

그래프 컷을 이용한 강인한 인체 실루엣 추출 (Robust Human Silhouette Extraction Using Graph Cuts)

안 정 호[†] 김 길 천[†] 변 혜 란^{††}

(Jung-Ho Ahn) (Kil Cheon Kim) (Hyeran Byun)

요약 본 논문에서는 실내 환경에서 동적 스테레오 카메라(active stereo camera)를 이용한 새로운 인체 실루엣 추출 방법을 제안한다. 제안한 알고리즘의 주된 응용분야는 이동 로봇 플랫폼에서의 인체 실루엣을 이용한 제스처 인식이다. 먼 거리에서 움직이는 객체를 분할(segmentation)하는 데에는 저해상도, 그림자, 스테레오 정합의 불확실성, 배경과 객체의 색 분포의 불안정성 등과 같은 다양한 문제를 내포한다. 우리는 먼저 이미지 분할 기법과 스테레오 정보를 이용하여 신뢰도 높은 객체와 배경 영역을 추정하였다. 이렇게 추정된 영역을 적절히 그래프 컷(graph cut)에 활용하는 방식을 고안함으로써 주변 환경의 변화에 강인한 인체 실루엣 추출을 가능하게 하였다. 제안한 방식은 실내에서 팬-틸트(pan-tilt) 스테레오 카메라로 획득된 비디오 데이터를 대상으로 실험하였으며, 색, 색과 스테레오, 색과 대비 정보를 기반으로 한 방법들과 비교 실험한 결과 정확도가 많이 향상된 것을 확인할 수 있었다.

키워드 : 실루엣 추출, 스테레오 비전, 그래프 컷, 객체 분할, 제스처 인식, 이동 로봇, 동적 스테레오 카메라

Abstract In this paper we propose a new robust method to extract accurate human silhouettes indoors with active stereo camera. A prime application is for gesture recognition of mobile robots. The segmentation of distant moving objects includes many problems such as low resolution, shadows, poor stereo matching information and instabilities of the object and background color distributions. There are many object segmentation methods based on color or stereo information but they alone are prone to failure. Here efficient color, stereo and image segmentation methods are fused to infer object and background areas of high confidence. Then the inferred areas are incorporated in graph cut to make human silhouette extraction robust and accurate. Some experimental results are presented with image sequences taken using pan-tilt stereo camera. Our proposed algorithms are evaluated with respect to ground truth data and proved to outperform some methods based on either color/stereo or color/contrast alone.

Key words : silhouette extraction, stereo vision, graph cut, object segmentation, gesture recognition, mobile robot, active stereo camera

1. 서론

본 논문에서는 스테레오 비디오로부터 실시간으로 인체 실루엣을 추출하는 방법을 제안한다. 제안하는 인체 실루엣 추출(human silhouette extraction) 알고리즘을 이용한 응용 분야로는 인체 실루엣을 이용하여 특징

(feature)을 추출하는 제스처(gesture) 및 걸음걸이(gait) 인식 등이 있다. 본 논문에서는 실내 환경에서의 동적 스테레오 카메라를 이용하여 획득된 비디오를 대상으로 하였다. 카메라 움직임과 배경 및 조명 변화가 있고 등장인물들이 카메라와의 3~5미터 정도의 거리에서 임의의 제스처와 포즈(pose)를 취하는 비디오 데이터를 대상으로 실험하였다.

객체 분할에 대한 연구는 현재까지 오랜 시간동안 활발하게 진행되어 왔으나 정확하고, 효율적인 객체 분할은 여전히 해결 되지 않은 문제로 남아 있다. 과거의 연구 중에 가장 활발하게 진행되어 온 연구로서 고정 카메라(static camera) 환경에서 배경 모델링(background modeling)을 통하여 객체를 분할하는 방법이 있다[1].

· 이 논문은 2005년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2005-204-D00041)

† 학생회원 : 연세대학교 컴퓨터과학과

jungho@cs.yonsei.ac.kr

kimkch@cs.yonsei.ac.kr

†† 종신회원 : 연세대학교 컴퓨터과학과 교수

hrbyun@cs.yonsei.ac.kr

논문접수 : 2006년 2월 1일

심사완료 : 2006년 11월 14일

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 Stella X[2] 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[3-5]과 GMMRF(Gaussian Markov Random Field)[6] 방법이 있다. 이 두 방법은 정지 영상에 대해서 사용자가 지정한 객체와 배경 영역에 대한 색과 대비 정보를 이용하기 때문에 사용자와의 인터랙션(interaction)이 허용되지 않는 로봇 비전 분야에서 처리가 불가능한 단점을 가지고 있다. 그래프 컷은 이후, 사용자 인터페이스가 없는 환경에서의 비디오 데이터에 대해 실시간으로 높은 성능을 보인 색, 대비, 스테레오 정보를 결합한 이중(bi-layer) 분할 방법으로 발전하였다[7]. 이는 고정된 스테레오 웹캠(webcam)을 이용하여 원격 회의(teleconferencing)를 위한 객체 분할을 응용분야로 삼았기에 사람이 카메라 가까이 위치함으로써 고해상도 영상에서 비교적 정확한 깊이 정보 획득이 가능한 경우에 대한 실험이 이루어졌다. 우리는 이 논문과는 달리 이동 로봇 환경에서 객체 분할을 통한 인체 실루엣 추출을 시도한다. 다시 말해, 객체와 카메라와의 거리 차로 인한 저해상도 영상 입력에 대해, 카메라 이동 환경에서 여러 사람이 영상에 등장과 사라짐을 반복함으로써 초래되는 객체 및 배경 색의 불안정성을 고려한 다층(multi-layer) 객체 분할을 시도하고 있다.

로봇 비전 분야에서의 인체 실루엣 추출은 동적 카메라 환경, 조명 변화, 저해상도(low resolution), 복잡한 배경, 처리 속도 등의 다양한 문제점을 가진다. 본 논문에서 고려한 로봇은 시야에 들어오는 다수의 사람을 검출하고 추적하지만, 실루엣 추출은 중앙의 작업 관리자(task manager)가 지정하는 한 명의 사람에 대해서만 이루어진다. 실루엣 추출 방법은 추적에 의해 입력된 외곽사각형(bounding box)과 스테레오 비디오의 좌측 영상과 깊이 영상(depth image)을 입력받아¹⁾, 외곽사각형 내의 영상에 대해 깊이 정보와 평균-이동 분할 방법(mean-shift segmentation)을 이용하여 신뢰도 높은 객

체와 배경 영역을 추정한다. 이전 프레임으로부터 추정된 객체와 배경 영역의 화소들을 이용하여 객체와 배경 영역의 색 분포를 추정한다. 이렇게 추정된 색 분포와 추정된 영역의 각각의 화소들로 노드(node)를 구성한 그래프의 에지(edge)에 대한 가중치를 결정하게 되고, 최종적으로 max-flow 방법에 의해 배경으로부터 객체를 정확하게 분리함으로써 인체 실루엣을 추출한다.

2. 그래프 컷을 이용한 객체 분할 방법

그래프 컷을 이용한 객체 분할의 선구적인 방법론은 Boykov 등에 의해 제안되었다. 이는 사용자가 직접 객체 영역과 배경 영역에 해당하는 화소들(seeds)을 지정하도록 하는 방식으로 객체와 배경에 대한 정확한 색 분포를 얻을 수 있게 하여 매우 정확하고 안정적인 결과를 보일 수 있었다. 유사한 사용자 인터랙션 방법인 GrabCut[5]에 의한 분할 방법에서는 정지 영상에 대해 사용자가 원하는 객체 영역의 외곽사각형을 설정하도록 하여, 입력된 외곽사각형 내에서 배경과 객체를 분할하였다.

본 논문에서 제안한 그래프 컷을 이용한 객체 분할 방법은 Boykov 등[3]이 제안한 방식과 유사성이 있다. 그들은 사용자가 객체와 배경 영역을 지정하도록 하는 사용자 인터랙션을 가정하지만 우리는 이러한 사용자 인터랙션 없이 객체와 배경 영역을 추정한다는 점이 커다란 차이점이다. 다시 말해, 본 논문은 사용자 인터랙션 기반 객체 분할 방식을 자동 비디오 분할(automatic video segmentation) 방식으로 확장했다는 점에서 큰 의미가 있다. 또한 제안된 추정 방식은 저해상도, 조명 변화, 동적 카메라 환경 등으로 발생하는 객체 분할의 많은 어려움을 해결하는 데 매우 효과적이다. 영역 추정 방법은 다음 절에서 소개한다. 우선 추정된 객체 화소들의 집합을 O , 추정된 배경 화소들의 집합을 B 라 하자.

본 논문에서 제안하는 그래프 컷을 이용한 인체 실루엣 추출 방법을 살펴보면 다음과 같다. 우선 이미지 내의 화소의 집합을 P 라 하고, 두 인접 화소 (p, q) 의 전형적인 4-이웃 시스템(standard neighborhood system)을 N 이라고 하자. $z = (z_1, \dots, z_{|P|})$ 를 주어진 이미지라 할 때, z_n 는 k 번째 화소의 RGB값을 표현하는 3차원 벡터에 해당한다. 그리고 집합 P 의 화소 p 에 대한 0과 1의 이진 레이블(binary label)을 f_p 라 하자. 여기서 1은 “객체”를, 0은 “배경”을 의미한다. 이때, $f = (f_1, f_2, \dots, f_{|P|})$ 는 객체 분할을 나타내게 된다. 우리는 수식 (1)과 같이 정의된 에너지 함수(energy function)를 최소로 하는 레이블 벡터 f 를 찾는다.

$$E(f) = \gamma D(f) + V(f) \quad (1)$$

1) 앞으로 별도의 언급이 없는 한 본 논문에서 말하는 영상들은 주어진 입력 영상을 외곽 박스 정보에 의해 자른 영상을 의미한다.

여기서 계수 γ 는 데이터 항목(data term) $D(f)$ 의 중요도를 나타내며, 데이터 항목 $D(f)$ 와 평활 항목(smoothness term) $V(f)$ 는 수식 (2)와 같이 정의한다.

$$D(f) = \sum_{p \in P} D_p(f_p) \quad (2)$$

$$V(f) = \sum_{p, q \in N} \delta(f_p, f_q) V_{p, q}(f_p, f_q)$$

수식 (2)에서 델타 함수 $\delta(f_p, f_q)$ 는 $f_p \neq f_q$ 일 경우에 1을 할당하고, 나머지 경우에는 0을 할당한다. 입력된 이미지는 대부분 객체와 배경 사이에는 높은 대비(contrast)가 나타나는 특징을 가진다. 이러한 대비의 특징을 평활 항목을 통해서 수식 (3)과 같이 정의하였다.

$$V_{p, q}(f_p, f_q) = \exp(-\|z_p - z_q\|^2 / \beta) \quad (3)$$

수식 (3)에서 β 는 상수 값으로 모든 $\{p, q\} \in N$ 에 대해서 $2\|z_p - z_q\|^2$ 의 기대값(expectation)으로 결정된다. 데이터 항목 D_p 는 레이블 f_p 와 화소 p 가 어울리는가의 정도를 나타낸다. 우리는 이를 위해 이전 프레임의 분할 결과에 따라 해당 화소의 색값 z_p 들과 가우시안 혼합 모델(GMM: Gaussian Mixture Model)을 이용하여 객체와 배경에 대한 색 확률 분포 $P(\cdot|1)$ 와 $P(\cdot|0)$ 을 구한다. 추정된 영역에 속하지 않는 화소들의 집합을 $U = P - (O \cup B)$ 로 정의할 때, D_p 는 식 (4)와 같이 정의된다.

$$D_p(f_p) = \begin{cases} -\log P(z_p | f_p) & \text{if } p \in U \\ (K-c)f_p + c & \text{if } p \in O \\ (c-K)f_p + K & \text{if } p \in B \end{cases} \quad (4)$$

여기서 $K = \max_{p, q \in N} V(f_p, f_q)$ 이고, c 는 작은 상수로 실험에서 보통 1을 할당하였다. 최종적으로 에너지 함수 (1)를 최소화시키는 분할 f 는 대표적인 min-cut/max-flow 알고리즘에 의해 그래프 표현을 이용하여 구할 수 있다[8]. Boykov[3]의 논문과는 달리 상수 c 를 0이 아닌 값으로 설정함으로써 추정된 객체와 배경 영역오류를 보완할 수 있다. 그림 1은 추정된 객체와 배경 영역에 의해 최종적으로 분할된 결과를 보여준다. 그림 1(a)는 좌측 영상을 주어진 외곽 박스에 의해 자른 영

상이고, (b)는 변위 영상을 외곽 박스에 의해 자른 영상이다. (c)의 흰색과 검은색 영역은 추정된 객체 및 배경 영역을 의미하며 회색 영역은 추정되지 않은 영역이다. (d)는 최종 분할된 객체를, (e)는 객체의 실루엣을 보여준다.

3. 객체 및 배경 영역의 추정

일반적으로 색, 대비 또는 스테레오 정보 중 어느 하나에만 의존하는 방식들은 오류를 유발하는 경우가 많다. 추출하고자하는 객체 주변에 비슷한 색이나 변위(disparity) 값을 가지는 물체가 존재할 경우 그러하다. 본 논문에서는 색, 대비(영상 분할 기법), 스테레오 정보를 적절히 결합하는 방식을 취하였다. 지금까지 컴퓨터 비전 분야에서는 객체 분할을 위한 많은 연구가 이루어졌지만, 객체 영역을 정확하게 추정할 수 있는 객체 특징(feature)은 거의 없다. 다시 말해, 모든 화소들에 대해서 정확하게 “객체” 또는 “배경”이라고 레이블하기란 매우 어려운 일이다. 하지만 객체나 배경의 일부만을 정확하게 추정하는 것은 상대적으로 쉬운 일이다. 따라서 우리는 객체나 배경의 일부 영역을 추정하고 이에 속하는 화소들에 대해 색 정보와 무관하게 D_p 에 K 또는 c 의 상수 값을 할당함으로써, 카메라 이동이나 조명 변화와 같은 주변 환경의 변화에도 강인한 객체 분할을 가능하도록 하였다.

3.1 깊이 영상에서의 객체 분할

본 논문에서는 우선 주어진 깊이 영상으로부터 간단한 방법으로 객체 영역의 깊이 정보만을 분할한다. 로봇 플랫폼에서는 얼굴, 표정, 제스처 등을 인식하는 많은 알고리즘이 포함되어야 하기 때문에 깊이 영상을 생성하는 방법은 성능과 속도 면에서 효율적인 알고리즘을 공통으로 사용하는 것이 일반적이다. 우리는 Videre사의 스테레오 카메라에서 제공하는 SVS(Small Vision System)를 사용하여 깊이 영상을 생성하여 모든 모듈들이 공유하도록 하였다. SVS는 영역 기반 상관관계(area-based correlation)에 기반하여 깊이 영상을 생성

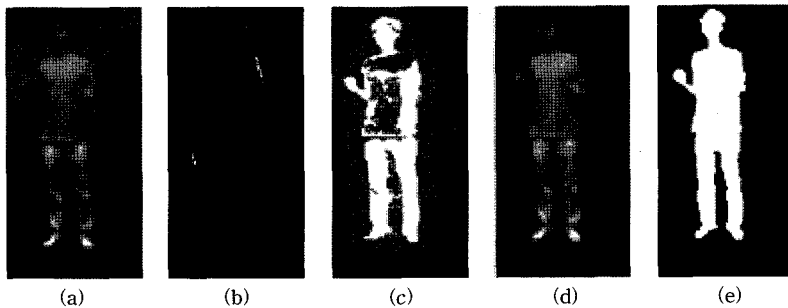


그림 1 제한한 그래프 컷을 이용한 인체 실루엣 추출

한다. 이후 후처리 과정으로 좌우 체크(left-right check), 관심 작용(interest operator), 부-화소 변위 추정(sub-pixel disparity estimation)을 수행한다.

변위 영상(disparity map)에서 객체 영역을 추정하기 위해서 우선은 t 번째 프레임의 객체 변위 평균 $\mu_{D,t}$ 와 분산 $\sigma_{D,t}$ 을 $t-1$ 번째 프레임의 객체 분할 결과에서의 변위 값이 0보다 큰 영역에 대해서 평균과 표준편차를 계산하여 추정한다. 이전 프레임의 결과를 이용하여 평균과 표준편차를 계산함으로써 카메라로부터 앞뒤로 움직이는 객체에 대해서도 효과적으로 객체를 분할할 수 있다. 분할된 객체 변위 영상 O_D 는 식 (5)와 같이 구해진다.

$$O_D(p) = \begin{cases} 1 & m_{D,t} - K_D \sigma_{D,t} < d_p < m_{D,t} + K_D \sigma_{D,t} \\ 0 & otherwise \end{cases} \quad (5)$$

여기서 d_p 는 화소 p 의 변위를 말하며 $\sigma_{D,t}$ 가 보통 작은 값이기 때문에 K_D 는 실험에서 보통 10으로 두었다. 또한, 평균과 표준편차에 의해 분할된 객체 영역에 대하여 미디언 필터링과 모폴로지 닫힘 연산(morphological closing)을 수행하여 객체 내에 발생하는 구멍(hole)과 같은 오류를 보정한다. 이때, 일반적으로 깊이 영상에서의 구멍은 텍스처가 없는 영역에서 발생하는데 이는 색 정보를 이용하는 그래프 컷에 의해 메워지기 때문에 모든 구멍을 강제로 메우지는 않는다. 그림 2의 (a)는 주어진 깊이 영상인 그림 1의 (b)로부터 분할된 객체의 깊이 영상 O_D 를 보여준다.

3.2 배경 영역 추정

깊이 영상에서 객체 영역을 분할한 후, 분할된 객체 변위 영상에 $d \times d$ 정사각형 구조 원소(square structuring element)를 이용하여 모폴로지 불림 연산(morphology dilation)을 수행하여 바깥 영역을 취함으로써 간단히 배경 영역을 추정한다. 그림 1(c)의 검은색 영역은 추정된 배경 영역을 보여준다. 실험에서 d 는 10으로 설정하였다. 영역 기반 상관관계를 이용하여 구한 변위 지도는 보통 부풀림이 크기 때문에 이러한 방식으로 추정된 배경 영역은 객체를 포함할 확률이 거의 없어 배경이라 추정해도 무리가 없다.

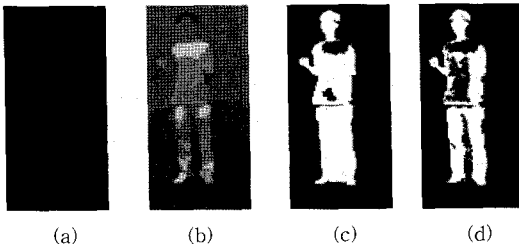


그림 2 객체 영역 추정

3.3 객체 영역 추정

신뢰도가 매우 높은 객체 영역 추정은 분할된 객체 깊이 영상 O_D 와 평균-이동 분할(mean-shift segmentation) [9]을 이용하였다. 이는 객체 분할이 아닌 영상 분할 방식으로 RGB 색 공간을 $L^*u^*v^*$ 색 공간으로 변환하여, 3차원 색 공간에 2차원 화소 좌표를 추가한 5차원 공간에서 주어진 화소의 색을 가장 가까운 최빈값(local mode)의 색 좌표로 대치시킴으로써 주어진 영상을 유사한 조각으로 분할하는 방식이다.

평균-이동 분할에 의해 얻어진 조각들을 $P_i, i=1, \dots, m$ 이라 하자. 각각의 영상 조각 P_i 에 대하여 객체의 평균 변위 값 μ_{D,P_i} 을 식 (6)과 같이 계산한다.

$$\mu_{D,P_i} = \sum_{p \in P_i} d(p) / |P_i| \quad (6)$$

여기서 $d(p)$ 는 화소 p 의 변위값이며, $|P_i|$ 는 P_i 에 속하는 화소들의 개수이다. 이렇게 구해진 μ_{D,P_i} 가 위에서 계산된 평균 $\mu(D,t)$ 의 표준편차 $\sigma_{D,t}$ 범위 내에 있다면, 해당 조각 P_i 를 후보 객체 조각으로 선택한다. 그림 2의 (c)는 평균-이동 분할 영상 (b)로부터 객체 깊이 영상 (a)를 이용하여 구해진 후보 객체 조각을 보여 준다.

마지막으로 추정된 배경 영역의 공간-색 히스토그램(spatial color histogram) H_b 을 이용하여 유사도가 낮은 화소들을 최종 객체 영역을 추정한다. 여기서 H_b 는 추정된 배경 영역의 화소 p 의 R, G, B의 값 r_p, g_p, b_p 과 영상 좌표 x_p, y_p 를 정규화 시킨 원소들 $(\lfloor r_p/\sigma_c \rfloor, \lfloor g_p/\sigma_c \rfloor, \lfloor b_p/\sigma_c \rfloor, \lfloor x_p/\sigma_s \rfloor, \lfloor y_p/\sigma_s \rfloor)$ 에 의해 만들어진 5차원 히스토그램을 말한다. 여기서, σ_c 와 σ_s 는 상수값으로 색과 좌표를 위한 정규화 상수이고, $\lfloor \cdot \rfloor$ 는 내림 함수(floor function)를 의미한다. 구해진 H_b 을 이용해서 위에서 추정된 후보 객체 조각들 내의 화소 p 에 대해 유사도 $-\log H_b(p)$ 를 계산하고, 이들의 평균 μ_C 와 표준편차 σ_C 를 계산한다. 이때, 화소 p 의 유사도가 적당한 상수 K_C 에 대해 $\mu_C + K_C \sigma_C$ 보다 크면 최종적으로 객체 영역으로 추정한다. 그림 2의 (d)는 최종 객체 영역을 보여 준다. 우리는 실험에서 K_C 를 1로 설정하였다.

본 논문에서 제안한 바와 같이 객체 깊이 정보를 이용하여 화소 단위가 아닌 영상 조각 단위로 인체 영역을 추정하면 인체 영역의 일부 외곽 정보를 추정할 수 있는 장점이 있다. 하지만 이러한 영역 기반 방법은 깊이 정보의 불완전함으로 인해 겨드랑이나 바지 가랑이와 같은 부분은 객체 영역으로 추정되는 단점이 있다. 이러한 문제점은 배경 영역의 공간적 색 분포를 이용하여 효과적으로 해결 할 수 있었다.

4. 실험 및 결과

본 논문에서 제안한 그래프 컷을 이용한 인체 실루엣 추출 방법은 5개 각각의 스테레오 비디오 데이터에 대해서 매 10 번째 프레임 마다 왼쪽 영상을 기준으로 추출된 참 영상(ground truth)을 이용하여 평가 하였다. 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.

4.1 실루엣 추출 평가

본 논문에서 제안한 방법은 K1, S1, S2, J3, K3 등 5개의 스테레오 비디오 데이터에 대해서 실험하였다. 데이터 K1과 S1은 고정된 스테레오 카메라 환경에서 로봇이 따라하기 데모를 수행할 목적으로 만들어 졌으며, 한명의 등장인물이 오른쪽/왼쪽 팔 다리 들기, 인사하기, 정면 및 측면으로 걷기 등의 다양한 제스처를 취하였다. 데이터 S2에서는 동적 카메라 환경에서 두 명의 등장인물이 주위를 걸으면서 팔, 다리 들기, 기지개 켜기 등의 제스처를 취하였다. 데이터 J3과 K3은 팬-틸트(Pan-tilt) 스테레오 카메라를 이용하여 촬영하였으며, 3명의 등장인물이 인사하기, 소파에 앉아 얘기하기, 음료수 가져다주기 등의 시나리오를 가지고 있다.

본 논문에서 제안하는 방법(SIG)은 색/대비(CC), 색/깊이 영상(CD) 그리고 GMM 대신 간단히 RGB의 색 히스토그램을 이용한 방법(HG) 등과 비교하여 실험 하였다. HG 방법은 제안한 방법과 비교하여 평균 2 fps (frame per second) 빠르다. CC 방법은 색 분포만을 이용하여 그래프 컷을 수행한 방법이고, CD는 영상 분할 후 객체 깊이 영상을 이용하여 객체 영역을 추정하는 방법으로 위에서 설명한 후보 객체 조각들을 이용한

실루엣을 추출한 방법이다. CC 방법의 경우 대부분 인체의 머리카락 등과 같은 색 분포가 작은 영역에 대해서 실루엣이 추출되지 않는 오류와 배경의 일부 영역이 객체의 색 분포와 유사하여 실루엣이라고 잘못 추출되는 오류를 자주 보였다. CD 방법의 경우는 깊이 영상 자체가 어느 정도 신뢰도가 높지 않은 데이터에 대해 안정적인 실루엣 추출 결과를 보이지 못했고 거드랑이나 바지 가랑이와 같은 구멍을 제대로 추출하지 못하는 단점을 보였다. 본 논문에서 제안한 SIG 방법과 HG 방법의 경우 CC와 CD에 비해 향상된 결과를 보였으나, HG 방법은 실루엣 추출 결과가 안정적이지 못했다.

표 1은 5개의 데이터에 대해서 위의 4개의 방법론의 오류율의 평균과 표준편차를 보여준다. 표 1은 제안한 방법이 오류율이 가장 낮을 뿐만 아니라 표준 편차도 작아 주변 환경의 변화에 많은 영향을 받지 않음을 보여 준다. 그림 3은 K3 데이터에 대해서 위의 4가지 방법의 오류를 매 10 프레임마다 측정한 결과이다. 그림 4는 4개의 방법에 대해 결과와 오류영역을 비교 결과를 보여준다. 또한 그림 5는 실험 데이터에 대한 다양한 실루엣 추출 결과를 보여준다.

4.2 구현 및 환경

본 논문에서 제안한 실루엣 추출 방법은 정확도의 적

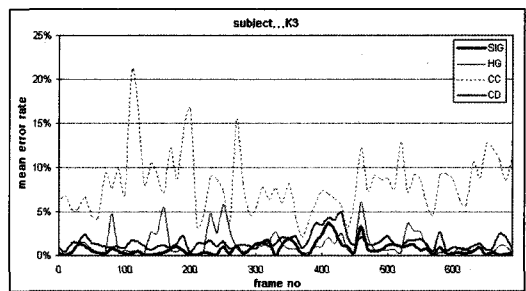


그림 3 K3 비디오 데이터에 대한 실루엣 추출 성능 비교

표 1 실루엣 추출 결과의 오류율(표준편차)(%)

	K1	S1	S2	J3	K3
SIG	0.72 (0.70)	0.29 (0.36)	1.32 (1.49)	0.99 (1.01)	0.76 (0.75)
HG	0.76 (0.69)	0.66 (1.09)	1.26 (1.66)	2.10 (1.97)	1.31 (1.42)
CC	7.27 (3.49)	1.63 (2.38)	4.09 (4.23)	6.71 (4.39)	8.22 (3.40)
CD	1.19 (0.59)	1.86 (0.65)	2.53 (1.11)	2.10 (1.03)	1.53 (0.89)

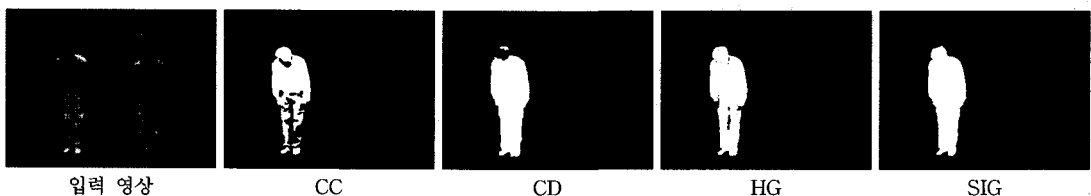


그림 4 각 방법론에 의한 실루엣 추출 결과 비교

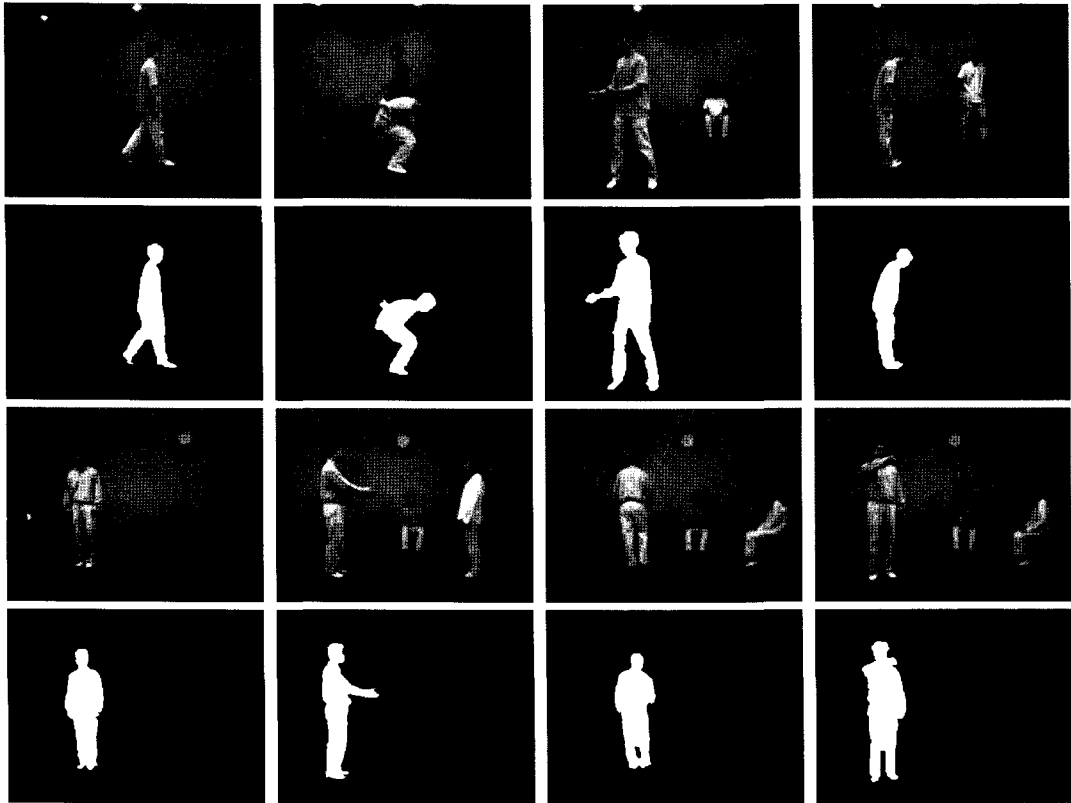


그림 5 동적 스테레오 카메라 환경에서의 인체 실루엣 추출 결과

은 손실과 빠른 수행 속도를 고려하여 구현되었다. 예를 들어, 그래프 컷에서 그래프를 구성함에 있어 8방향이 아닌 4방향 인접화소를 고려하였으며, max-flow 방법을 구현함에 있어 forward star 방법 대신 adjacent list 방법을 사용함으로써 속도를 향상 시켰다. 또한 지수 연산을 Taylor 급수로 대체하여 계산하였으며, GMM 연산에서 모든 화소를 입력하지 않고 3 화소에 1 개씩 샘플링하여 사용하였다. 제안한 알고리즘은 CPU 2.8 GHz와 1 GB RAM의 하드웨어 상에서 평균 90 X 190 크기의 외곽 박스 영상을 처리하는 데 평균 12 fps의 속도를 보였다. 동일한 조건하에서 히스토그램을 이용하여 컬러 모델링을 한 HG 방법은 제안한 방법과 비교하여 평균 2 fps(frame per second) 빨랐으나 안정적인 성능을 보여주지 못했다.

5. 결론

본 논문에서는 동적 스테레오 카메라 환경에서 인체의 실루엣을 추출하는 문제를 다루었다. 우리는 기존의 사용자 인터랙션을 이용한 분할 방식을 비디오 분할 시에도 적용할 수 있는 적절한 방식을 고안함으로써 스레

레오 비디오 데이터로부터 그래프 컷을 이용한 새로운 객체 추출 방법을 제안하였다. 스테레오 정보와 평균-이동 영상 분할 방법을 이용하여 신뢰도 높은 객체와 배경 영역의 일부를 미리 추정함으로써 주변 환경의 변화 시에도 안정적인 실루엣 추출이 가능하도록 그래프 컷의 성능을 대폭 향상시킬 수 있었다. 그러나 이동 로봇 비전을 위해서는 수행 속도의 향상을 위한 추가적인 연구가 필요하다.

참고 문헌

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 246-252, 1999.
- [2] Stella X. Yu, R. Gross and J. Shi, "Concurrent object recognition and segmentation by graph partitioning," Proc. Neural Information Processing Systems(NIPS'02), 1383-1390.
- [3] 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

- (ICCV), Canada, 2001.
- [4] Y. Li, J. Sun, C.-K. Tang and H.-Y. Shum, "Lazy Snapping," Proceeding ACM SIGGRAPH, 2004.
- [5] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: interactive foreground extraction using iterated graph cuts," ACM Trans. Graph. 23(3), 309-314, 2004.
- [6] 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(ECCV), pp. 428-442, 2004.
- [7] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross and C. Rother, "Bi-layer segmentation of binocular stereo video," IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2005.
- [8] 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 (PAMI), Vol. 26, pp. 1124-1137, 2004.
- [9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," IEEE Transaction on Pattern Analysis and Machine Intelligence(PAMI), pp. 603-619, 2002.

조교수. 1998년~2003년 연세대학교 컴퓨터과학과 부교수. 2003년~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 인공지능, 영상인식, 영상처리



안 정 호

1996년 연세대학교 수학과 졸업(이학사)
1998년 연세대학교 대학원 수학과 졸업
(이학석사). 2001년 Texas A&M Uni-
versity, Statistics(M.S.). 2006년 연세대
학교 대학원 컴퓨터과학과 졸업 (공학박
사). 2006년~현재 연세대학교 컴퓨터과

학과 박사후 연구원. 관심분야는 컴퓨터 비전, 패턴인식, 인
공지능



김 길 천

2001년 경기대학교 전자계산학과 졸업
(이학사). 2001년~현재 연세대학교 대학
원 컴퓨터과학과 석사과정 재학중. 관심
분야는 영상처리, 비디오처리, 패턴인식,
문자추출, 인공지능



변 헤 관

1980년 연세대학교 수학과 졸업(이학사)
1983년 연세대학교 대학원 수학과 졸업
(이학석사). 1987년 University of Illinois,
Computer Science(M.S.). 1993년 Pur-
due University, Computer Science
(Ph.D.). 1994년~1995년 한림대학교 정

보공학과 조교수. 1995년~1998년 연세대학교 컴퓨터과학과