

자세인식을 위한 정확한 깊이정보에서의 3차원 다중 객체검출 및 추적

3D Multiple Objects Detection and Tracking on Accurate Depth Information for Pose Recognition

저자 (Authors)	이재원, 정지훈, 홍성훈 Jae-Won Lee, Jee-Hoon Jung, Sung-Hoon Hong
출처 (Source)	멀티미디어학회논문지 15(8) , 2012.08, 963-976(14 pages) Journal of Korea Multimedia Society 15(8) , 2012.08, 963-976(14 pages)
발행처 (Publisher)	한국멀티미디어학회 Korea Multimedia Society
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02005336
APA Style	이재원, 정지훈, 홍성훈 (2012). 자세인식을 위한 정확한 깊이정보에서의 3차원 다중 객체검출 및 추적. 멀티미디어 학회논문지, 15(8), 963-976
이용정보 (Accessed)	서강대학교 163.***.1.208 2020/05/30 18:22 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

자세인식을 위한 정확한 깊이정보에서의 3차원 다중 객체검출 및 추적

이재원^{*}, 정지훈^{**}, 홍성훈^{***}

요 약

‘제스처’는 음성을 제외한 가장 직관적인 인간의 의사표현 수단이다. 그에 따라 제스처를 이용하여 컴퓨터를 제어하는 방법에 대한 많은 연구가 진행되고 있다. 이러한 연구에서 사용자를 검출하고 추적하는 방법은 매우 중요한 과정 중의 하나이다. 기존의 2차원 객체 검출 및 추출 방법은 조명이나 주변 환경의 변화에 민감하고, 2차원과 3차원 정보의 혼합사용 방법은 연산량이 많다는 단점이 있다. 또한 3차원 정보를 이용한 기존 방법들은 유사한 깊이의 객체 분할이 불가능하다. 따라서 본 논문에서는 깊이 정보의 누적 값인 Depth Projection Map (DPM)과 움직임 정보를 이용하여 객체를 검출하고 추적하는 방법을 제안한다. 실험 결과 제안 방법은 조명이나 환경변화에 강인하고, 연산속도가 빠르며, 유사한 깊이의 물체도 잘 검출하고 추적할 수 있음을 확인하였다.

3D Multiple Objects Detection and Tracking on Accurate Depth Information for Pose Recognition

Jae-Won Lee^{*}, Jee-Hoon Jung^{**}, Sung-Hoon Hong^{***}

ABSTRACT

‘Gesture’ except for voice is the most intuitive means of communication. Thus, many researches on how to control computer using gesture are in progress. User detection and tracking in these studies is one of the most important processes. Conventional 2D object detection and tracking methods are sensitive to changes in the environment or lights, and a mix of 2D and 3D information methods has the disadvantage of a lot of computational complexity. In addition, using conventional 3D information methods can not segment similar depth object. In this paper, we propose object detection and tracking method using Depth Projection Map that is the cumulative value of the depth and motion information. Simulation results show that our method is robust to changes in lighting or environment, and has faster operation speed, and can work well for detection and tracking of similar depth objects.

Key words: Object Extraction(객체 추출), Object Tracking(객체 추적), Multiple Objects(다중 객체)

※ 교신저자(Corresponding Author): 홍성훈, 주소: 광주광역시 북구 용봉동 300번지 전남대학교 전자컴퓨터공학부(500-757), 전화: 062)530-1758, FAX: 062)530-1759, E-mail: hsh@chonnam.ac.kr

접수일: 2011년 12월 30일, 수정일: 2012년 3월 22일

완료일: 2012년 6월 15일

^{*} 준회원, 전남대학교 전자컴퓨터공학과

(E-mail: 777kamja@naver.com)

^{**} 준회원, LG이노텍 부품소재연구소

(E-mail: hootybaby@hanmail.net)

^{***} 정회원, 전남대학교 전자컴퓨터공학부(전남대학교 정보통신연구소)

※ 본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임.

1. 서 론

객체검출 및 추적이란 입력 영상으로부터 의미 있는 단일 혹은 다중 객체나 영역을 배경과 따로 분리하여 각각에 의미를 부여하는 과정을 의미한다. 이러한 객체 검출 및 추적 방법은 지능형 감시시스템, 내용기반 검색 시스템, 자세인식 시스템 등의 많은 분야에서 중요하게 사용된다. 본 논문에서는 자세인식 시스템에 적용하기 위해 다중 사용자 객체를 검출하고 각각의 검출된 객체를 추적하는 방법을 제안한다. 초기의 객체 검출 및 추적 연구는 2차원 정보만을 이용하는 방향으로 진행되었다[1-4]. 하지만 2차원 정보만을 이용한 객체 검출 및 추적 방법은 조명, 그림자, 잡음 등의 다양한 주변 환경의 변화에 매우 민감한 문제가 있어, 이를 해결하기 위하여 깊이정보를 이용하는 방법이 등장하였다. 깊이정보는 카메라로부터의 거리(깊이)정보를 가지고 있기 때문에 객체 검출 및 배경과의 분리, 객체들 간의 겹침 등에 대해서 2차원 정보에서는 불가능했던 것들을 가능하게 하기 때문에 객체 검출 및 추적에 매우 중요한 요소로 사용되어지고 있다. 깊이정보를 이용한 객체 분할, 검출 및 추적연구로는 평균 배경 깊이정보를 사용한 객체 검출[5], 깊이의 확률밀도함수(probability density function)를 이용한 객체 검출 및 추적[6], mean-shift를 이용한 객체 검출[7], Graph-Cut을 이용한 객체 검출[8], 그리고 2차원과 3차원 정보를 혼합하여 사용한 객체 검출 및 추적[9,10] 등의 다양한 연구가 소개되어지고 있다. 2차원 정보와 3차원 깊이 정보를 혼합 사용한 방법은 2차원 정보만을 사용한 것보다 주변 환경에 덜 민감하고 정확도가 높아지지만 연산량이 많고 깊이정보만을 이용한 객체 검출 결과와 성능이 크게 다르지 않아 최근에는 깊이정보만을 이용하여 객체를 검출하는 방법이 주로 연구되어지고 있다. 하지만 기존의 깊이정보를 이용한 방법들은 객체를 검출하고 추적하는 과정에서 깊이정보의 유사성에 의하여 객체가 아닌 영역까지 객체로 검출하는 문제가 생긴다. 본 논문에서는 이와 같은 단점을 해결하기 위하여 깊이정보의 유사성과 움직임 정보를 혼합하여 사용하는 방법을 제안한다. 본 논문에서 제안하는 객체 검출 및 추적 방법은 깊이정보 유사성을 이용하기 위해 깊이정보의 누적 값인 Depth Projection Map을 이용하여 주변 환경 변화에

강인하고, 적은 연산량으로 정확한 객체 검출 및 추적을 지원하며, 검출된 객체와 주변 배경의 깊이가 유사한 경우에도 다수의 객체를 정확하게 검출 및 추적할 수 있는 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 객체 검출 및 추적 방법에 대해 전체적인 설명을 하고, 3장에서는 제안하는 방법의 세부적인 내용에 대해서 설명한다. 그리고 4장에서는 실험 결과에 대한 분석을 하고, 5장에서 결론을 맺는다.

2. 제안하는 알고리즘의 구성

그림 1은 본 논문에서 제안하는 다중 객체추출 및 추적 방법의 전체 구성도를 보여준다. 우선, 입력되는 깊이 영상의 3프레임에 대하여 Frame Difference Mask (FDM)를 구하여 움직임 정보를 파악하는데, FDM로부터 일정 임계치 이상의 움직임이 관측되었을 때 사용자가 나타났다고 판단하고 객체 검출을 수행한다. 이 과정에서 깊이에 대한 누적 분포도를 나타내는 Depth Projection Map (DPM)을 생성하고 DPM에서 누적 값에 대한 라벨링을 통하여 깊이 영상에서의 초기 사용자 객체 후보를 추출한다. 그리고 각 사용자 객체 후보에 대해서 사람인지 아닌지에 대한 판단을 하는 타당성 검증 과정을 통해 사람이라고 판단되면 신규 사용자 객체로 등록한다. 그리고 검출된 객체가 사전에 등록되어 있으면 추적 과정을 수행한다. 추적은 DPM 영역에서 이미 검출된 이전

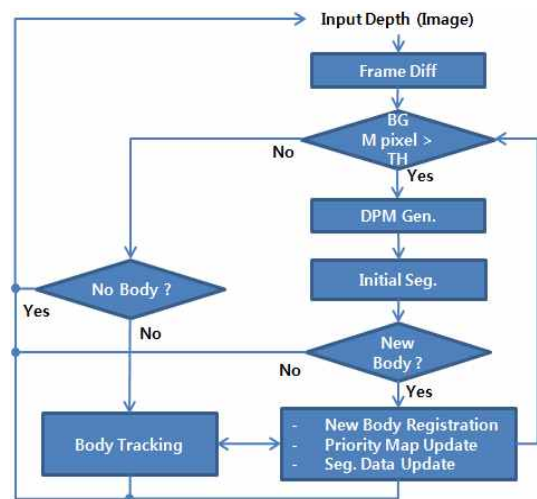


그림 1. 제안하는 다중 객체추출 및 추적 전체 구성도

프레임의 사용자 객체와 겹침 비율이 가장 큰 클래스 영역의 움직임이 없는 픽셀에 대한 등록과 움직임 영역에 대한 등록과정을 통하여 객체를 실시간으로 추적한다. 이와 같은 과정을 통하여 사용자 객체를 검출하고 각 객체에 대한 추적을 수행하여 객체정보를 자세 추정을 위한 정보로 넘겨줄 수 있다.

3. 다중 객체 검출 및 추적의 세부 알고리즘

3.1 입력 깊이 영상

깊이 영상은 카메라로부터 입력받은 2차원 영상의 각 화소단위에 색상정보나 명암정보 대신 깊이 정보를 부여한 영상이다. 이러한 깊이 영상은 크게 2개의 방법을 이용하여 얻을 수 있다. 하나는 스테레오 매칭으로 인간이 두 눈과 뇌에서의 시각피질을 이용하여 입체 공간을 인식할 수 있듯이 이와 유사하게 한 쌍의 카메라로부터 들어오는 영상을 이용하여 관측 공간상의 거리 및 관측 물체의 삼차원 형태를 감지하는 원리를 사용하는 방법이다. 하지만 이 방법은 연산량이 많아 실시간으로 정확한 깊이 정보를 얻기에 적합하지 않은 단점이 있는데, 실시간 처리를 위해 연산 속도를 올릴 경우 깊이 정보의 정확도가 떨어질 수 있다. 그림 2의 (a)는 스테레오 매칭을 적용한 VisionST사의 VSTC-P5V 카메라를 이용하여

얻은 깊이 영상으로 30fps를 지원하지만 깊이정보를 128 level로 표현하기 때문에 깊이에 대한 정확도가 떨어진다. 또한 매칭 방법을 사용하기 때문에 깊이정보가 가로로 퍼지는 오류도 나타난다. 깊이 영상을 얻는 또 다른 방법은 카메라 전방에 레이저를 방사하고 레이저가 반사하여 돌아오는 속도를 계산하여 깊이정보를 얻는 방법이다. 이와 같은 방법은 실시간으로 사용이 가능하며 정확도가 높다는 장점이 있다. 예전에는 장비가 비싸서 쉽게 사용하기 힘들었지만 최근에 Microsoft사에서 KINECT라는 저가의 카메라를 내놓으면서 쉽게 사용이 가능해졌다. KINECT를 이용하여 얻은 깊이 영상은 30fps를 지원함으로써 실시간 연산이 가능하고, 스테레오 비전 방법에 비해 정확도가 매우 높은 깊이정보를 제공한다. 또한 레이저를 이용하기 때문에 조명이 어두운 곳에서도 사용이 가능하다. KINECT는 11bits 깊이 영상을 제공하는데, 본 논문에서는 범용성을 고려하여 그림 2의 (b)에 나타난 바와 같이 8bits 깊이 영상으로 변환하여 입력영상으로 사용한다. 기존의 TOF sensor 카메라의 깊이영상 또한 정확한 깊이정보를 제공하며 KINECT 카메라와 동일하게 제안하는 방법에 사용 가능하지만 가격이 높다는 단점이 있다.

3.2 Frame Difference Mask(FDM)

고정된 카메라를 활용한 객체의 움직임 검출은 일반적으로 현재 프레임과 이전 프레임과의 화소 값의 차를 나타내는 프레임 차이(Frame difference) 정보를 이용한다[1,2]. 현재 프레임의 화소와 동일한 위치의 이전 프레임의 화소의 밝기 값 차이가 일정 임계치보다 크면 프레임 간에 어떠한 움직임이 발생하였다고 판단할 수 있다. 본 논문에서는 깊이 영상의 프레임 차이 정보를 이용하여 움직임 검출을 수행하는데, 이전과 현재의 프레임 차이를 수행하여 얻은 움직임 영역 검출 결과와 현재와 이후의 프레임 차이를 수행하여 얻은 움직임 영역 검출 결과를 AND 연산하여 움직임 영역을 검출한다. 3 프레임간의 차이를 이용한 움직임 영역 검출은 식(1)과 같이 나타낼 수 있고, 결과는 그림 3과 같다.

$$FDM_n = TH_{n-1}(FD_{n-1}) \wedge TH_{n+1}(FD_{n+1}) \quad (1)$$

여기서 FDM_n 은 현재 프레임에 대한 움직임 영역을 나타내고, FD_{n-1} 과 FD_{n+1} 은 각각 이전과 현재, 현

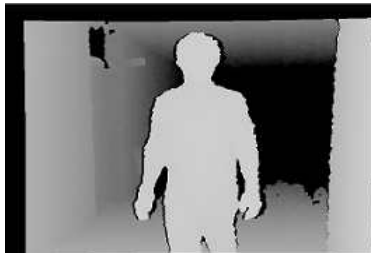


(a) 부정확한 깊이 영상

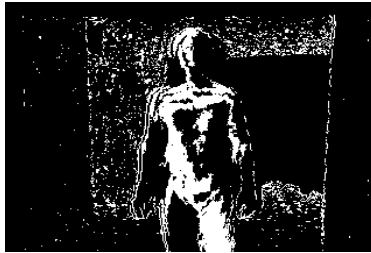


(b) 정확한 깊이 영상

그림 2. 부정확한 깊이 영상과 정확한 깊이 영상의 예



(a) 깊이 영상



(b) FDM 영상

그림 3. 3 프레임간의 차이(FDM) 수행 예

재와 이후 프레임간의 차이를 나타내며, TH_{n-1} 과 TH_n 은 각각 이전과 현재, 현재와 이후 프레임간의 차이에 임계치를 적용한 것을 의미한다. 본 논문에서 식(1)과 같이 3 프레임을 이용하여 움직인 영역검출을 수행한 이유는 그림 4에 나타난 바와 같이 2 프레임을 이용할 경우 움직임이 없는 배경영역이나 미세한 움직임을 갖는 영역을 움직임 영역으로 판정하는 것을 억제하기 위해서 이다.

FDM으로부터 구한 움직임 픽셀 수의 갑작스런 변화는 영상에서 움직이는 객체가 포착되었다는 것을 알려준다. 매 프레임마다 움직임 픽셀 수를 검사하여 임계치보다 많은 움직임 픽셀 수가 나타났을 경우 사용자 객체 검출을 수행한다. 본 논문에서는 임계치로 (전체 화소수/128)를 설정하였는데, 이는 그림 3이나 그림 4에 나타난 바와 같이 깊이 정보의

오류로 정지된 배경영역에서 나타나는 움직임 픽셀에 대한 실험을 통해서 결정한 것이다. 임계치만큼의 움직임은 사용자가 없어도 지속적으로 나타난다고 가정하고, 임계치 이상의 움직임 픽셀이 발생하면 사용자가 나타났다고 판단하여 객체검출을 수행한다.

3.3 Depth Projection Map (DPM)

객체 검출 과정은 연산량과 검출의 강인성을 고려하여 Depth Projection Map 정보를 사용한다[11]. DPM은 깊이에 대응하는 픽셀의 개수를 나타내는데, 본 논문에서는 수평축을 기준으로 깊이에 대응하는 픽셀의 개수를 사용하였다. 결국 수평축 기준 DPM은 입력 깊이 영상에서 동일한 수평 위치에 있는 모든 픽셀에 대해 각 깊이 값에 해당하는 픽셀의 개수를 기록한 것으로, $DPM(x, d)$ 는 입력 깊이 영상의 수평좌표 x 에 있는 모든 픽셀 중 깊이 값이 d 인 픽셀 수이다. 그림 5의 (b)는 입력 깊이 영상에 대한 수평축 기준 $DPM(x, d)$ 의 예를 나타낸 것으로, 수평축은 입력 깊이 영상의 수평좌표 x 이고, 수직축은 깊이값 d 이며, $DPM(x, d)$ 에 기록된 값은 입력 깊이 영상에서 수평좌표 x 에 있는 픽셀의 깊이값이 d 인 픽셀 수를 나타낸다.

DPM에서 연결된 영역에 대하여 라벨링(labeling)을 수행하여 얻어진 각 클래스는 수평축을 기준으로 유사한 깊이 정보를 갖는 하나의 객체라고 판단할 수 있다. 여기서 각 클래스는 그림 5(b)에 나타난 바와 같이 수직방향으로 프로젝트된 깊이정보를 Connected component analysis를 통해 얻어진다. 라벨링된 DPM 영상은 입력 깊이 영상의 x 축과 깊이에 대한 정보를 함축적으로 담고 있어서 원래의 깊이 영상을 그림 5의 (c)와 같이 유사한 깊이의 영역으로 분할할 수 있다.



(a) 입력 깊이 영상



(b) 3FDM(depth)



2FDM(depth)

그림 4. 3 프레임간의 차이(3FDM)와 2 프레임간의 차이(2FDM)의 수행 예

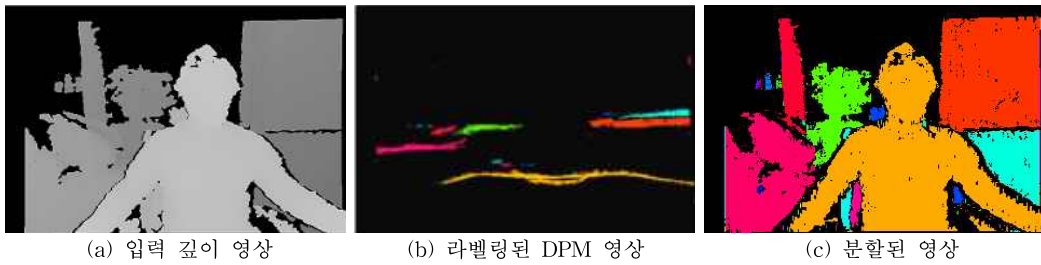


그림 5. 기본 깊이 영상과 라벨링된 DPM 영상

그림 6은 DPM, dpeth pdf[6], mean-shift[12] 방법을 이용하여 깊이영상을 분할한 결과와 처리시간을 보여준다. 가장 널리 알려진 분할 방법인 mean-shift 방법은 연산속도가 매우 느려서 실시간 연산에는 적합하지 않다. Depth pdf는 연산속도는 빠르지만 유사한 깊이영역을 하나의 영역으로 분할하므로 추가적인 라벨링 연산이 더 필요하여 결국 연산속도가 느려진다. 반면 DPM을 이용하여 분할을 수행한 결과를 확인하면 연산속도도 매우 빠를 뿐만 아니라 유사한 깊이에 대해서도 정확하게 영역을 분할함을 확인할 수 있다. 따라서 본 논문에서는 DPM을 이용하여 라벨링된 클래스를 하나의 후보 객체 단위로 보고 후보 객체가 사람인지에 대한 타당성 여부를 결정하여 최종적으로 초기 사용자 객체를 검출하는 방법을 사용한다.

3.4 초기 사용자 객체 검출(Initial Segmentation)

DPM에서 라벨링을 이용하여 구해진 후보 객체 단위 중에서 타당성 검사를 통해 초기 사용자 객체로의 등록 여부를 판단하여 객체를 검출한다. 우선 사용자 객체의 등록은 카메라에서 가까운 사용자 객체 순으로 등록하도록 설계하였다. 본 논문의 취지는 자세 인식을 위해 사용자 객체를 검출하고 추적하는 것이기 때문에 카메라에서 가장 가까운 객체에 높은

우선순위를 부여한다.

전체 영상에 대해서 일정 수준 이상의 움직임이 일어나면 초기 사용자 객체 검출을 시작하는데, 앞서 구한 DPM에서 라벨링된 각 클래스에 대해 움직임 픽셀 수가 일정 임계치보다 많고 클래스의 면적이 일정 크기 이상인 경우에만 사람 객체인지 아닌지에 대한 타당성 검사를 수행한다. 이와 같은 과정을 통해서 연산량의 감소 효과도 얻을 수 있다.

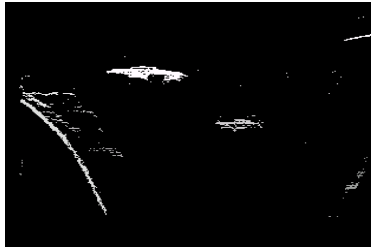
타당성 검사는 사람 객체의 형태에 대한 사전 정보를 이용하는데, 동일한 사람이 가질 수 있는 깊이 정보의 최대범위와 신체에 대한 수평, 수직 비율을 고려한다. 본 논문에서는 사용자가 서있거나 허리를 구부린 상태에서 손을 앞이나 뒤로 뻗는 경우를 동일한 사람 객체에 대한 깊이 값의 최대범위라고 정하고 검사하였다. KINECT 카메라의 깊이 정보는 거리에 대해 선형적인 값을 가지고 있기 때문에 실험을 통해 카메라로부터의 거리와 관계없이 동일한 사람이 가질 수 있는 깊이값 차이의 최대범위는 40이 타당함을 확인하였다. 예를 들면 그림 7의 (a)에서 배경에 해당하는 옆면의 벽에 포함된 픽셀이 깊이 연산 오류로 인해 움직임이라고 결정되어 사람객체 후보로 포함될 가능성이 있다. 따라서 이와 같이 움직임이 있는 배경 영역을 사용자 객체 후보에서 제외하는 과정이 필요한데, 깊이 정보의 최대범위를 판정기준으로 한

구분	원본	DPM	depth pdf	mean-shift
결과 영상				
속도	0.007 sec/frame	0.013 sec/frame	0.020 sec/frame	0.660 sec/frame

그림 6. 깊이 유사성을 이용한 영상 분할 방법 비교



(a) 움직임 영역



(b) DPM 영역

그림 7. 움직임 영역과 DPM 영역의 예

다. 그림 7의 (b)에 나타난 DPM 영역에서 좌측에 길게 분포한 클래스는 그림 7의 (a)에서 좌측 벽면에 포함된 픽셀 중 움직임 픽셀에 대응하는 클래스로써 깊이 정보가 매우 넓은 범위를 가지고 있다. 사용자 객체가 가질 수 있는 깊이 범위는 한정되어 있기 때문에 DPM영역에서 깊이 정보의 최대범위를 넘어서는 객체를 사용자 객체 후보에서 제외시킨다.

신체에 대한 수평, 수직 비율 기준에 의한 타당성 검사는 우선 미리 정해진 최소기준사용자의 크기보다 작거나 최대기준사용자 크기보다 큰 클래스를 제거하는 과정을 수행한다. 여기서 최소기준사용자의 크기는 전체 영상 크기의 1/32의 크기를 사용하고 최대기준사용자의 크기는 전체영상사이즈의 1/2의 크기를 사용한다. 이는 실험을 통해서 얻은 절대적으로 사용자라고 볼 수 없는 크기로 이를 벗어나는 영역을 미리 제외함으로써 정확도와 연산속도의 증가 효과를 가져올 수 있다. 그리고 사용자 객체 후보 영역에 대해서 그림 8과 같이 X, Y축 프로젝션을 통해서 높이와 너비의 비율을 구하고, 인체의 비율에 맞는지 여부를 검사한다. 그림 8의 빨간선은 수평, 수직 각 최고점에 대한 1/4 지점으로 사용자가 손을 뻗었을 경우에 손끝까지 계산하면 인체비율과 맞지 않기 때문에 손과 같은 얇은 부분은 제외하고 계산하기 위하여 1/4지점까지 계산하여 적용하였다. 본 논문에서는 사람의 상체만 검출되었을 경우도 고려하여 비율

을 1.2~10.0으로 정하여 수행하였다. 이 때 사용자의 비율은 카메라 앞에서 여러 사용자들의 상체, 정면, 측면 등의 자세의 객체를 검출하여 얻은 비율들을 고려하여 카메라 앞에서 움직이는 모든 사용자가 포함될 수 있는 비율로 설정한 것이다.

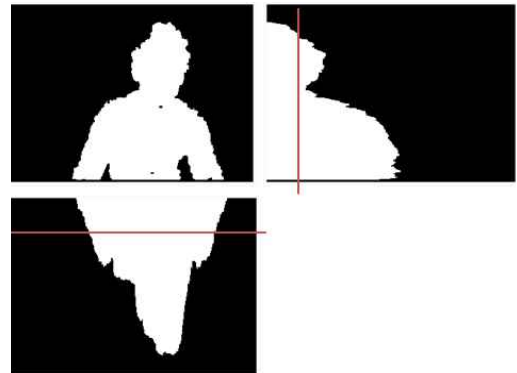


그림 8. X, Y 프로젝션 비율 검사를 통한 후보객체 타당성 검사

3.5 신규 객체 등록(New Body Registration) 및 객체 정보 업데이트

타당성 검사를 통과한 신규 사용자 객체는 사용자 객체 정보가 신규 등록되고, 이 객체는 등록된 정보를 기준으로 사용자 객체 추적 과정에 의해 추적된다. 한편, 이미 등록된 객체 영역은 초기 사용자 객체 검출 과정에서 제외시킴으로써 연산량을 줄이고 정확도도 높게 된다. 그림 9는 각 사용자 객체에 대한 정보를 등록한 결과로 각 객체의 추적을 위해 등록된 초기 사용자 객체 영상을 보여준다.

이때 각 사용자 객체 정보는 객체의 영역 정보, 객체 고유 번호, 객체의 상태(Detected: 검출됨, Hide: 영상 내에서 객체 숨김, Miss: 영상 밖으로 객체가 완전히 사라짐), 현 객체의 가장 높은 깊이 값과 가장 낮은 깊이 값(즉 객체의 깊이 범위), 객체의 바운딩 박스 정보(객체의 상, 하, 좌, 우의 범위), 객체의 영역 크기의 정보로 구성되어있다. 여기서 객체의 상태와 객체의 깊이 범위, 바운딩 박스 정보를 이용하여 그림 10의 (d), (e), (f)와 같이 현 영상 내에서 객체의 위치를 알 수 있고, 다른 객체 또는 배경과의 겹침에 의해 사라진 객체에 대한 정보를 유지함으로써 재추적이 가능하다. 그림 10의 (b)와 (c)처럼 겹쳐진 사용자 객체는 완전히 사라진 것이 아니라 바운딩 박스



그림 9. 각각 사용자 초기 객체 검출 후 등록 된 결과 예

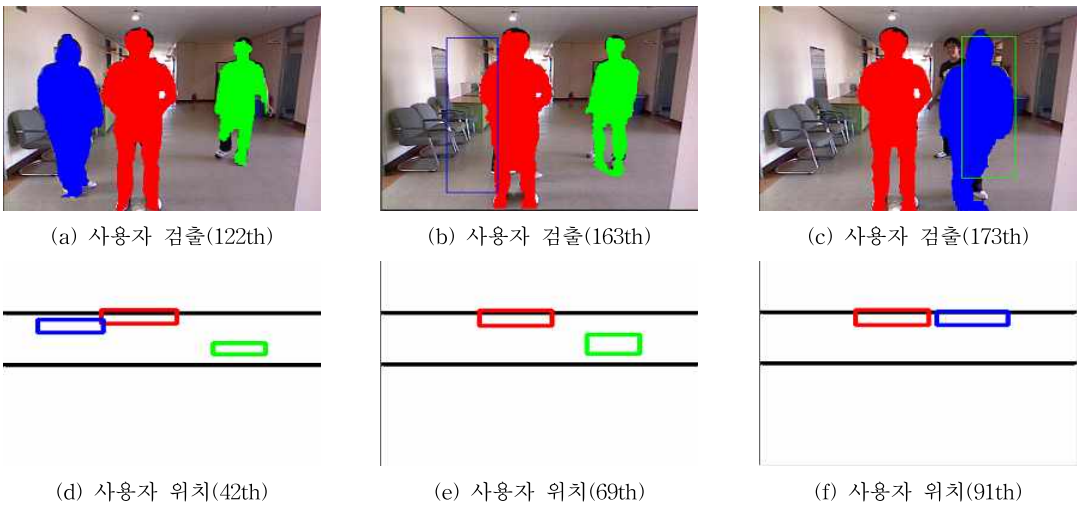


그림 10. 각각 사용자 객체 검출 및 위치 정보 파악

정보를 보유하고 있음으로써 영상 내에서 안 보이는 걸로 정보(Hide)를 갱신하고 객체가 다시 나타나면 재 추적을 수행한다.

3.6 객체 추적(Object Tracking)

초기 사용자 객체가 등록되면 우선순위가 높은 객체부터 독립적으로 객체를 추적하고 객체 정보를 업데이트 한다. 객체 추적 과정은 먼저 검출된 초기 사용자 객체 영역에 대해서 탐색범위(Search Range)를 정하는데, 탐색 범위는 사용자가 1프레임 사이에 움직일 수 있는 최대 범위를 의미한다. 탐색 범위는

사용자의 깊이(카메라로부터의 거리)와 관계가 있으므로 사용자 객체 외곽점들의 깊이에 해당하는 범위를 선택적으로 지정한다. 그림 11은 탐색 범위에 대한 예이다.

사용자 객체 추적에서도 초기 객체 검출과 같은 방법으로 DPM 영역에서의 후보 객체를 구하고 객체를 추적하는데, 이때 탐색범위 밖의 영역은 제외시킴으로써 연산량 감소 효과를 가질 수 있다. 그리고 초기 객체 검출과 같이 움직임 정보를 이용하는데, 추적 과정에서는 초기 객체 검출 과정과는 다르게 2 FDM을 사용하여 현재 프레임의 사용자 객체를 이전

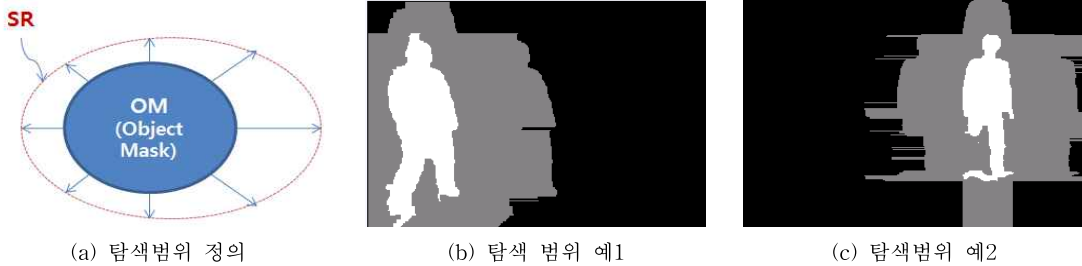


그림 11. 사용자 객체에 대한 탐색범위 정의 및 예

프레임의 사용자 객체로부터 추적을 수행한다. 객체의 검출과정에서의 3프레임 FDM은 잡음으로 인한 오류를 줄이기 위해 사용되었으나 객체의 추적과정에서의 FDM은 이전 객체 영역정보와 함께 탐색범위 내에서 일어나는 움직임을 이용하므로 잡음과 관계없이 2FDM을 통해 나타나는 모든 움직임을 파악한다. 따라서 탐색범위 내에서 일어나는 모든 움직임을 검출하기 위하여 임계치를 1로 설정하여 수행한다. 그림 12의 (b)는 3프레임 FDM을 수행한 결과이고 (c)는 2프레임 FDM을 수행한 결과인데, 2프레임에 대해 FDM을 수행한 결과가 3프레임 FDM 수행한 결과보다 움직임 영역이 확실하게 나타나는 것을 확인할 수 있다.

현재 프레임의 객체는 이전 프레임의 객체와 형태와 크기가 유사하므로 객체를 추적할 때 탐색범위에 대응하는 현재 프레임의 DPM 영역에 있는 클래스들 중 이전 프레임에서 구한 현재 추적중인 객체의 영역과 포함 비율이 가장 높은 클래스를 사용자 객체 후보 영역으로 선택한다. 하지만 이와 같이 구한 사용자 객체 후보 영역을 사용자 객체로 바로 사용하면 현재 추적 중인 객체가 다른 어떤 사람이나 사물과 접촉하는 경우 DPM영역에서 2개의 물체가 동일한 클래스로 인식되어 추적하던 객체뿐만 아니라 달라

붙은 객체까지 하나의 물체로 인정해 버리는 오류를 범하게 된다. 따라서 본 논문에서는 정지 영역(SM)과 움직임 영역(MM)을 구분하여 추적할 후보 객체 정보를 갱신한다.

정지 영역(SM)과 움직임 영역(MM)에 대한 정의는 그림 13의 (a)에 나타난 바와 같다. 만일 객체가 깊이의 변화 없이 오른쪽으로 움직였다면 이전 프레임의 객체(O_{n-1})와 오른쪽으로 움직인 현재 프레임의 객체(O_n) 간의 프레임 차이(FDM)는 객체가 겹쳐진 영역을 정지영역으로 판정하고 겹쳐지지 않은 영역은 움직임 영역으로 판정하게 된다. 따라서 현재 프레임에서 객체는 움직임 영역과 정지영역의 합으로 구할 수 있다. 그림 13의 (b)는 실제 깊이 영상 내에서의 정지 영역과 움직임 영역을 보여주는데, 흰색 부분이 움직임 영역이고 검은색 부분이 정지 영역이다. 이와 같은 처리에 의해 사용자 객체와 접한 정지되어 있는 사물은 사용자 객체에서 제외된다. 제안하는 방법을 이용하여 깊이의 유사성에 의해서 사용자영역이 아닌 다른 영역이 사용자로 검출되는 오류를 줄일 수 있다. 그림 14는 움직임정보를 사용하지 않아서 오류가 나타나는 경우와 움직임정보를 사용하여 오류가 나타나지 않는 제안하는 방법에 대한 결과를 보여준다. 움직임정보를 사용하지 않고 DPM을 통한

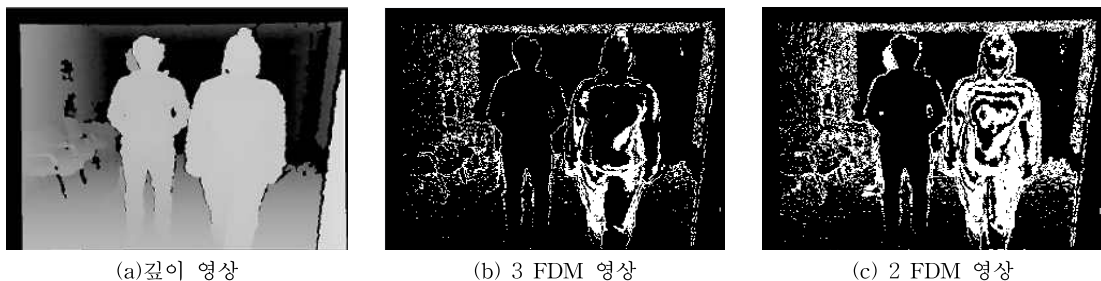
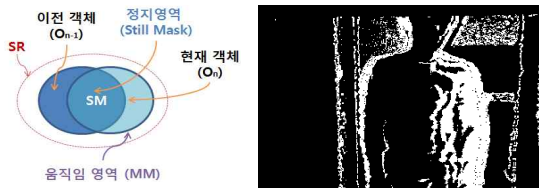


그림 12. 깊이 영상에 대한 3 FDM과 2 FDM 비교



(a) SM과 MM의 정의 (b) 실제 영상에서의 SM, MM
그림 13. 정지 영역(SM)과 움직임 영역(MM)의 정의와 예

깊이의 유사성만을 사용하여 객체를 추적한 결과 그림 14의 (a), (b)와 같이 추적하던 사용자가 접촉한 배경영역까지 함께 사용자로 검출되는 오류가 나타났다. 또한 (c)와 같이 몸과 손이 유사성의 차이로 분리되는 경우 손과 같은 작은 영역들을 함께 검출하지 못하는 단점도 나타났다. 반면 본 논문에서 제안하는 움직임정보를 함께 사용하는 방법은 그림 14의 (d), (e), (f)와 같이 정확한 추적이 가능함을 확인할 수 있다.

사용자 객체 추적을 통해서 검출된 객체는 모폴로지 필터 처리를 수행하고 추적된 사용자 객체로 등록되며 사용자 객체 정보는 업데이트 된다. 여기서 모폴로지 필터는 최종 후처리 과정으로 객체 추적과정의 최종 객체의 영역의 주변에 작은 잡음들이나 영역 중간에 작은 구멍들이 생길 수 있는데 이를 해결하기 위하여 사용하였다. 이와 같은 추적과정은 초기 객체로 등록된 모든 객체에 대해서 개별적으로 수행되고, 각 수행과정에서 타 객체와의 겹침 부분은 앞선 객체

에게 우선순위를 부여하여 판단하게 된다. 각 객체의 추적에 대해 만약 전체 객체 후보를 다 검색하여도 해당 객체에 대한 후보가 없을 경우에는 사용자 객체를 놓친 것으로 판단한다. 사용자 객체의 놓침 현상에 대해서는 영상 밖으로 나가서 놓친 경우에는 추적이 끝난 것으로 파악하고 등록 객체를 제거한다 (Miss). 다른 물체에 의해서 가려지거나 너무 멀리 떨어져서 객체 검출 범위 밖으로 벗어난 경우는 사용자 객체 정보를 숨겨짐(Hide)으로 등록하고 최종 검출되었던 객체정보와 위치 정보를 지니고 있다가 숨겨진 객체와 가까운 거리에서 새로운 객체가 나타나는 경우 동일한 객체로 판단하고 추적을 수행한다. 앞서 객체 검출 방법에서 언급한 것과 같이 각 객체는 객체의 상태(Detected: 검출됨, Hide: 영상 내에서 객체 숨김, Miss: 영상 밖으로 객체가 완전히 사라짐) 정보를 가지고 있고, 이와 함께 객체의 위치정보를 이용하여 객체의 새로운 등록, 또는 재 추적을 수행할 수 있다. 그림 15는 3명의 객체의 검출 및 추적의 결과를 보여준다. (a)~(c)까지의 영상에서 각 객체는 Hide되어 있는 객체가 없기 때문에 검출 시 마다 신규 등록이 되었다. 그리고 (d)는 객체의 겹침에서의 추적 과정을 보여주고, (e)는 겹침에 의한 객체의 Hide 결과를 보여준다. (f)는 Hide된 객체가 있었기 때문에 새로 검출된 객체를 Hide된 객체로 판단하고 재 추적함을 보여준다.

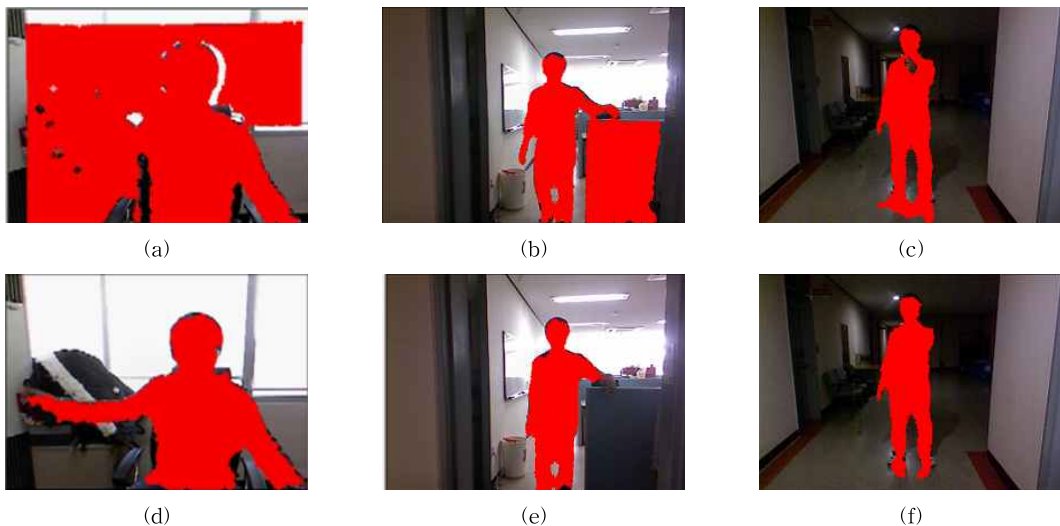


그림 14. 움직임정보를 이용하지 않은 경우와 사용한 경우의 결과 비교

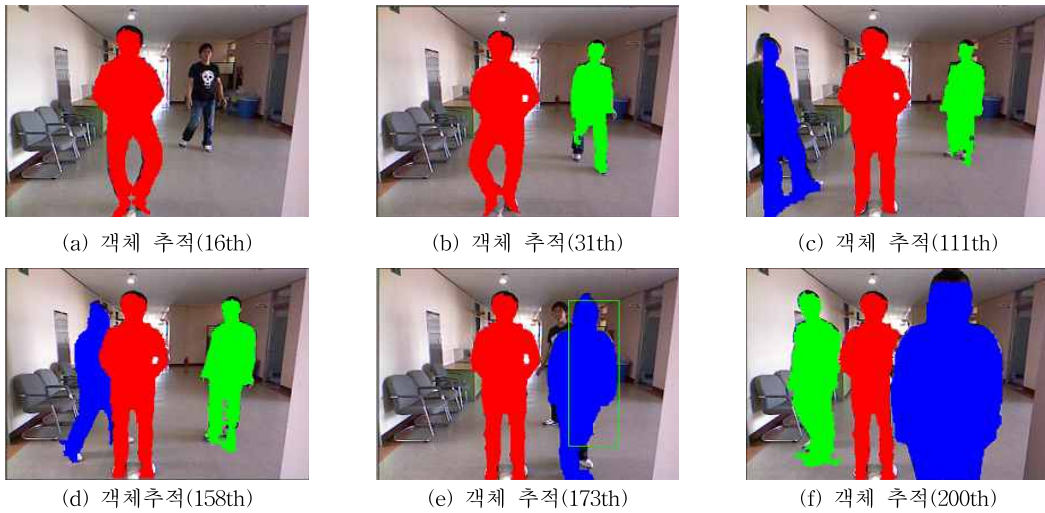


그림 15. 사용자 객체 추적 및 업데이트 결과

4. 실험 결과 및 분석

제안하는 다중 객체 검출 및 추적 알고리즘의 성능을 평가하기 위해 다양한 상황에 대해 실험하였다. 실험영상은 MS사의 KINECT를 이용하여 획득한 영상으로 본 논문에서 언급하는 객체 검출 및 추적은 자세인식을 위해 수행함을 전제하에 실내 장소인 연구실, 복도 등에서 상체만 나온 영상, 복잡한 배경에서의 영상, 물체와의 겹침 영상, 사용자간의 겹침 영상 등의 다양한 깊이 영상들을 사용하였다. 제안하는 알고리즘의 처리 속도는 AMD Phenom II ×6 2.80GHz CPU에서 640×480 해상도의 30fps의 입력영상을 사용하여 1명의 사용자 검출 및 추적 시에는 23fps(0.044s/f), 2명의 사용자에는 21fps(0.048s/f), 3명의 사용자에는 19fps(0.053s/f)의 연산 속도로 측정되었다. 제안 알고리즘은 실험결과 6~7명까지 인식 되지만 영상 내에서 4명 이상의 사용자가 영상 내에 존재 할 경우 너무 복잡하고 자세인식의

입력으로는 부적합하다고 판단하여 Microsoft Kinect와 마찬가지로 4명까지 검색하는 것으로 제한하였다.

그림 16은 객체가 상체만 있을 때 검출 및 추적한 결과를 나타내고, 그림 17은 복잡한 배경에서의 객체 검출 및 추적 결과를 보여주는데, 상체만 나타난 경우나 복잡한 배경에서도 정확히 사용자 객체 검출 및 추적이 수행됨을 확인할 수 있다.

그림 18은 사용자가 카메라에서 너무 멀어 Hide되는 경우를 보여주는데, 너무 멀 경우에는 객체가 너무 작아서 자세 인식이 부정확해질 사용자 객체는 불필요하기 때문에 추적을 하지 않도록 설계하였다. 그림 19는 객체간의 겹침에 의한 Hide 등록과정을 나타낸 것으로 신규 객체로 인해 기존 객체가 가려진 경우에도 지속적으로 추적이 가능함을 보여준다.

그림 20은 다중 사용자 객체에 대한 검출 및 추적 과정을 보여주는데, 여러 명의 사용자 객체를 동시에 추적하고, 겹침에 의해 사라진 객체는 Hide로 등록하며 사라진 객체가 다시 나타날 경우 재 추적을 수행

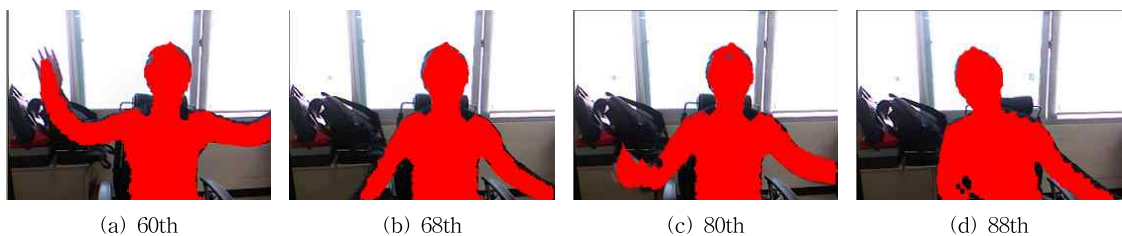


그림 16. 상체만 나온 사용자 검출 예

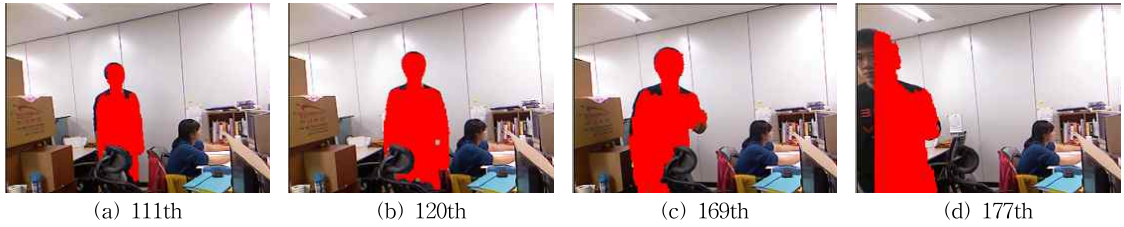


그림 17. 복잡한 배경에서의 사용자 검출 예

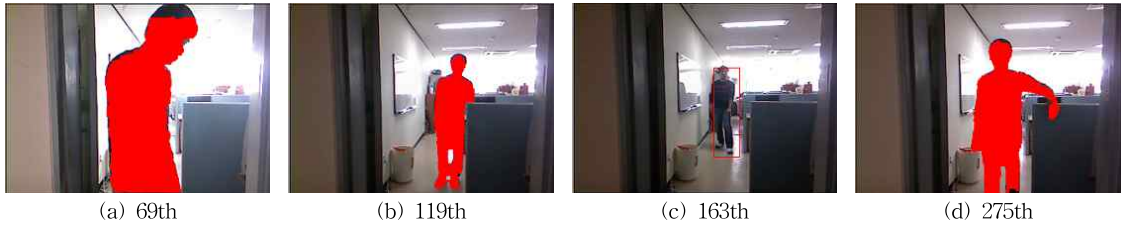


그림 18. 사용자 객체 너무 먼 경우 Hide 되는 예



그림 19. 사용자 객체간 겹쳐서 Hide 되는 예

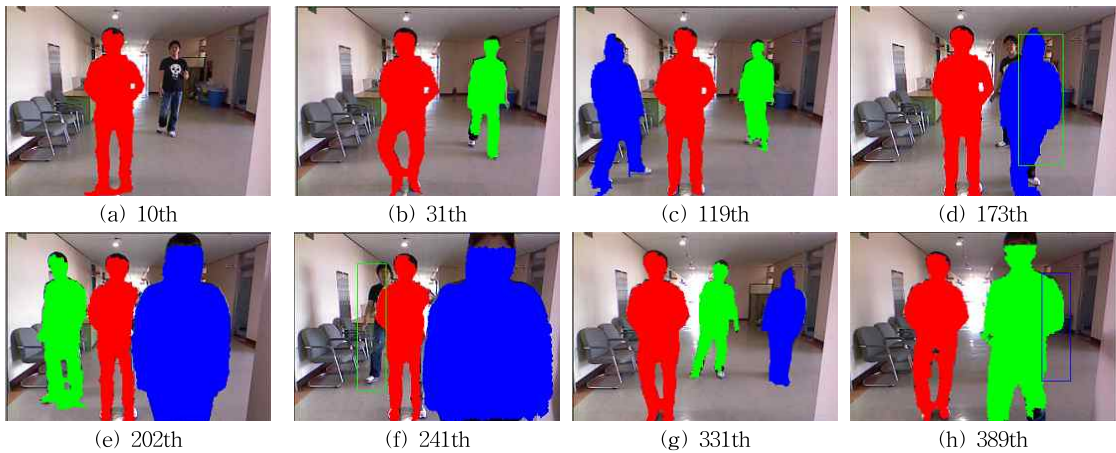


그림 20. 다중 사용자 객체 검출 및 추적 예

함을 알 수 있다. 이때 객체의 색을 보면 객체 고유번호가 바뀌지 않고 추적하고 있음을 확인할 수 있다.

본 논문에서 제안하는 방법의 정확성 판단 방법으로 프레임 단위로 사용자 검출 및 추적을 성공한

True와 사용자가 아닌데 사용자라고 검출하는 오류인 False Positive, 사용자가 있는데도 검출하지 못한 경우인 False Negative에 대한 결과를 이용한다. 그림 21은 False Negative의 예로 (a)와 (b)는 영상의

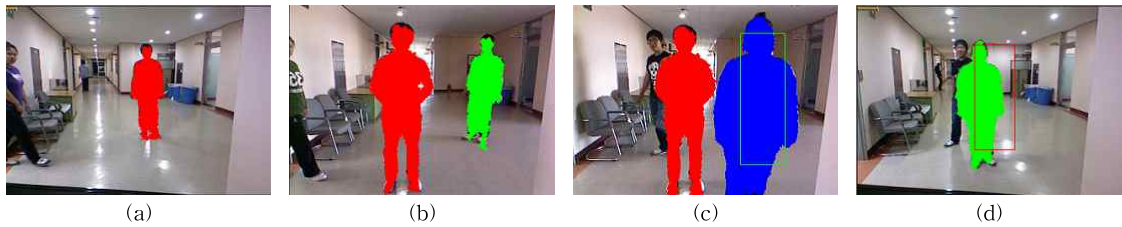


그림 21. False Negative의 예

왼쪽에 사용자가 나타났지만 비율의 오류로 검출하지 못한 결과이고, (c)와 (d) 역시 사용자 간의 겹침에 의하여 맨 뒤의 사용자가 사라졌다가 다시 나타나는 경우에 비율의 오류로 검출이 되지 않은 경우로 이런 경우와 그 밖의 영상 내에 사용자가 존재하지만 검출하지 못한 결과를 검출 오류인 False Negative로 지정하였다. False Positive는 영상에서 움직이는 객체 외의 다른 영역이 객체로 검출되는 경우인데 본 논문에서 제안하는 알고리즘의 결과에 나타나지 않았다.

본 논문에서는 앞서 언급한 바와 같이 1명부터 3명에 대한 사용자가 여러 가지 환경에서 임의의 방향으로 서로 겹치면서 움직이는 영상을 실험영상으로 사용하였다. 1명인 영상 4set, 2명인 영상과 3명인 영상 6set에 대해 실험을 하였고, 각 영상은 1명인 영상

은 400 프레임 내외이고, 2명과 3명인 영상은 500프레임 내외이다. 표 1은 위의 조건으로 제안 알고리즘을 실험한 결과로써 위의 그림들에서 주관적인 성능을 보여준 것과 함께 객관적인 정확성 근거를 보여준다. 1명인 상황에서는 매우 높은 정확성을 보여주고, 2명이나 3명일 때 역시 충분히 사용 타당성이 있는 정확성 결과를 보여준다. 이 때 1명의 4set와 2명, 3명의 6set는 물체와의 겹침, 사람간의 겹침, 다양한 이동방향, 이동속도, 포즈, 다양한 설정에 대해서 실험을 하였기 때문에 충분한 정확도 측정 타당성이 있다고 판단된다.

그림 22는 parvizi[6]의 논문의 방법과 본 논문에서 제안하는 방법을 같은 영상에 적용한 결과를 보여준다. parvizi의 논문은 depth pdf를 이용하여 깊이의

표 1. 제안 알고리즘 성능 평가

Situation	Correct	False Positive	False Negative	Accuracy
1명, 4set	1555/1566	0	11	99.2%
2명, 3명 6set	3060/3152	0	92	97.1%
Total	4615/4718	0	103	97.8%

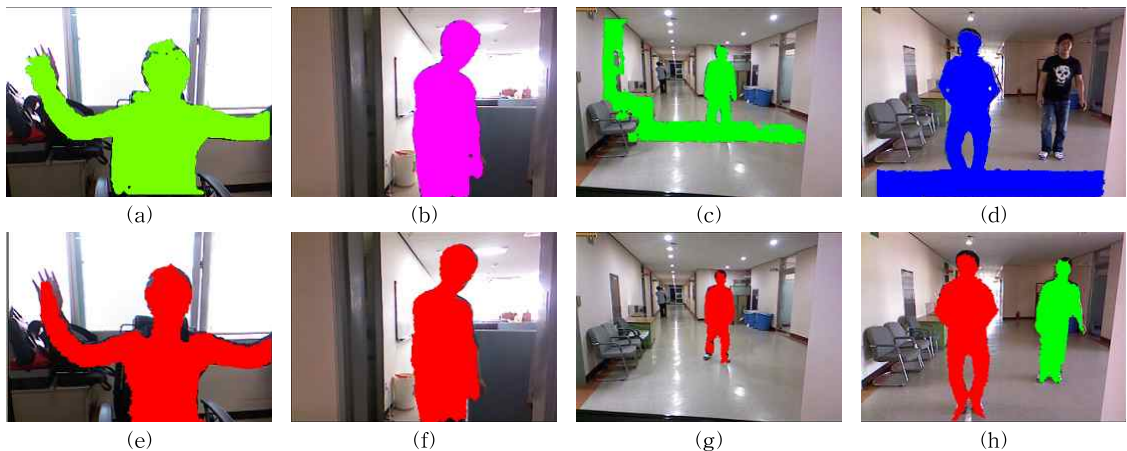


그림 22. depth pdf 방법[6]과 제안 방법의 비교

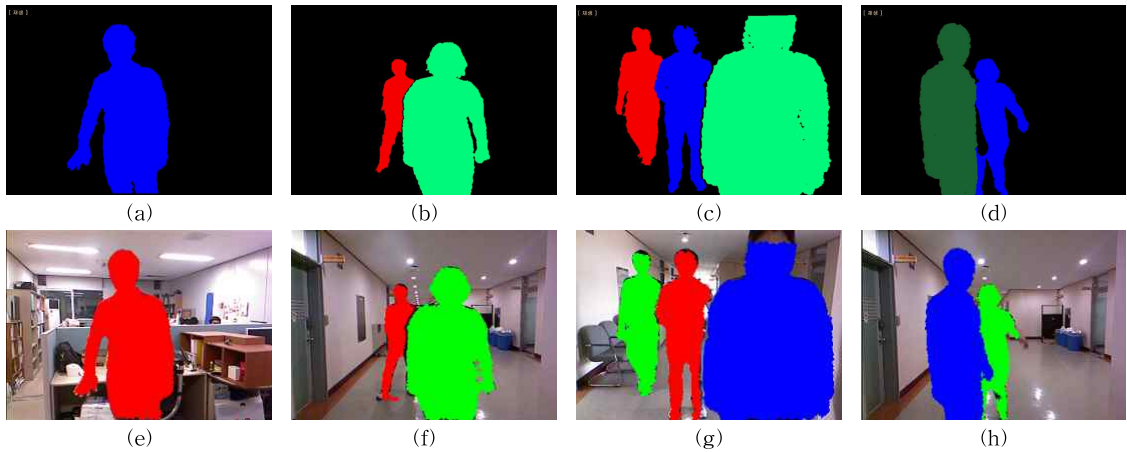


그림 23. MS사의 OpenNI 방법과 제안 방법의 비교

유사도를 통한 사용자 검출 및 추적을 수행하는데, depth pdf에 의한 영역분할 과정에서 그림 22의 (a), (b)와 같이 영상 내에 바닥이 없는 경우에는 제안방법과 같이 정확한 검출 및 추적 결과를 보였으나, (c), (d)와 같이 바닥이 있는 경우에는 객체와 지면의 깊이 유사도를 구분하지 못하여 객체 검출 및 추적이 불가능한 단점을 보였다. 반면 제안하는 방법은 그림 22의 (e), (f), (g), (h)와 같이 카메라의 시점과 관계없이 정확한 결과를 보임을 확인할 수 있다. 또한 그림 23은 Microsoft사의 OpenNI에서 제공하는 방법으로 사용자를 검출한 결과와 제안하는 방법의 결과를 비교하여 보여준다. 그림 23의 (a)와 (e), (c)와 (g)는 유사한 결과를 보여주고, (b)와 (f)에서는 제안하는 방법에서 왼쪽 사람 다리 쪽에 지면이 약간 함께 검출되는 오류가 나타났다. (d)와 (h)에서는 제안하는 방법에서 오른쪽 사람의 손이 약간 덜 검출되는 오류가 나타났고, 대신 OpenNI방법에서 사람의 겹침 부위에서 왼쪽사람의 팔이 검출되지 않는 오류는 제안하는 방법에서 나타나지 않았다. 그림 23의 MS의 OpenNI방법과 비교를 통해 제안하는 방법의 타당성을 충분히 확인할 수 있다.

5. 결 론

본 논문에서는 자세인식을 위한 3차원 다중 객체 검출 및 추적을 위해 깊이정보에 대한 움직임과 DPM을 이용하여 객체를 검출하고, DPM과 이전 검출 객체와 움직임을 이용한 객체에 대한 추적

방법을 제안하였다. 제안하는 방법은 깊이 정보를 이용함으로써 조명이나 그림자, 잡음, 복잡한 배경 등 다양한 환경 변화에 강인하고, 깊이 정보만을 사용함으로써 실시간에 적합한 빠른 연산을 수행한다. 또한 깊이정보만을 사용하는 기존의 방법들의 깊이 정보의 유사성의 의한 추적 실패의 단점을 깊이정보와 움직임 정보를 사용하여 추적하는 방법을 제안함으로써 사물과의 겹침이나 사용자 간의 겹침 상황에서도 정확한 검출 및 추적을 수행하는 방법을 제안하고 그 결과에 대해 비교 분석을 통해 타당성을 보였다. 이와 더불어 제안하는 방법은 사용자 객체를 정확하게 검출하고 추적하며 빠른 연산속도로 수행되기 때문에 비접촉식 인터페이스 시스템의 개발 시에 사용자 객체의 자세를 인식하는 알고리즘의 입력으로 충분히 사용 가능하다고 판단된다.

참 고 문 헌

- [1] Shao-Yi Chien, Yu-Wen Huang, Bing-Yu Hsieh, Shyh-Yih Ma, and Liang-Gee Chen, "Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques," *IEEE Transaction on Multimedia*, Vol. 6, No. 5, pp. 732-748, 2004.
- [2] Chia-Feng Juang, Chia-Ming Chang, Jiu-H-Rou Rou, and Lee D, "Computer Vision-Based Human Body Segmentation and Posture Esti-

- mation," *IEEE Transactions on Systems, Man and Cybernetics, Part A, Systems and Humans*, Vol. 39, No. 1, pp. 119-133, 2009.
- [3] Haritaoglu, I, Harwood, D, and Davis, L.S, "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.
- [4] 전재덕, 이미정, 김종호, 김상균, 강병두, "Eigen-background와 Clustering을 이용한 객체 검출 시스템," 멀티미디어학회논문지, 제13권, 제1호, pp. 47-57, 2010.
- [5] Ismail Haritaoglu, David Harwood, and Larry S. Davis, "W4S: A Real-Time System for Detecting and Tracking People in 2.5D," *ECCV'98 In Computer Vision*, pp. 877-892, 1998.
- [6] Parvizi, E, Wu, and Q.M.J, "Multiple Object Tracking Based on Adaptive Depth Segmentation," *Canadian Conference on Computer and Robot Vision 2008*, pp.273-277, 2008.
- [7] Yinghua Shen, Chaohui Lu, and Pin Xu, "Stereoscopic Video Object Segmentation Based on Disparity Map," *Measuring Technology and Mechatronics Automation 2010*, Vol. 3, pp. 493-495, 2010.
- [8] Boykov, Y and Kolmogorov, V, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, pp. 1124-1137, 2004.
- [9] 안정호, 김길전, 변혜란, "그래프 컷을 이용한 강인한 인체 실루엣 추출," 정보과학회논문지, 제34권 제1호, pp. 52-58, 2007.
- [10]곽수영, 변혜란, "휴머노이드 로봇을 위한 사람 검출, 추적 및 실루엣 추출 시스템," 한국통신학회논문지, 제34권, 제6호, pp. 561-669, 2009.

- [11] Y. Ma, S. Worrall, and A. M. Kondoz, "Automatic Video Object Segmentation using Depth Information and an Active Contour Model," *Proc. IEEE 10th Workshop on Multimedia Signal Processing*, pp. 910-914, 2008.
- [12] Comaniciu D and Meer P, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603- 619, 2002.



이 재 원

2004년 3월~2009년 2월 전남대학교 전자컴퓨터공학부 학사
2009년 3월~2012년 2월 전남대학교 전자컴퓨터공학과 석사

2012년 3월~현재 전남대학교 전자컴퓨터공학과 박사과정
관심분야 : 영상처리, 컴퓨터 비전, 멀티미디어시스템



정 지 훈

2003년 3월~2010년 2월 전남대학교 전자컴퓨터공학부 학사
2010년 3월~2012년 2월 전남대학교 전자컴퓨터공학과 석사

2012년 3월~현재 LG이노텍 부품소재연구소 연구원
관심분야 : 영상통신시스템, 영상신호처리, 스테레오 매칭



홍 성 훈

1988년 2월 영남대학교 전자공학과 공학사
1991년 2월 한국과학기술원 전기및전자공학과 공학석사
1999년 2월 한국과학기술원 전기및전자공학과 공학박사

1991년~2000년 7월 LG전자 DTV 연구소 책임연구원
2007년~현재 전남대학교 전자컴퓨터공학부 교수
관심분야 : 의료영상 처리, 영상 부호화, 멀티미디어 시스템