

# 방향성 2차원 타원형 필터를 이용한 스테레오 기반 포즈에 강인한 사람 검출

(Stereo-based Robust Human Detection on Pose Variation Using Multiple Oriented 2D Elliptical Filters)

조 상 호 <sup>†</sup>      김 태 완 <sup>\*\*</sup>      김 대 진 <sup>\*\*\*</sup>  
(Sangho Cho)      (Taewan Kim)      (Daejin Kim)

**요약** 이 논문은 방향성 2차원 타원형 필터(Multiple Oriented 2D Elliptical Filters;MO2DEFs)를 사용하여 스테레오 영상으로부터 포즈에 강인한 사람 검출을 제안한다. 기존의 물체 지향 크기 적응 필터(Object Oriented Scale Adaptive Filter;OOSAF)는 정면을 보고 있는 사람만을 검출하는 단점을 지니고 있는데 반해 제안한 방향성 2차원 타원형 필터는 사람의 크기나 포즈에 관계없이 사람을 검출하고 추적한다. 2D 공간-깊이 히스토그램에 특정 각도로 향하는 4개의 2차원 타원형 필터들을 적용하고, 필터링된 히스토그램에서 임계값을 통해서 사람을 검출한 다음, MO2DEFs 중 승적 결과가 가장 큰 2차원 타원형 필터의 방향을 사람의 방향으로 판단한다. 사람 후보들은 얼굴을 검출하거나 검출된 사람의 선택된 방향의 머리-어깨 형태를 정합함으로써 검증한다. 실험 결과는 (1) 포즈 각도 예측의 정확도는 약 88%이고, (2) 제안한 MO2DEFs를 사용한 사람 검출의 성능이 OOSAF를 사용한 사람 검출의 성능보다 15~20%만큼 향상되었으며, 특히 정면이 아닌 사람의 경우에 더 향상이 있었다.

**키워드** : 스테레오 영상, 사람 검출, 방향성 2차원 타원형 필터, 포즈 각도 예측, 사람 검증, 얼굴 검출, 형태 정합

**Abstract** This paper proposes a robust human detection method irrespective of their pose variation using the multiple oriented 2D elliptical filters (MO2DEFs). The MO2DEFs can detect the humans regardless of their poses unlike existing object oriented scale adaptive filter (OOSAF). To overcome OOSAF's limitation, we introduce the MO2DEFs whose shapes look like the oriented ellipses. We perform human detection by applying four different 2D elliptical filters with specific orientations to the 2D spatial-depth histogram and then by taking the thresholds over the filtered histograms. In addition, we determine the human pose by using convolution results which are computed by using the MO2DEFs. We verify the human candidates by either detecting the face or matching head-shoulder shapes over the estimated rotation. The experimental results showed that the accuracy of pose angle estimation was about 88%, the human detection using the MO2DEFs outperformed that of using the OOSAF by 15~20%, especially in case of the posed human.

**Key words** : Stereo Image, Human Detection, Multiple Oriented 2D Elliptical Filter, Pose Angle Estimation, Human Verification, Face Detection, Shape Matching

· 본 연구(논문)는 산업자원부 지원으로 수행하는 21세기 프론티어 연구개발사업 (인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었음  
· 본 연구(논문)는 지식경제부 지자제연구소육성사업(RTI04-02-06) 지원으로 수행되었음

<sup>†</sup> 비 회 원 : 포항공과대학교 컴퓨터공학과  
scho@postech.ac.kr

<sup>\*\*</sup> 학생회원 : 포항공과대학교 컴퓨터공학과  
taey16@postech.ac.kr

<sup>\*\*\*</sup> 정 회 원 : 포항공과대학교 컴퓨터공학과 교수  
dkim@postech.ac.kr

논문접수 : 2007년 8월 2일

심사완료 : 2008년 9월 2일

Copyright©2008 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제35권 제10호(2008.10)

### 1. 서론

사람 검출은 사람 로봇 상호작용이나 영상 감시, 사람 움직임 추적, 제스처 인식, 사람 행동 분석 등 여러 분야에서 필수적인 작업이다. 본 연구는 이러한 응용분야 중, 특히 사람 로봇 상호작용(Human-Robot Interaction; HRI) 분야에서의 사람 검출을 수행하고자 한다. 사람 친화적 환경에서 삶과 공존하기 위해, 지능형 로봇은 주변의 사람을 인식하는 것이 필수적이다.

사람 검출하는 방법은 여러 가지가 있다. 우선 적은 비용과 손쉽게 다룰 수 있다는 점에서 주로 단일 고정 카메라를 사용한 사람 검출이 있다[1-4]. 그러나 이동형 로봇의 경우, 로봇과 사람이 서로 움직이기 때문에 조명 조건이 변하고, 배경이 시간에 따라 변하므로 사람 검출이 어렵다. 이런 경우에는 사람을 검출하기 위해 스테레오 카메라에서 얻는 깊이 정보를 사용하면 효과적이다 [5-11]. 다중 카메라를 사용하는 경우[12]도 있으나 이는 사람 감시용과 같은 고정형 카메라에 주로 사용하는 방식이다. 본 연구는 사람 로봇 상호작용 분야에서의 사람 검출을 수행하고자 하므로 스테레오 기반 영상을 사용하여 사람 검출을 수행하고자 한다.

스테레오를 기반으로 한 사람 검출에 관한 다양한 연구가 이뤄졌다. Ghidary 등[5]은 사람의 머리와 얼굴의 3차원 위치를 찾아서, 위치, 형태, 색 정보를 사용하여 사람을 검출하였다. Beymer 등[6]은 사람을 검출하기 위해 스테레오 카메라로부터 얻어진 시각차 맵의 배경 차분(background subtraction)과 시각차에서의 틀 정합을 사용하였다. 이러한 방법은 움직이는 스테레오 카메라에는 적합하지 않다. Salinas 등[7]은 정적인 배경 정보의 맵을 만들고 맵에서 움직이는 물체를 분리했다. 그들은 사람 검증에 얼굴 검출을 사용하였다. 그러나 그들은 사람이 정면으로 서있다고 가정하였다. Ran 등[8]은 움직임 예측을 통한 스테레오 연관 관계를 계산하여 배경 분리를 하였으나, 배경 분리 방법은 고정된 카메라 환경에서 효과적이다. Abd-Elmageed 등[9]은 깊이 정보로 배경을 분리하고 Adaboost 기반의 검증기를 사용하여 사람을 찾았으나, Adaboost 기반의 검증기는 학습 시간과 노력이 많이 소모된다. Li 등[10,11]은 물체 지향 크기-적응 필터(Object Oriented Scale Adaptive Filter; OOSAF)를 구상하고, 카메라와 사람 사이의 거리에 따라 필터의 인자가 변하는 물체-지향 크기-적응 필터를 적용함으로써 사람 후보를 분리했다. 그들은 사람의 머리-어깨의 틀 정합을 사용하여 사람 후보를 검증했다. 그들의 접근 방법은 높은 사람 검출률을 보이며, 배경 차분을 사용하지 않으므로 움직이는 로봇 플랫폼에 적합하다. 그러나 사람이 정면이 아닌 경우, 낮은 사

람 검출률을 나타낸다.

우리는 카메라와 사람이 움직이고 조명 조건이 변하는 복잡한 환경에서의 스테레오 영상에서부터의 포즈에 강인한 사람 검출을 제안하고자 한다. 따라서 움직이는 플랫폼에서도 높은 검출률을 보이는 OOSAF를 개량하여 다양한 포즈의 사람을 검출하는 방법을 제안한다. 그림 1은 제안하는 사람 검출 시스템의 흐름을 보인다.

제안하는 방법은 사람 후보 검출과 사람 검증, 이렇게 두 가지 모듈로 구성된다. 사람을 검출하기 위해 우리는 2차원 공간-깊이 히스토그램에서 4가지 특정 방향의 방향성 2차원 타원형 필터(Multiple Oriented 2D Elliptical Filters; MO2DEFs)를 적용하고, 필터링한 히스토그램에서 임계값을 넘는 부분을 찾음으로서 사람 후보를 분리한다. 그 후, 사람을 찾는 것을 더 향상시키고, 사람의 포즈를 결정하기 위해서 추출된 사람 후보를 합친다. 사람의 포즈는 MO2DEFs 중 승적(convolution) 결과가 최대치인 2차원 타원형 필터의 방향을 취함으로써 결정한다. 사람 검증은 얼굴을 검출하거나 머리-어깨 형태 정합을 통해서 수행된다. 형태 정합은 추출된 사람 후보의 머리-어깨 형태와 그 사람 후보의 예상되는 각도에 해당하는 사람의 머리-어깨 형태를 정합함으로써 수행한다.

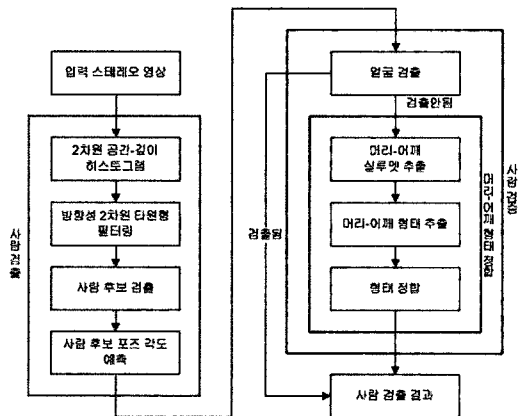


그림 1 제안하는 사람 검출 시스템

본 논문의 구성은 다음과 같다. 2장에서는 MO2DEFs를 사용한 사람 검출을 설명한다. 3장에서는 얼굴 검출과 머리-어깨 형태 정합을 사용해 사람 검증을 설명한다. 4장에서는 예측된 사람 포즈와 사람 검출의 정확성에 대한 실험 및 평가를 기술하였고, 마지막으로 5장에서는 본 논문의 결론을 기술하였다.

## 2. 사람 후보 검출

### 2.1 물체 지향 크기 적응 필터(OOSAF)

Li 등[10,11]은 사람과 카메라간의 거리에 따라 적합한 크기로 변환하는 크기-적용 필터를 사용하여 사람을 추출하는 물체-지향 크기-적용 필터(OOSAF)를 선보였다. 이것은 3차원 정보인 시각차(Disparity) 맵  $D(x,y)$ 에서 사람의 키 방향인  $Y$ 축 방향으로 투상(project)함으로써, 얻게 되는 2차원 공간-시각차(Spatial-Disparity;  $X-D$ ) 히스토그램  $H'(x_d, d)$ 에 크기 적용 필터를 승적한 결과  $\tilde{H}'(x_d, d)$ 를 사용하여 사람 후보 검출을 수행한다. 여기서  $\tilde{H}'(x_d, d)$ 는 다음과 같이 정의 된다.

$$\tilde{H}'(x_d, d) = (G_x^s * (G^d * H'))(x_d, d)$$

$$G_x^s(x_d) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x_d^2}{\sigma_x^2}\right)$$

$$G^d(d) = \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{d^2}{\sigma_d^2}\right)$$

여기서  $G_x^s(x_d)$ 와  $G^d(d)$ 는 각각 공간(Spatial;  $X$ )축과 시각차(Disparity;  $D$ )축 방향으로의 1차원 가우시안 유사 커널 함수이다. 이 커널 함수들은 공간(Spatial;  $X$ )축과 시각차(Disparity;  $D$ )축에 대한 크기 적용 변수  $\sigma_x^2$ 와  $\sigma_d^2$ 를 통해 카메라와의 거리에 따라 커널의 크기가 가변적인 성질을 갖게 된다. 즉, 이런 커널들을 공간-시각차 히스토그램에서 카메라를 바라보는 사람의 모습을 근사화한 모델이라고 생각할 수 있다. 그림 2의 (c)와 같은 공간-시각차 히스토그램에 OOSAF를 승적하여, 유사도가 특정 임계값보다 높은 영역을 사람의 후보 영역으로 판단 할 수 있다. 그리하여, 고정된 카메라 환경이 아닌 로봇과 같은 이동형 환경에서도 스테레오 정보를 이용하여, 3차원 공간 상의 사람 위치를 추정할 수 있다.

**2.2 공간-깊이 히스토그램**

기존의 물체-지향 크기-적용 필터의 문제는 2차원 공간-시각차( $X-D$ ) 공간에서 승적 연산을 수행한다는 점이다. 이는 여러 문제를 야기한다. 첫 번째, 두 좌표축은 서로 다른 특성을 가진다. 왜냐하면 두 좌표축은 각각 픽셀 크기 단위로 나타나는 위치와 시각차를 나타내기 때문이다. 두 번째, 크기 적용 필터의 크기 인자는 사람

과 카메라 사이의 거리에 따라 변해야 한다. 세 번째, 시각차는 사람과 카메라간의 거리에 반비례하므로 다루기 어렵다. 이 문제들을 피하기 위해 적절한 변환을 통해 얻어지는 2차원 공간-깊이 히스토그램(spatial-depth histogram)  $H(x_z, z)$ 을 제안한다.

그림 2는 2차원 공간-시각차 히스토그램을 2차원 공간-깊이 히스토그램으로 변환하는 일반적인 예를 보여준다. 그림 2의 (c)와 같은 공간-시각차 히스토그램은 그림 2의 (d)와 같은 공간-깊이 히스토그램으로 변환이 가능하다. 시각차는 카메라와 사람과의 거리인 깊이와 반비례 관계이므로 세로축이 뒤집혀져 나왔으며, 카메라에서 사람이 멀어질수록 영상에 나타나는 양이 많아지므로 (d)의 위쪽이 좁아지게 변환된다.

$x_d$ 를 공간-시각차 히스토그램  $H'(x_d, d)$ 에서 픽셀 크기 단위로 나타나는 공간 좌표라고 하고,  $x_z$ 를 공간-깊이 히스토그램  $H(x_z, z)$ 에서 실제 크기 단위로 나타나는 공간 좌표라고 하자. 그러면 2차원 공간-깊이 히스토그램은 2차원 공간-시각차 히스토그램에서 다음과 같이 계산하여 얻을 수 있다.

$$H(x_z, z) = H\left(\frac{C_F}{z} x_z + c_{x_d}, \frac{C_B C_F}{z}\right), \quad (1)$$

여기서  $c_{x_d}$ 는 2차원 공간-시각차 히스토그램에서 공간 좌표축의 중앙이며,  $C_F$ 는 스테레오 카메라의 초점길이(focal length),  $C_B$ 는 스테레오 카메라의 기선(baseline)이라고 한다.

기존의 OOSAF의 또 다른 문제점은 사람이 정면이 아닌 경우 잘 동작하지 않는 점이다. 이 문제를 해결하기 위해, 포즈에 강인한 MO2DEF를 새로 제안한다. 이 MO2DEF는 2차원 공간-깊이 히스토그램에 승적을 수행하여 결과를 얻는다. 이 방법은 다음과 같은 특성이 있다. 우선 형태는 2차원 공간-깊이 공간상에서의 사람의 몸의 윤곽과 유사한 2차원 타원 형태이다. 그리고 가중치는 몸의 윤곽의 수직 방향으로 서서히 감소한다. 또, 0°, 45°, 90°, 135°와 같은 특정한 방향을 향한다. 4개의 방향성 2차원 타원형 필터들 중 사람 후보의 포즈에 잘

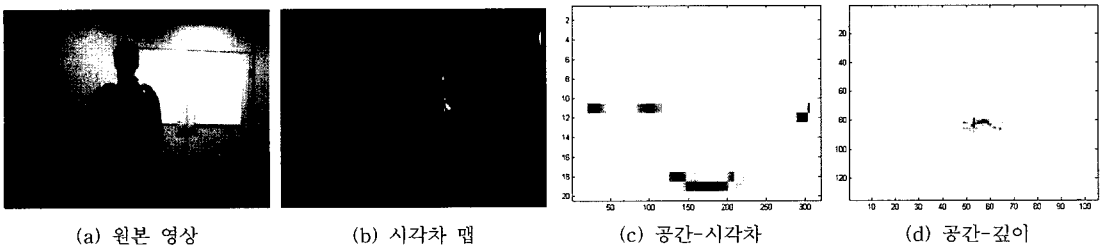


그림 2 2차원 공간-시각차 히스토그램과 2차원 공간-깊이 히스토그램

정합하는 방향의 필터를 통해 사람 후보가 검출된다. 또 사람 포즈의 예측된 방향은 사람 검출에의 단서로 사용된다. 제안하는 MO2DEF는 어떻게 만드는지에 대한 자세한 설명은 아래에 서술되어 있다.

**2.3 방향성 2차원 타원형 필터**

제안하는 MO2DEF는 사람 몸의 윤곽과 닮은 타원 형태이며, 사람의 회전을 감안한 방향성 2차원 필터를 가진다. 각도  $\theta$ 만큼 회전한 방향성 2차원 타원형 필터의 타원형 형태는 다음과 같이 정의할 수 있다.

$$\frac{(x \cos \theta + z \sin \theta)^2}{\left(\frac{W}{2}\right)^2} + \frac{(-x \sin \theta + z \cos \theta)^2}{\left(\frac{T}{2}\right)^2} = V_\theta, \quad (2)$$

여기서  $W$ 와  $T$ 는 사람 몸의 평균 너비와 두께이다.

방향성 2차원 타원형 필터의 2차원 커널 함수  $F_\theta(x, z)$ 는 다음과 같은 형태로 정의한다.

$$F_\theta(x, z) = \begin{cases} V_\theta & \text{if } 0 \leq V_\theta \leq 1, \\ 2 - V_\theta & \text{if } 1 < V_\theta \leq 2, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

필터 가중치는 타원의 윤곽에서 최대치(-1)이고,  $(x, z)$ 가 타원의 윤곽으로부터 멀어질수록 감소한다. 그림 3은  $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 로 회전된 방향성 2차원 타원형 필터들의 2차원 커널 함수들을 나타낸 것이다. 그림 3에서 보듯이, 카메라는 사람의 앞쪽만 볼 수 있으므로 2차원 커널 함수의 앞쪽 절반만 사용하게 된다. 이 때 축을 기준으로 자르지 않고 앞쪽 면을 기준으로 자른다.

그 후, 만들어진 2차원 커널 함수는 필터링 결과의 합을 동일하게 하기 위해 정규화를 수행한다. 즉, 2차원 커널 필터의 합을 1로 맞춰준다.

필터링된 결과는 공간-깊이 히스토그램(그림 2(d))와 4개의 방향성 2차원 타원형 필터와의 승적으로 얻어지며 다음과 같이 계산한다.

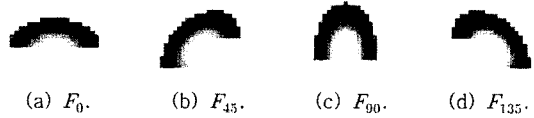


그림 3 방향성 2차원 타원형 필터

$$\Psi_\theta(x, z) = (H(x, z) * F_\theta)(x, z), \quad (\theta = 0, 45, 90, 135), \quad (4)$$

여기서  $*$ 는 승적 연산자이며,  $F_\theta(\cdot, \cdot)$ 는 2차원 커널 함수이다. 그림 4는  $\theta$ 에 따른 각 필터링된 결과를 나타낸다.

**2.4 사람 후보 분리**

Li 등[10]은 사람 후보 분리를 제안했다. 그들은 필터링된 공간-깊이 히스토그램들  $\Psi_\theta(x, z)$ 로부터 사람 몸의 평균 넓이를 사용하여 사람 후보의 위치를 얻고, 사람 몸의 평균 너비와 두께에 의하여 사람 후보의 영역을 결정한다. 이 방법에서는 사람을 찾기 위해 공간-시각차 공간을 사용한다. 그러나 우리는 MO2DEFs를 사용하기 위해 공간-깊이 공간을 사용한다. 따라서 공간-시각차 공간을 공간-깊이 공간에 맞도록 사람 분리의 모든 과정을 전체적으로 수정하였다. 회전에 관계없이 사람을 분리하기 위해서, MO2DEFs에 의해 필터링된 4개의 각 결과에서 사람 후보를 분리해야 한다. 그림 5는 필터링된 결과  $\Psi_\theta(x, z)$ 에서 사람 후보를 분리하는 과정을 보인다.

**2.5 사람 포즈 각도 예측**

사람 후보를 검출하기 위하여 방향성 2차원 타원형 필터들을 사용할 때 대략적으로 사람 포즈의 각도를 예측할 수 있다.

필터링된 공간-깊이 히스토그램  $\Psi_\theta(x, z)$ 는 필터의 각도가 사람의 각도와 유사할 때 더 높은 값을 가진다. 사

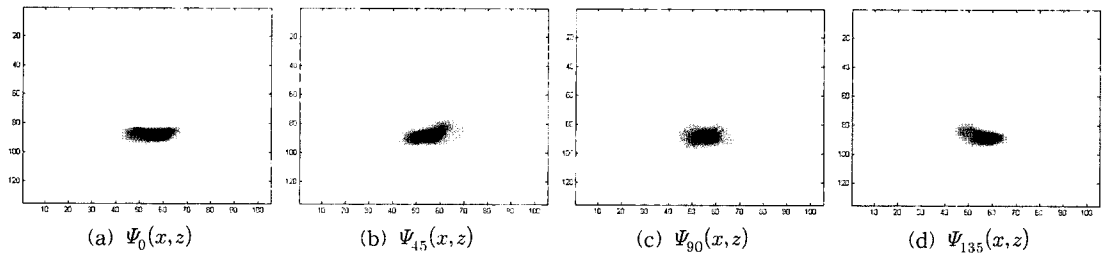


그림 4 4개의 각 2D 필터를 사용하여 필터링한 결과

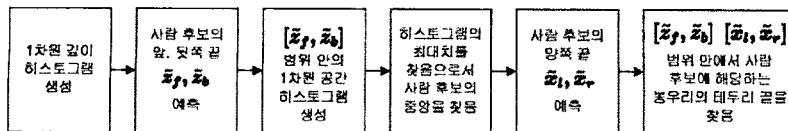


그림 5 필터링된 결과에서의 사람 후보 분리 흐름도

람 각도 예측은 다음과 같이 수행한다.

첫 번째로, 필터링된 2차원 공간-깊이 히스토그램들  $\Psi_\theta(x, z)$  ( $\theta=0^\circ, 45^\circ, 90^\circ, 135^\circ$ )에서 찾아진 사람 후보 영역을 둘러싸는 사각형들을 얻음으로서 사람 후보를 분리한다. 그 후, 각 히스토그램들에서 얻어진 사각형의 위치가 일치하는 것끼리 합친다. 관련된 사각형 전체를 둘러싸는 영역이 되므로 영역이 더 커지게 된다.

두 번째로, 분리된 사람 후보에 대한 합쳐진 영역을 기준으로 각 필터링된 2차원 공간-깊이 히스토그램들  $\Psi_\theta(x, z)$ 의 최고치를 찾는다. 이 과정은 각각의 필터링 결과인 4개의 히스토그램에 대해서 모두 계산된다.

세 번째로, 4개의 최고치 값 중 가장 큰 최고치를 가지는 2차원 타원형 필터가 해당하는 각도를 사람 포즈 각도로 예측한다.

### 3. 사람 검증

사람 후보 검출에 의해 얻어진 분리된 사람 후보들은 실제 사람일 수도 있고 아닐 수도 있다. 사람 후보가 실제 사람인지를 알기 위해서는 사람 검증 방법이 필요하다. 이 논문에서는 얼굴 검출과 머리-어깨 형태 정합의 두 가지 검증 방법을 사용한다.

#### 3.1 얼굴 검출을 사용한 사람 검증

얼굴 검출을 수행하기 전에 Li 등[10]이 제안한 방법을 사용하여 검출된 사람 후보에 해당하는 영역 안에서 머리 후보 영역 검출을 수행한다. 그 후 Adaboost 기반의 얼굴 검출[13]을 사용하여 앞에서 검출된 머리 후보 영역에서 얼굴 검출을 수행한다. 얼굴을 찾으면 검출된 사람 후보는 실제 사람이라고 판단한다. 하지만 사람이 정면을 향하지 않으면 얼굴 검출을 할 수 없다. 이 문제를 해결하기 위하여 머리-어깨 형태 정합 방법을 사용할 것을 제안한다.

#### 3.2 머리-어깨 형태 정합을 사용한 사람 검증

머리-어깨 형태 정합을 사용한 사람 검증은 머리-어깨 실루엣 추출, 머리-어깨 형태 추출, 머리-어깨 형태 정합의 과정으로 수행된다. 그림 6은 머리-어깨 형태 정합을 사용한 사람 검증의 흐름을 보인다.

가장 먼저, 움직임과 시각차 정보를 사용하여 머리-어깨 실루엣을 추출한다. 실루엣은 입력 이미지에서 2중 시간적 차분(double temporal differencing)[14]의 결과와 시각차 맵에서의 사람 후보 검출의 결과로부터 추출되며, 실루엣의 구멍은 점들로 채우게 된다.[15]

그 다음, 머리-어깨 형태가 얻어진다. 형태는 추출된 머리-어깨 실루엣의 테두리에서 점을 표집(sampling)하여 추출된다. 형태 정합을 위해 표집하는 점의 수는 항상 고정되어 있어야 하며 방향 역시 같아야 하므로 표집은 항상 시계방향으로 수행하기로 약속한다.

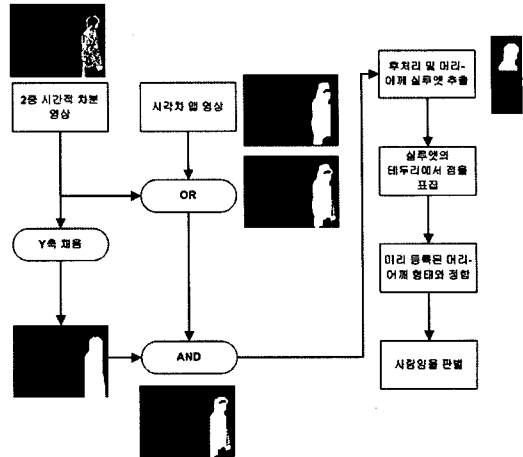


그림 6 머리-어깨 형태 정합을 사용한 사람 검증 흐름도

마지막으로, 추출된 머리-어깨 형태는 등록된 머리-어깨 형태와 정합된다. 추출된 머리-어깨 형태는 아직 정렬(align)되지 않았으므로, 형태 정합에 바로 적용할 수 없다. 형태 정합의 전처리로서, 정렬 과정[16]은 유사성 변환 요소(similarity transformation factors)를 제거하기 위해 수행된다. 정렬된 입력 형태가 미리 정의된 머리-어깨 형태와 유사하느냐에 따라 사람인지를 판단하게 된다. 이 판단은 추출된 머리-어깨 형태의 형태 인자 벡터(shape parameter vector)인  $\mathbf{p}$ 와 등록된 머리-어깨 형태의 형태 인자 벡터  $\mathbf{p}_0$ 간의 마할라노비스 거리(Mahalanobis distance)를 통해서 결정한다. 여기서 회전각  $\theta$ 는 사람 포즈 각도 예측에서 얻어진 사람 포즈 각도이다. 만약 거리가 임계치  $D_{th}$ 보다 작으면 그 형태는 사람의 것이라고 판단한다.

### 4. 실험 결과

실험에는 바닥에서 1.2m 높이에 범블비(BumbleBee) 스테레오 카메라[17]를 설치하여 사용하였다. 또한 제안하는 사람 검출 시스템은 3.4 GHz Pentium-4 CPU와 RAM 2GB의 PC에서 구현되었다. 식 (2)에서 사용하는 사람의 평균 너비  $W$ 는 0.6, 평균 두께  $T$ 는 0.4로 정하고 실험하였다.

총 3가지 종류의 연속 영상(image sequence)을 사용하였다. 각각을 TYPE1, TYPE2, TYPE3이라고 한다. TYPE1 연속 영상은 4개의 연속 영상으로 이루어져 있으며, 한 사람이 각 4가지 각도에 대해서 앞뒤로 이동한 내용이 들어 있다. TYPE2 연속 영상은 또한 4개의 연속 영상으로 되어 있으며, 각 연속 영상마다 영상의 수는 430, 550, 600, 419개로 이루어졌다. 이 연속 영상들에서는 한 사람이 임의의 각도로 앞뒤로 이동하였다.

TYPE3 연속 영상은 두 연속 영상으로 되어 있으며, 각 연속 영상마다 영상의 수는 500, 565개로 이루어졌다. 이 연속 영상들에서는 2~3명의 사람이 임의의 각도로 앞뒤로 이동하는 내용이 들어있고, 서로 겹쳐지는 것을 허용하였다. TYPE3 연속 영상의 연속 영상 1에서 두 사람은 5번 겹쳐지고, 연속 영상 2에서는 7번 겹쳐진다.

4.1 포즈 각도 예측

포즈 각도 예측은 정확도를 높은 사람 검출을 달성하기 위한 매우 중요한 부분이다. 예측된 포즈 각도는 사람 검증 단계의 형태 정합에서 정합할 등록된 머리-어깨 형태의 각도를 결정할 때 사용된다. TYPE1 연속 영상을 사용하여 포즈 각도 예측의 정확도를 측정하였다. 표 1에서 보이는 대로, 0°, 90°의 포즈 각도에서 완벽하게 예측했으며, 45°, 135°의 포즈 각도에서는 각각 81%와 71%의 정확도로 예측되었다. 45°, 135°의 포즈 각도는 때때로 0°나 90°로 예측되었다.

표 1 포즈 각도 예측의 정확도

	0°	45°	90°	135°
Sequence 1	100%	64.71%	100%	75%
Sequence 2	100%	85%	100%	66.67%
Sequence 3	100%	86.96%	100%	69.57%
Sequence 4	100%	88.68%	100%	72.09%
Average	100%	81.34%	100%	70.83%

4.2 사람 검출

TYPE2, TYPE3 연속 영상을 사용하여 사람 검출 성능을 측정하였다. 검출 정확도  $f_d$ 는 검출된 사람의 수  $N_d$ 와 전체 사람 수  $N_t$ 의 비율로 다음과 같이 정의하여 계산하였다.

$$f_d = \frac{N_d}{N_t} \quad (5)$$

표 2는 TYPE2와 TYPE3 연속 영상을 사용하여 기존의 OOSAF와 제안하는 MO2DEFs의 사람 검출 성능

을 비교하였다. TYPE2 연속 영상의 결과를 보자시피, (1) 방향성 2차원 타원형 필터를 사용함으로써 적합한 포즈 강인성을 얻을 수 있으므로 제안하는 MO2DEFs의 사람 검출률이 기존의 OOSAF보다 약 20% 높다는 점과, (2) 연속 영상에 사람의 정면이 많지 않아서 얼굴 검출이 사람 검출률에 큰 영향을 주지 못한다는 점을 알 수 있다. TYPE3 연속 영상의 결과를 보면, (1) MO2DEFs를 사용한 사람 검출 성능이 OOSAF를 사용한 사람 검출 성능보다 더 높고, (2) TYPE3의 검출 성능은 TYPE2와는 달리 두 사람이 겹쳐지는 경우도 발생하여 검출 성능이 TYPE2보다 더 떨어지는 것을 알 수 있었다. (3) 또, 연속 영상 1보다 연속 영상 2가 더 성능이 낮은데, 이는 연속 영상 2가 연속 영상 1보다 사람이 보다 자주 겹쳐졌기 때문이다. (4) 얼굴 검출이 사람 검출에 주는 영향은 TYPE2보다 TYPE3가 더 크게 미쳤다. 왜냐하면 사람이 서로 겹쳐질 때 사람 검증 방법으로 얼굴 검출이 형태 정보보다 더 효과적이기 때문이었다.

그림 7은 TYPE2와 TYPE3 연속 영상을 사용하여 기존의 OOSAF 방법과 제안하는 MO2DEF 방법의 사람 검출 결과에 대한 몇 가지 예이다. 그림 7에서 TYPE2에 대한 결과를 통해 (1) 기존의 OOSAF 방법은 사람이 포즈 각도 45°(첫 번째 행의 215번 프레임)나 90°(첫 번째 행의 197번 프레임)로 회전할 때 잘 찾지 못했고, (2) 제안하는 MO2DEFs는 포즈에 강인한 방향성 2차원 타원형 필터에 의해 사람 포즈에 관계없이 사람 몸을 찾을 수 있다는 것을 알 수 있었다 (TYPE2에 대한 예 중 세 번째와 네 번째 행).

그림 7에서의 TYPE3에서의 결과는 (1) 기존의 OOSAF는 사람의 포즈 각도가 90°(첫 번째 행의 157번 프레임)로 돌거나 서로 겹쳐졌을 때(490번 프레임) 검출 성능이 좋지 않았고, (2) 제안하는 MO2DEFs는 포즈에 강인한 방향성 2차원 타원형 필터에 의해 90°(세 번째 행의 157번 프레임)로 회전해도 사람 검출이 가능했다.

표 2 TYPE2, TYPE3 연속 영상을 사용하여 OOSAF와 MO2DEFs간의 사람 검출

		OOSAF		MO2DEFs	
		Without face detection	With face detection	Without face detection	With face detection
TYPE2	Sequence 1	73.25%	75.58%	96.05%	96.05%
	Sequence 2	75.27%	75.27%	90.63%	90.63%
	Sequence 3	75.00%	75.50%	96.50%	97.00%
	Sequence 4	77.57%	79.00%	91.41%	93.08%
	<b>Average</b>	<b>75.27%</b>	<b>76.34%</b>	<b>93.65%</b>	<b>94.19%</b>
TYPE3	Sequence 1	70.70%	73.18%	92.09%	94.19%
	Sequence 2	65.16%	66.04%	88.28%	89.38%
	<b>Average</b>	<b>67.93%</b>	<b>69.61%</b>	<b>90.19%</b>	<b>91.79%</b>

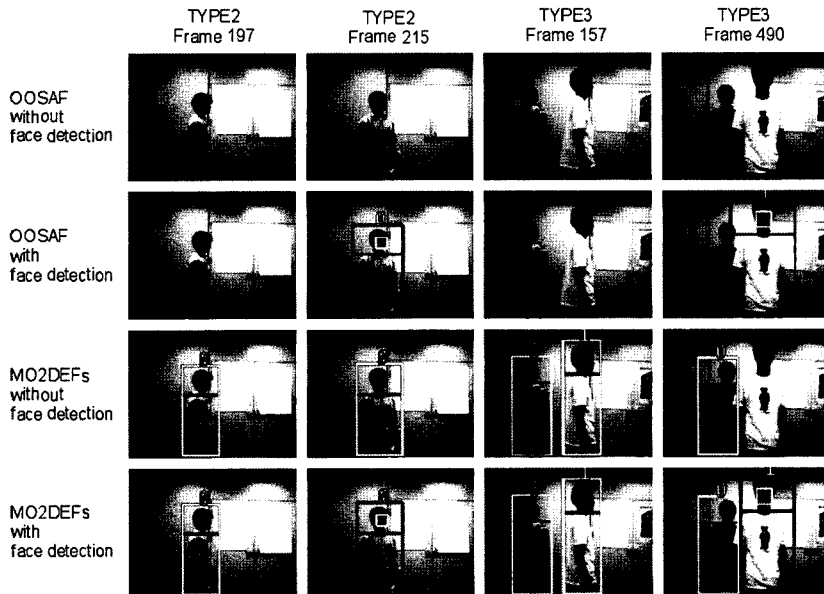


그림 7 TYPE2, TYPE3 연속 영상을 사용한 사람 검출 결과의 몇 가지 예

(3) 얼굴 검출은 검출된 사람이 겹쳐졌을 때에도 효과적으로 사람 검증을 했다. (네 번째 행의 frame 490)

## 5. 결론

다양한 방향을 바라보는 사람을 검출하기 위해 우리는 방향성 2차원 타원형 필터들(MO2DEFs)을 사용하여 스테레오 연속 영상에서의 포즈에 강인한 사람 검출 방법을 제안하였다. 기존의 OOSAF는 사람 검출 성능이 우수하나, 돌아간 사람의 경우 잘 동작하지 않는다.

이 문제를 해결하기 위해서, 0°, 45°, 90°, 135°로 돌아간 타원의 앞쪽 절반의 형태를 지니는, 포즈에 강인한 2차원 타원형 필터(MO2DEFs)를 만들었다. 타원의 크기는 사람의 실제 평균 너비와 두께이다. MO2DEFs의 형태는 돌아간 사람의 형태와 유사하므로 사람의 각도와 상관없이 사람을 찾을 수 있다. MO2DEFs를 만들기 위해서는 타원을 회전시켜야 하고, 이를 위해 공간-시각 차 공간(spatial-disparity space)을 공간-깊이 공간(spatial-depth space)으로 변환하였다.

사람 검출은 사람 후보를 추출하고 사람 후보의 포즈 각도를 예측하며, 다음과 같은 순서로 수행된다. 가장 먼저, 입력 스테레오 영상에서 2차원 공간-깊이 히스토그램을 계산하고, 계산된 히스토그램에 각 4가지 각도로 구성된 MO2DEFs를 적용한다. 그 후 4개의 필터링 결과 히스토그램으로부터 사람 후보를 검출하고, 각 사람 후보는 그들 간의 최대치를 찾아서 사람 포즈 각도를 예측한다. 예측된 각도는 머리-어깨 형태 정합에서 사람

후보를 검증할 때 사용된다.

사람 검증은 검출된 사람 후보가 실제 사람인지를 판별하기 위해서 수행한다. 사람 검증을 위해서 얼굴 검출과 형태 정합을 수행한다. Adaboost 기반의 얼굴 검출은 빠르고 높은 정확도 때문에 적절한 검증기이다. 따라서 사람 검증 시에 얼굴을 먼저 검출한다. 그러나 이는 정면인 사람만 찾을 수 있다. 이 문제를 해결하기 위해, 추출된 사람 후보의 머리-어깨 형태와 미리 등록된 머리-어깨 형태 중 예측된 각도의 머리-어깨 형태를 비교함으로써 머리-어깨 형태 정합을 수행한다.

실험은 사람 포즈 각도 예측 부분과 사람 검출 부분으로 나뉘어서 수행되었으며, 사용한 영상의 연속 시퀀스는 혼자서 일정한 각도로 걸어가는 TYPE1, 혼자서 다양한 각도를 돌아다니는 TYPE2, 여러 명이 다양한 각도로 돌아다니는 TYPE3를 사용하였다. 실험 결과는 TYPE1을 사용한 사람 포즈 각도 예측의 정확도는 약 88%였다. 또, TYPE2와 TYPE3로 구한 사람 검출의 정확률을 보면, MO2DEFs를 사용한 사람 검출률은 OOSAF를 사용했을 때보다 약 20% 더 나은 성능을 보였으며, 특히 정면이 아닌 사람의 경우 OOSAF보다 MO2DEFs가 보다 월등한 성능을 보여주는 것을 확인했다. 정면이 아닌 다양한 각도로 걷는 경우에도 높은 검출률을 보여주었으므로 MO2DEFs를 이용한 검출이 포즈에 강인한 사람 검출이라는 목표에 만족할 수 있었다. 마지막으로, 사람 검증 방법으로 얼굴 검출은 사람이 겹쳐질 때, 그리고 머리-어깨 형태 정합은 사람이 정

면이 아닐 때 각각 효과적인 것으로 나타났다.

**참 고 문 헌**

[1] M. Hussein, W. A. Almageed, Y. Ran, and L. Davis, "Real-Time System for Human Detection, Tracking and Verification in Uncontrolled Camera Motion Environment," Proc. of IEEE International Conf. on Computer Vision Systems, pp. 41, 2006.

[2] S. S. Ghidary, Y. Nakata, T. Takamori, and M. Hattori, "Localization and Approaching to the Human by Mobile Human Robot," Proc. of 9th IEEE International Workshop on Robot and Human Interactive Communication, pp. 63-68, 2000.

[3] J. Zhou, and J. Hoang, "Real time robust human detection and tracking system," Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 3, pp. 149, 2005.

[4] C. Wren, A. Azarbayejani, T. Darell, "Pfinder: Real-time tracking of human body," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, pp. 780-785, 1997.

[5] F. Xu, and K. Fujimura, "Human detection using depth and gray images," Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance, pp. 115-121, 2003.

[6] D. Beymer, and K. Konolige, "Real-Time Tracking of Multiple People Using Continuous Detection," Proc. of International Conf. on Computer Vision Frame-Rate Workshop, 1999.

[7] R. M. Salinas, E. Aguirre, M. G. Silvente, and A. Gonzalez, "People detection and tracking through stereo vision for human-robot interaction," Proc. of 4th Mexican international conf. on artificial intelligence, pp. 337-346, 2005.

[8] Y. Ran, and Q. Zheng, "Multi moving people detection from binocular sequences," Proc. of the 2003 IEEE International Conf. on Acoustics, Speech, and Signal Processing, vol. 3, pp. 37-40, 2003.

[9] W. Abd-Almageed, M. Hussein, M. Abdelkader, and L. Davis, "Real-Time Human Detection and Tracking from Mobile Vehicles," Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, pp. 149-154, 2007.

[10] L. Li, Y. T. Koh, S. S. Ge, and W. Huang, "Stereo-Based Human Detection For Mobile Service Robots," Proc. of 8th International Conf. on Control, Automation, Robotics, and Vision, pp. 74-79, 2004.

[11] L. Li, S. S. Ge, T. Sim, Y. T. Koh, and X. Hunag, "Object-Oriented Scale-Adaptive Filtering For Human Detection From Stereo Images," Proc. of the 2004 IEEE Conf. on Cybernetics and Intelligent Systems, pp. 135-140, 2004.

[12] K. Kim, L. S. Davis, "Multi-camera Tracking and Segmentation of Occluded People on Ground Plane

Using Search-Guided Particle Filtering," Proc. of the 9th European Conference on Computer Vision, vol. 3, pp. 98-109, 2006.

[13] B. Jun, and D. Kim, "Robust Real-Time Face Detection Using Face Certainty Map," Proc. of International Conference on Biometrics, pp. 29-38, 2007.

[14] Y. Kameda, and M. Minoh, "A human motion estimation method using 3-successive video frames," Proc. of the International Conf. on Virtual Systems and Multimedia, 1996.

[15] C. Kim, and J. N. Hwang, "A Fast and Robust Moving Object Segmentation in Video Sequences," Proc. of International Conf. on Image Processing, vol. 2, pp. 131-134, 1999.

[16] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models - Their Training and Application," Computer Vision and Image Understanding, vol. 61, num. 1, pp. 38-59, 1995.

[17] Point Gray Research Inc. <http://www.ptgrey.com/>



조 상 호

2005년 2월 숭실대학교 컴퓨터학부(학사). 2007년 8월 포항공과대학교 컴퓨터공학과(석사). 2007년 8월~현재 삼성전자주식회사 정보통신총괄 네트워크사업부 근무. 관심분야는 사람검출, 제스처 인식



김 대 완

2005년 2월 한국해양대학교 제어컴퓨터정보공학(학사). 2007년 3월~현재 포항공과대학교 정보통신대학원 정보통신학과(석사). 2005년 3월~2006년 10월 LG 전자 에어컨 연구소 연구원. 주관심분야는 물체검출 및 인식, 특칭매칭, 이미지

검색



김 대 진

1982년 2월 연세대학교 전자공학과(학사). 1984년 2월 KAIST 전기 및 전자공학과(석사). 1992년 8월 Syracuse University, Electrical and Computer Engineering(박사). 1984년 3월~1986년 12월 한국방송공사 기술연구소 1992년 3월~1999년 6월 동아대학교 컴퓨터공학과 부교수. 1999년 7월~현재 포항공과대학교 컴퓨터공학과 교수. 주관심분야는 지능시스템, 인간 컴퓨터 상호작용