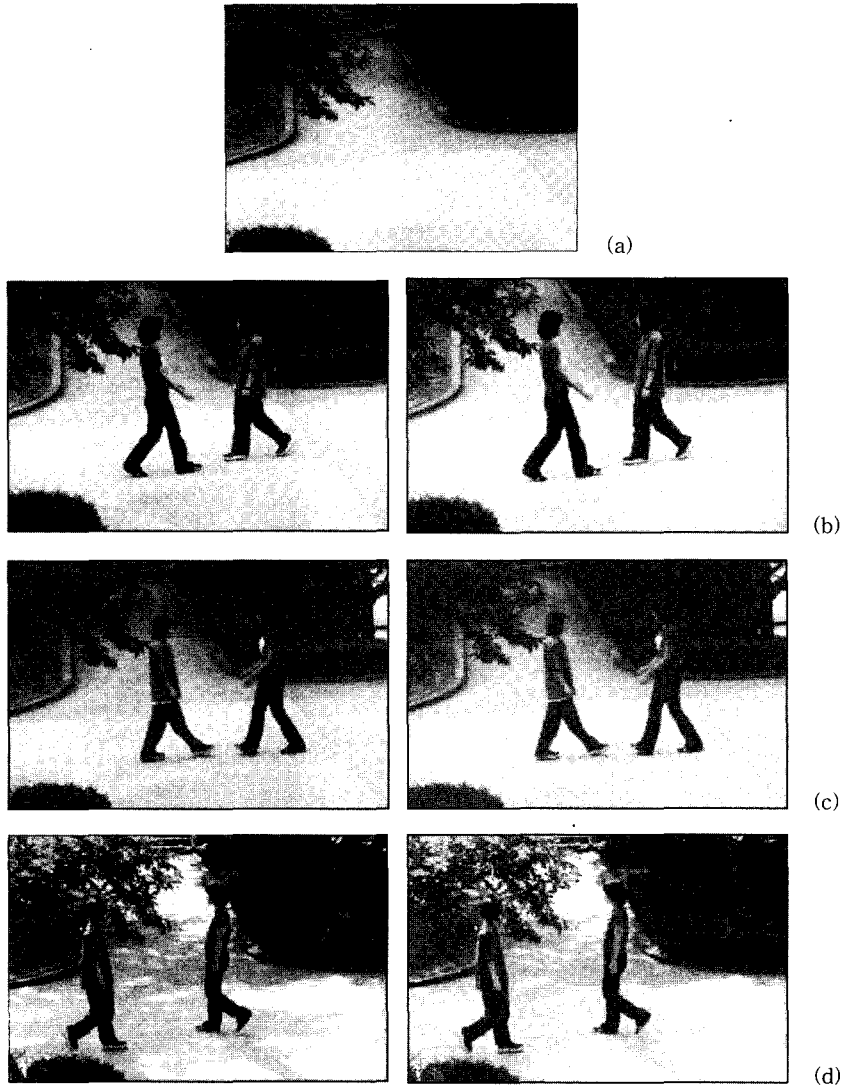


그림 3 조명보정: (a) 본래영상 R (실외), (b) 보통 영상, (c) 어두운 영상, (d) 밝은 영상. 그림 (b,c,d)의 왼쪽 줄은 각각 그림 2 (a,b,c)에 식 (1)을 적용한 결과이고, 오른쪽 줄은 제안된 조명 보정을 적용한 결과이다.

으로 분리된다. 분리된 배경과 배경모델의 차이가 임계치보다 적다면 배경모델(평균(μ'_B)과 표준편차(σ'_B))은 식 (5)를 이용하여 갱신된다.

그림 4는 실내(상단), 실외(하단) 환경에서의 배경제거 영상을 보여주고 있다. 그림 4(a)는 조명보정 없이 배경을 제거시킨 결과이고, 그림 4(b)는 식 (1)을 적용한 영상한 후 배경을 제거시킨 영상, 그림 4(c)는 적응적 조명 모델링을 적용한 후 배경을 제거시킨 결과이다. 식 (1)만을 적용한 그림 4(b)영상을 살펴보면 그림 4(a)

에 비해 잡음과 그림자 영역이 다소 감소한 결과를 보였으나, 여전히 잡음이 남아 사람 모델링과 추적(3장)을 어렵게 만든다. 하지만, 적응적 조명 모델링을 적용한 그림 4(c)영상은 앞의 두 결과영상에 비해 조명과 음영의 잡음이 효과적으로 제거된 모습을 확인 할 수 있을 뿐 아니라 색상의 선명도가 높아졌음을 볼 수 있다. 따라서 시간 중속 본래영상을 이용하여 조명보정을 한 결과에 비해, 적응적 조명 모델링이 조명과 잡음에 강건함을 보여준다.

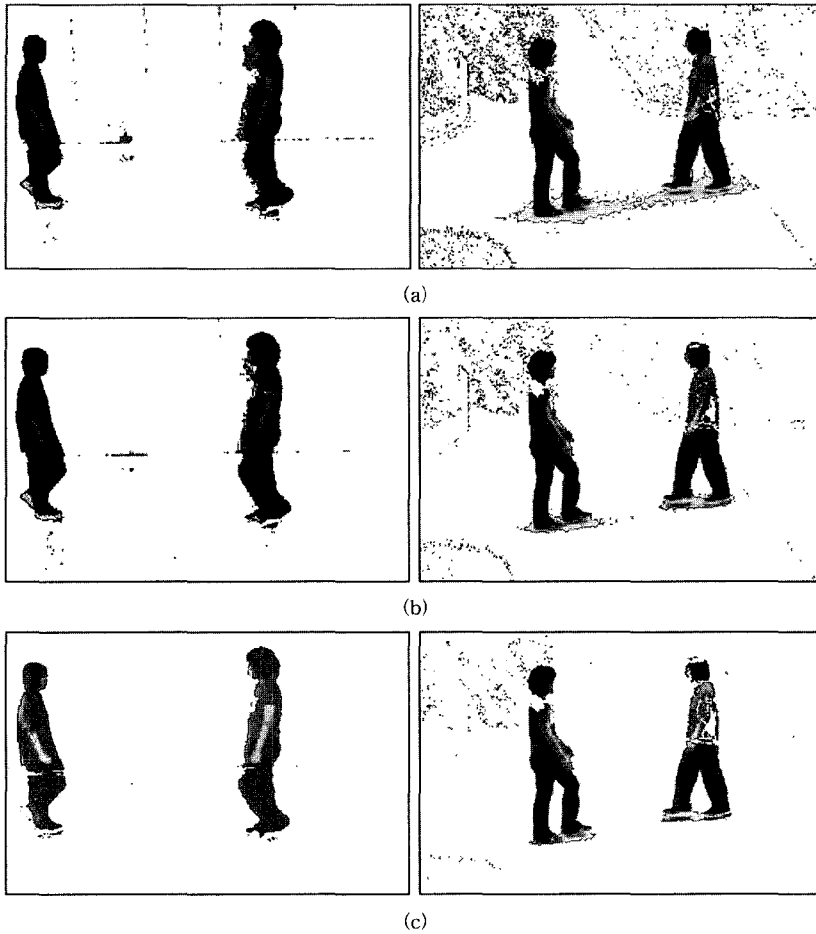


그림 4 배경제거: (a) 조명보정 없이 배경제거, (b) 식 (1)을 적용한 후 배경제거, (c) 식 (3), (4),(5)를 적용하여 조명 보정한 후 배경제거. 왼쪽 줄은 실내 영상, 오른쪽 줄은 실외 영상의 결과이다.

3. 블랍 기반 다중 사람 추적

3.1 사람 모델 초기화

사람추적은 추적되는 사람이 비디오에 처음 나타나기 시작할 때 사람모델을 초기화하는 것으로 시작된다. 본 논문에서는 전경영상으로부터 유사 색상의 영역을 구성하고, 계층적 사람 모델을 설계하여 추적과 신분 증명에 이용할 것이다. 분리된 전경영상의 화소는 각 화소와 인접화소 간의 밝기 유사도를 계산하는 연결요소 알고리즘을 사용하여 블랍으로 모아진다. 각 블랍 B_i 는 블랍의 색상, 크기, 중심점, 경계 상자, 인체를 구성하는 경계 등의 정보를 포함한다. 그러나, 이렇게 밝기를 이용한 군집화는 색상과 빛의 변화로 자동 블랍 형성에 많은 문제를 야기하므로 개별적인 인체 부위를 완벽하게 구성하는 것은 매우 어렵다. 특히, 초기 군집화는 거의

밀착된 블랍의 표면에 존재하는 상당한 조명 변화로 과잉 분할(over-segment)되기 쉬운데, 이를 해결하기 위해서 비슷한 색상을 가진 이웃 블랍들과 함께 합쳐질 것이다. 만약 생성된 블랍이 너무 작거나, 길거나, 이질적으로 비조밀하면 지워진다.

사람은 인체 부위에 대응되는 블랍의 부분집합으로 정의될 수 있으므로, 프레임 내의 블랍들은 다수의 사람들을 추적하기 위해 대응되는 인체에 할당되어야 한다. 인체를 구성하는 알고리즘은 우선 블랍 인접 그래프와 각 블랍의 경계상자사이의 최소거리를 계산한다. 경계상자의 꼭지점들 사이의 거리는 화소 단위에서의 인접도보다 훨씬 빠르게 계산되어진다[17]. P_0 가 블랍 B_i 의 부분집합이라고 하자. 잠재적인 인체 영역집합은 P_0 와 인접그래프에서 시작하여 반복적으로 만들어 진다. 집합

P_1 은 P_0 의 적절한 인접 블랍과 합쳐짐으로써 얻어질 수 있다. 다음은, 합쳐진 블랍 P_k 의 각 새로운 집합은 초기 집합인 P_0 와 합쳐진 블랍 P_{k-1} 의 집합을 합침으로써 얻어진다. 결국 후보 인체 CP_n 은 합쳐진 블랍 P_k 의 집합들을 포함한다. 즉, $CP_n = \cup_{k=0}^n P_k$ 이고, n 은 사람의 숫자이다.

블랍을 인체 부위에 매핑시키기 위해서, 본 논문에서는 계층적 사람모델을 이용하였다(그림 5). 우선, 비디오의 모든 사람이 바로 서 있거나 약간 기울어져 서 있다고 가정한다. 입력영상(그림 5(a))에서 전경영상(그림 5(b))이 구해지면, 모델의 상위 계층은 사람 전체와 전체에 대한 정보를 포함한다(그림 5(c)). 모델에서의 상대적인 위치에 따라 각 블랍은 중간 계층 내에 머리, 상체와 하체의 세 인체 부위 중 하나로 할당된다(그림 5(d)). 만일 중간 계층 내 인체 부위가 두 개 이상의 블랍을 포함한다면, 이 블랍들은 하위 계층에서 피부색 영역과 아닌 영역으로 분류하기 위하여 피부색 유사도를 계산한다(그림 5(e)). 사람모델은 세 인체부위와 중간 계층에서의 그들 간의 기하학적인 관계와 블랍, 하위 계층에서 그들간의 기하학적 (위치와 상대적 크기 등), 색상적인(평균 색상과 표준편차 등) 관계로 정의된다.

$$CP_n = \{R_n^0, C_n^1, R_n^1, C_n^2, R_n^2, C_n^3, R_n^3\}$$

여기서 R^0 는 세 부위 사이의 관계를 나타낸다. C_n^j 와 R_n^j ($j=1,2,3$)는 CP_n 의 j 번째 신체 부위의 블랍과 그들간의 관계를 나타낸다.

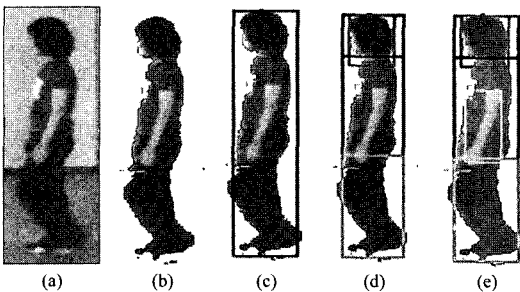


그림 5 계층적 인체모델링 : (a) 입력영상, (b) 전경영상, (c) 상위 계층, (d) 중간 계층, (e) 하위 계층

3.2 적응적 사람 모델링

사람은 사람모델을 이용하여 추적되기 때문에, 사람모델의 정보는 다수의 사람을 추적하기 위해 저장되어야 한다. 사람의 모든 동작이 프레임마다 상대적으로 작기 때문에, 색상기반 모델에서의 큰 변화는 추적을 쉽게 실패하게 만드는 원인이 된다. 이런 색상모델 추적의 민감성을 해결하기 위해서, 프레임의 블랍들을 인체 참조모

델과 비교한다. 적응적 조명모델(2장)처럼 사람 참조모델은 조명 변화에 따른 왜곡을 보상시켜준다.

사람 CP_n^{t-1} 은 $t-1$ 까지 계산된 평균 (μ_n^{t-1})과 표준편차 (σ_n^{t-1})에 의해 표현되고, t 번째 프레임은 새로운 블랍 B_n^t 과 그들 간의 관계 Br_n^t 로 구성된다고 하자. 인체 모델 CP_n^{t-1} 과 새 블랍 B_n^t 사이의 최소 거리는 다음과 같이 계산된다.

$$d_n^t = \min_{j=1 \dots 3} \left(\frac{\|B_n^t - \mu_n^{t-1, C^j}\|_A}{\sigma_n^{t-1, C^j}} \right) + \min_{j=0 \dots 4} \left(\frac{\|Br_n^t - \mu_n^{t-1, R^j}\|_A}{\sigma_n^{t-1, R^j}} \right) \quad (6)$$

여기서 μ_n^{t-1, C^j} 와 μ_n^{t-1, R^j} 는 시간 $t-1$ 에서 j 번째 신체 부위 내 블랍과 관계들의 평균을, σ_n^{t-1, C^j} 와 σ_n^{t-1, R^j} 는 표준 편차를 나타낸다. 최소거리가 미리 정해진 임계치보다 적으면, 제안된 모델링 알고리즘은 블랍 B_n^t 와 관계 Br_n^t 를 적응적 사람모델(μ_n^{t-1, C^j} 와 μ_n^{t-1, R^j})에 추가하고, 그들의 중심과 불확실성을 다시 계산함으로써 적응적 모델을 갱신한다[15]. 유사도 임계치는 실험적으로 정해지고, 사용자에게 의해 조절될 수 있다.

3.3 모델 기반 다중 사람 추적

사람의 신체는 고정된 형태를 가지고 있지 않기 때문에 사람을 추적하는 것은 여러 어려운 문제들을 포함하고 있다. 블랍을 구성한 후에, 블랍 기반 사람 추적은 이전 프레임에서 현재영상까지 연속된 프레임들 사이의 거리를 계산함으로써 블랍을 배치한다. 그러나 이런 블랍 기반 방법은 각 프레임에 서로 다른 개수의 블랍으로 인하여 다수의 사람을 추적하는 것을 어렵게 만든다. 블랍은 분할 또는 합병되거나, 심지어 사라지거나 새로 만들어질 수 있다. 이런 문제를 해결하기 위해서 다대다 (many-to-many) 블랍 배치가 적용될 수 있다[4]. Y. Huang와 T.S. Huang는 시간 $t-1$ 에서의 블랍이 시간 t 에서의 블랍과 관계되는 것을 피하기 위해 동시에 다수의 블랍을 연계하는 다양한 다중 에이전트 추적 프레임워크를 적용하였다.

본 논문에서는 사람 CP_n^{t-1} 이 $t-1$ 번째 프레임까지 이미 추적되었고, 프레임 t 에 새로운 블랍 B_n^t 가 구성되었다고 가정한다. 이후 다수의 사람은 아래의 방식으로 추적된다.

case 1. 만일 B_n^t 가 CP_n^{t-1} 에 포함된다면, CP_n^{t-1} 내 대응되는 블랍은 B_n^t 으로 추적된다.

case 2. 만일 CP_n^{t-1} 내 블랍이 t 번째 프레임에서 여러 블랍으로 분리되었다면, CP_n^{t-1} 내 블랍은 t 번째 프레임내 한 블랍으로 추적되고 t 번째 프

레이프 내 다른 블랍들은 CP_n^{t-1} 에 추가된다.

case 3. 만일 CP_n^{t-1} 내 블랍이 t 번째 프레임에서 블랍 B'_i 로 합쳐진다면, CP_n^{t-1} 내 블랍은 t 번째 프레임내 한 블랍으로 추적되고 t 번째 프레임 내 다른 블랍들은 CP_n^{t-1} 에 제거된다.

case 4. 만일 B'_i 가 CP_n^{t-1} 에 포함되지만 대응되는 블랍이 존재하지 않으면, B'_i 를 CP_n^{t-1} 에 추가한다.

case 5. 만일 B'_i 가 CP_n^{t-1} 에 포함되지 않으면, 블랍은 새로 생성된 블랍으로 여겨져 사람 리스트에 새로운 사람로 추가된다.

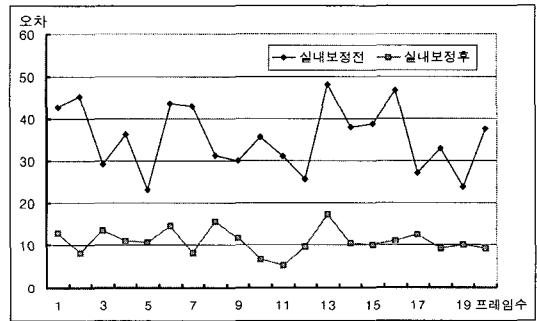
여기서 영역이 경계상자에 의한 사람에 포함된다는 것은 그 영역이 경계상자의 90%이상 겹침을 말한다. 블랍과 적응적 사람모델의 대응은 식 (6)에 의해 계산된다. 제안된 방법은 사람들과 블랍들의 리스트 처리를 간단하게 할 뿐만 아니라, 기존의 사람이 비디오 프레임에서 나가는가, 새로운 사람이 비디오 프레임 안으로 다시 들어오는가, 또 이전에 비디오 프레임에서 나갔던 사람이 다시 들어오는가를 계속해서 확인할 수 있게 한다. 제안된 방법의 또 다른 장점은 옮겨 블랍화하는 것에 대한 부담을 덜어준다. 블랍이 조명 변화에 의해 놓쳐져도 모델에 기반한 추적은 다른 존재하는 블랍을 이용하여 개별적인 사람을 유지할 수 있게 한다. 인체가 구성된 후, 인체 내 블랍의 개수는 유연하게 변경가능하다. 제안된 방법은 과도한 블랍화(over-blobbing, case 2)와 부족한 블랍화(under-blobbing, case 3)을 처리할 수 있다.

4. 실험결과

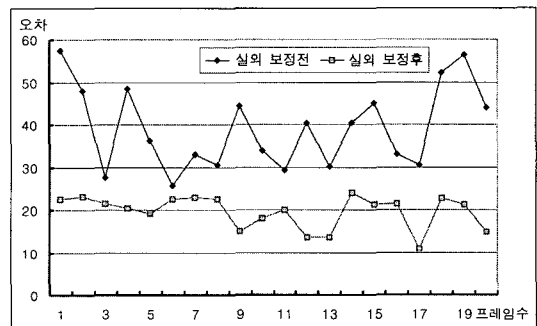
제안된 논문은 운영체제 Windows 2000XP 상에서 JAVA (JMF)를 이용하여 구현되었으며, 실험은 Pentium-IV 1.8GHz인 CPU와 512 MB 메모리 사양의 Windows 2000 XP 환경에서 수행되었다. Sony DSC-P10에서 획득된 동영상(420×316)은 실내·외에서 12개의 서로 다른 조명 환경과 시나리오로 사람의 움직임을 포함한다.

표 2는 서로 다른 조명 환경의 실내 동영상 4개, 실외 동영상 4개에서 각각 20프레임 동안 한 사람을 추적했을 경우 원래 사람이 가지고 있는 색상정보와 얼마나

일치하는지를 나타내는 평균 밝기 값에 대한 도표이다. 그림 6은 표 2에 대한 그래프로, 가로축은 프레임의 수를 나타내며 세로축은 본래영상에서 사람이 가지고 있는 색상정보와 서로 다른 조명환경에서 추적된 사람과의 밝기의 오차를 나타낸다. 추적된 사람의 밝기정보는 추적된 사람의 전체 밝기의 평균값으로 추출 하였고, 원래 사람이 가지고 있는 정보는 원래 영상에서 추출된 사람의 20프레임동안의 밝기의 평균값을 기준으로 하였다. 그림 6(a)는 실내 환경에서 조명보정 전과 조명보정 후의 밝기의 오차를 비교한 그래프이며, 그림 6(b)는 실외 환경에서 조명보정 전과 조명보정 후의 밝기의 오차를 비교한 결과이다. 두 그래프의 결과에서 조명보정을 한 그래프의 조명보정 전의 그래프에 비하여 그래프의



(a) 실내



(b) 실외

그림 6 실내·외 환경에서의 조명 보정전과 조명보정 후의 밝기차이의 평균(표 2 참조)

표 2 실내·외 환경에서의 조명 보정전과 조명보정 후의 밝기차이의 평균

프레임		환경																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
실내	보정전	42.7	45.2	29.2	36.2	23.2	43.5	43.0	31.2	30.0	35.7	31.0	25.7	48.2	38.0	38.7	46.7	27.0	33.0	23.7	37.5
	보정후	13.0	8.2	13.5	11.0	10.7	14.5	8.2	15.7	11.7	6.7	5.2	9.5	17.2	10.5	10.0	11.0	12.5	9.2	10.1	9.2
실외	보정전	57.2	48.0	27.5	48.5	36.2	25.7	33.0	30.5	44.5	34.0	29.2	40.5	30.2	40.5	45.0	33.0	30.5	52.5	56.5	44.0
	보정후	22.5	23.2	21.5	20.5	19.2	22.5	23.0	22.5	15.2	18.2	20.0	13.7	13.7	24.0	21.2	21.5	11.0	22.7	21.2	15.0

기복이 거의 일정하고 오차의 차이가 적음을 알 수 있다. 이는 제안하는 적응적 조명 모델링을 이용해서 예상되는 잡음과 조명을 제거함으로써, 영상이 가지고 있는 본래의 색상정보가 조명이나 잡음에 의해 왜곡이나 손실을 최소화했음을 보여준다. 따라서 본 논문에서 제안한 적응적 조명보정에 의해 보정된 영상이 조명보정 전의 영상에 비하여 본래의 색상정보를 큰 변화와 손실 없이 유지하고 있음을 알 수 있으며, 제시한 조명보정 방법이 다양한 조명의 변화에 대하여 강건함을 확인할 수 있었다.

보통, ROC(Receiver Operating Characteristics)는 정확도와 민감도 사이의 관계로서 성능을 나타내는데 이용된다. 본 논문에서는 사람의 추적률을 측정하기 위해서 추적민감도(tracking sensitivity)와 추적 정확도(tracking accuracy)로 구성된 ROC 그래프를 이용하였다. 추적민감도는 사람이 모델에 의해 옳게 추적될 확률을 의미하고, 추적정확도는 추적된 모델이 사람일 확률을 의미한다. 각 추적 단계에서 사람으로 추적된 모델의 수를 세고 전체 사람 수에 대한 비율을 계산하였다.

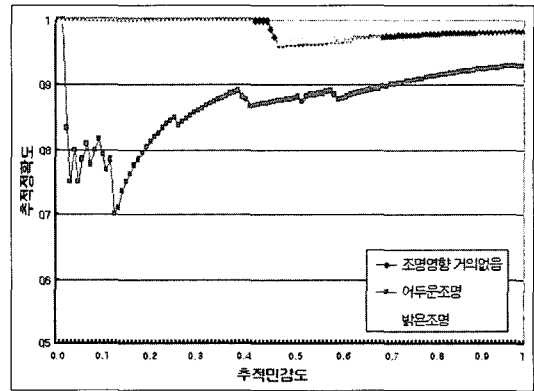
$$\text{추적민감도} = \frac{\text{추적된 사람수}}{\text{전체 추적될 사람수}}$$

$$\text{추적정확도} = \frac{\text{옳게 추적된 누적 사람수}}{\text{추적된 누적 사람수}}$$

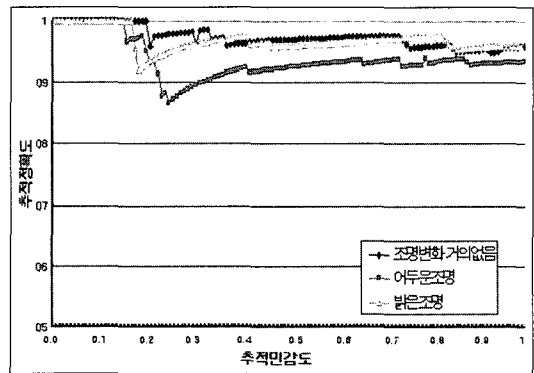
그림 7은 조명보정 후 동영상에서 계층적 사람 모델링을 이용하여 사람의 추적률에 관한 ROC 그래프이다. 그림 7(a)는 실내 영상에서 각각 조명의 영향을 거의 받지 않은 영상과 어두운 조명영상 그리고 밝은 조명영상에서 나타난 사람을 추적한 ROC 그래프이고, 그림 7(b)는 실내 영상에서 각각 조명의 영향을 거의 받지 않은 영상과 어두운 조명영상 그리고 밝은 조명영상에서 나타난 사람을 추적한 ROC 그래프이다. 그림 7(a)에서는 실내 환경에서 조명의 영향을 거의 받지 않은 조명영상 128 프레임에 대하여 98.3%의 추적률을 나타냈고, 어두운 조명영상 118 프레임에 대하여 93.8%의 추적률을 나타냈으며, 밝은 조명영상 139 프레임에 대하여 98% 추적률을 보여주었다. 그림 7(b) 실외 환경에서는 조명의 영향을 거의 받지 않은 조명영상 144 프레임

에 대하여 96.3%의 추적률을 나타내었고, 어두운 조명영상 126 프레임에 대하여 93.4%의 추적률을 보여주었다. 밝은 조명영상에서는 125프레임에 대하여 96% 추적률을 보여주었다.

그림 7(a)과 (b)의 그래프를 보면, 밝은 조명영상에 대해 실내·외 환경 모두에서 사람추적률이 높음을 보여주었다. 하지만, 그림 7(a),(b)의 어두운 조명영상에서의 ROC그래프에서는 초기의 실내 그래프가 실외 영상보다 기복이 심한 것을 볼 수 있다. 특히 실험에 사용된 실내



(a) 실내



(b) 실외

그림 7 사람추적에 대한 ROC(표 3 참조)

표 3 조명보정 후 사람추적에 대한 ROC

조명환경		프레임												
		10	20	30	40	50	60	70	80	90	100	110	120	
실내	보통 영상	1	1	1	1	1	0.97	0.97	0.97	0.98	0.98	0.98	0.98	
	밝은 영상	1	1	1	1	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.98	
	어두운 영상	0.80	0.78	0.85	0.88	0.87	0.88	0.89	0.89	0.91	0.92	0.93	0.93	
실외	보통 영상	1	1	0.98	0.99	0.96	0.97	0.97	0.97	0.97	0.96	0.95	0.96	
	밝은 영상	1	1	0.95	0.96	0.97	0.96	0.96	0.97	0.97	0.98	0.96	0.96	
	어두운 영상	1	0.97	0.88	0.90	0.92	0.92	0.93	0.94	0.94	0.93	0.93	0.93	

영상의 어두운 조명환경을 구현하기 위해 조명이 거의 없는 환경에서 영상을 획득했기 때문에, 실외영상보다 상대적으로 조명이 거의 없었다. 모델이 체계화 되지 않는 초기에는 그래프의 기복이 심했지만, 처리된 프레임 수가 늘어남에 따라 모델이 체계화되고 추적률이 조명변화가 거의 없었던 영상의 결과에 다가가는 좋은 결과를 나타냈다. 그림 7(a),(b)의 ROC 그래프는 사람이 여러 사람이 들어오고, 겹쳐지고, 헤어지고, 나가는 시나리오에 따른다. 그림 7(a),(b) 그래프에서 초기 구간에서 1을 유지하다가 약간 하향되는 것을 볼 수 있는데, 이것은 여러 사람이 시간의 차이를 두고 들어오기 때문에 나중에 들어오는 사람의 모델이 체계화되지 않았기 때문에 순간적으로 사람 추적률이 떨어졌음을 의미한다. 그리고, 그래프의 중간 구간에서 그래프가 순간적으로 하향되었다가 바로 복원되었음을 볼 수 있는데, 이는 적응적 사람모델들이 서로 겹쳐지는 순간을 의미한다. 하지만 그래프가 바로 복원되어, 겹쳐진 사람모델들이 헤어진 후에도 올바르게 추적되고 있음을 확인할 수 있었다.

그림 8과 9는 3장에서 설계한 계층적 인체모델을 이용하여 조명조건이 각각 다른 실내·외 환경에서 다수의 사람을 추적한 결과를 보여준다. 그림 8의 첫째, 둘째 줄은 실내 환경에서의 밝은 조명영상, 셋째, 넷째 줄은 실내 환경에서의 조명의 영향이 거의 없는 영상, 다섯째 줄은 실내 환경에서의 어두운 조명영상에서의 결과이다. 그림 9의 첫째, 둘째 줄은 실외 환경에서의 밝은 조명영상, 셋째, 넷째 줄은 실외 환경에서의 조명의 영향이 거의 없는 영상, 다섯째 줄은 실외 환경에서의 어두운 조명영상에서의 결과이다. 실내·외 환경에서의 조명조건이 각각 다른 10개의 영상에서 사람들이 들어오고(그림 8(a),(b), 9(a),(b)), 겹쳐지고(그림 8(c), 9(c)), 헤어지고(그림 8(d), 9(d)), 나감(그림 8(e), 9(e))의 순서에 따라 사람추적을 순차적으로 보여주고 있다. 그림 8과 9의 결과를 보면 빛의 영향이 거의 없는 환경(다섯째, 열째 줄 결과)인, 과도한 조명의 영향을 받는 환경(첫째, 둘째, 다섯, 여섯째 줄 결과)에서도 다수의 사람이 정확히 추적되었음을 볼 수 있다.

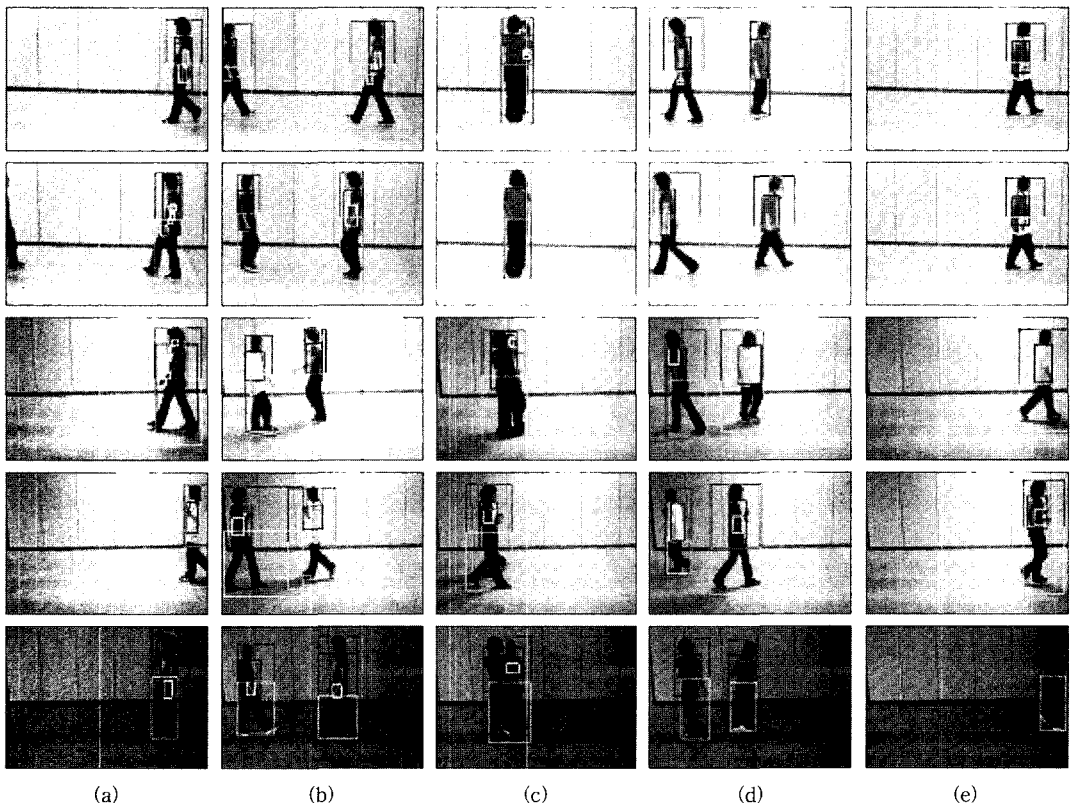


그림 8 실내 영상에서 사람추적 결과: (a) 들어옴, (b) 다수의 사람 들어옴, (c) 겹침, (d) 헤어짐, (e) 나감. 첫째와 둘째 줄은 밝은 조명, 셋째와 넷째 줄은 조명의 변화가 거의 없는 조명, 다섯째 줄은 어두운 조명의 영상에서의 추적 결과이다.



그림 9 실외영상에서의 사람추적 결과: (a)들어옴, (b) 다수의 사람 들어옴, (c) 겹침, (d) 헤어짐, (e) 나감. 첫째와 둘째 줄은 밝은 조명, 셋째와 넷째 줄은 조명의 변화가 거의 없는 조명, 다섯째 줄은 어두운 조명에서의 사람 추적 결과이다.

5. 결론 및 향후 과제

본 논문에서는 고정된 카메라에서 서로 다른 실내·외 환경을 적응적 조명 모델링을 이용하여 다수의 사람을 인식하고, 인식된 다수의 사람을 추적하는 알고리즘을 제안하였다. 일반적으로 비디오 영상은 주위 잡음이나 조명의 영향을 많이 받기 때문에 조명과 잡음에 강건하게 물체 및 사람을 추적하기란 매우 어려운 일이다. 따라서 여러 가지 조명환경에 강건하게 사람을 추적하기 위하여 본 논문에서는 매 프레임마다 이전 프레임에서 계산된 조명으로 현재영상을 보정하는 적응적 조명 모델링을 이용하였다. 본 논문 4장에서 그림 7의 결과를 살펴보면 실내·외 환경의 다양한 조명영상에서 적응적 조명보정 후의 영상이 조명보정 전의 영상보다 원래영상의 색상정보를 큰 변화와 손실 없이 유지하고 있음을 보여주었다. 또한, 사람의 겹침으로 인한 왜곡에도 완전하게 사람을 추적할 수 있는 방법이 필요하다. 그림 8과 9에서는 실내영상에서 평균 96.8%, 실외영상 95.2%의 사람 추적 추적률의 결과를 얻어, 본 논문에서 제안한

제출적 인체모델의 방법이 효과적임을 보여주었다.

본 논문에서 제안한 방법은 앞으로 다수의 카메라의 서로 다른 다양한 조명과 환경에서 조명문제를 해결하기 위한 연구로의 확장이 향후 과제로 남아 있다. 본 논문에서 제안하는 인체모델 알고리즘은 다수의 비겹침 카메라 네트워크에서의 사람추적 연구의 일부분이다. 한 카메라에서 추적된 사람의 정보를 FOV(Field of view)가 겹치지 않은 다른 카메라로 전달하여 그 사람의 출현을 검출하고 계속 추적하기 위한 인체모델로 사용될 것이다.

참고 문헌

[1] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE trans. on PAMI*, 19(7):780-785, 1997.
 [2] S. Park and J. K. Aggarwa, "Segmentation and tracking of interacting human body parts under occlusion and shadowing," in *Proc. of International Workshop on Motion and Video Computing*, pp.

- 105-111, 2002.
- [3] Y. Ricquebourg and P. Bouthemy, "Real-time tracking of moving persons by exploiting spatio-temporal image slices," *IEEE trans. on PAMI*, 22(8):797-808, 2000.
- [4] Y. Huang and T. S. Huang, "Model-based human body tracking," in *Proc. of International Conference on Pattern Recognition*, pp. 552-555, 2002.
- [5] M. B. Capellades, D. Doermann, D. DeMenthon and R. Chellappa, "An appearance based approach for human and object tracking," in *Proc. of International Conference on Image Processing*, pp. 85-88, 2003.
- [6] P. D. O'Malley, M. C. Nechyba and A. A. Arroyo, "Human activity tracking for wide-area surveillance," in *Proc. of Florida Conference on Recent Advances in Robotics*, 2002.
- [7] O. Javed, Z. Rasheed, O. Alatas and M. Shah, "KNIGHTM: A real-time surveillance system for multiple overlapping and non-overlapping cameras," in *Proc. of the fourth International Conference on Multimedia and Expo*, 2003.
- [8] S. Khan and M. Shah, "Consistent labeling of tracked objects in multiple cameras with overlapping fields of view," *IEEE trans. on PAMI*, 25(10):1355-1360, 2003.
- [9] I. Haritaolu, D. Harwood and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Trans. on PAMI*, 22(8): 809-830, 2000.
- [10] S. J. McKenna, S. Jabri, Z. Duric and H. Wechsler, "Tracking interacting people," in *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp.348-353, 2000.
- [11] J. K. Aggarwal, and Q. Cai, "Human motion analysis: A review," *Computer vision and image understanding*, 73(3):428-440, 1999.
- [12] D. Gavrilu, "The visual analysis of human movement: A survey," *Computer vision and image understanding*, 73(1):82-98, 1999.
- [13] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," *IEEE trans. on PAMI*, 25(7): 918-923, 2003.
- [14] Y. Matsushita, K. Nishino, K. Ikeuchi and M. Sakauchi, "Illumination normalization with time-dependent intrinsic images for video surveillance," *IEEE trans. on PAMI*, 26(10):1336-1347, 2004.
- [15] K. M. Lee, and W. N. Street, "Model-based detection, segmentation and classification using on-line shape learning," *Machine vision and application*, 13(4):222-233, 2003.
- [16] K. M. Lee, and Y. M. Lee, "Tracking multi-person robust to illumination changes and occlusions," in *Proc. of International of Conference on Artificial reality and Telexistence*, pp. 429-432, 2004.
- [17] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE trans. on Multimedia*, 1(3):264-276, 1999.



이 경 미

1993년 덕성여대 전산학과(학사). 1996년 연세대 전산학과(석사). 2001년 미국 Univ. of Iowa 전산학과(박사). 현재 덕성여대 컴퓨터과학부 조교수. 관심분야는 멀티미디어처리 및 검색, 컴퓨터시각, 패턴인식, 기계학습



이 윤 미

2004년 덕성여자대학교 인터넷정보공학과 학사졸업. 2004년~현재 덕성여자대학교 전산및정보통신 대학원 석사과정. 관심분야는 영상처리, 컴퓨터비전