

논문 2014-51-9-14

# 적응적 UV-histogram과 템플릿 매칭을 이용한 거리 영상에서의 고속 인간 검출 방법

(Fast Human Detection Method in Range Data using Adaptive  
UV-histogram and Template Matching)

윤 범 식\*, 김 회 율\*\*

(Bumsik Yoon and Whoi-Yul Kim<sup>©</sup>)

## 요 약

본 논문에서는 이전 연구 방법에서의 UV-histogram을 확장하여 적응적 UV-histogram을 제시함으로써, 복잡한 구성의 장면에서 사람의 검출율을 높이는 방법을 제시한다. 제안 방법은 먼저 U-histogram에서 사람 영역을 1차 추출하고, 각각의 레이블링된 U에서 V-histogram을 생성함으로써, 이전 방법에서 구분할 수 없었던 사람 후보 영역을 정확하게 추출한다. 또한 제안 방법은 사람 판정시, 초점거리와 거리에 따라 적응적인 크기를 가지는 오메가 모양의 템플릿을 이용하여 검출의 정확도를 높였으며, 누적 영상을 이용하여 오검출을 템플릿 재매칭 함으로써, occlusion에도 강인한 특성을 가진다. 실험 결과는 Bae의 연구방법에 비하여 복잡한 환경에서 약 15%의 정확도 향상, 80%의 재현율 향상을 보이며, Xia의 연구방법에 비하여 20배 빠른 수행속도를 보여, 제안 방법의 성능이 우수함을 입증한다.

## Abstract

In this paper, a fast human detection method using adaptive UV-histogram and template matching is proposed. The proposed method improves the detection rate in the scene of complex environment. The method firstly generates U-histogram to extract human candidates and adaptively generates V-histogram for each labeled U-histogram, thus it could extract humans correctly, which was impossible in the previous method. The method tries to match the human candidates with the adaptively sized omega shape template to the focal length and distance in order to improve the detection accuracy. It also detects false positives by rematching the template with accumulated foreground images and hence is robust to the occlusion. Experimental results showed that the proposed method has superior performance to the Bae's method in the complex environment with about 15% improvement in precision and 80% in recall and has 20 times faster processing time than Xia's method.

**Keywords** : human detection, UV-histogram, template matching

\* 정회원, 삼성전자  
(Samsung Electronics)

\*\* 정회원, 한양대학교 전자컴퓨터통신공학과  
(Dept. of Electronic and Computer Engineering,  
Hanyang University)

<sup>©</sup>Corresponding Author(E-mail: wykim@hanyang.ac.kr)  
접수일자: 2014년4월24일, 수정일자: 2014년8월27일  
수정완료: 2014년9월11일

## I. 서 론

카메라를 이용한 실시간 사람 검출 방법은 차량 시스템, 지능형 로봇, 보안 및 감시 시스템, HCI (Human Computer Interaction) 등 다양한 분야에서 응용되고 있다. 앞서 설명한 분야에서 사람 검출 방법은 이러한 응

용 분야에서 전처리 과정으로 이용되기 때문에, 이후 적용되는 알고리즘의 성능에 큰 영향을 미친다. 따라서 빠르고 정확한 사람 검출 방법에 대한 연구가 활발히 진행되고 있으며, 크게 RGB 영상을 이용하는 방법과 거리 영상을 이용하는 방법으로 나눌 수 있다.

RGB 영상을 이용하는 방법<sup>[1,2]</sup>은 사람의 윤곽선에 대한 특징값을 추출하고 이를 이용하여 사람을 검출한다. Dalal<sup>[1]</sup>은 영상에서 HOG (Histogram of Oriented Gradient)를 이용하여 특징값을 추출하고, SVM (Support Vector Machine) 분류기를 통해 사람을 검출하였다. Zhu<sup>[2]</sup>는 RGB 영상을 다양한 크기와 위치의 블록들로부터 Dalal의 HOG를 계산하고, 사람 검출에 적합한 블록을 Boosting으로 학습하고 이를 Cascade 형태로 결합하여 사람을 검출하였다. RGB 영상을 이용하는 방법은 연산속도가 느리고 사람이 속한 배경이 복잡할 경우 검출이 어려운 단점이 있다.

거리 영상을 이용하는 방법<sup>[3-4]</sup>은 사람의 머리와 어깨를 연결하는 오메가 ( $\Omega$ ) 모양의 실루엣을 이용하여 사람을 검출한다. Xia<sup>[3]</sup>는 오메가 모양의 실루엣을 Chamfer 매칭으로 사람 후보를 검출하고, 사람의 3차원 머리 모델과 일치하는 후보들을 사람으로 검출하였다. Hyun<sup>[4]</sup>은 오메가 모양의 실루엣을 템플릿 형태로 구성하고, 이를 거리 영상에 템플릿 매칭하여 사람을 검출하였다. 이 방법은 템플릿 매칭으로부터 발생하는 알고리즘의 속도 저하를 방지하기 위해 사람이 영상의 중앙 부근에 위치한다고 가정하여 ROI를 설정하고 연산속도를 향상시켰다. 거리 영상을 이용한 방법들은 배경에 영향을 받지 않는다는 장점이 있지만, 다양한 크기, 다양한 거리에 있는 사람 검출을 위해서는 여전히 일일이 거리 영상 전체에 템플릿 매칭을 수행해야 한다.

Bae<sup>[7]</sup>는 오메가 모양의 템플릿을 이용한 사람 검출 방법의 속도를 UV-histogram을 통해 개선하는 방법을 제안하였다. 이는 UV-histogram을 이용하여 사람 후보 영역을 획득하고 후보 영역에서만 템플릿 매칭을 수행함으로써 수행속도를 크게 향상하였다.

본 연구에서는 UV-histogram 방법을 확장한다. Bae의 방법은 비교적 잡음에 강인하나, 다소 복잡한 환경에서는 좋은 성능을 기대하기 어려웠다. 예를 들어, 사람 근처에 다른 물체나 벽이 근접해 있는 경우, 사람에 대한 거리값 histogram이 물체나 벽과 혼합되어 구분이 어렵게 된다. 사람이 비교적 후방에 존재할수록 이러한

현상은 두드러진다. 거리가 멀어질수록 사람에 의해 가려져 있던 배경에 대한 점구름이 histogram을 많이 차지하게 되기 때문이다. 일반적으로, 사람들과 사람 외의 물체들은 수평 평면상에서 서로 다른 부분을 차지하므로 U-histogram에서는 잘 구분되지만, 수직 평면을 표현하는 V-histogram에서는 카메라와 비슷한 거리에 있는 물체들의 경우는 같은 histogram 간격에 포함되어 구분이 되지 않거나, 거리 축에 대하여 넓게 분포하는 물체들의 경우에는 사람과 물체를 분간하기 어렵게 된다. 또한 이전 방법은 실세계 좌표를 토대로 균등 간격 구간을 나누어 histogram을 생성했기 때문에 거리가 멀수록 histogram 간격(bin)당 쌓이는 점의 개수가 희소해진다는 점도 판별력을 저하시키는 요인이 되었다.

제안하는 적응적 UV-histogram 방법은 기본적으로 U-histogram 생성 후에 레이블링된 물체의 각각에 대하여 적응적으로 V-histogram을 생성하는 방법을 제시함으로써, U와 V histogram를 한 번에 생성하여 적용했을 때 발생하는 위 문제들에 대하여 해결책을 제시한다. 또한 ToF (Time-of-Flight) 잡음 특성을 고려한 필터링 방법을 추가하였으며, 템플릿 매칭에 있어서도 ROI 설정 및 매칭 방법의 개선하여 사람 검출의 성능을 향상시켰다.

본 논문의 구성은 다음과 같다. II장에서는 관련 연구를 간략히 서술하고, III장은 제안하는 방법에 대하여 서술하며, IV장은 제안하는 방법에 대한 실험 결과, 마지막으로 V장에서 연구에 대한 결론을 맺는다.

## II. 관련 연구

이 장에서는 이전 방법에서 제시한 거리 영상에서 UV-histogram를 이용하여 사람 여부를 판단 방법에 대한 연구를 고찰한다. 자세한 기술은 Bae<sup>[7]</sup>에서 찾을 수 있으며, 본 논문에서는 주요 부분인 UV-histogram 방법과 오메가 템플릿 매칭에 대하여 간략히 서술을 한다.

### 2.1 UV-histogram 방법

UV-histogram는 개선된 거리 영상을 기반으로 만들어진다. V-histogram은 거리 영상에서 같은 세로축에 위치하는 거리 값들의 histogram을 쌓음으로써 획득한다. 반대로 U-histogram은 거리 영상에서 같은 가로축

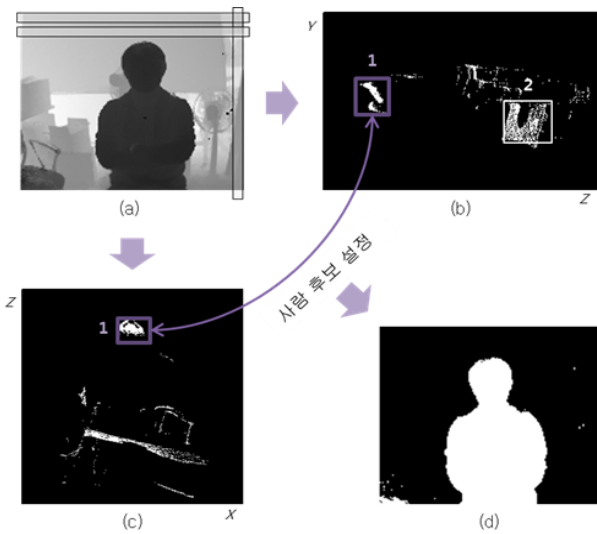


그림 1. UV-histogram 생성 및 레이블링의 예  
Fig. 1. Example of UV-histogram generation and labelling.

에 위치하는 거리 값들의 histogram을 쌓음으로써 획득한다. 획득한 UV-histogram에서 일정한 개수 이상으로 누적된 거리 값만 이용하여 이진화된 UV-histogram을 획득한다. 이 과정을 통하여, 잡음에 강인한 UV-histogram을 획득할 수 있다.

여기서 UV-histogram을 사용하는 주된 이유는 거리 영상에서 평면으로 표현되는 물체들이 UV-histogram에서 가로선 혹은 세로선으로 표현되어 물체 추출이 용이하기 때문이다. UV-histogram에서 선으로 표현된 물체를 사람의 후보 영역으로 가정하고 이를 획득하기 위하여 레이블링을 수행한다. U-histogram에서 획득한  $n$ 개의 레이블 집합을  $L_U = \{L_U^1, \dots, L_U^n\}$ 라고 하고, V-histogram에서 획득한  $m$ 개의 레이블 집합을  $L_V = \{L_V^1, \dots, L_V^m\}$ 라고 하자. 이때  $L_U$ 의  $i$ 번째 레이블과  $L_V$ 의  $j$ 번째 레이블이 동일한 평균 거리 값을 가지고 있다면, 이를 집합  $P = \{(L_U^i, L_V^j), \dots\}$ 에 포함시킨다. 집합  $P$ 에 속한 레이블 쌍에서 획득할 수 있는  $x, y, z$  범위를 이용하여 거리 영상을 이진화 하고 사람의 후보 영역을 획득한다. 그림 1은 이진화된 UV-histogram의 생성 및 레이블링 예를 나타낸다. 그림 1의 (a)는 중간값 필터를 적용하여 개선된 거리영상을 나타내며, (b)와 (c)는 각각 이진화된 사람 후보 영역을 나타낸다.

## 2.2 오메가 템플릿 매칭

다양한 크기의 사람을 검출할 때 발생하는 추가적인 매칭 과정을 제거하기 위하여, 템플릿의 크기를 근사적으로 나타내는 곡선을 사용한다. 거리에 대한 템플릿 크기의 곡선은, 60cm부터 300cm까지 30cm 간격값  $X$ 로부터 사람의 상반신 길이 화소수  $Y$ 를 구한 9개의 데이터 집합을 가지고 수식 (1)에 최소자승법 (least square method)으로 맞추어 구한다.

$$Y = \frac{a}{X} + b \quad (YX = bX + a) \quad (1)$$

측정 데이터를 이용한 곡선 맞춤 결과로  $a = 6.7223 \times 10^3$ ,  $b = 1.6167$ 의 값을 얻었다.

일반적으로, 다양한 사람의 크기에 대응하기 위한 오메가 모양의 템플릿은 그림 2와 같이 영상 피라미드로 구성된다<sup>[6]</sup>. 피라미드에 속한 오메가 모양의 템플릿을 거리 영상에 반복적으로 매칭하여 다양한 크기의 사람을 검출한다. 하지만, 본 방법은 거리에 따른 템플릿의 크기를 곡선으로 모델링하여 사람 후보가 위치한 거리에 알맞은 크기의 템플릿을 생성하여 매칭하므로, 연산 시간을 줄일 수 있다.

최종적으로, 구해진 사람 후보 영상과 오메가 모양의 템플릿 이미지를 이용하여 템플릿 매칭을 수행한다. 사람의 후보 영상  $I$ 에서 오메가 템플릿  $T$ 를 슬라이딩시키면서 매칭을 수행한다. 매칭시 발생하는 계산을 최소화하기 위해, 수식 (2)과 같은 간단한 픽셀값의 제곱차 (squared difference) 매칭 방법을 사용한다. 수식 (2)은 완벽하게 템플릿과 일치하면 0을 반환하고, 템플릿과 일치하지 않을수록 큰 값을 반환한다.

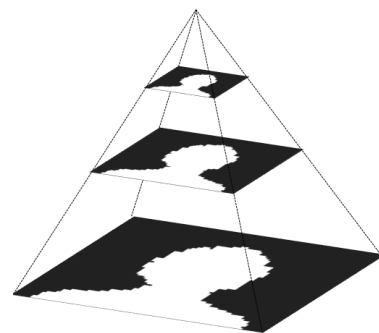


그림 2. 오메가 템플릿 피라미드 예  
Fig. 2. Example of Omega Template Pyramid.

$$M(x, y) = \sum_{x', y'} [T(x', y') - I(x + x', y + y')]^2 \quad (2)$$

### III. 제안하는 방법

본 장에서는 제안하는 방법에 대하여 기술한다.

그림 3은 제안 방법의 흐름도이다. 먼저 입력 거리 영상에 대하여 ToF 잡음 특성을 고려한 거리 영상에 대한 개선 방법을 적용하고, UV 도메인에서 사람 영역 후보를 추출한다. UV 도메인에서는 먼저 측정 거리를 고려해 U-histogram을 생성하고 후보 영역의 크기를 고려해 레이블링을 하며, 영역 구별을 위하여 각 레이블링된 U blob로부터 적응적으로 V-histogram을 생성 및 레이블링 한다. 그리고, 검출된 후보 영역에 대하여 occlusion에 강인한 보다 개선된 템플릿 매칭을 수행하여 최종 사람을 검출한다.

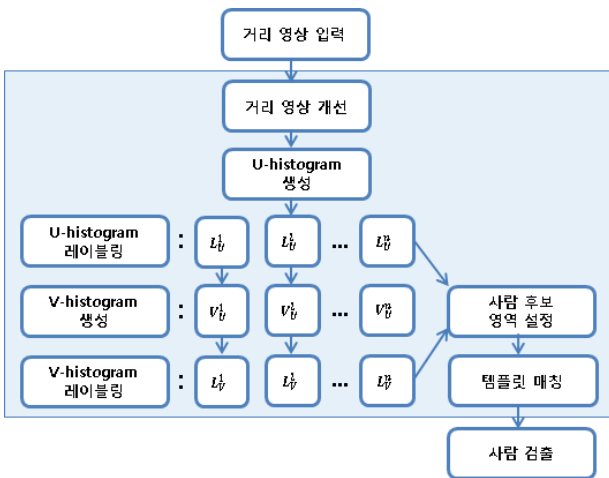


그림 3. 적응적 UV-histogram를 이용한 사람검출 흐름도  
 Fig. 3. Flowchart for the people detection by adaptive UV-histogram.

#### 3.1 거리 영상 개선

본 연구에서는 이전 방법에서의 거리 영상의 개선 방법을 강화하였다. 이전 방법은 비교적 가까운 거리의 사람을 타겟으로 하였고 때문에 중간값 필터를 이용하는 것만으로도 충분한 효과를 기대할 수 있었다. 그러나, 타겟 거리가 멀어질수록 거리 측정은 부정확해지고, 사람 주위에 물체가 많거나 사람 자체가 움직일 경우,

잡음의 영향이 강해지므로 이에 대한 고려가 필요하다. 그리하여 본 연구에서는 ToF 카메라의 전경과 배경 사이의 날림화소(Flying Pixel) 잡음의 필터링 방법을 적용하였다. 날림화소란 전경과 배경 사이의 경계 부근에서 이상 거리측정값이 발생하는 것을 지칭한다. 이상값은 전경과 배경 사이의 값뿐만 아니라 전경보다 가까운 값일 수도 있다.

날림 화소가 발생하는 이유는 첫째 ToF 카메라 화소의 각 해상도 (Angular resolution)가 작아 한 화소에서 전경과 배경이 동시에 측정됨으로 인하여 발생하는 것이고, 두 번째는 거리 측정을 위한 신호 축적 시간 (Integration Time) 중 전경 물체의 움직임이 발생할 경우, 수신 IR 신호에 위상차가 생겨 발생하는 것이다. 이러한 이상값은 UV-histogram에서 histogram 생성시 어울지 않은 bin에 거리값이 쌓이는 결과를 야기할 수 있다. 간단한 날림 화소의 제거 방법은 저 IR강도(low IR intensity) 영역을 무시하는 것이나, 본 연구 방법은 IR 강도가 낮은 사람의 머리카락 부분을 포함하여 오메가 템플릿을 적용하기 때문에 적용할 수 없다.

본 연구에서는 이상값을 제거하기 위하여 중간값 필터된 거리영상  $D$ 에 대하여 식(3)과 같이 날림화소  $D_f$ 를 추정하고, 해당 화소는 거리 영상에서 제거한다.

$$D_f = D_{f_h} \cup D_{f_v}$$

$$D_{f_u} = \left\{ \begin{array}{l} |D(u, v) - D(u-1, v)| > L_{th}, \\ |D(u, v) - D(u+1, v)| > L_{th} \end{array} \right\} \quad (3)$$

$$D_{f_v} = \left\{ \begin{array}{l} |D(u, v) - D(u, v-1)| > L_{th}, \\ |D(u, v) - D(u, v+1)| > L_{th} \end{array} \right\}$$

이는 인접한 두 화소의 거리차가  $L_{th}$  이상인 것은 histogram bin에서 분리하는 효과가 있다. SR4000을 이용한 실험으로부터,  $L_{th}$ 는 0.3m로 두었을 때 좋은 효과를 보였다.



그림 4. 날림화소 제거 전과 후 거리 영상  
 Fig. 4. Depth images of before and after the removal of Flying Pixel.

그림 4는 필터링의 결과이다. 윤곽선 부근의 이상 거리값이 많이 제거된 것을 알 수 있다.

### 3.2 적응적 UV-histogram의 생성

가로축은 거리영상과 같은 길이를 가지며 세로축은 histogram bin의 개수  $K$ 의 길이를 갖는다.  $K$ 는 관심 거리 영역이  $d_{\min}$ 와  $d_{\max}$ 로 주어지면 (4)와 같이 정해진다.

$$K = \left\lceil \frac{(d_{\max} - d_{\min})}{\text{depth\_resolution}} \right\rceil \quad (4)$$

$\text{depth\_resolution}$ 은 최소한 앞뒤 사람을 구분할 수 있도록 설정되어야 한다. 즉, 동일 인물에 대한 depth histogram은 연속적인 bin에서 값을 가져야 하고, 다른 사람이면  $\text{depth\_resolution}$  이내에 불연속 구간이 존재해야 한다. 사람이 앞뒤로 최소 0.1m 이상 떨어져 있어야 있다고 가정하여 본 실험에서는 해당 값으로 고정하였다.

그리고 거리영상의 거리  $D(u, v)$ 는 식 (5)에 의하여 정수 값을 갖도록 양자화된다.

$$Q(u, v) = \left\lfloor \frac{D(u, v) - d_{\min}}{(d_{\max} - d_{\min})} K \right\rfloor \quad (5)$$

#### 3.2.1 U-histogram 생성

U-histogram 좌표계에서 각  $u$ 에 대하여  $v$ 축을 따라 양자화된 거리로 histogram을 생성하는 것은 식 (6)으로 표현된다.

$$U_{hist}(u, k) = \sum_{v=0}^{V-1} [Q(u, v) = k] \quad (6)$$

최종적으로,  $U_{hist}$ 는 식 (7)과 같이  $H_{th}$  이상의 개수를 문턱값으로 하여 이진화 된다.

$$U_b(u, k) = [U_{hist}(u, k) > H_{th}] \quad (7)$$

$H_{th}$ 의 최대값은  $d_{\max}$ 의 거리에서도 사람의 상반신이  $U_b$ 에서 값을 가질 수 있는 정도이어야 한다. 실험에서 사용된  $H_{th}$  값은, 3.3절의 거리에 따른 오메가 템플릿 크기의 분석 결과에 따라, 15의 값으로 설정되었다. (그림 6 참조)

#### 3.2.2 U-histogram 레이블링

8방향 Connected Component 방법을 이용하여 blob들을 추출한 후, blob의 폭  $w_{blob}$ 와 높이  $h_{blob}$ 가 식 (8)의 조건을 만족하는 blob만 추출한다.

$$\begin{aligned} 0.2m < w_{blob} \frac{U_b^{-1}(u, k + h_{blob}/2)}{f} < 1.5m \\ 0.2m < h_{blob} \frac{d_{\max} - d_{\min}}{K} < 1.0m \end{aligned} \quad (8)$$

여기서  $U_b^{-1}$ 은 histogram의 역투영된 거리값이고,  $f$ 는 초점거리이다. 식 (8)에서  $w_{blob}$ 은 blob이 차지하는 좌우방향 크기로, 설정 범위는 사람이 옆방향으로 서 있을 때의 좌우방향 크기(몸통 두께)가 최소 0.2m, 2-3명의 occlusion이 있을 때를 고려해 차지할 수 있는 좌우방향 크기가 최대 1.5m 라고 상정한 것이며,  $h_{blob}$ 은 blob이 차지하는 앞뒤방향 크기로, 사람이 정면을 바라보고 있을 때의 앞뒤방향 크기가 최소 0.2m, 앞뒤로 차지하는 최대 크기를 1.0m로 상정한 것이다. 범위값을 넓힐 경우, 검출율의 영향은 거의 없으나, 템플릿 매칭 시도 회수가 증가하여 수행시간이 증가한다.

이후 U 레이블링 방법은 기존의 방법과 같다.

#### 3.2.3 U에 적응적인 V-histogram 생성

U-histogram 레이블링된 blob의  $(u, k)$ 를 만족하는 범위에 대하여 식 (9)을 이용하여 histogram을 구한다.

$$\begin{aligned} V_{hist}(k, v) = \sum_{u=v_{\min}}^{u_{\max}} [Q(u, v) = k], \\ k_{\min} \leq k \leq k_{\max} \end{aligned} \quad (9)$$

그리고, 이진화된 V-histogram  $V_b$ 는 식 (10)로 표현된다.

$$V_b(k, v) = [V_{hist}(k, v) > 0] \quad (10)$$

추가적으로, 식 (11)과 같이 V에서도 U에서와 같이 물체의 높이가 관심 영역에 있는 blob만을 추출할 수 있다.

$$0.4m < h_{blob} \frac{V_b^{-1}(u, k + w_{blob}/2)}{f} < 2.2m \quad (11)$$

이는 사람의 상반신 크기와 최대 키 사이만을 추출함을 의미한다.

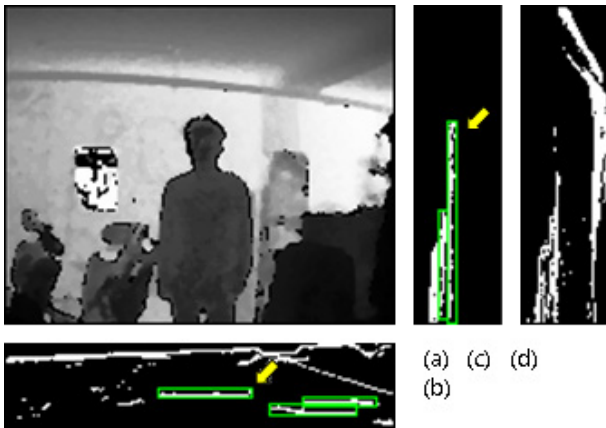


그림 5. 적응적 UV-histogram의 결과  
Fig. 5. Result of adaptive UV-histogram.

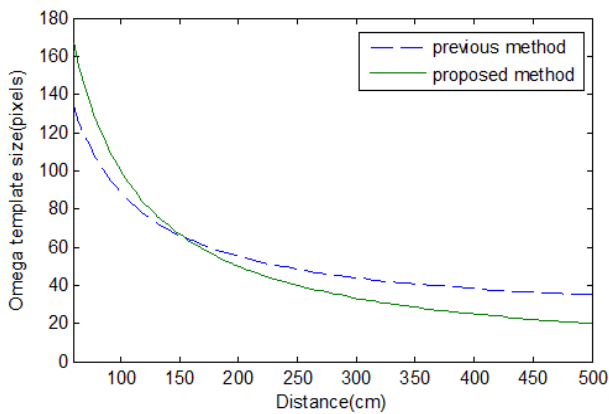


그림 6. 템플릿 크기에 대한 추정  
Fig. 6. Estimation of template size.

최후로, 추출된 레이블은 U 레이블에서의 최대 및 최소  $u$ 와 V 레이블에서의 최대 및 최소  $v$ , UV에서의 공통 최대 및 최소  $k$ 를 가지게 되고, 이는 이후 레이블의 이진화에 사용된다.

그림 5는 레이블링 결과를 보여준다. (a)는 거리 영상이고, (b)는 U-histogram에서 레이블링된 U이며, (c)는 (b)로부터 추출된 레이블에 대하여 V-histogram에서 레이블링된 U를 나타낸다. 그리고, 비교를 위하여 이전 방법에서의 U 없이 홀로 V-histogram을 생성한 결과를 (d)에 나타내었다. (b)에서는 화살표 부근과 같이 앞 또는 옆 사람과 사람이 구분되지 않는다. 또한, V 제작시, U에서와 같은  $H_{th}$ 를 사용하였을 때, 사람의 머리 부분이 제대로 클러스터링되지 않는 경우가 많다.

### 3.2.4 레이블의 이진화

기본적 이진화 방법은 이전과 동일하고, 각 레이블에

서 최대 최소 영역을 완화하여 histogram에서 무시되었을 수 있는 영역을 고려한다. 예를 들어 사람이 몸이 기울어지거나 좌우로 몸을 흔들 경우, 몸의 가장자리 영역은 축적되는 거리값 밀도가 작아 무시되었을 수 있는데, 이러한 상태에서 레이블이 이전 거리 영상으로 변환될 경우, 머리카락이 잘려 사람 검출이 되지 않는다. 이에 대비하여 좌우 10cm, 위 방향으로 30cm에 해당하는  $u$ 와  $v$  값만큼 레이블의 영역을 완화해준다.

## 3.3 사람 검출 과정

### 3.3.1 오메가 모양의 템플릿 크기 조절

본 연구에서는 이전 연구방법에서 보다 원거리의 사람을 검출할 수 있도록 템플릿 크기를 조절하는 방법을 제안한다. 이전 연구에서는 약 3m 거리까지의 측정 데이터를 이용하여 (1)과 같이 least square 방법을 써서 오메가 템플릿의 크기를 구하였다. 이 방법을 원거리 그대로 적용할 경우, 화소 크기의 측정이 부정확하게 되고 결과적으로 거리가 더 멀어질수록 실제 크기와 오차가 커져 정확한 매칭의 결과를 기대하기 어렵게 된다. 제안하는 방법은 사용되는 카메라의 초점거리를 이용하여 템플릿의 크기를 구한다. 핀홀 카메라 모델에 의하여 거리 영상에 적용되는 템플릿의 크기는 어깨부터 머리 위까지의  $\Omega$  모양에 해당하는 물리적 크기인 가로  $x_T$ , 세로  $y_T$ (단위: m)를 이용하여 식 (12)와 같이 구할 수 있다.

$$w_T(d) = \frac{f_x}{d} \times x_T$$

$$h_T(d) = \frac{f_y}{d} \times y_T$$
(12)

$d$ 는 레이블링된 점구름의 중간 거리값이다. 제작된 기준 템플릿은  $x_T, y_T$  모두 약 0.4m의 크기를 가진다. SR4000 스펙<sup>[5]</sup>에 의하면 해당 카메라의 초점거리는 10 mm이고, 픽셀의 pitch는 40 $\mu$ m이다. 그러므로 픽셀 단위의 초점 거리는 250의 크기를 가져야 한다. 실제로 카메라 내부에서 사용되는 정확한 초점거리를 측정하기 위하여, 임의의 측정된 점구름과 OpenCV<sup>[8]</sup>를 이용하여 내부 파라미터를 계산하여 보았다. 측정된 초점거리는 식 (13)과 같다.



$$\begin{aligned} f_x &= 250.4011365322923 \\ f_y &= 250.4096510870294 \end{aligned} \quad (13)$$

계산된 량은 스펙상의 값과 근소한 차이를 보였고,  $f_x$ 와  $f_y$ 는 거의 같고 통일하는 것이 간편하므로, 모두  $f = 250.4$ 로 사용하였다. 그림 6은 기존 방법의 최소자승법을 이용한 템플릿 크기와 제안방법의 초점거리를 이용한 크기의 차이를 보여준다.

### 3.3.2 템플릿 매칭 (Template Matching)

템플릿 매칭 범위는 UV-histogram에서 각각의 레이블링 경계에서 좌우 방향으로  $0.25 \times w_T(d)$ , 상하 방향으로  $0.25 \times h_T(d)$  만큼 확장된다. 템플릿과 거리영상은 각각  $\{-1, 1\}$ 의 값을 가지는  $T_b$ 와  $D_b$  로 이진화하여 식 (14)의 템플릿 크기로 정규화된 상호상관도 (cross correlation)  $R$ 에 의하여 전경 매칭과 배경 매칭이 모두 의미를 가질 수 있도록 한다. 이러한 상호상관도 방법을 사용하면, FFT를 이용할 수 있게 되어 수행속도는 더 빨라질 수 있다.

$$R(x, y) = \frac{\sum_{x', y'} T_b(x', y') D_b(x + x', y + y')}{w_T h_T} \quad (14)$$

$R$ 은  $-1 \leq R \leq 1$ 의 범위를 가지며, 매칭 스코어는  $R$ 의 최대값이 되고,  $R > R_{th}$  일 때 1차 사람 후보로 판정한다.  $R_{th}$ 는 정확도와 재현율에 있어서 트레이드오프의 관계가 있으나, 정확도 100%를 기준으로 최대 재현율을 가지도록 책정하여 0.7의 값으로 설정하였다. 이는 템플릿 영역의 85%가 일치할 때의 스코어에 해당한다.

1차 사람 후보로 판정된 것 중 오메가 모양 주위가 앞의 occlusion이 제거됨에 의해 오검출된 것은, 그림 7과 같이, 앞의 물체들을 누적시킨 것을 더하여 해당 위



그림 7. occlusion에 의한 오검출 재매칭  
Fig. 7. Re-matching of miss-detected by occlusion.

치에서 재매칭 판정함으로써 가려낼 수 있다. 이에 더해, 여러 사람이 같은 거리상에 위치할 경우를 대비하여  $R$ 은  $R_{th}$ 가 넘는 모든 곳을 순차적으로 지워가며 검출한다.

## IV. 실험

제안하는 방법의 성능 검증을 위해, 두 가지 데이터셋으로 그 성능을 측정하였다. 첫 번째 데이터셋(이하 DS1)은 SR4000을 사용한 ToF 데이터셋으로써, 같은 ToF 데이터를 사용하는 Bae의 방법과 비교하기 위하여 사용되었으며, 두 번째 데이터셋(이하 DS2)은 Xia<sup>[3]</sup>의 저자가 제공하는 Kinect for XBOX 360<sup>[9]</sup>(이하 Kinect)을 사용한 데이터셋으로써, Xia의 방법과 비교하기 위하여 사용되었다.

### 4.1 DS1

DS1에서 거리영상은  $176 \times 144$ 의 해상도를 가지며, SET A, B, C, D, E, F로 구성된다. SET A, B는 Bae에서 사용된 SET로, 사람들 사이에 장애물이 거의 없다. SET C, D, E, F는 복잡하고 움직임이 많은 환경에서의 성능을 비교/판단하기 위하여 추가 제작되었다. 해당 SET들은 카메라로부터 최대 거리가 약 7m인 강의실 내에서 촬영되었으며, SET C는 한명이 앞뒤로 움직이고, SET D는 3명이 앉았다 일어서며, SET E는 3명이 앞뒤로 서서 좌우로 움직이고, SET F는 3명이 자유롭게 실내를 걸어 다니는 것을 실시간 촬영하였다. 그림 8은 SET C, D, E, F에서의 결과 장면들 중 하나씩을 제시한 것이다. SR4000의 설정은 modulation frequency를 15Mhz (최대측정거리: 10m), 축적 시간은 설정 최소값으로 하였다.

성능을 정량적으로 측정하기 위하여, 각 SET 별로 정확도(Precision)와 재현율(Recall)을 측정하였다. 정확도는 정확하게 검출된 사람 수를 (정확하게 검출된 사람 수+사람으로 오검출된 수)로 나눈 것이며, 재현율은 각 데이터셋에서 총 검출된 사람의 수를 (전체 프레임 수×프레임별 사람 수)로 나눈 것이다. 프레임별 사람 수는 오메가 부분이 1/3 이상 occlusion 되지 않고 완전하게 프레임 내에 들어오는 사람만 카운트하였다.

표 1에서 알 수 있듯이, SET A, B에서는 두 방법의 성능이 유사하나, SET C, D, E, F에서는 성능 차이가

확연하다. Bae에서의 재현율이 현저히 낮은 이유는, 각 프레임의 UV-histogram 구성시, 장애물과 근접한 인물이 한 blob으로 인식되어 장애물로 오인하는 경우가 많기 때문이며, 제안 방법에서는 U-histogram에 따라 적응적으로 UV-histogram을 구성하였기 때문에 이러한 문제가 개선되었음을 알 수 있다. 또한 Bae에서의 정확도가 상대적으로 낮은 것은, 원거리에 대한 오메가 템플릿의 크기 추정의 오차가 커지기 때문이며, SET C, F에서 정확도가 크게 떨어지는 이유는, 이러한 이유로 인하여 정적인 장애물이 사람으로 오검출되는 경우가 지속적으로 발생했기 때문이다. 제안방법에서는 초점거리를 이용한 템플릿 크기 추정 및 오검출 재매칭 등의 방법으로 이러한 점이 개선되었음을 알 수 있다.

DS1에서 사람이 미검출된 주요 원인은 SET A, B의 경우, 사람이 머리에 손을 올리는 동작을 취했을 때 오메가 템플릿 매칭 스코어가 작게 나왔기 때문이며,

표 1. DS1의 정확도와 재현율  
Table 1. Precision and Recall for DS1.

	Bae		Proposed	
	정확도	재현율	정확도	재현율
SET A	100	94.5	100	91.8
SET B	99.7	91.3	100	92.1
SET C	71.8	10.7	100	99.8
SET D	98.4	15.4	100	94.0
SET E	95.1	17.2	100	96.8
SET F	73.2	9.9	100	92.4



그림 8. DS1의 사람 검출 결과  
Fig. 8. Result of human detection for DS1.

SET D와 E에서는 멀리 있는 사람의 머리 부분 거리값의 분산이 커 히스토그램 안에 들어오지 못한 경우가 있었기 때문이다. SET F에서는 사람이 좌우로 방향을 틀어 움직일 때 오메가 템플릿 매칭 스코어가 작았다. 정확도에 있어서는 모든 SET에서 100%의 성능을 보였으나, 혹 사람이 아닌 오메가 모양을 가지는 물체가 영상 내 존재하게 되는 경우에는 오검출될 가능성이 있다.

DS1에서의 수행 시간은 Intel(R) Core(TM) i5-3550 CPU @3.30GHz, RAM 8GB에서 장면 당 평균 4-5ms 걸렸다.

4.2 DS2

DS2는 2명의 인물이 실내를 드나들며 자유롭게 거니는 장면을 Kinect를 이용하여 촬영한 데이터셋이다. Xia와 제안방법에 동일한 성능판단기준을 적용하기 위하여, Xia로부터 제공되는 지상검증자료(Ground Truth)를 기준으로 정확도와 재현율을 측정하였다. 지상검증자료는 각 프레임별로 사람의 수를 기록하고 있다. DS2는 Kinect를 이용하므로, DS1에서의 SR4000의 거리 영상과 해상도 및 초점거리가 각기 다르다. 거리 영상은 640×480의 해상도를 가지며, 거리센서의 초점거리는 Kinect SDK<sup>[10]</sup>에 정의된 531.15의 값을 가진다. 또한, DS2에는 3.1절의 ToF를 위한 거리영상 개선 알고리즘

표 2. DS2의 정확도와 재현율  
Table 2. Precision and Recall for DS2.

	정확도	재현율
Xia	100	96.0
Proposed	100	94.9

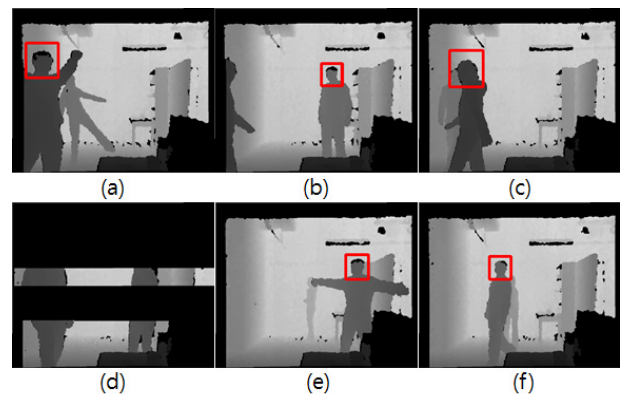


그림 9. DS2의 미검출 사례  
Fig. 9. Undetected case for DS2.



표 3. DS2의 수행시간  
Table 3. Process time for DS2.

		수행시간 (ms)
uv-hist 생성	u-hist 생성	0.81
	u-labeling	0.17
	v-hist 생성	3.20
	이진영상 생성	1.75
템플릿 매칭		11.54
기타		1.30
총합		18.77

은 적용하지 않는다.

표 2는 DS2에 대한 Xia의 방법과 제안 방법의 정확도와 재현율의 비교 결과이다. Bae의 방법은 템플릿 매칭에 있어서 고정된 크기변환 파라미터를 사용하므로, 다른 카메라 파라미터를 가지는 DS2에는 적용이 불가능하다. Xia와 성능을 비교해 볼때, 정확도는 완전히 같은 성능을 보이나, 재현율에서 약간의 차이를 보였다.

그림 9는 DS2에서 미검출된 사례들이다. DS1의 경우와 비슷하게, 사람의 상반신 또는 머리 부분이 사람에게 의하여 가려져 있는 경우(a,c,e,f)와 사람이 영상 가장자리에 걸쳐있는 경우(b), 오메가 템플릿 매칭 스코어가 낮아 결과적으로 미검출되었다. 이러한 결과는 Xia에서도 동일하게 리포트 되었다. 한편, DS2에는 (d)와 같이 시퀀스 중간에 몇 개의 프레임이 크랙되어 저장되어 있는데, Xia의 방법은 프레임간의 움직임 추정을 통한 트래킹을 추가하였고, 이로 인하여 재현율을 소폭 향상시켰다.

Xia의 방법과 제안방법의 성능의 주요한 차이점은 알고리즘의 수행시간이다. Xia의 수행시간은 프레임당 0.4초인데 반하여 제안 방법은 평균 18.77ms의 속도를 보여, 약 20배의 속도 차이를 나타내며, 이로써 실시간 수행이 가능하다.

표 3에서는 DS2에 대한 세부 수행 시간을 나타내었다. DS1에 비하여 템플릿 매칭에서 시간이 더 걸렸으며, 이는 DS2의 거리 영상 해상도가 약 12배 크기 때문이다.

## V. 결 론

본 논문에서는 적응적 UV-histogram과 오메가 모양의 템플릿을 이용하여 빠르게 사람을 검출하는 방법을

제안하였다. 제안하는 방법은 적응적 UV-histogram를 이용하여 먼저 U-histogram에서 사람 후보 영역을 획득하고, 이에 대하여 V-histogram에서 사람에 해당할 수 있는 부분만 추출하여 템플릿 매칭을 수행함으로써 사람 후보영역 추출 재현율을 높였으며, 후보영역 추출 이후에도 occlusion에 강한 템플릿 매칭 방법을 사용함으로써 정확도를 크게 향상시켰다. 실험 결과, 거리 영상과 템플릿 매칭을 이용하는 기존의 사람 검출 방법과 비교하였을 때, 제안하는 방법은 장애물이 많은 복잡한 환경 (SET C, D, E, F)에서의 재현율은 약 80%, 정확도는 약 15%의 성능 향상을 나타내었다.

오메가 모양의 템플릿 매칭의 단점은 사람이 오메가 모양을 벗어나게 되면 사람 후보가 검출되지 않는다는 데에 있다. 향후 연구는 이러한 단점을 극복할 수 있는 방안에 대한 연구이다. 예를 들면 템플릿 매칭은 머리 부분만 본다면지 하여 조금 느슨하게 하고, 추출된 후보 영역에 대하여는 강인한 feature 매칭을 하는 방법이 있을 수 있다.

처리속도에 있어서는, DS1과 같은 저해상도 거리영상에 대하여 사람 검출력 성능을 확인한 만큼, 고해상도의 거리영상인 경우 UV 도메인에서의 합당한 스케일링을 등의 방법을 연구하면 더욱 고속으로 수행할 수 있을 것으로 생각된다.

## REFERENCES

- [1] N. Dalal B. Triggs, and C. Schmind, "Histogram of oriented of gradients for human detection" in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, 2005.
- [2] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," CVPR 2006.
- [3] L. Xia, C. C. Chen, and J. K. Aggarwal, "Human Detection Using Depth Information by Kinect," Workshop on Human Activity Understanding from 3D Data in conjunction with CVPR, 2011.
- [4] J. M. Hyun, J. H. Kwak, H. J. Ji, K. K. Kim, and W. Y. Kim, "Human Detection in Range data using Human Shape Features," Korean Signal Processing Conference, 2011.
- [5] <http://www.mesa-imaging.ch/prodview4k.php>

- [6] R. Benenson, M. Mathias, R. Timofte and L. Van Gool, "Pedestrian detection at 100 frames per second," CVPR 2012.
- [7] H.J. Bae, M. S. Na, W. Y. Kim, "Fast Human Detection Method in Range data using UV-disparity and Template Matching," Proceedings of the IEEK Fall Conference, pp. 512-515, 2013.
- [8] <http://www.opencv.org>
- [9] <http://www.xbox.com/en-US/xbox360/accessories/kinect>
- [10] <http://msdn.microsoft.com/en-us/library/hh855347.aspx>

---

 저 자 소 개
 

---



윤 범 식(정회원)

1997년 연세대학교 전기공학과  
학사 졸업

2000년 연세대학교 전기전자  
공학과 석사 졸업

2000년~현재 삼성전자  
VD사업부 책임연구원

2013년~현재 한양대학교 전자통신컴퓨터공학과  
박사과정

<주관심분야 : 영상처리, 컴퓨터비전, 패턴인식  
등>



김 회 울(정회원)

1980년 한양대학교 전자공학과 학사  
졸업

1983년 Pennsylvania State  
University 전기공학과  
석사 졸업

1989년 Purdue University  
전기공학과 박사 졸업

1989년~1994년 University of Texas 조교수

1994년~현재 한양대학교 융합전자공학부  
정교수

<주관심분야 : 영상처리, 컴퓨터비전, 패턴인식,  
머신비전, 지능형 자동차 등>