휴머노이드 로봇을 위한 사람 검출, 추적 및 실루엣 추출 시스템

정회원 곽 수 영*, 변 혜 란*

Human Tracking and Body Silhouette Extraction System for Humanoid Robot

Sooyeong Kwak*, Hyeran Byun* ***Regular Members***

요 약

본 논문은 스테레오 카메라가 이동하는 환경에서 카메라 움직임을 보정하여 새로운 다수의 사람을 검출하는 방법과 검출된 사람을 추적하고, 실루엣을 추출하는 통합된 시스템을 제안한다. 제안하는 시스템은 사람 검출, 추적, 실루엣 추출 3가지 모듈로 구성되어 있으며 3가지 모듈은 카메라가 이동하는 환경을 고려한 것이다. 사람 검출 모듈에서는 카메라 움직임(egomotion) 보정을 이용한 움직이는 영역 추출 결과와 스테레오 정보를 결합하여 움직이는 객체를 검출하였으며, 추적모듈은 변위 정보가 가중된 히스토그램 알고리즘으로 검출된 객체를 추적한다. 실루엣을 추출하는 모듈은 트라이맵(trimap)을 이용하여 사람의 실루엣 부분을 대략적으로 추정하는 단계와 그래프 컷(graph cut)을 적용하여 정교하게 실루엣 추출하는 단계로 이루어져 있다. 본 논문에서 제안하는 방법을 실내 환경에서 팬-틸트(pan-tilt) 스테레오 카메라로 획득한 실험데이터를 대상으로 실험한 결과 다수의 사람의 검출 및 추적, 정교한 실루엣 추출이 가능한 것을 확인하였다. 본 논문의 실루엣 추출결과는 제스처 인식이나 걸음걸이 인식 등의 다양한 분야에도 적용가능하다.

Key Words : Human Detection, Human Tracking, Silhouette Extraction, Camera Ego-Motion Compensation, Mean-Shift, Graph Cut, Disparity Map

ABSTRACT

In this paper, we propose a new integrated computer vision system designed to track multiple human beings and extract their silhouette with an active stereo camera. The proposed system consists of three modules: detection, tracking and silhouette extraction. Detection was performed by camera ego-motion compensation and disparity segmentation. For tracking, we present an efficient mean shift based tracking method in which the tracking objects are characterized as disparity weighted color histograms. The silhouette was obtained by two-step segmentation. A trimap is estimated in advance and then this was effectively incorporated into the graph cut framework for fine segmentation. The proposed system was evaluated with respect to ground truth data and it was shown to detect and track multiple people very well and also produce high quality silhouettes. The proposed system can assist in gesture and gait recognition in field of Human-Robot Interaction (HRI)

※ 본 연구는 학술진흥재단 여성과학자 (KRF-2007-204-D00038), 지식경제부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구 (IITA-2009-(C1090-0902-0046)) 부분 지원으로 수행되었음

* 연세대학교 컴퓨터과학과(ksy2177@yonsei.ac.kr, hrbyun@yonsei.ac.kr)

논문번호 : KICS2008-11-512, 접수일자 : 2008년 11월 18일, 최종논문접수일자 : 2009년 6월 8일

Ⅰ. 서 론

다수의 사람을 검출하고 추적하는 기술은 이동로봇, 감시 시스템, 제스처 인식 시스템, 이벤트 인식 시스템 등 많은 응용 분야에서 연구되고 있다. 현재의 비전 기술을 이용하여 실시간으로 객체를 검출하고, 추적하는 일은 어려운 일임에도 불구하고 컴퓨터 성능의 향상과 영상처리 기법의 발전으로 인하여 현재 연구가 활발히 진행되고 있다. 본 논문에서는 이동로봇 환경에서 다수의 사람을 검출 및 추적하고, 사람의 실루엣을 추출하는 방법을 제안한다.

사람을 검출하고 추적하는 방법은 카메라 환경에 따라 고정 카메라 환경에서의 방법론과 이동 카메라 환경에서의 방법론으로 구분 할 수 있다. 전자의 경우에는 대부분의 비디오 감시 시스템과 같이 고정된 배경으로부터 객체를 분리하여 검출된 객체가 사람인지 움직이는 객체인지를 판단하는 방법이 가장 많이 사용되고 있다. W4시스템에서는[1] 각 화소(pixel)의 최소/최대 밝기값과 표준편차를 이용하여 고정된 배경을 일정 프레임 동안 학습시켜 화소 단위로 배경을 모델링하여 모델링된 배경을 이용하여 객체를 검출하고, 실루엣을 추출하여 추적하는 방법을 제안하였다. 그러나 위의 방법들은 카메라가 이동하는 로봇 환경에서는 대부분 고정된 배경을 획득할 수 없기 때문에 이동로봇 환경에는 적합하지 않다. 이동 카메라 환경에서 움직이는 객체를 검출하고 추적하는 방법으로 Yilmaz 등은[2] 인위적으로 주어진 초기 윤곽(contour) 정보를 이용하여 객체의 윤곽을 추적하는 레벨 셋(level set) 방법을 제안하였다. 하지만 레벨 셋 방법은 정확한 초기값을 주어야 한다는 단점이 있으며, 계산양이 많아 실시간 로봇 시스템에 적용하기 어려운 문제가 있다.

또한 Davis 등은[3] 이동 카메라 환경에서 컨덴세이션(condensation) 방법을 이용하여 보행자를 추적하는 방법을 제안하였다. 컨덴세이션 방법은 보행자라는 한정된 모양만을 고려하였기 때문에 사람이 다양한 제스처를 취할 때에는 실루엣 추출이 어려운 단점이 있다.

사람 실루엣 추출은 객체 분할(object segmentation)의 한 분야로써 현재까지 오랜 시간동안 활발하게 진행되어 왔으나 정확하고, 효율적인 객체 분할은 여전히 해결 되지 않은 문제로 남아 있다. 과거의 연구 중에 가장 활발하게 진행되어 온 연구로써 고정 카메라(static camera) 환경에서 배경 모델링(background modeling)을 통하여 객체를 분할하는 방법이 있다[4]. 배경 모델링 방법은 고정된 카메라 환경만을 고려하였기 때문에 이동 로봇과 같이 동적 카메라 환경에는 적합하지 않고, 조명 변화에 매우 민감하다는 단점이 있다. 동적 카메라 환경에서의 객체 분할 방법으로는 Yu[5] 등이 제안한 그래프 분할 방법(graph partitioning method)이 있다. 제안된 그래프 분할 방법은 인간의 다양한 포즈 변화를 수용할 수 없고, 처리 속도가 매우 느리기 때문에 실시간 응용에도 적합하지 않다. 이러한 단점들을 보완하기 위해서 최근에는 객체의 우도(likelihood)와 영상의 화소간의 대비(contrast)에 기반한 객체 분할 방법들이 제안되고 있다. 그 중 효율적인 방법에는 그래프 컷 방법[6, 7]과 GMRF (Gaussian Markov Random Field)[8] 방법이 있다. 이 두 방법은 정지 영상에 대해서 사용자가 지정한 객체와 배경 영역에 대한 색과 대비 정보를 이용하기 때문에 사용자와의 상호작용(interaction)이 허용되지 않는 로봇 비전 분야에서 처리가 불가능한 단점을 가지고 있다. 본 논문에서는 위와 같은 고정 카메라 환경과 실시간 응용 등의 문제점을 해결하기 위하여 로봇 시스템에 적합한 이동 카메라 환경에서 다수의 사람을 검출, 추적하고 다양한 자세에 대해 정확한 실루엣을 추출하는 방법을 제안한다.

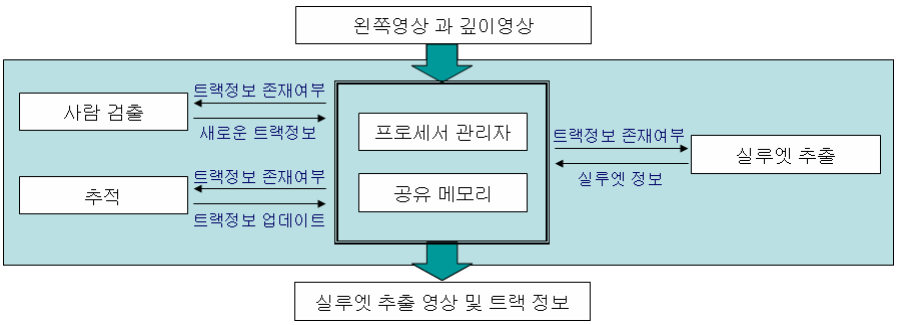


그림 1. 시스템 흐름도

Ⅱ. 시스템 개요

본 논문에서 제안하는 방법은 사람 검출, 추적, 실루엣 추출 그림 1에서 보는바와 같이 크게 3가지 모듈로 나누어져 있다. 3개의 모듈은 공유메모리(shared memory)를 이용하여 각 모듈의 결과물을 공유하도록 설계하였다. 사람 검출 모듈에서는 스테레오 카메라로 영상이 입력되면 공유메모리에 사람의 추적 결과인 트랙(track) 정보가 있는지 확인한 후, 트랙 정보가 존재하지 않는 경우 독립적인 움직임 추출 방법을 이용하여 사람을 검출하는 모듈이 실행된다. 검출 모듈에서는 새롭게 등장하는 사람만을 검출하는 역할을 담당하며, 검출된 사람 수, 사람의 위치 등의 정보를 공유메모리에 저장한다. 사람의 위치 정보는 외곽사각형(bounding box)로 표현되어 진다. 새로운 사람이 검출되면 추적 모듈에서는 공유메모리에 존재하는 외곽사각형을 입력받아 추적을 시작한다. 추적 모듈에서는 매 프레임 추적되는 사람의 위치 정보(외곽사각형 정보)를 공유메모리에 갱신시켜준다. 추적 모듈을 거치고 나면 실루엣 추출 모듈이 실행된다. 이 모듈에서도 공유메모리에 존재하는 트랙 정보를 확인한 다음 추적되고 있는 사람이 존재하게 되면 추적에 의해 입력된 외곽사각형과 스테레오 비디오의 좌측영상, 변위지도(disparity map)를 입력받아 외곽사각형 내부영역에서만 실루엣을 추출하는 방법을 실행하게 된다. 사람 검출 및 추적 모듈은 사람의 수에 제한이 없이 화면에 꽉 차지 않을 만큼의(카메라 높이에 따라 인원수는 달라짐) 사람을 검출, 추적 할 수 있지만. 실루엣 추출 모듈은 프로세서 관리자가 명령을 내려 결정된 한 사람에 대해서만 정교한 실루엣을 추출하도록 시스템을 설계하였다.

Ⅲ. 사람 검출

본 논문에서는 이동 카메라 환경에서 사람을 검출하기 때문에 카메라 움직임을 보정하여 독립적인 사람의 움직임만을 추출하여 사람을 검출하는 방법을 제안한다. 먼저, 카메라 움직임을 보정하기 위하여 연속적인 두 영상의 대응점을 찾아야 한다. 연속적인 두 영상이 주어졌을 때 이전 영상에서 해리스 코너 추출(Harris Corner Detection) 방법을[9] 이용하여 특징점을 추출하고, 현재 영상에서 KLT(Kanade-Lucas-Tomasi Feature Tracker) 방법을 이용하여 이전 영상에서 추출된 특징점과 대응되는 대응점을 찾는다. 연속적인 두 영상의 대응점이 결정되면, 카메라의 움직임을 찾기 위하여 변환 모델(transformation model)을 사용한다. 본 논문에서는 변환 방법 중 선형(linear)인 어파인 모델(affine model)을 사용하지 않고 비선형(nonliner)인 바이리니어 모델(bilinear model)을 사용하였다. 선형 모델은 로봇의 측면 이동과 회전은 보정할 수 있지만 카메라의 전진/후진의 움직임을 보정하지 못한다는 단점이 있다. 이에 비해, 비선형 모델은 카메라 전/후진의 문제점 까지도 보정해주는 장점이 있기 때문에 비선형 모델을 선택하였으며, 수식 (1)과 같은 바이리니어 모델은 다른 비선형 방법보다 연산이 간단하여 계산 속도가 빠르다는 장점이 있어 실시간 응용 분야에 적합하다고 판단하여 이를 변형 모델로 선정하였다. 수식 (1)에서 는 시간일 때의 추출된 특징점이며 는 의 대응점으로 추출된 특징점을 나타낸다.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a\ b\ c\ d \\ e\ f\ g\ h \end{pmatrix} \begin{pmatrix} x \\ y \\ xy \\ 1 \end{pmatrix} \tag{1}$$

카메라 움직임을 보정 할 때 고려해야 할 문제는 추출된 특징점들이 카메라의 움직임인지 이동하는 사람의 움직임인지를 판단하여야 한다. 사람의 영역에서 추출된 특징점은 카메라 움직임을 보정할 때 부정확한 보정 결과를 유발하는 원인이 됨으로 제거하여야 한다. 본 논문에서는 카메라 움직임을 가지는 특징점을 인라이어(inlier), 움직이는 사람에서 추출된 특징점을 아웃라이어(outlier)라고 정의하고 카메라 움직임을 보정하는 전 단계로서 수식 (2)의 함수를 이용하여 아웃라이어를 제거한다. 즉, 시간에 추출된 특징점을 바이리니어 모델을 사용하여 변환한 결과와 대응점의 차를 함수 라고 정의하였으며 의 값이 임계값 보다 작으면 대응점을 인라이어로 분리()하고, 그렇지 않으면 아웃라이어로 분리()하여 배경에서만 나타나는 대응점만을 분리해 낸다. 실험적으로 임계값 은 2로 결정하였다.

$$F = \left| \begin{pmatrix} x' \\ y' \end{pmatrix} - \begin{pmatrix} a\ b\ c\ d \\ e\ f\ g\ h \end{pmatrix} \begin{pmatrix} x \\ y \\ xy \\ 1 \end{pmatrix} \right| \tag{2}$$

$$\begin{cases} (x,y),(x'y') \in S_{Inliers} & \text{if } F < \epsilon \\ (x,y),(x'y') \in S_{Outliers} & otherwise \end{cases}$$

수식 (2)를 이용하여 아웃라이어와 인라이어를 정확히 분리한 다음, 인라이어만을 이용하여 정확한 변환 모델 계수 a, b, c, d, e, f, g, h를 계산하고 카메라 움직임을 보정한다. 그림 2의 (a), (b)는 수식 (2)을

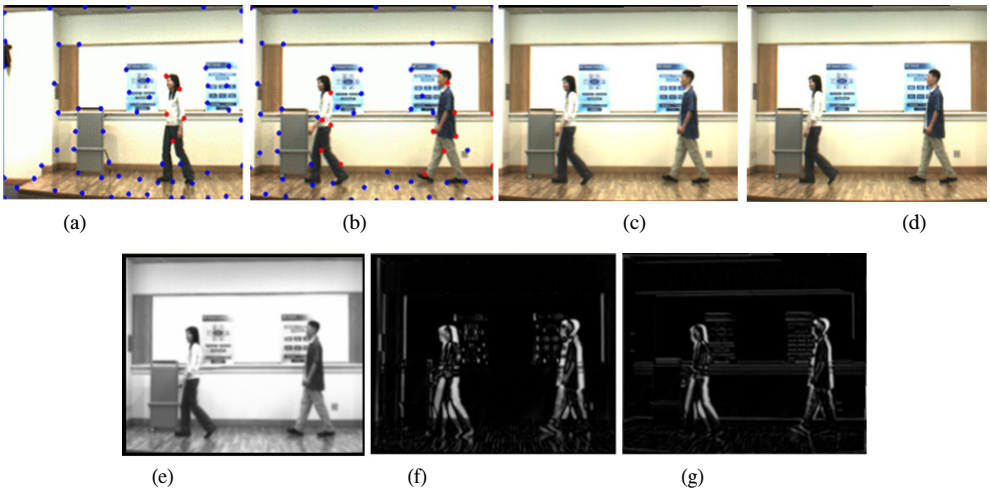


그림 2. (a)(b) 인라이어와 아웃라이어 추출 결과 (동그라미: 인라이어, 사각형: 아웃라이어) (c) t시간때의 영상 (d)t+1시간때의 영상 (e)t+1시간때의 영상을 t시간때의 영상으로 보정한 보정 영상 (f) 카메라 움직임이 보정되지 않은 두 영상의 차 (g) 카메라 움직임이 보정된 두 영상의 차

이용하여 인라이어와 아웃라이어를 추출한 영상이다.

카메라 움직임이 보정되면 보정된 영상과 현재 영상의 차를 이용하여 사람의 대략적인 위치를 추정할 수 있다. 보정된 영상과 현재 영상과의 차 영상을 구하면 사람의 움직임만을 정확하게 추출 할 수 있다. 그림 2-(f)와 그림 2-(g)는 카메라 움직임을 보정하지 않은 영상과 보정한 영상과의 차 영상을 나타낸 것이다. 카메라 움직임을 보정하지 않은 영상의 경우 그림 2-(f)에서 보듯이 카메라의 움직임으로 인해 나타나는 부분(배경의 움직임 부분)과 사람이 이동하면서 발생한 움직임 부분 두 영역이 명확하게 드러나지만 카메라의 움직임을 보정한 영상의 경우는 카메라가 이동하면서 나타난 칠판의 움직임 부분이 나타나지 않은 것을 볼 수 있다. 하지만, 카메라 움직임 보정하는 단계에서 영상의 좌표계를 변환하다 보면 미세한 오차들이 발생할 수 있기 때문에 카메라 움직임이 보정된 영상에서도 약간의 배경 흔들림은 존재할 수 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 카메라 움직임 보정으로 인해 발생한 오차보다도 사람의 움직임이 두드러진다는 특징을 이용하여 차영상에서 일정이상의 값을 이용하여 필터링 함으로써 대략적인 사람의 움직임의 위치를 추정하도록 하였다.

카메라 보정을 통한 이전프레임과 현재 프레임의 차영상으로 사람의 대략적인 위치는 파악이 가능하나 사람의 크기를 정확히 추정하기는 어렵다. 이전 프레임과 다음 프레임과의 차는 아주 미세하기 때문에 사람의 움직임이 다리에서만 나타나기도 하고, 팔에서만 나타나기도 한다. 이러한 문제점을 해결하기 위해서 본 논문에서는 카메라 보정을 통해 구해진 프레임 차(그림 3-(a))와 변위 지도 정보(그림 3-(b))를 결합하여 최종적으로 사람 영역을 검출하였다. 본 논문에서는 스테레오 카메라를 사용하였기 때문에 변위지도를 획득 할 수 있었으며 실험에 사용된 카메라의 API를 이용하여 변위지도를 획득하였다. 사람의 움직임이 발생한 위치정보를 그림 3-(c)에서 보듯이 차영상을 통해 획득하고 이 영역에 변위지도 값을 계산하여 이와 유사한 값을 가지는 영역을 사람의 영역이라 판단하여 외곽사각형을 결정하게 된다. 그림 3은 제안된 사람

그림 3. 사람 검출 결과 (a)입력으로 사용된 왼쪽영상 (b) 변위지도 (c)카메라 움직임 보정을 통해 계산된 프레임 차 영상 (d) 움직임이 존재하는 영역의 영상 분할 결과 (e) 최종 움직이는 객체 검출 결과

검출단계를 단계별로 나타낸 그림이다.

Ⅳ. 사람 추적

본 논문에서는 위에서 검출된 사람 추적을 위하여 평균 이동 추적 방법에 기반한 변위 정보가 가중된 히스토그램(disparity weighted histogram) 기법을 제안한다. 다수의 사람 간에 발생하는 가려짐을 판단하고, 가려짐 후 성공적인 복원을 위해 각 단계마다 후보 객체의 히스토그램 유사도(similarity)와 각 객체간의 거리 정보를 유지한다. 평균 이동 방법에 기반한 컬러 추적 방법은 국지적 최적의 후보 객체를 찾기 위해 확률 밀도 함수의 기울기를 이용하는 간결하고 강인한 방법으로 목표 객체와 후보 객체의 표현, 유사도의 측정, 평균 이동 방법을 이용한 최적화 등 3 단계로 나누어 질 수 있다.

- 목표 객체와 후보 객체의 표현 : 목표 객체와 후보 객체의 PDF 모델은 아래와 같이 정의 될 수 있다[10]. 수식 (3)에서 $\hat{q_u}$는 목표 객체, $\hat{p_u}$는 후보 객체를 나타낸다. k는 공간적 가중을 위한 프로파일 커널(kernel)이며, h는 목표 객체 모델의 대역폭(bandwidth)에 비례한 커널 대역폭, C_q와 C_p 는 정규화 요소들이며, δ는 크로네커 델타 함수(kronecker delta function), u는 히스토그램 빈의 색인이다.

$$\begin{aligned}\hat{q_u} &= C_q\sum_{i=1}^{n}k\left(\|x_i\|^2\right)\delta[b(x_i)-u],\\ \hat{p_u} &= C_q\sum_{i=1}^{n_h}k\left(\|\frac{y-x_i}{h}\|^2\right)\delta[b(x_i)-u]\end{aligned} \tag{3}$$

- 유사도 측정 : Bhattacharyya 계수를 이용하여 두 히스토그램의 유사도를 수식 (4)와 같이 측정할 수 있다. 여기서 테일러(Talyor)전개를 이용하면 수식 (5)는 아래와 같이 정리될 수 있다. 수식 (5)에서 m은 히스토그램 빈의 수, n_h는 공간 커널 속에 포함되는 화소의 수이며, y_0는 이전 객체의 위치를 나타낸다.

$$\hat{\rho} = \sum_{u=1}^{m}\sqrt{\hat{p_u}(y)\hat{q_u}} \tag{4}$$

$$\begin{aligned}\hat{\rho} &\approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p_u}(y_0)\hat{q_u}}+\frac{C_p}{2}\sum_{i=1}^{n_h}w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\\ w_i &= \sum_{u=1}^{m}\delta[b(x_i)-u]\sqrt{\frac{\hat{q_u}}{\hat{p_u}(y_0)}}\end{aligned} \tag{5}$$

- Mean shift를 이용한 최적화 : 객체의 추적 문제는 유사도 평면상에서의 최대화 문제로 생각될 수 있고, 이 최대화 문제는 평균 이동 방법에 의해 효율적으로 계산될 수 있다. 이전 위치에 기반한 새로운 위치 좌표는 수식 (6)과 같이 계산된다. 여기서, $g(x) = -k'(x)$이다.

$$\hat{y_1} = \frac{\sum_{i=1}^{n_h}x_i w_i g\left(\left\|\frac{\hat{y_0}-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h}w_i g\left(\left\|\frac{\hat{y_0}-x_i}{h}\right\|^2\right)} \tag{6}$$

4.1 변위정보가 가중된 히스토그램을 이용한 사람 추적

평균 이동(meanshift) 추적 방법은 유사한 컬러 분포를 가진 객체에 대해서는 효율적으로 분리해 내기 힘들다는 단점을 가지고 있다. 이러한 단점을 극복하기 위해서 본 논문에서는 변위 정보가 가중된 히스토그램 기법을 제안한다. 사람 검출 모듈에 의해 검출된 영역의 사람 영역에 대한 변위 정보의 분포는 배경이나 다른 객체와 비교할 때 분산이 작다고 가정 할 수 있다. 카메라의 움직임에 의해 배경이 동적으로 변하더라도 카메라의 초당 프레임 수가 크다면 변위값의 프레임 당 차이도 작다.

여기서, 우리는 변위 정보와 컬러 정보가 결합된 간결한 변위 정보가 가중된 히스토그램 기법을 제안한다. 목표 영역의 정규화된 히스토그램 표현과 히스토그램의 최대 확률 값을 각각 $\left\{\hat{v_v}\right\}_{v=1\cdots k}$ 와 v_M이라고 하자. 만약 객체 영역의 크기를 충분히 제한되게 선택할 수 있다면, v_M의 확률 값을 가지는 빈 M에 포함되는 화소들은 객체에 해당하는 화소일 가능성이 가장 높다고 할 수 있다. 여기서 $\{W_v\}_{v=1\cdots k}$ 을 변위 정보 히스토그램 빈을 위한 가중치라 하면, 수식 (7)과 같이 정의 될 수 있다. 여기서, v는 히스토그램의 빈 색인이다. 이러한 가중치를 이용해서 새로운 목표 객체($\widehat{Q}_u$)와 후보 객체($\widehat{P}_u$)를 수식 (8)과 같이 표현할 수 있다.

$$\left\{W_v = 1-\frac{|M-v|}{M}\right\}_{v=1\cdots k} \tag{7}$$

$$\begin{aligned}\widehat{Q}_u &= C_{q,w}\sum_{i=1}^{n}W_{i,v}k\left(\|x_i\|^2\right)\delta[b(x_i)-u]\\ \widehat{P}_u &= C_{q,w}\sum_{i=1}^{n_h}W_{i,v}k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta[b(x_i)-u]\end{aligned} \tag{8}$$

(a) (b) (c)

그림 4. 제안된 추적 방법의 단계별 결과: (a) 객체 정보 획득 단계 (b) 공간 커널과 변위 정보의 혼합 단계 (c) 평균 이동 방법 기반 최적화 단계

수식 (8)에서 $C_{q,w}$와 $C_{q,w}$는 정규화 요소이며, $W_{i,v}$는 목표 객체 영역 속에 포함되는 화소를 위한 변위 가중치이다. 수식 (8)에 대한 Bhattacharyya 계수는 수식 (9)와 같이 유도될 수 있다.

$$\hat{\rho}_u \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{P}_u(y_0)\hat{Q}_u} + \frac{C_{p,w}}{2}\sum_{i=1}^{n_h} w_{i,D}\, k\left(\left\| \frac{y-x_i}{h} \right\|^2\right) \quad (9)$$

$$w_{i,D} = \sum_{u=1}^{m} W_{i,v}\,\delta[b(x_i)-u]\sqrt{\frac{\hat{Q}_u}{\hat{P}_u(y_0)}}$$

그러므로 유사도 측정의 최대값을 얻기 위해서 $w_{i,D}$로 가중된 데이터의 커널 밀도 추정을 의미하는 $\hat{\rho}_u$의 두 번째 항이 최대화 되어야 하고, 이전 객체 위치에 기반한 새로운 객체 위치는 평균 이동 방법으로 유도될 수 있다. 프로파일 커널을 Epanechnikov 프로파일[11]로 선택하면, 프로파일 커널의 미분 커널인 $g(x)$가 상수값을 가지게 되어 새로운 객체 위치는 식 (10)과 같이 보다 간단한 형태가 된다. 그림 4는 제안된 추적 방법의 단계별 결과를 보여준다.

$$\hat{y}_1 = \sum_{i=1}^{n_h} w_{i,D}\, x_i \Big/ \sum_{i=1}^{n_h} w_{i,D} \quad (10)$$

4.2 가려짐 판단과 복구

객체의 공간적 정보와 히스토그램 유사도 정보를 이용하면 가려짐을 판단하고 복구할 수 있다. 만약, 객체간의 거리가 어떤 경계값에 가까워지면, 각 객체의 히스토그램의 유사도는 가려지는 객체의 히스토그램의 유사도가 현저히 변하는 특성을 가지게 된다. 이를 이용하면 가려지는 객체와 가려짐을 발생시키는 객체를 찾을 수 있다. 가려짐 후의 복원을 위해서, 본 논문에서는 가려짐을 발생시키는 객체의 주변에서 탐색을 수행하였다. 그림 5는 가려짐 판단과 복원 결과에 대한 결과를 보여준다.

또한, 본 논문의 경우 컬러와 변위정보를 결합하여 특징정보로 사용하고 있기 때문에 그림 5의 실험에서

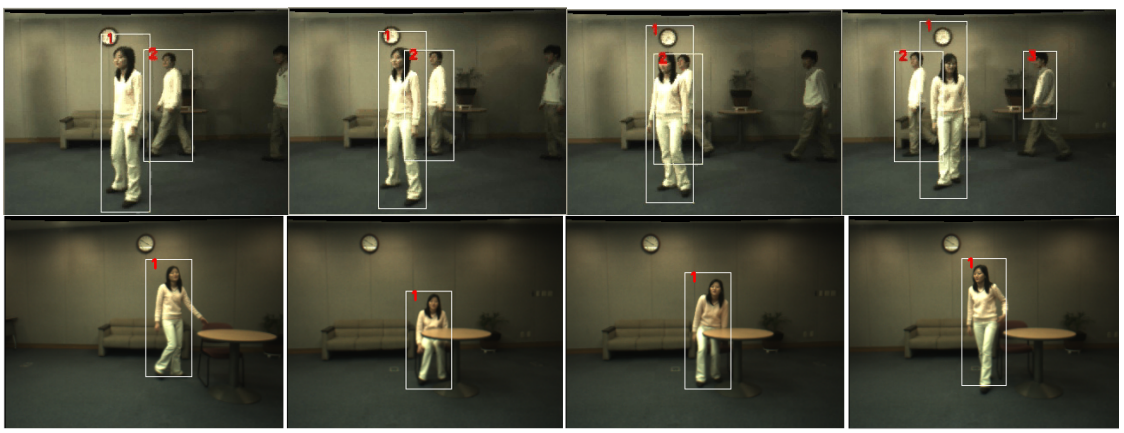


그림 5. (위)동일한 색상 분포를 가진 두 물체에 가려짐이 발생한 경우의 추적 결과 (아래)사람에 신체 일부분이 다른 물체에 의해 가려진 경우의 추적 결과

보듯이 동일한 색상의 옷을 입은 두 사람이 가려지더라도 깊이감이 다르기 때문에 일시적인 가려짐에도 계속적인 추적이 가능하며 사람의 신체 일부분이 다른 물체에 의해 가려지더라도 사람의 신체 일부분의 색 분포와 변위지도를 통해 추적이 가능하다는 것을 볼 수 있다.

Ⅴ. 실루엣 추출

본 논문에서는 추적되는 사람의 정교한 실루엣을 추출하기 위해서 객체 영역, 배경 영역, 비결정 영역 등의 트라이맵 추정을 통한 그래프 컷을 이용하였다. 또한, 처리 속도의 효율성을 위해 추적에서 얻어진 외곽사각형 정보를 이용하여 외곽사각형 내의 화소들만을 고려하였다. 본 논문에서 제안하는 이동 카메라 환경에서의 실루엣 추출의 문제는 객체 분할(object segmentation)문제로 볼 수 있다. 이동 카메라 환경에서의 객체 분할은 기존의 많은 연구가 이루어진 배경 모델링을 이용하는 방법과는 달리 고정된 배경을 얻을 수 없고, 저해상도, 그림자, 부정확한 변위 정보, 컬러 변화 등의 다양한 문제로 인하여 매우 어려운 문제이다. 본 논문에서는 이러한 문제를 해결하여 정교한 실루엣을 추출하기 위해서 컬러, 대비, 변위 지도 등을 이용하여 트라이맵을 정확하게 추정함으로써 그래프 컷의 성능을 향상시키는 방법을 제안한다.

본 논문에서 제안하는 그래프 컷을 이용한 실루엣 추출 방법을 살펴보면 다음과 같다. 우선 영상 내의 화소의 집합을 P라 하고, $z=(z_1,\cdots,z_{|P|})$를 주어진 영상이라 할 때, z_n는 k번째 화소의 RGB 컬러 벡터이다. 그리고 집합 P의 각 화소 p에 대한 0과 1의 이진 레이블(binary label)을 f_p라 하자. 여기서 1은 "객체"를, 0은 "배경"을 의미한다. 이때, $f=(f_1,f_2,\cdots,f_{|P|})$는 객체 분할을 나타내게 된다. 이때, 위에서 설명한 다양한 문제를 해결하기 위해서 본 논문에서는 각 화소에 "비결정 영역"에 해당하는 -1 레이블을 추가한 {0, 1, -1}의 3 영역에 대한 트라이맵을 추정하여 최종적으로 비결정 영역에 대해서만 그래프 컷을 적용하여 빠르고 정교한 실루엣을 추출한다.

이동 카메라 환경에서의 실루엣 추출의 문제는 객체 분할(object segmentation)문제로 볼 수 있다. 이동 카메라 환경에서의 객체 분할은 기존의 많은 연구가 이루어진 배경 모델링을 이용하는 방법과는 달리 고정된 배경을 얻을 수 없고, 저해상도, 그림자, 부정확한 변위 정보, 컬러 변화 등의 다양한 문제로 인하여 매우 어려운 문제이다. 본 논문에서는 이러한 문제를 해결하여 정교한 실루엣을 추출하기 위해서 컬러, 대비, 변위 지도 등을 이용하여 트라이맵을 정확하게 추정함으로써 그래프 컷의 성능을 향상시키는 방법을 제안한다.

본 논문에서 제안하는 그래프 컷을 이용한 실루엣 추출 방법을 살펴보면 다음과 같다. 우선 영상 내의 화소의 집합을 P라 하고, $z=(z_1,\cdots,z_{|P|})$를 주어진 영상이라 할 때, z_n는 k번째 화소의 RGB 컬러 벡터이다. 그리고 집합 P의 각 화소 p에 대한 0과 1의 이진 레이블(binary label)을 f_p라 하자. 여기서 1은 "객체"를, 0은 "배경"을 의미한다. 이때, $f=(f_1,f_2,\cdots,f_{|P|})$는 객체 분할을 나타내게 된다. 이때, 위에서 설명한 다양한 문제를 해결하기 위해서 본 논문에서는 각 화소에 "비결정 영역"에 해당하는 -1 레이블을 추가한 {0, 1, -1}의 3 영역에 대한 트라이맵을 추정하여 최종적으로 비결정 영역에 대해서만 그래프 컷을 적용하여 빠르고 정교한 실루엣을 추출한다.

5.1 트라이맵 추정

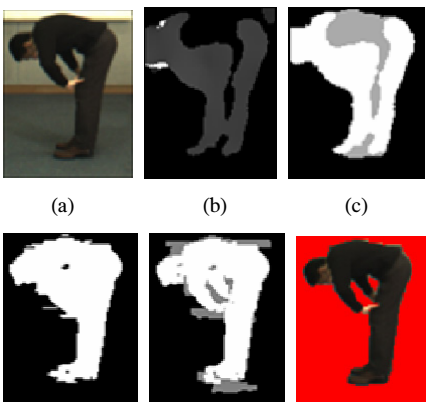
본 논문에서 제안하는 트라이맵 추정 방법은 Boykov 등에[6] 의해 제안된 사용자 상호작용(user-interaction)을 통한 객체 분할 방법을 보완하여 객체, 배경, 비결정 영역을 결정하였다. Boykov 등은 사용자가 직접 마우스를 이용하여 영상 내에 분할을 원하는 객체와 배경을 구분하여 적당한 화소(seed)들을 지정하도록 가정하였으나, 본 논문에서는 사용자에 의한 입력 없이 자동으로 트라이맵을 추정하여 객체를 분할하는 방법을 제안하였다.

정확한 트라이맵을 추정하기 위해서 우선 입력된 왼쪽 컬러 영상을 평균 이동 분할(mean shift segmentation)[12] 방법을 적용하여 영상을 조각으로 분할하고, 분할된 조각들과 이전 프레임의 실루엣 결과 그리고 변위 지도를 이용하여 최종적으로 트라이맵을 추정하였다. 이때, t번째 프레임에서의 평균 이동 분할 방법에 의해 분할된 영상 조각의 집합을 $R^t=\{R_i^t\}$로 표현한다.

- 배경 영역(seed) 추정 : 본 논문에서는 배경 화소를 추정하기 위해서 일단 사람은 이전 프레임과 현재 프레임 사이에 d 화소 이상 움직이지 않는다고 가정하였다. 따라서 이전 프레임에서의 실루엣 추출 결과에 $d\times d$ 크기의 구조연산자를 이용하여 모폴로지의 불림연산(dilation)을 수행하여 불려진 실루엣의 바깥쪽 화소들을 배경 영역으로 가정하

였다. 또한, 영상 조각 R_i^t 중에서 현재 프레임의 외곽사각형의 테두리와 닿는 조각 역시 배경 영역으로 가정하였다. 영상 조각과 이전 프레임의 부풀려진 실루엣에서 추정된 영역을 OR 연산하여 최종적으로 배경 영역을 추정하였다.

- 객체 영역(seed) 추정 : 객체 영역을 O^t라 할 때 이를 추정하기 위해서 변위 지도와 이전 프레임의 실루엣 결과를 이용한다. 우선, $t-1$번째 프레임의 객체에 해당하는 변위값의 평균 m_D^{t-1}와 표준편차 s_D^{t-1}를 이용하여 후보 객체 화소들의 집합 $O_D^t = \{p \in P | m_D^{t-1} + K_D s_D^{t-1} < d_p < m_D^{t-1} + K_D s_D^{t-1}\}$를 정의한다. 이때, d_p는 p화소에서의 변위값을 나타낸다. 다음으로 후보 객체 화소 집합 O_D^t와 이전 실루엣 결과 S^{t-1}을 이용하여 화소 p에서의 객체 영역 우도 L_R을 $p \in O_D^t$ 이면 1, $p \in S^t - O_D^t$이면 w_s, 그리고 다른 경우에는 0으로 정의한다. 또한, 위에서 설명한 분할된 영상 조각들 $R \in R^t$ 중에서 $\sum_{p \in R} L_R(p)/n_R > w_s$ 조건을 만족하는 R을 역시 후보 객체 영역으로 추정한다. 이때, n_R은 영상 조각 R 내에 존재하는 화소의 개수를 나타낸다. 다음으로는 정확한 객체 영역을 추정하기 위해서 후보 객체 영역 중에서 확실한 배경 화소들을 제거해야 하는데 이를 위해서 R_s^t를 위에서 선택된 후보 객체 영역들의 집합이라고 할 때, 본 논문에서는 공간-색 확률(spatial-color probability)를 이용하여 배경 화소들을 제거한다. 이를 위해 위에서 추정된 배경 영역 B^t를 이용하여 화소의 위치 정보(2차원)과 RGB 컬러 정보(3차원)으로 구성된 5차원 히스토그램을 구성한다. 5차원으로 구성된 히스토그램을 이용하여 계산된 배경 확률을 P_B라 하고, 추정된 객체 영역을 $p \in O^t$라 할 때, $-\log P_B(p) > m_c^t + K_C \sigma_C^t$와 $p \in R^t$ 조건을 만족하면 최종적인 O^t로 선택한다. 그림 6은 (a)와 (b)로부터 추정된 트라이맵을 보여준다.



(a) (b) (c)

(d) (e) (f)

그림 6. 추정된 트라이맵을 이용한 단계별 실루엣 추출 결과: (a) 입력 영상, (b) 변위 지도, (c) 객체 우도 L_R (d) 영상 조각R^t 선택, (e) 트라이맵: 흰색(O^t), 검정색(B^t), 회색(U^t), (f) 제안된 방법의 실루엣 추출 결과

5.2 추정된 트라이맵을 이용한 그래프 컷

본 논문에서는 실루엣 추출을 위하여 최종적으로 수식 (11)과 같은 에너지 함수를 최소화하는 레이블의 집합 f를 찾는 것이라 할 수 있다.

$$E(f) = \sum_p D_p(f_p) + \sum_{\{p,q\} \in N} \delta(f_p, f_q) V_{p,q}(f_p, f_q) \quad (11)$$

수식 (11)에서 N은 표준 4-이웃 시스템을 나타내며, $\delta(f_p, f_q)$는 $f_p \neq f_q$일 경우 1이고, 나머지 경우에는 0으로 표현되는 델타 함수이다. $V_{p,q}$는 평활 항목으로 $V_{p,q} = \exp(-\| z_p - z_q \|^2/\beta)$로 정의되고, β는 상수값으로 모든 $\{p,q\} \in N$ 에 대해서 $2\|z_p - z_q\|^2$의 기대값(expectation)으로 결정된다. 데이터 항목 D_p는 레이블 f_p와 화소 p가 어울리는가의 정도를 나타낸다. 이를 계산하기 위해 이전 프레임의 분할 결과에 따라 해당 화소의 컬러값 z_p들과 가우시안 혼합 모델(GMM: Gaussian Mixture Model)을 이용하여 객체와 배경에 대한 컬러 확률 분포 $P(\cdot|1)$ 와 $P(\cdot|0)$을 구한다. 비결정 영역 화소들의 집합을 $U = P - (O \cup B)$로 정의할 때, D_p는 식 (12)와 같이 정의 된다.

$$D_p(f_p) = \begin{cases} -\log P(z_p|f_p) & \text{if } p \in U \\ (K-c) f_p + c & \text{if } p \in O \\ (c-K) f_p + K & \text{if } p \in B \end{cases} \quad (12)$$

여기서 $K = \max_{p,q \in N} V(f_p, f_q)$이고, c는 작은 상수로 실험에서 보통 1 또는 2를 할당하였다.

최종적으로 에너지 함수(수식 (11))를 최소화시키는 레이블의 집합 f는 대표적인 min-cut/max-flow 방법을 이용하여 구할 수 있다[13]. 하지만 위에서 추정된

트라이맵은 단지 특징들에 의한 우도에 의해 결정되기 때문에 몇몇 화소들에서 오류가 발생할 수 있다. 이때, 상수 c가 트라이맵으로 추정된 영역들에 대해서 다른 레이블을 가질 수 있도록 하여 추정에 의한 오류를 보완 할 수 있는 기능을 한다. 그림 6은 이러한 효과를 보여준다. 본 논문에서는 추정된 영역의 화소들에 대해서 데이터 항목 D_p는 컬러와 변위값에 전혀 영향을 받지 않고, 상수 K 또는 c에 의해서면 결정되기 때문에 트라이맵 추정 방식의 그래프 컷 방법은 조명 변화와 이동 카메라 환경에 강인한 방법론이라고 할 수 있다.

Ⅵ. 실험 및 결과

본 논문에서 제안한 시스템은 Windows XP에서 Visual C++ 6.0을 이용하여 구현하였으며, 펜티엄-IV의 CPU 3.0 GHz와 1GB RAM의 하드웨어에서 실험하였다. 실험데이터는 Videre사의 STH-MDCS2 본 논문에서 제안한 시스템은 Windows XP에서 Visual C++ 6.0을 이용하여 구현하였으며, 펜티엄-IV의 CPU 3.0 GHz와 1GB RAM의 하드웨어에서 실험하였다. 실험데이터는 Videre사의 STH-MDCS2 스테레오 카메라를 이용하여 실내 환경에서 촬영된 스테레오 비디오 CM2, CM3, JH3, KC1, KC3, SS1, SY3 등 7개의 데이터를 이용하였다. 각각의 비디오는 700 프레임의 길이를 가지도록 영상을 획득하였고 영상의 크기는 320x240이다. 실험환경은 로봇의 위치는 고정되어 있는 상태에서 좌우 머리를 이동하는 환경에서 실험하였기 때문에 카메라의 움직임은 좌우이동이 대부분이며 사람의 전신이 나타나는 거리에서 실험데이터가 촬영되었다. 실험환경에 사용된 카메라의 움직임 정도는 사람이 걷는 속도와 유사하다고 볼 수 있다. 실험데이터 KC1과 KC3는 카메라가 고정되어 있는 환경에서 1명의 사람이 등장하여 다양한 자세를 취하는 데이터이며 나머지 5개의 실험데이터 CM2, CM3, JH3, SS1, SY3 는 팬-틸트 스테레오 카메라환경에서 등2-3명의 사람이 등장하여 인사하기, 소파에 앉아 얘기하기, 음료수 가져다주기 등의 시나리오를 가진다.

또한 CM3, JH3는 다수의 사람이 등장하기 때문에 가려짐이 발생하기도 하며 SY3에서는 급격한 조명변화가 발생한다. 본 논문에서 실험한 환경은 초당 30 fps로 영상을 획득하고 사람검출, 추적, 실루엣을 추출하는 결과를 획득하는데 소요되는 속도는 9fps이다. 검출, 추적 및 실루엣 추출에 대한 결과는 표 1과 같으며 그림 7은 본 논문에서 제안한 시스템에 의해 추출된 결과들을 보여준다.

6.1 사람 검출 및 추적 결과 평가

본 논문에서 제안한 시스템에 대한 평가를 위해서 각각의 실험데이터에 대해서 매 10 프레임 마다 왼쪽 영상을 기준으로 추출된 참 영상(ground truth)을 이용하였다. 이때, 참 영상은 실험자가 직접 영상 편집 도구(Photoshop)를 이용하여 실루엣을 추출하였다. 실험 결과는 표 1과 같다. 표 1에서 사람 검출 방법을 평가하기 위한 기준으로 각각의 데이터마다 총 사람이 등장하는 횟수를 등장 횟수로 표현하였고, 총 사람이 검출된 횟수를 검출 횟수, 올바르게 검출된 횟수와 올바르지 않게 검출된 횟수를 True+D와 False+D로 나타내었다. 추적 방법에 대해서는 참 영상에서 추출된 외곽사각형 영역과 제안한 방법에서 추출된 외각사각형 영역을 비교하여 90퍼센트 이상 겹치게 되면 올바르게 추적되었다 판단하고 추적률은 매 프레임 사람의 등장 횟수와 올바르게 추적된 횟수의 비율을 나타낸다. 본 논문에서 제안한 방법의 추적률은 평균

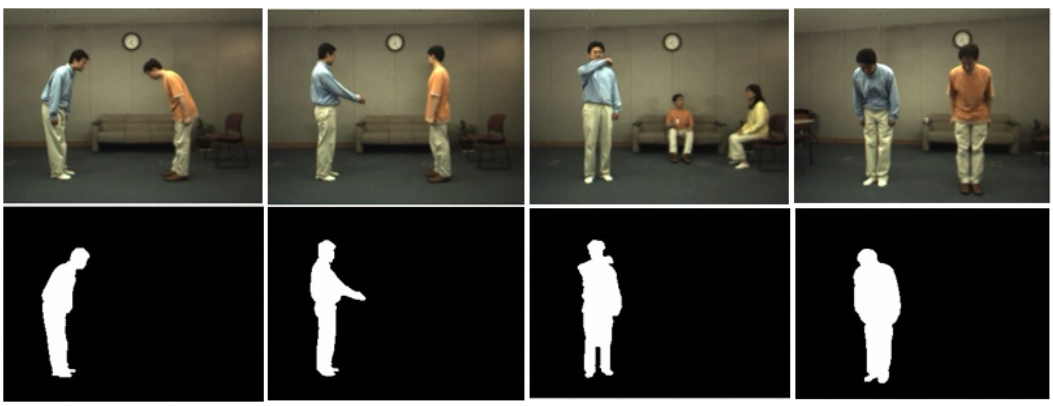


그림 7. 그래프 컷을 통한 최종 획득된 실루엣 추출 결과(중심되는 사람 1명의 실루엣 추출 결과)

표 1. 사람 검출 및 추적/ 실루엣 추출 실험 결과

		CM2	CM3	JH3	KC1	KC3	SS1	SY3
사람 검출 및 추적	등장횟수	4	6	9	1	10	1	7
	검출횟수	4	8	10	1	10	1	7
	True +D	4	6	9	1	10	1	7
	False +D	0	2	1	0	0	0	0
	추적률(%)	98.96	99.63	98.83	99.71	99.88	99.91	91.72
실루엣 추출	Error Rate(%)	1.06	2.55	1.25	1.36	1.26	0.61	0.42

98.31%로 대부분의 데이터에서 매우 우수한 성능을 보였다.

6.2 실루엣 추출 결과 평가

추정된 트라이맵을 이용한 그래프 컷 기반의 실루엣 추출 방법을 평가하기 위해서 기본적으로 7개의 데이터에서 매 10 프레임 마다 추출한 참 영상을 이용하였다. 참 영상은 객체, 배경 그리고 고려하지 않을 영역(don't care pixel)으로 구분하였다. 이때, 고려하지 않을 영역은 참 영상에서 객체로 레이블 된 영역의 외곽을 따라 ±1 화소 영역으로 설정하였다. 오류율은 고려하지 않을 영역은 무시하고 잘못 추출된 화소 개수 대비 전체 화소수로 계산하였다.

Ⅶ. 결 론

본 논문에서는 스테레오 이동 카메라 환경에서 사람 검출, 추적과 실루엣 추출 방법을 결합한 새로운 비전 시스템을 제안하였다. 본 논문에서 제안한 시스템은 제스처 인식, 이벤트 인식 등의 다양한 비전 관련 응용분야에 적용이 가능하다. 기존의 배경 모델링 기반의 객체 검출, 추적 및 실루엣 추출 방법은 구현이 쉽고, 간단하지만 이동 카메라 환경에서는 고정된 배경을 얻을 수 없기 때문에 적용이 불가능하다. 본 논문에서 제안한 시스템은 이동 카메라 환경에서도 히스토그램 기반의 빠른 추적이 가능하며 새롭게 제안한 트라이맵을 이용한 그래프 컷 기반의 실루엣 추출 방법은 아주 정교한 실루엣 추출이 가능한 방법을 제안하였다. 제안한 시스템은 실내 환경에서 획득된 다양한 종류의 스테레오 비디오 데이터로 실험한 결과, 안정적이면서 우수한 성능을 보임을 확인할 수 있었고 사람검출, 추적, 실루엣 추출의 모든 과정을 거쳐서 최종 결과를 얻는데 걸리는 시간은 초당 9프레임으로 측정되었다. 하지만 본 시스템을 실생활에 적용하기 위해서는 약간의 속도 향상이 필요로 하다. 이를 위해서 앞으로의 연구에서는 실루엣 추출 시 처리 속도의 60% 이상을 차지하는 평균 이동 분할 방법의 대안 및 효율적인 분할 방법을 강구함으로써 속도와 정확도를 동시에 향상시킬 수 있는 방법에 대해서 연구할 계획이다. 또한 로봇 비전에 적용하기 위해서는 팬-틸트 되는 카메라의 움직임 외에 로봇이 자율주행할때 처리할 수 있는 방법도 연구할 계획이다.

참 고 문 헌

〔1〕 I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-Time Surveillance of People and Their Activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No. 8, pp. 809-830, 2000.

〔2〕 A. Yilmaz, X. Li and M. Shah, "Object Contour Tracking Using Level Sets", Asian Conference on Computer Vision, 2004.

〔3〕 L. Davis, V. Philomin and R. Duraiswami "Tracking humans from a moving platform", The 15th International Conference on Pattern Recognition. Vol. 4, pp. 171-78, 2000.

〔4〕 C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", Proceeding IEEE International Conference on Computer Vision and Pattern Recognition, pp. 246-252, 1999.

〔5〕 S. Yu, R. Gross and J. Shi, "Concurrent object recognition and segmentation by graph partitioning", Proceeding Neural Information Processing Systems, pp. 1383-1390. 2002.

〔6〕 Y. Boykov, and M. Jolly, "Iterative graph cuts for optimal boundary and region segmentation of objects in N-D Images," Proceeding IEEE 8th International Conference on Computer

Vision, Canada, 2001.
〔7〕 Y. Li, J. Sun, C.-K. Tang and H.-Y. Shum, “Lazy Snapping”, Proceeding ACM SIGRAPH, Vol 23, No. 3, 2004.
〔8〕 A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. “Interactive image segmentation using an adaptive GMMRF model,” Proceeding European Conference on Computer Vision, pp. 428-442, 2004.
〔9〕 C. Harris and M. J. Stephens, “A combined corner and edge detector,” In Alvey Vision Conference, pp. 147 – 152, 1988.
〔10〕 D. Comaniciu, and V. Ramesh.: Kernel-Based Object Tracking. IEEE Transaction on Pattern Analysis and Machine Intelligence Vol. 25. pp. 564-577, 2003.
〔11〕 M. P. Wand and M. C. Jones.: Kernel Smoothing. Chapman & Hall. 1995
〔12〕 D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 603-619, 2002.
〔13〕 Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/ max-flow algorithms for energy minimization in vision, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 26, pp. 1124-1137, 2004.

곽 수 영 (Sooyeong Kwak) 정회원

2003년 2월 계명대학교 컴퓨터공학과 졸업
2005년 2월 연세대학교 컴퓨터과학과 석사
2005년 3월~현재 연세대학교 컴퓨터과학과 박사과정
<관심분야> 영상처리, 컴퓨터 비전 및 패턴 인식

변 혜 란 (Hyeran Byun) 정회원

1980년 2월 연세대학교 수학과 졸업
1987년 6월 Univ. of Illinois at Chicago Computer Science, M.S.
1993년 12월 Purdue Univ. Computer Science, Ph. D.
2004년 3월~현재 연세대학교 컴퓨터과학과 정교수
<관심분야> 패턴 인식, 컴퓨터 비전 시스템