일반논문 (Regular Paper)
방송공학회논문지 제21권 제2호, 2016년 3월 (JBE Vol. 21, No. 2, March 2016)
http://dx.doi.org/10.5909/JBE.2016.21.2.252
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

실내 환경에서 검출 속도 개선을 위한 2D 영상에서의 사람 크기 예측

길 종 인[a)], 김 만 배[a)‡]

Estimating Human Size in 2D Image for Improvement of Detection Speed in Indoor Environments

Jong In Gil[a)] and Manbae Kim[a)‡]

요 약

사람 검출의 성능은 카메라의 위치 및 각도 등에 큰 영향을 받는다. 이로 인해 획득한 2D 영상에서 사람은 위치에 따라 각기 다른 크기를 갖는 형태로 나타난다. 이렇게 다양한 크기를 갖는 사람들을 모두 검출하는 것은 실시간 시스템의 구현을 어렵게 만드는 요인이 된다. 그러나 만일 영상의 특정 위치의 사람의 크기를 예측할 수 있다면, 해당 위치의 사람 검출을 위한 연산량이 크게 감소될 수 있을 것이다. 본 논문에서는 실내 공간의 구조를 깊이맵으로 구성하고, 실내 공간에 존재하는 사람의 영상을 3D 공간에 재구성함으로써 크기를 예측하는 기법을 제안한다. 3D 공간에서는 어느 위치에서든지 사람의 크기가 일관되므로 이를 2D 영상으로 투영하게 되면 2D 영상의 좌표에 따른 정확한 사람의 크기를 추정할 수 있다. 실험 결과로부터 제안 방법이 효과적으로 사람의 크기를 예측할 수 있고, 기존이 기계학습 기반 사람 검출 방법들의 처리속도가 감소됨을 증명하였다.

Abstract

The performance of human detection system is affected by camera location and view angle. In 2D image acquired from such camera settings, humans are displayed in different sizes. Detecting all the humans with diverse sizes poses a difficulty in realizing a real-time system. However, if the size of a human in an image can be predicted, the processing time of human detection would be greatly reduced. In this paper, we propose a method that estimates human size by constructing an indoor scene in 3D space. Since the human has constant size everywhere in 3D space, it is possible to estimate accurate human size in 2D image by projecting 3D human into the image space. Experimental results validate that a human size can be predicted from the proposed method and that machine-learning based detection methods can yield the reduction of the processing time.

Keyword : human size estimation, camera calibration, image projection, depth map

a) 강원대학교 컴퓨터정보통신공학과(Dept. of Computer and Communications Engineering, Kangwon National University)
‡ Corresponding Author : 김만배(Manbae Kim)
E-mail: manbae@kangwon.ac.kr
Tel: +82-33-250-6395
ORCID: http://orcid.org/0000-0002-4702-8276
※2015년도 강원대학교 대학회계 학술연구조성비로 연구하였음 (관리번호-520150466).
· Manuscript received January 14, 2016; Revised March 3, 2016; Accepted March 3, 2016.

Ⅰ. 서 론

실내 환경에서 사람 검출 방법은 많은 연구가 이루어지고 있고, 나날이 발전하고 있다. 주로 기계학습 및 특징기반 방법을 통해 생성한 분류기를 이용하여 사람인지 아닌지를 판단한다. 그러나 이러한 판단 방법들은 많은 연산을 수행해야 하는데, 이중 한 가지 문제는 카메라 센서의 기하학적 원근투영(perspective projection)으로 발생하는 영상 내 좌표에 따른 사람 객체 크기의 변화이다. 예를 들어 객체가 카메라에 근접하면 크게 보이고, 반대이면 작게 나타나는 현상이 발생한다. 이러한 문제를 해결하기 위해 사람 검출기는 영상 해상도를 조절하는 이미지 피라미드를 이용한다[1-3]. 영상을 1/2, 1/4, 1/8 등으로 스케일링하여 다중 해상도 기반 검출을 수행한다. 따라서 처리속도는 사용하는 해상도에 비례하여 증가하게 된다. 이러한 관점에서 보면 만일 영상의 각 픽셀 좌표에 있는 객체의 크기를 미리 알고 있다면, 수행시간을 크게 절약할 수 있게 된다. 본 논문에서는 단일 카메라로 촬영한 영상의 각 좌표에서의 사람의 크기를 예측할 수 있는 반자동(semi-automatic) 방법을 제안한다. 이 기법은 주로 실내 환경에서 유용하게 사용될 수 있고, 실내에 카메라를 설치할 때에 전처리 과정으로 미리 사람의 크기 정보를 검출기에 전달함으로써 검출 속도의 향상을 줄 수 있다.

Park 등은 peak-weighting 기법을 이용하여 동영상에서 사람이 이동하여도 거리와 객체의 면적을 이용하여 객체로 판단된 영역의 모든 픽셀에 가중치를 계산하고, 가중치 합을 계산하여 객체의 크기를 추정하도록 하였다[4]. 이를 위해 카메라의 화각, 카메라의 설치 높이 등과 같은 다수의 파라메타 정보를 미리 알고 있어야 한다. 그러나 객체가 서로 인접해 있는 경우 두 객체를 분리해낼 수 없으므로 크기를 예측할 수 없고 또한 영상에서 사람이 원거리에 있고 작은 물체가 카메라에 근접하면 크기가 유사하게 얻어지는 단점이 있다. 반면 Kispal 등은 영상으로부터 객체의 실제 크기를 추정하는 방법을 제안하였다[5]. 감시 카메라는 상당한 왜곡을 가지고 있다는 가정하에, 카메라 캘리브레이션을 통해 카메라의 왜곡을 보정하고, 보정된 영상으로부터 크기를 예측하였다. 단점은 영상에서 사람의 전체 외형이 보여져야 하고, 또한 이 외형이 정확하게 얻어져야 한다. Li 등은 2D 영상에 투영된 정보를 이용하여 3D 공간을 구성하고 여기서 사람을 탐색하는 방법을 제안하였다[6]. 2D 영상을 다수의 그리드 영역으로 분리하고 각 그리드에 사람이 존재하는 지를 분류하는 방식을 사용한다. 이러한 방법은 카메라의 포즈(pose) 및 렌즈 왜곡 등에 영향을 받기 때문에 3D 공간으로부터 왜곡 및 기울기가 보정되도록 투영변환을 수행하였다. Zeng 등은 이를 다중 카메라로 확장하였다[7]. Li의 방법과 유사한 방법을 이용하여 장면을 다중 카메라의 영상으로 투영시켰고, 각 시점에서의 검출 결과를 융합하여 검출의 정확도를 높였다. Park과 Kispal의 연구에서는 주로 객체 구별 및 왜곡 보정이 목적이고, Li와 Zeng의 연구는 검출의 정확도를 높이기 위한 전처리 과정에 초점이 맞추어져 있었다. 이와 달리 본 논문의 목적은 2D 영상에 존재하는 사람의 검출 속도를 개선하기 위해 2D 영상에 투영된 사람의 크기를 미리 추정하는 것이다. 2D 영상을 3D 공간에 역 투영함으로써 2D 영상에 존재하는 사람의 크기를 측정한다. 비록 좌표에 따라 사람의 크기가 달라지지만 3D 공간에서는 어느 위치에서든지 객체의 실제 크기는 변하지 않는다. 이를 다시 2D 영상으로 투영한다면 정확한 사람의 크기를 예측할 수 있다. 또한 구현 단계에서는 사람의 일부가 가려져 있더라도 활용이 가능하다. 스테레오 비전과 달리 단일 카메라을 사용하기 때문에 효율적인 반자동 기법을 활용한다.

본 논문의 구성은 다음과 같다. 먼저 Ⅱ장에서는 제안하는 방법의 전체 흐름을 설명한다. Ⅲ장에서는 사람 크기 추정을 위한 전처리로써 깊이맵을 추출하는 방법을 소개하고, Ⅳ장에서 추출한 깊이맵을 이용하여 사람의 크기를 추정하는 방법을 제안한다. Ⅴ장에서 수행한 실험의 결과를 분석하고, 마지막으로 Ⅵ장에서 결론을 맺는다.

Ⅱ. 제안방법

그림 1은 제안 방법의 전체 흐름도를 보여준다. 먼저 2D 영상으로부터 기본적인 실내 구조에 대한 정보가 필요하다. 먼저 바닥면(ground plane)을 결정한 후에 3D 공간을 설정

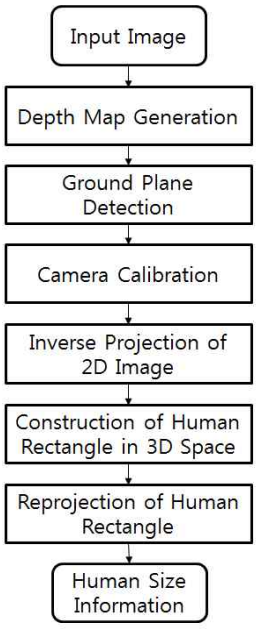

그림 1. 제안하는 방법의 전체 흐름도
Fig. 1. Overview of the proposed method

하고, 깊이맵(depth map)을 생성한다. 그리고 2D 영상을 3D 공간에 역투영 하기 위해 카메라 캘리브레이션(camera calibration)이 수행한 후에 획득한 내부 파라메터(intrinsic parameter)와 외부 파라메터(extrinsic parameter)를 이용하여 바닥면을 3D 공간에 대응시킨다. 3D 공간에 구성된 바닥면으로부터 사람을 직육면체의 형태로 생성하고 직육면체의 8개 정점(vertex)을 다시 2D 영상으로 재투영한다. 투영된 정점의 좌표를 관찰함으로써 사람의 크기를 추정하는 것이 가능하다. 각 블록은 다음장에서 자세히 설명한다.

Ⅲ. 실내 공간의 깊이 구조 예측

사람의 크기를 예측하기 위해서는 3D 공간에 대한 깊이맵을 필요로 한다. TOF 카메라와 같은 깊이 카메라를 이용하여 깊이맵을 획득할 수도 있지만, 깊이 카메라를 이용하게 되면 내부에 존재하는 책상 및 의자 등의 물체들로 인해 정확한 깊이맵 및 실내 구조를 획득하는 것은 어려움이 많다. 사람의 크기 예측을 위해서는 이러한 불필요한 객체는 모두 제거하고, 내부의 공간정보만 획득하는 것이 필요하다.

깊이맵을 생성하기 위해 삼각 분할(triangle decomposition) 기법을 이용한다. 삼각 분할 기법은 딜러니 삼각화[8](Delaunay triangulation)와 구로 쉐이딩[9](Gouraud shading)의 두 단계로 이루어져 있다. 실내 환경은 다수의 벽과 천장 및 바닥이 모여서 공간을 구성한다. 이러한 평면은 폴리곤(polygon)의 형태로 나타낼 수 있다. 따라서 K개의 폴리곤 P_k (k=[1,K])을 생성하고, 각 폴리곤 정점에 수작업으로 깊이값을 할당하면 전체 구조의 깊이맵를 생성할 수 있다. 먼저 영상의 내부 구조를 파악한 후, 천장, 바닥 및 벽이 이루는 교차점을 제어점으로 선택한다. 즉, 제어점은 폴리

(a) (b)

그림 2. 삼각 분할 및 깊이맵, (a) 입력 영상의 10개의 정점(yellow point) 및 분할된 삼각형(white line), 및 (b) 생성된 깊이맵
Fig. 2. Triangle decomposition and depth map. (a) ten vertices (yellow point) and decomposed triangle (white line) of input image and (b) output depth map

곤의 정점에 대응한다. 각 제어점에 깊이를 할당하면 폴리곤의 내부는 모두 자동으로 깊이를 계산할 수 있다. 이를 위해 각 제어점에 대해 카메라로부터 거리를 직접 측정하여 깊이를 할당하였다. 이로부터 각 제어점들의 좌표 및 깊이 $P_k=(x_k,y_k,D_k)$를 할당한 후에 딜러니 삼각화를 이용하여 삼각형으로 분할된다. 마지막으로 구로 쉐이딩으로 삼각형 내부의 모든 픽셀에 대해 깊이를 계산한다. 그림 2는 입력 영상으로부터 폴리곤을 구성하고 각 폴리곤을 구성하고 있는 정점을 노란색으로 표시하였고, 이로부터 획득한 깊이맵을 보여준다.

Ⅳ. 사람 크기 추정

사람의 크기 추정의 첫 단계는 바닥면(ground plane) 추출이다. 바닥면을 추출하기 위해 몇 가지 가정이 존재한다. 먼저 사람은 누워있거나 앉아있을 수 없고 항상 바닥에 지지한 상태로 서있어야 하고, 공중에 떠 있을 수는 없다. 이를 위해 영상으로부터 바닥면을 추출해야 하고, 바닥면을 기초로 하여 사람의 크기를 예측한다. 즉, 바닥면과 인접해 있지 않은 위치에서는 사람의 크기가 결정되지 않는다. 기존에 바닥면을 자동으로 검출하는 몇몇 기존 알고리즘이 연구되고 있으나[10,11], 본 논문에서는 정확도를 높이기 위해서 바닥면을 수작업으로 지정한다.

바닥면 검출 이외에 카메라 캘리브레이션(camera calibration)에서 내부 파라메터(intrinsic parameter) 값을 구하고, 또한 3D 공간을 구성하는 월드 좌표계와 2D 영상 좌표 사이의 변환 관계와 관련된 외부 파라메터(extrinsic parameter) 값을 구한다. 내부 파라메터는 초점거리(focal length), 주점(principal point) 및 왜곡 계수(skew coefficient)와 같은 카메라 내부의 요인을 나타내고, 외부 파라메터는 3D 좌표계에서 카메라의 위치 및 방향과 같은 기하학적인 정보를 포함한다.

카메라 캘리브레이션을 위해 가장 많이 이용되는 방법은 Zhang의 방법이다[12]. 실제 공간에 패턴 영상을 설치해놓고 이를 다양한 각도로부터 촬영한다. 이를 위해 주로 그림 3과 같은 체스보드 패턴이 사용된다. 이렇게 획득한 다수의 영상으로부터 특징점을 추출한 후에, 패턴의 월드 좌표와 이미지 좌표 사이의 호모그래피(homography)를 측정함으로써 파라메터를 계산한다.

그림 3. 내부 파라메터 추정을 위해 사용된 체스보드 마커를 다양한 각도에서 촬영한 영상
Fig. 3. The chessboard image obtained from different angle for estimating intrinsic parameter

외부 파라메터는 OpenCV를 이용하여 구한다. 외부 파라메터를 추정하기 위한 함수는 4개의 파라메터를 입력으로 받는다. 먼저 3D 공간을 구성하려면 원점을 설정해야 하므로 카메라가 촬영할 장소에서 임의의 위치를 원점으로 설정한다. 임의의 마커를 하나 두고 원점으로부터 마커의 네 모서리와의 거리를 측정한다. 그리고 마커의 네 모서리에 대응하는 이미지 좌표를 탐색한다. 이와 같이 마커의 4개의 월드 좌표와 이미지 좌표, 그리고 내부 파라메터 추정 단계에서 획득한 내부 파라메터와 왜곡 계수를 입력으로 설정하면, 외부 파라메터의 3x3 회전행렬 R과 3x1 이동벡터 t를 얻을 수 있다. 내부 및 외부 파라메터가 구해지면 월드 좌표계를 구성된다. 그림 4는 깊이맵과 깊이맵으로부터 추출한 바닥면, 그리고 구성된 월드 좌표계의 예를 보여준다.

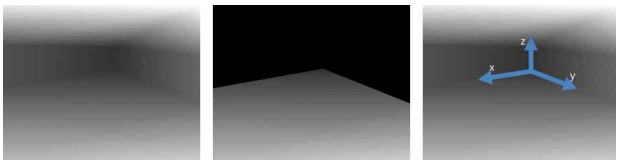

그림 4. 바닥면과 월드 좌표계. (a) 생성된 깊이맵, (b) ground plane 및 (c) 월드 좌표계
Fig. 4. Ground plane and world coordinate. (a) generated depth map, (b) ground plane and (c) world coordinate

좌표계의 원점과 x축, y축 그리고 z축이 결정되면, 3D 공간에서 자유자재로 객체를 위치시키는 것이 가능하다. 바

닥면으로부터 3D 공간상에 사람을 위치시키기 위해서 먼저 2D 영상의 바닥면에 해당하는 픽셀의 좌표로부터 3D 공간상의 좌표를 계산해야 한다. 3D 점 (X, Y, Z)을 다음 식 (1)을 통해 모델링한 변환 관계를 이용함으로써 2D 영상 (u,v)로 투영시킬 수 있다. 그리고 내부 파라메터 K와 외부 파라메터 R, t를 식 (2)와 같이 하나로 결합하여 투영 행렬 (projection matrix) M으로 표현한다면 식 (3)과 같이 변경할 수 있다.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|-Rt] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

$$M = K[R|-Rt] = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

위와는 반대로 2D 영상의 픽셀을 3D 공간상의 한 점으로 변환하기 위해서는 식 (3)으로부터 역투영(inverse projection)을 수행함으로써 월드 좌표계의 좌표를 찾을 수 있다. 역투영은 식 (4)와 같이 표현될 수 있는데, 이때, M^+는 의사 역행렬(pseudo-inverse matrix)로써 식 (5)를 이용하여 측정 가능하다. 또한 u, v는 영상의 좌표이고, z는 해당 좌표의 깊이로써, Ⅲ장에서 획득한 깊이값을 이용한다.

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M^+ \begin{bmatrix} u \\ v \\ d \\ 1 \end{bmatrix} \quad (4)$$

$$M^+ = M^\top (MM^\top)^{-1} \quad (5)$$

만일 바닥면의 어떠한 한 픽셀이 3D 공간상의 한 점으로 매핑되면, 이 점을 바닥점(ground point)으로 설정하고, 이를 기준으로 하여 사람의 볼륨을 직육면체 $(\Delta x, \Delta y, \Delta z)$로 표현한다. 이 점을 그림 5에서 붉은색으로 표시하였다. Δz는 사람의 높이이다. 실험에서는 $(\Delta X, \Delta Y, \Delta Z)$=(30,30,180)(cm)을 사용하였다.

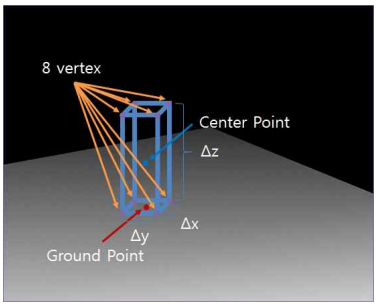

그림 5. 바닥면으로부터 생성한 사람 볼륨의 직육면체의 예
Fig. 5. The example of parallelepiped shape of human volume generated from the ground plane

설정된 직육면체로부터 2D 영상의 각 픽셀의 사람 크기를 계산한다. 직육면체의 8개의 정점 V와 중심점 P_c (그림 5의 center point)를 구한다. 3D 공간상에서 바닥점의 좌표 P_G는 Z=0이므로, $P_G = (X, Y, 0)$ 이다. 중심점의 좌표 $P_c = (X, Y, \Delta z/2)$ 이다. 나머지 8개의 정점 V에 대해서도 유사한 방식으로 식 (6)을 이용하여 좌표 계산이 가능하다.

$$V = (X \pm \frac{\Delta x}{2}, Y \pm \frac{\Delta y}{2}, \frac{\Delta z}{2} \pm \frac{\Delta z}{2}) \quad (6)$$

획득한 P_c 및 8개의 정점 V_i 에 대해서 다시 식 (3)의 투영 행렬 M을 이용하여 2D 공간상에 투영한다. 2D 영상의 투영점을 (u_i, v_i)라 했을 때, 식 (7)과 같이 8개의 2D 점으로부터 최소, 최댓값을 얻으면 식 (8)에서 사람의 크기를 나타내는 경계 상자(bounding box) HS를 생성할 수 있다.

$$\begin{aligned} u_{\min} &= \min_i u_i, \; u_{\max} = \max_i u_i \\ v_{\min} &= \min_i v_i, \; v_{\max} = \max_i v_i \end{aligned} \quad (7)$$

$$HS_{u,v} = [u_{\min}, v_{\min}] \times [u_{\max}, v_{\max}] \quad (8)$$

이는 다음과 같이 해석할 수 있다. 3D 공간상의 중심점을 Pc, 2D 공간상의 중심점을 p_c라 했을 때, 2D 영상에서 사람이 p_c에 위치했을 경우 계산한 경계 상자의 크기를 갖는 사

람을 추출해 낼 수 있고, 중심점은 바닥면으로부터 계산되었으므로, P_c로부터 P_G의 위치가 예측 가능하고, 따라서 해당 사람의 깊이를 계산할 수 있다. 즉, p_c의 좌표로부터 위의 역과정을 통해 깊이를 계산하는 것이 가능해진다.

그림 6은 이러한 과정으로부터 각 픽셀에 대한 사람의 크기를 계산한 결과를 보여준다. 크기는 세로 너비와 가로 너비가 다르기 때문에 Euclidean 거리를 측정하여 값을 할당하였다. 이러한 사람의 크기는 바닥면으로부터 계산하였으므로, 바닥면으로부터 멀리 위치한 천장과 같이 사람이 존재할 수 없는 곳에서는 사람의 크기가 정의되지 않음을 알 수 있다. 즉 사람의 크기가 정의되지 않는 곳에는 사람이 존재할 수 없음을 의미한다.

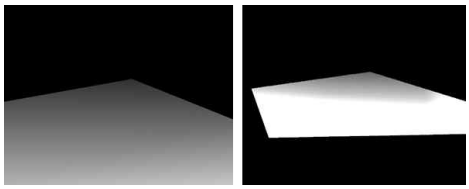

그림 6. 중심점으로 할당된 사람의 크기. (a) 바닥면 및 (b) 사람 크기 맵
Fig. 6. The human size assigned by center point. (a) ground plane and (b) human size map

Ⅴ. 실험 결과

제안하는 방법의 성능을 검증하기 위해서 연구실, 세미나실 및 강의실에서 영상을 직접 촬영하여 실험 영상으로 사용하였다. 실험을 위해 사용된 RGB 카메라 센서의 해상도는 1024x1090이며, 실험이 수행된 PC 환경은 3.4GHz Intel Core i7-3770, 16GB RAM이다.

다음 그림 7은 카메라로부터 입력된 영상으로부터 획득한 중간 결과들을 보여준다. 첫 번째 열에서는 입력 영상으로부터 깊이맵을 생성하기 위해 제어점을 생성하고 각 제어점으로부터 분할된 폴리곤을 보여준다. 두 번째 열에서는 제어점에 할당된 깊이값으로부터 생성된 깊이맵을 보여주고 있다. 깊이맵으로부터 세 번째 열과 같이 바닥면을 쉽게 분리해낼 수 있다. 마지막 열에서는 최종적으로 결정된 바닥면에 존재하는 사람의 크기를 이미지로 나타내었다. 이때 2D 영상에 투영된 사람의 크기는 경계 상자(bounding box)로 표현할 수 있는데, 경계 상자는 다시 수직 길이와 수평 길이로 구성되어 있다. 그림 7의 네 번째 열에서 표현된 사람의 크기는 수직과 수평의 L2 norm을 계산하여 영상으로 표현하였다.

내부 파라메터 추정을 위한 도구로써 GML C++ Camera Calibration Toolbox를 이용하였다[14]. 체스보드 마커를 적

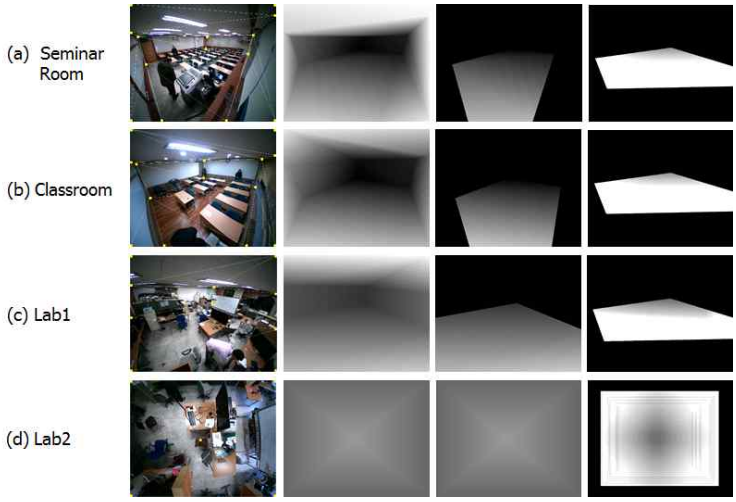

그림 7. 제안 방법을 이하여 획득한 결과. 왼쪽부터 차례로 삼각 분할 결과, 깊이맵, 바닥면 및 각 픽셀에 할당된 사람의 크기맵을 보여준다
Fig. 7. Intermediate results of four test images. The triangle decomposition, dept map, ground plane and human size map in the column order

표 1. 측정된 카메라 파라메터 및 투영 행렬
Table 1. Measured camera parameters and projection matrix

Test image	K	R	t	M
Seminar Room	$\begin{bmatrix} 428.5 & 0.0 & 611.18 \\ 0.0 & 327.74 & 496.08 \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$	$\begin{bmatrix} 0.88 & 0.27 & -0.36 \\ 0.45 & -0.65 & 0.60 \\ -0.06 & -0.70 & -0.70 \end{bmatrix}$	$\begin{bmatrix} -260.91 \\ 3.03 \\ 528.18 \end{bmatrix}$	$\begin{bmatrix} 338.47 & -309.90 & -588.75 & 210989.9 \\ 159.46 & -628.56 & -92.44 & 263319.80 \\ -0.069 & -0.702 & -0.708 & 528.18 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$
Classroom		$\begin{bmatrix} 0.55 & 0.82 & -0.105 \\ 0.17 & -0.23 & -0.958 \\ -0.81 & 0.51 & -0.278 \end{bmatrix}$	$\begin{bmatrix} -350.8 \\ 61.6 \\ 566.8 \end{bmatrix}$	$\begin{bmatrix} -259.3 & 666.1 & -215.1 & 196046.0 \\ -328.2 & 150.8 & -546.3 & 307570.0 \\ -0.81 & 0.51 & -0.27 & 566.80 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$
Laboratory1		$\begin{bmatrix} -0.90 & 0.20 & -0.36 \\ 0.39 & 0.12 & -0.91 \\ -0.14 & -0.97 & -0.19 \end{bmatrix}$	$\begin{bmatrix} 554.75 \\ -163.12 \\ 777.51 \end{bmatrix}$	$\begin{bmatrix} -477.8 & -504.4 & -272.9 & 712972.4 \\ 96.3 & -430.2 & -484.3 & 315939.3 \\ -0.14 & -0.97 & -0.19 & 777.51 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$
Laboratory2		$\begin{bmatrix} -0.99 & -0.11 & 0.13 \\ -0.11 & 0.99 & -0.05 \\ -0.01 & -0.05 & -1.00 \end{bmatrix}$	$\begin{bmatrix} 677.79 \\ -325.96 \\ 255.44 \end{bmatrix}$	$\begin{bmatrix} -430.6 & -77.9 & -604.8 & 446616.7 \\ -51.9 & 400.2 & -516.0 & -12709.4 \\ -0.0 & -0.0 & -1.0 & 255.4 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$

절한 곳에 두고, 카메라를 이용하여 여러 각도로 다수의 체스보드 영상을 촬영하였다. 획득한 영상들을 GML Toolbox 소프트웨어에 입력하면 내부 파라메터 K를 얻을 수 있다. 또한 카메라 외부 파라메터의 획득을 위해 OpenCV 라이브러리를 이용하였다. 3D 공간상의 4개의 월드 좌표와 그에 대응하는 이미지상의 4점의 이미지 좌표를 측정하여 입력으로 넣어주면 카메라 외부 파라메터 R, t를 얻을 수 있다. 표 1은 4가지 실험영상의 카메라 내부 및 외부 파라메터를 보여주며 이로부터 계산된 투영 변환 행렬을 보여주고 있다.

그림 8에서는 각 실험 영상으로부터 측정된 사람 크기

그림 8. 연구실 환경에서 다양한 위치에서 제안방법으로 추정된 사람의 크기 (적색 경계 상자로 표시됨). (a) 세미나 룸, (b) 강의실 및 (c), (d) 연구실
Fig. 8. Estimated human size according to each location in laboratory (marked with red bounding box). (a) seminar room, (b) classroom and (c), (d) laboratory

결과를 보여주고 있다. 실험 결과로부터 제안하는 방법은 사람이 어느 위치에 있건 적절한 크기의 경계 상자를 형성하고 있음을 보여주고 있다. 적절한 크기의 경계 상자가 형성되었다면 이제 이 하위 영상을 분류기에 넣고 사람인지 아닌지 판단하는 과정만 남게 된다. 반면 그림 8에서는 카메라가 설치된 장소에 다수의 사람을 모아놓고 실험한 결과를 보여주고 있다. 그림 7의 결과와 마찬가지로 어느 위치에 존재하더라도 적절한 경계 상자가 형성됨을 확인할 수 있다.

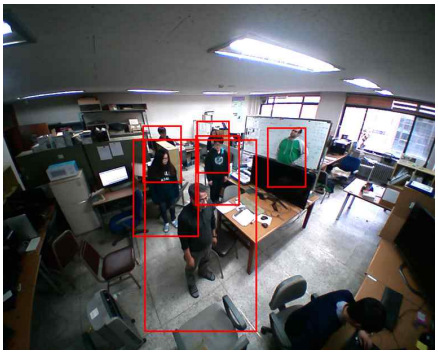

그림 9. 연구실에 위치한 다수의 사람의 추정된 크기
Fig. 9. Estimated size of a large number of people in laboratory

다음 표 2에서는 기존의 기계학습을 이용하여 사람을 검출할 때, 이미지 피라미드를 이용할 경우와 제안하는 사람 크기 추정 방법을 이용하였을 때, 수행속도를 비교한 결과를 보여주고 있다. 사용한 기계학습 기법은 SVM[15]과 AdaBoost[16]를 이용하였고, 분류기 학습을 위한 특징으로는 HOG[17]와 LBP[18]를 이용하여 실험하였다. 실험에서는 Matlab을 사용하였으며, HOG의 특징은 72 차원, LBP의 특징은 59차원을 갖는다. 표 2는 수행 속도는 최소 1.8배에서 최대 31.93배까지 속도가 절약됨을 보여준다. 영상의 모든 픽셀에 대해 하위 영상을 생성하여 분류하게 된다면 상당한 시간이 소요되므로 본 실험에서는 32 픽셀마다 한번씩 하위 영상을 생성하여 분류하도록 하였다.

표 2. 이미지 피라미드를 이용한 검출과 제안 방법을 이용한 검출의 수행 속도 비교 (단위: 초)
Table 2. Running time comparison between image pyramid and proposed method (unit: sec.)

Methods	Image Pyramid	Using Estimated Human Size
SVM+HOG	45.327133	4.500318
SVM+LBP	169.585957	5.310709
AdaBoost+HOG	62.834575	12.097457
AdaBoost+LBP	25.249331	13.953557

Ⅵ. 결 론

본 논문에서는 영상에 존재하는 사람의 크기를 미리 예측함으로써 기계 학습을 이용한 사람 검출의 수행시간을 단축 할 수 있는 방법을 제안하였다. 이를 위해서 반자동 기법을 이용하여 깊이맵을 생성하였고, 깊이맵으로부터 바닥면을 분리해내었다. 사람은 바닥면에 지지하고 있다는 전제하에 2D 영상을 3D 공간으로 역투영 시켰으며, 이로부터 사람의 볼륨 직육면체를 형성함으로써 사람을 나타내도록 하였다. 3D 공간에 생성된 직육면체를 다시 2D 영상으로 투영시킴으로써 2D 영상에 투영될 수 있는 사람의 크기를 예측하였다.

실험 결과로부터 카메라가 설치된 장소 어느 곳에서든지 2D 영상에 투영된 사람의 크기가 예측 가능하였음을 증명하였다. 이러한 사람 예측 기법에서는 무엇보다 카메라 캘리브레이션이 중요하다. 카메라 파리메터 및 깊이맵과 같은 다수의 파라메터는 월드 좌표를 형성하는데 중요한 역할을 하기 때문에 오차가 발생하게 된다면 원하는 월드 좌표계를 형성할 수 없게 되고, 이는 2D 영상에 올바르게 투영되지 못할 것이다.

추후 SVM과 AdaBoost를 제외한 다양한 객체 추적 기법에 본 논문에서 제안하는 사람 크기 추정 방법을 적용함으로써 큰 속도 향상을 가져다 줄 것으로 기대된다.

참 고 문 헌 (References)

[1] V. A. Topkar, A. K. Sood and B. Kjell, ″Object Detection Using Contrast Based Scale-space,″ in Proc, IEEE Conf. Computer Vision and Pattern Recognition, pp. 700-701, June 1999.

[2] P. Dollar, R. Appel, S. Belongie and P. Perona, ″Fast Feature Pyramids for Object Detection,″ IEEE Trans, Pattern Analysis and Machine Intellingence, Vol. 36, No. 8, pp. 1532-1545, Aug. 2014.
[3] N. He, J. Cao and L. Song, ″Scale Space Histogram of Oriented Gradients for Human Detection,″ IEEE Intl. Symposium on Information Science and Engineering, pp. 167-170, Dec. 2008.
[4] M. Park, N. Moon, S. Ryu, J. Kong, Y. Lee and W. Mun, ″A Pixel-Weighting Method for Discriminating Objects of Different Sizes in an Image Captured from a Single Camera,″ in Proc. IEEE 3rd Canadian Conf. Computer and Robot Vision, pp. 36-36, 2006.
[5] I. Kispal and E. Jeges, ″Human Height Estimation Using a Calibrated Camera,″ in Proc, CVPR, 2008.
[6] Y. Li, B. Wu and R. Nevatia, ″Human Detection by Searching in 3D Space Using Camera and Scene Knowledge,″ IEEE 19th Intl. Conf. Pattern Recognition, pp. 1-5, Dec. 2008.
[7] C. Zeng and H. Ma, ″Human Detection Using Multi-camera and 3D Scene Knowledge,″ IEEE 18th Intl. Conf. Image Processing, pp. 1793-1796, Sep. 2011.
[8] P. Cignono, C. Montani and R. Scopigno, ″DeWall: A fast divide and conquer Delaunay triangulation algorithm,″ in Computer-Aided Design, 30(5), 1988, pp, 333-341, 1980.
[9] H. Gouraud, ″Continuous shading of curved surfaces,″ IEEE Trans. Computer, Vol. C-20, Issue. 6, pp. 623-629, 1971.
[10] S. Choi, J. Park, J. Byun and W. Yu, ″Robust ground plane detection from 3D point clouds,″ IEEE 14th Intl, Conf. Control, Automation and System, pp. 1076-1081, Oct. 2014.
[11] J. Arrospide, L. Salgado, M. Nieto and R. Mohedano, ″Homography-based ground plane detection using a single on-board camera,″ IEEE Intelligent Transport Systems, Vol 4, Issue 2, pp. 149-160, June 2010.
[12] Z. Zhang, “A Flexible New Technique for Camera Calibration,″ IEEE Trans. Pattern Analysis and Machine Intellingence, Vol 22, No 11, pp. 1330-1334, Nov. 2000.
[13] J. Moore, ″The Levenberg-Marquardt Algorithm, Implementation, and Theory,″ Numerical Analysis, G.A. Watson, ed., Springer-Verlag, 1977.
[14] http://graphics.cs.msu.ru/en/node/909
[15] C. Chang and C. Lin, ″LIBSVM :a library for support vector machines,″ ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011.
[16] P. Viola and M. Jones, ″Robust Real-time Object Detection,″ In Proc. 2nd Int’l Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling, Vancouver, Canada, July 2001.
[17] N. Dalal and B. Triggs, ″Histograms of oriented gradients for human detection″, IEEE Conference on Computer Vision and Pattern Recognition, Vol 1, pp. 886-893, 2005.
[18] Ojala. T, Pietikainen, M and Harwood. D, ″A Comparative Study on Texture Measures with Classification Based on Feature Distributions,″ Pattern Recognition, 29(1), pp. 51-59, 1996.

저 자 소 개

길 종 인

- 2010년 8월 : 강원대학교 컴퓨터정보통신공학과 학사
- 2012년 8월 : 강원대학교 컴퓨터정보통신공학과 석사
- 2012년 9월 ~ 현재 : 강원대학교 iT대학 컴퓨터정보통신공학과 박사과정
- 주관심분야 : 3D영상처리, 얼굴 검색, 객체 트랙킹, 관심도생성

김 만 배

- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington, Seattle 전기공학과 공학석사
- 1992년 : University of Washington, Seattle 전기공학과 공학박사
- 1992년 ~ 1998년 : 삼성종합기술원 수석연구원
- 1998년 ~ 현재 : 강원대학교 IT대학 컴퓨터정보통신공학과 교수
- ORCID : http://orcid.org/0000-0002-4702-8276
- 주관심분야 : 3D영상처리, 모션관심도, 객체인식 및 트랙킹