

팬/틸트 제어기반의 스테레오 카메라의 기하학적 정보를 이용한 새로운 높이 추정기법

정회원 고 정 환*, 종신회원 김 은 수*

A New Height Estimation Scheme Using Geometric Information of Stereo Camera based on Pan/tilt control

Jung-Hwan Ko* *Regular Member*, Eun-Soo Kim* *Lifelong Member*

요 약

본 논문에서는 지능적인 보안 시스템 구현의 새로운 접근 방식으로, 인공지능형 시각 시스템에 기반한 팬/틸트 탑재형 스테레오 카메라의 기하학적 정보를 이용한 사람 키 높이 추정기법을 제안하였다. 제안된 시스템에서는 먼저, 스테레오 카메라에서 입력된 영상으로 부터 YCbCr 컬러 모델과 위상형 상관 기법을 사용하여 표적 얼굴영역의 중심좌표를 검출한 다음, 지능형 시각 시스템에 기반한 팬/틸트 탑재형 스테레오 카메라의 능동적인 제어각도 산출을 통해 표적의 다양한 변화에 관계없이 표적의 3차원 좌표를 실시간적으로 검출하고, 이를 통한 표적 대상의 적응적인 신체 크기 추정을 수행하였다. 각기 다른 신장을 가진 3사람의 다양한 이동경로를 가진 480 프레임의 테스트용 스테레오 영상을 사용한 실험 결과, 표적의 추정된 신체 크기값 역시 계산치 및 측정치 사이의 표준편차가 평균 1.03의 작은 값으로 유지됨은 물론, 제안된 시스템에서 계산된 좌표값과 실제 실험대상으로 참여한 사람의 정확한 실제 위치와의 차이 역시 평균 1.2cm 미만의 오차와 전체적으로 1.18 %의 오차를 보임으로써 이를 통한 새로운 실시간 스테레오 표적 감시추적 시스템의 구현 가능성을 제시하였다.

Key Words : surveillance, height estimation, pan/tilt, stereo camera, face detection

ABSTRACT

In this paper, a new intelligent moving target tracking and surveillance system basing on the pan/ tilt-embedded stereo camera system is suggested and implemented. In the proposed system, once the face area of a target is detected from the input stereo image by using a YCbCr color model and phase-type correlation scheme and then, using this data as well as the geometric information of the tracking system, the distance and 3D information of the target are effectively extracted in real-time. Basing on these extracted data the pan/tilted-embedded stereo camera system is adaptively controlled and as a result, the proposed system can track the target adaptively under the various circumstance of the target. From some experiments using 480 frames of the test input stereo image, it is analyzed that a standard variation between the measured and computed the estimated target's height and an error ratio between the measured and computed 3D coordinate values of the target is also kept to be very low value of 1.03 and 1.18% on average, respectively. From these good experimental results a possibility of implementing a new real-time intelligent stereo target tracking and surveillance system using the proposed scheme is finally suggested.

※ 본 연구는 2004년 광운대학교 교내학술연구비 지원사업으로 수행되었습니다.

* 광운대학교 차세대 3D 디스플레이 연구센터 (misagie@kw.ac.kr, eskim@daisy.kw.ac.kr)

논문번호 : KICS2005-11-461, 접수일자 : 2005년 11월 14일

I. 서론

최근, 세계각지에서 발생하고 있는 일련의 테러 사건들 이후, 공공 건물에 대한 보안감시 요구가 증대되고 있으며, 사회가 발전함에 따라 범죄의 유형도 다양화되고 다양화되는 경향을 보이고 있어 기존의 아날로그 CCTV 감시 시스템을 대체할 수 있는 최첨단 영상처리 기술을 이용한 보안감시 시스템의 요구가 증대되고 있다. 또한, 감시 카메라로부터 획득된 비디오 영상의 분석을 통해 표적 대상인의 얼굴 인식과 보다 정확한 형체 정보의 검출을 통한 표적의 실시간 감시 및 모니터링 기술은 날로 증가하고 있는 범죄와 사건을 해결하기 위한 실마리를 제공해 줄 수 있으며, 이미 널리 설치되어 있는 감시용 CCD 카메라에 보다 지능적인 인식 모듈의 추가로 가능하다는 점에서 활발한 연구가 진행되고 있다¹⁻²⁾. 일반적으로 보안감시 시스템은 건물내, 외곽으로 출입하는 표적에 대한 인증과 확인을 요구하는 표적 인식기술 및 표적의 돌발적인 행동을 감시하기 위한 추적 기술, 그리고 표적으로 지목된 사람의 사건 발생 이후의 상황을 재해석하는 모니터링 기술 등 다양한 형태로 발전되고 있다. 이러한 보안감시 기술 중, 사건 발생 직후 감시 카메라를 통해 녹화된 자료를 이용한 정확한 상황 파악 및 근거자료 확보와 개인의 인상착의 및 이동 경로를 파악하는 일은 사건을 해결하기 위한 실마리를 제공해 줄 수 있는 중요한 단서로 작용하고 있다. 그러나 이러한 보안감시 시스템이 설치되어있는 공간 안에서는 표적으로 지목된 용의자의 행동이 예측하기 힘든 돌발적인 상황으로 변할 수 있고, 더욱이 카메라 시야에서 멀어질 경우, 혹은 얼굴 안면이 은폐된 상황이라면 용의자의 얼굴 인식은 더욱 어려운 문제로 야기될 수 있다. 따라서 이러한 한계점을 극복하기 위해서는 용의자의 신체 크기나 외형상의 특징 및 용의자의 이동 경로를 추적함으로써 사건의 단서를 찾을 수 있다. 그러나 기존의 단안식 감시 카메라에 기반하여 개발된 대부분의 비전 알고리즘은 같은 시야의 카메라를 통해 입력되는 연속적인 영상들의 비교분석만을 통하기 때문에 표적의 자세, 회전, 크기, 그리고 조망 방향 변화 등에 능동적으로 대처하는데 많은 어려움이 있으며, 시스템의 강건성 및 감시범위의 제한 등의 문제점이 나타나게 된다²⁾. 또한, 표적으로 지목된 용의자의 전체 모습이 카메라 시야에 들어오지 않을 경우, 얼굴 인식에서 좀 더 포괄적인 외형적인 신체 모습을 추정하

는 데에는 많은 어려움이 존재하게 된다. 이러한 기존 보안감시 시스템의 한계를 극복하기 위한 새로운 접근방식으로 기존의 자동 물체추적 시스템에 인간의 시각계를 모방한 지능형 시각 시스템으로, 스테레오 카메라를 탑재하여 표적 대상을 정확히 감시 추적하여 오작동 및 사각지대(occlusion region)의 발생을 방지함은 물론, 침입위치 및 이동 경로에 대한 원격 모니터링의 가능성을 제시해주고 있다³⁻⁴⁾.

따라서, 본 논문에서는 이를 위한 새로운 접근 방법으로 지능형 이동표적 추적 및 감시 시스템의 구현을 위한 팬/틸트가 탑재된 스테레오 카메라 기반의 표적 높이 추정 기법을 제안 하고자 한다. 먼저, 스테레오 카메라에서 입력된 영상으로 부터 YCbCr 칼라 임계치 기법과 위상형 디지털 상관 기법을 사용하여 표적의 얼굴영역과 좌표계를 검출한 다음, 지능형 시각 시스템에 기반한 팬/틸트 탑재형 스테레오 카메라의 능동적인 제어를 통해 표적의 다양한 변화와 무관하게 표적과의 거리정보는 물론 움직이는 표적의 3차원적 정보를 검출함으로써 실시간으로 표적의 현재 위치 및 신체 크기를 동시에 검출할 수 있는 새로운 형태의 지능형 감시 추정 기법을 제안하였다. 그리고, 30 frame/s씩 실제 이동경로를 달리하며 각기 다른 3사람을 대상으로 획득한 240 프레임의 테스트용 스테레오 입력영상에 대한 표적영상의 실시간적 키 높이 검출 실험을 통해 표적에 대한 위치 정보 검출 및 신체 정보의 검출 성능을 분석함으로써 본 논문에서 새로이 제안한 실시간적 스테레오 표적 감시추적 시스템의 실질적인 구현 가능성을 제시하고자 한다.

II. 제안된 스테레오 카메라 기반의 높이추정 기법

본 논문에서는 실시간 처리 및 상황변화에 강건한 시스템의 구현을 목적으로 팬/틸트가 탑재된 스테레오 카메라를 이용한 새로운 지능형 사람 키 추정기법을 제안하고자 한다. 그림 1은 본 논문에서 새로이 제시한 스테레오 카메라 기반의 사람 키 추정기법의 전체 흐름도를 나타낸 것으로 크게 2단계로 이루어진다.

즉, 1 단계에서는 실시간으로 입력되는 스테레오 입력영상 중 좌영상으로부터 YCbCr 칼라 모델과 영상 차분필터를 이용하여 이동 중인 표적의 얼굴 성분만을 검출한 뒤, 기준영상으로 갱신된 좌영상과 우영상간에 위상형 상관 기법을 이용하여 우측 카

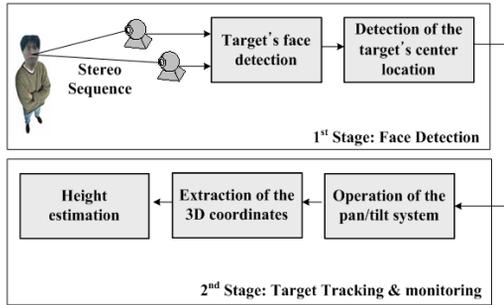


그림 1. 제안된 사람 높이 추정기법의 흐름도
Fig. 1. Operational flowchart of the proposed height estimation

메라로부터 입력된 얼굴 중심좌표를 검출하게 된다. 검출된 이들 좌, 우 얼굴영상의 중심좌표 값에 의해 스테레오 카메라의 팬/틸트가 제어되어 결과적으로 표적얼굴에 대한 주시각 제어를 수행하게 된다. 2 단계에서는 주시각이 제어된 스테레오 카메라의 팬/틸트 움직임 각도와 3차원 구면 좌표계(three dimensional spherical polar coordinate)를 이용하여 스테레오 카메라와 표적간의 3차원 좌표계상의 실제 위치값을 검출함으로써 표적 대상의 실제 키를 산출하게 된다.

2.1 표적얼굴 후보영역 및 중심좌표 검출 (1단계)

일반적으로 스테레오 영상은 좌, 우 카메라의 이격 거리가 현격이 떨어지지 않는 한 거의 유사하므로, 한쪽 영상에서 추적물체를 추출하여 상대적인 상관(correlation) 연산을 수행한 뒤 연산된 결과값의 일치된 한 점을 스테레오 영상의 대응점으로 도출하여 획득된 위치 좌표값에 따라 스테레오 카메라의 팬/틸트를 제어하게 된다. 그림 2는 본 논문에서 구현한 얼굴 후보영역 검출부터 중심좌표 검출까지의 알고리즘 흐름도를 나타낸 것이다. 즉, 본 논문에서 제안한 사람 키 높이 추정기법에서는 표적의 얼굴 영역을 검출하기 위해 YCbCr 체계의 칼라 모델을 이용하여 일반적인 조명상태에서의 피부색과 피부색이 아닌 영역으로 분리하게 된다.

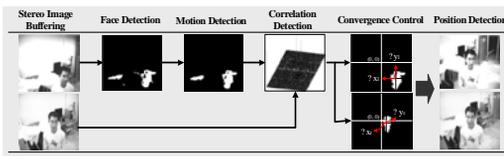


그림 2. 보행자의 얼굴 검출을 위한 흐름도
Fig. 2. Overall operational flowcharts for the person's face detection

여기서, 광도는 Y로 기호화되고, 푸른 정보는 Cb, 붉은 정보는 Cr로 기호화되며, 카메라로부터 입력되는 RGB 체계의 칼라 모델은 식 (1)과 같은 변환 행렬을 이용하여 얼굴영역을 분리하게 된다^[5].

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.16874 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

한편, 영상간의 공통된 움직임을 검출하기위한 방법으로 영상 차분필터를 사용하였으며, 그 결과는 식 (2)와 같이 두 영상간의 움직임 성분으로 나타나게 된다.

$$d_t(a, b) = |I_t(x, y) - I_{t-1}(x+i, y+j)| \quad (2)$$

다음으로, 영상 차분필터의 결과로부터 공통된 움직임을 구별해 내기 위하여 d_{1t} 과 d_{2t} 의 값을 서로 논리곱 연산(AND operation)을 취함으로써 표적영상을 검출할 수 있다. 이런 과정을 통하여 연속으로 입력되는 두 영상으로부터 운동성분을 찾아내게 된다. 즉, 표적 운동 성분값을 $T_m(x, y)$ 라고 하면 식 (3)과 같이 나타낼 수 있으며 B_t 는 이진화 과정을 의미한다.

$$T_m(u, v) = B_t\{d_{1t}(x, y)\} \text{ AND } \{B_t d_{2t}(x, y)\} \quad (3)$$

또한, 검출된 움직임 성분을 토대로 기타 불규칙적인 조명 변화로 인해 생길 수 있는 잡음영역을 제거하기 위해 모폴로지 오프닝(morphological opening) 연산을 수행하였으며, 최종적으로, 4-연결도를 사용한 순차 연결성분 알고리즘 (Sequential Connected Components Algorithm Using 4-connectivity)과 사이즈 필터링(size filtering)을 통해 각 영역별 픽셀의 평균값을 구하여 최대의 평균값을 갖는 영역부터 차례로 얼굴 후보영역으로 검출할 대상영역을 제한함으로써 최종적인 표적의 얼굴 영역만을 검출하게 되고, 그 외의 영역은 배경으로 간주하여 모두 제거하게 된다^[6]. 또한, 추출된 비대칭 형태의 얼굴의 후보영역에 대한 기하 무게 중심법을 이용하여 좌, 우 얼굴 영상의 중심좌표를 검출하게 된다. 식 (4)는 검출된 움직임 성분의 얼굴 후보영역을 각각 n 개의 x, y 화소값으로 나눈값으로 얼굴영역의 무게 중심값을 정의할 수 있으며, 이는 표적 얼굴영상의 중심좌표로 설정하여 스테레오 카메라

시야(FOV : field of view)의 중심으로 이동시키기는 주시각 제어를 위한 팬/틸트 제어값으로 변환된다.

$$\Delta x_l = \frac{\sum_{l=1}^{n_x} x_l}{n_x}, \quad \Delta y_l = \frac{\sum_{l=1}^{n_y} y_l}{n_y} \quad (4)$$

즉, 좌측 영상에서 추적 물체의 위치좌표 $(-\Delta x_l, +\Delta y_l)$ 를 추출하여 중앙으로 위치하도록 $(+\Delta x_l, -\Delta y_l)$ 만큼 이동시켜 새로운 영상인 $I(x, y)$ 을 구한다. 이 영상은 위상형 상관 평면의 기준영상과 다음 프레임의 좌측 영상에서 갱신된 기준 영상으로 사용하게 된다. 그리고 이동값인 $(+\Delta x_l, -\Delta y_l)$ 은 좌측 카메라의 팬/틸트 제어값이 된다. 일반적으로, 정확한 주시각 제어를 위해서는 좌, 우측 카메라에서 추출된 얼굴 후보영역의 일치된 주시점 검출이 필수적이다. 따라서, 본 논문에서는 상관 판별력을 보다 개선하고 효율을 극대화한 시스템으로 발표된 광 BPEJTC(optical binary phase extraction joint transform correlator) 알고리즘[3]을 디지털적으로 변형한 위상형 상관(phase-type correlation) 기법을 이용하여 검출된 얼굴 후보 영역간의 상관을 통해 좌, 우 얼굴 영역의 일치점 및 영상간의 떨어진 상대거리를 구하여 우측 카메라의 팬/틸트 제어값으로 사용하였다.

일반적으로 좌측 기준영상 $I_l(x, y)$ 와 우측영상 $r_l(x, y)$ 간에 푸리에 변환한 주파수 세기 분포인 JTPS(joint transform power spectrum)를 역 푸리에 변환하여 얻은 상관 결과는 식 (5)와 같이 주어진다 [3].

$$\begin{aligned} C(x, y) &= \mathcal{F}^{-1} E(u, v) \\ &= [I'_l(x, y) \otimes r_l(x, y)] * \delta[x + (\Delta x_l - \Delta x_r), \\ &\quad y + (\Delta y_l - \Delta y_r + w)] + [r_l(x, y) \otimes I'_l(x, y)] \\ &\quad * \delta[x - (\Delta x_l - \Delta x_r), y - (\Delta y_l - \Delta y_r + w)] \end{aligned} \quad (5)$$

여기서, \otimes 는 컨벌루션(convolution)을, 그리고 $*$ 는 상관을 각각 나타낸 것이다. 식 (6)은 중심 좌표에 있는 기준영상과 우측영상 간에 떨어진 상대거리를 나타내는 상관 침두치로 식 (4)와 같이 각 영상에 대한 얼굴 중심 좌표값을 검출할 수 있다.

$$x_{peak} = \pm(\Delta x_l - \Delta x_r), \quad y_{peak} = \pm(\Delta y_l - \Delta y_r + w) \quad (6)$$

식 (4)에서 상관 침두치는 중심 좌표에 있는 좌, 우측 입력영상에서 얼굴간의 떨어진 상대거리로 침두치의 위치 좌표를 얻을 수 있으며, 이러한 좌표값으로부터 좌, 우측 얼굴 영상의 위치좌표를 식 (7), (8)과 같이 검출할 수 있다.

$$\Delta x_r = \Delta x_l - x_{peak}, \quad \Delta y_r = \Delta y_l + w - y_{peak} \quad (7)$$

$$\Delta x_l = x_{peak} + \Delta x_r, \quad \Delta y_l = y_{peak} - w + \Delta y_r \quad (8)$$

식 (7), (8)과 같이 검출된 얼굴 영상의 위치좌표는 스테레오 카메라의 팬/틸트 제어값으로 변환되며, 이를 통해 좌, 우측 얼굴 영상의 한 점으로 수렴하는 주시각 제어가 이루어지게 된다.

2.2 팬/틸트 제어각도 산출

본 논문에서는 Kanatani의 영상보정 알고리즘^[6]을 응용하여 2차원 영상평면에서 검출된 이동된 좌표값을 팬/틸트 시스템의 제어각도로 변환하였다. 그림 3은 공간상의 같은 점에 대해 대하여 카메라의 회전으로 인한 두 카메라 좌표의 차이로 인하여 두 영상 면에서의 점의 이동을 나타낸 것으로써 카메라 렌즈의 축을 중심으로 회전하는 카메라에서 획득한 영상은 그 정보가 변하지 않는다는 것을 보여주고 있다.

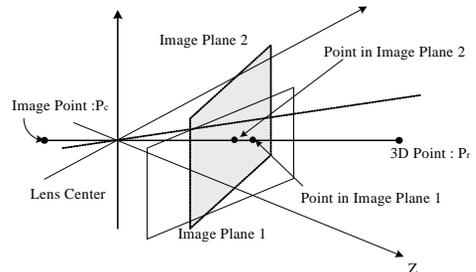


그림 3. 두 영상평면위에 맺힌 3차원 공간상의 점
Fig. 3. 3D point on two image plane with same lens center

그림 3에서, 실제 3차원 공간상으로 수렴되는 임의의 한 점을 $P_r(X, Y, Z)$ 이라 하고, 그 점이 2차원 영상 평면에 투영되는 현재 영상과 이전 영상에서의 점을 각각 $P_c(x_t, y_t)$ 와 $P_c(x_{t-1}, y_{t-1})$ 라고 하면, Kanatani의 영상보정 알고리즘^[6]에 의해 식 (9)와 (10)로 표현될 수 있다.

$$x_{t-1} = f \frac{x_t + \theta \sin \alpha y_t + f \theta \cos \alpha}{-\theta \cos \alpha x_t + \Phi y_t + f} \quad (9)$$

$$y_{t-1} = f \frac{y_t - \phi \sin \alpha x_t + f\theta}{-\phi \cos \alpha x_t + \theta y_t + f} \quad (10)$$

또한 식 (9)와 (10)을 카메라 렌즈의 초점거리 f , 이전영상과 현재영상의 위치 좌표값으로 표현하면 식 (11)과 (12)로 나타낼 수 있다.

$$\theta = \frac{\phi x_{t-1} x_t + f(x_{t-1} - x_t)}{x_{t-1} x_t \cos \alpha + f(y_t \sin \alpha + f \cos \alpha)} \quad (11)$$

$$\phi = \frac{\theta x_t (f \sin \alpha - y_{t-1} \cos \alpha) + f(y_{t-1} - y_t)}{f^2 - y_{t-1} y_t} \quad (12)$$

또한 식 (11)과 (12)를 카메라 렌즈의 초점거리 f , 이전영상과 현재영상의 위치 좌표값으로 표현하면 식 (13)과 (14)로 나타낼 수 있다.

$$\theta = \frac{\phi x_{t-1} x_t + f(x_{t-1} - x_t)}{x_{t-1} x_t \cos \alpha + f(y_t \sin \alpha + f \cos \alpha)} \quad (13)$$

$$\phi = \frac{\theta x_t (f \sin \alpha - y_{t-1} \cos \alpha) + f(y_{t-1} - y_t)}{f^2 - y_{t-1} y_t} \quad (14)$$

따라서, 스테레오 영상의 이전영상과 현재영상의 2차원 화소값으로 구성된 위치값을 검출하면 식 (11)과 (12)를 통해 팬과 틸트 각도가 산출되고 이는 팬/틸트 시스템의 엔코더를 통해 최종적으로 모터의 제어각도로 구동될 수 있다. 결국, 추적대상의 표적 물체를 카메라 시야(FOV)의 중앙에 놓게 하기 위한 팬/틸트 시스템은 영상 평면의 원점 (0, 0)을 중심으로 2차원 영상 평면에 사상된 표적물체의 위치 변이($\Delta x_r, \Delta y_r$)를 팬/틸트 시스템의 제어값인 각도값으로 환산하여 추적 및 감시 기능을 수행하게 된다.

2.3 표적거리 측정 및 표적 영상의 키 높이 추출

PID 제어기와 인코더를 통해 획득된 표적영상의 팬/틸트 제어값은 스테레오 카메라의 시야를 중심으로 좌, 우측 팬/틸트를 이동시키는 주시각 제어를 수행하게 된다. 이때 획득된 좌, 우 카메라의 팬/틸트값은 스테레오 카메라와 표적 대상간의 기하학적 인 관계에 의해 현재 위치한 3차원 좌표는 물론 신체 크기까지도 추정할 수 있다. 감시추적 대상인 표적 대상의 신체 크기는 스테레오 카메라로부터 입력된 좌, 우 입력영상에 대한 팬/틸트 이동각도 및 카메라로부터 얼굴 중심좌표까지의 광축 거리를 이

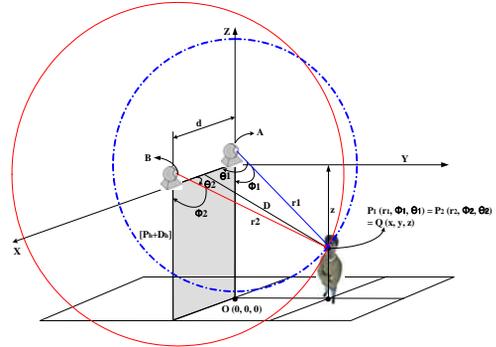


그림 4. 표적 영상의 3차원 구면 좌표계를 이용한 표적영상의 키 높이 검출
Fig. 4. Space model of the proposed stereo video surveillance system

용한 구면 좌표계에서 직교 좌표계로의 변환을 통해서 구할 수 있다. 우선, 그림 4와 같이 표적이 임의의 공간 안에 직립해 있다고 가정했을 때, 주시각이 제어된 스테레오 카메라의 팬/틸트의 각도는 각각 θ_1 과 θ_2 및 ϕ_1 과 ϕ_2 로 주어지게 된다.

따라서, 이를 이용하여 산출된 스테레오 카메라의 수평선상으로부터 표적 영상의 얼굴 중심좌표까지의 수직 거리는 식 (15)와 같이 D로써 표현할 수 있다⁴⁾.

$$D = \left(\frac{\tan(\theta_1) \times \tan(\theta_2)}{\tan(\theta_1) + \tan(\theta_2)} \right) \times d \quad (15)$$

또한 그림 4와 같이 임의의 공간 안에서 감시추적 시스템과 표적간의 기하학적 파라메타들은 표 1과 같이 정의할 수 있다.

표 1. 감시추적 시스템의 기하학적 파라메타
Table 1. Geometric parameters of the surveillance system

파라메타	파라메타의 정의
$\theta_{1,2}$	좌, 우측 팬 각도
$\phi_{1,2}$	좌, 우측 틸트 각도
d	스테레오 카메라 사이의 거리
D	스테레오 카메라와 표적얼굴 사이의 수직거리
P_h	팬/틸트 높이
D_h	팬/틸트를 올려놓은 단상 높이
M_h	표적의 눈높이까지의 키

그림 4에서 각각 θ_1 과 ϕ_1 의 각도를 갖은 팬과 틸트에 탑재된 우측 카메라의 위치가 점 O라고 했을 때, 이는 어떤 점의 위치를 하나의 길이와 두개

의 각으로 나타낼 수 있는 3차원 구면 좌표계로 표현될 수 있으며, 공간 안에서 얼굴영상의 중심좌표를 점 P라고 하면, 이는 (r_1, ϕ_1, θ_1) 의 좌표계로 표현될 수 있다. 여기서, r_1 은 좌, 우 팬/틸트와 표적 영상까지의 거리이므로 스테레오 카메라의 수평축과 표적 영상까지의 수직 거리인 D를 이용하여 식 (16)와 같이 쉽게 구할 수 있다.

$$r_1 = \frac{D}{\sin \theta_1}, \quad r_2 = \frac{D}{\sin \theta_2} \quad (16)$$

따라서, 표적 영상의 얼굴 중심좌표인 점 P(x, y, z)는 식 (17), (18), (19)와 같이 OX, OY, OZ가 서로 직교하는 3차원 직교좌표인 P(x, y, z)로써 변환될 수 있다. 또한, 표적 영상의 중심좌표인 점 P(x, y, z)에 수선을 내린 점 Q(x, y, o)는 표적 대상물이 현재 서 있는 X, Y 좌표계상의 점이므로 표적이 이동하는 움직임 경로를 추정하는데 이용되게 된다.

$$x = r_1 \cdot \cos \theta_1 \cdot \sin \phi_1 = \frac{D}{\sin \theta_1} \cos \theta_1 \cdot \sin \phi_1 = D \cdot \cot \theta_1 \cdot \sin \phi_1 \quad (17)$$

$$y = r_1 \cdot \sin \theta_1 \cdot \sin \phi_1 = \frac{D}{\sin \theta_1} \cdot \sin \theta_1 \cdot \sin \phi_1 = D \cdot \sin \phi_1 \quad (18)$$

$$z = r_1 \cdot \cos \phi_1 = \frac{D}{\sin \theta_1} \cdot \cos \phi_1 = D \cdot \frac{\cos \phi_1}{\sin \theta_1} \quad (19)$$

따라서, 식 (17), (18), (19)에서 검출된 직교 좌표계의 Q(x, y, o)는 실시간으로 표적 물체의 실제 위치를 파악할 수 있으며, 원격 모니터링을 통해 표적 물체의 실제 이동경로를 추정할 수 있는 정보로 이용될 수 있다. 결국, 식 (19)에서 변환된 z값은 식 (20)과 같이 전체 사각 공간안에서 팬/틸트 높이 (P_h) 와 팬/틸트를 올려놓은 단상의 높이 (D_h) 에서 z값을 뺀 후, 일반적인 사람의 머리끝부터 코까지의 길이 (f_h) 만큼 더해지게 되면 대략적인 표적 영상의 키를 추정할 수 있다.

$$M_h = (P_h + D_h) - z + f_h \quad (20)$$

그림 5는 얼굴의 중심좌표에 따라 제어된 틸트

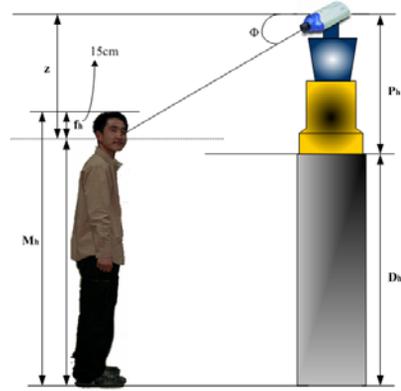


그림 5. 표적 영상의 키 높이 추정
Fig. 5. Height estimation of the target person

각도와 미리 산정된 팬/틸트 높이 (P_h) 및 팬/틸트를 올려놓은 단상의 높이 (D_h) , 그리고 직교 좌표계를 통해 최종 산출된 z값을 이용하여 식 (18)에서와 같이 표적 대상체의 키 높이에 대한 추정을 나타낸 것이다. 따라서, 식 (15), (16), (17)에서 검출된 직교 좌표계값으로 표적 영상의 실제 위치와 이동경로를 추정할 수 있으며, 식 (18)은 원격 모니터링을 통해 얼굴 정보와 함께 감시 대상인물의 키 높이를 추정하는데 이용될 수 있는 정보로 사용될 수 있게 된다.

III. 실험 및 결과분석

본 논문에서 새로이 제안한 스테레오 카메라 기반의 표적 키 높이 추정기법의 실험 시나리오는 그림 6과 같이 일정한 사각 공간 안에서 표적의 얼굴 인식을 통해 미리 설치된 스테레오 감시 카메라로부터 표적의 얼굴 영상까지의 거리 산출 및 추적된 팬/틸트 제어각도를 이용하여 표적의 구면 좌표계를 검출한 후, 직교 좌표계로의 변환을 통해 실제 표적 영상의 3차원 정보 및 키 높이를 검출하는 것이다.



(a) 실험 시스템도 (b) 팬/틸트 제어과정

그림 6. 팬/틸트 기반의 스테레오 표적 감시추적 실험 시스템도
Fig. 6. Experimental setup for the pan/tilt based stereo surveillance system

먼저, 스테레오 영상 획득용 카메라로는 320×240 픽셀의 해상도와 30 frame/sec의 촬영속도를 갖는 웹 카메라를 사용하였으며, 팬/틸트 제어장치는 한울 로보틱스사의 HWR-PT1 시스템 2개를 사용하였다. 또한, 좌, 우 카메라 사이의 이격 거리는 55cm로 고정하였으며, 디지털 시스템은 512MB 메모리의 펜티엄 1.8GHz를 사용하였다.

3.1 얼굴 중심좌표 검출 및 주시각 제어 실험

본 논문에서는 실시간으로 획득되는 스테레오 입력영상 시퀀스에 대해 감시 및 추적 성능을 검증하기 위해 각기 다른 신체 특징을 가진 세 사람의 서로 다른 이동경로를 320×240 픽셀의 해상도를 갖는 스테레오 입력영상을 초당 30 프레임의 속도로 처리하여 총 240프레임의 테스트 영상을 가지고 실험을 수행하였다. 그림 7은 전체 480 프레임의 테스트 실험영상 중, 서로 다른 3사람을 상대로 획득된 3 프레임(3, 98, 161 번째 프레임)에 대한 좌우 스테레오 입력영상 샘플을 나타낸 것이다.

표 2는 실험에 사용된 3개의 샘플 스테레오 입력영상에 대해 얼굴 피부색 임계값인 $77 < C_b < 127$, $133 < C_r < 173$ 의 범위를 적용하여 얼굴 후보영역을 추출하고, 위상형 상관기법을 이용하여 중심좌표



그림 7. 실험에 사용된 스테레오 입력영상
Fig. 7. The input stereo image pairs

표 3. 표적 중심좌표에 대한 구면좌표 파라미터
Table 3. spherical coordinates of target center location

프레임	좌측 카메라		수직거리 (D)	우측 카메라	
	광축거리 (r ₁)	팬/틸트 제어각도 (θ ₁ , φ ₁)		광축거리 (r ₂)	팬/틸트 제어각도 (θ ₂ , φ ₂)
3	204	[91°, 61°]	208	205.4	[60°, 69°]
98	181.1	[83°, 77°]	182	180.5	[84°, 82°]
161	274.1	[96°, 79°]	275	273.4	[69°, 82°]

[단위 : cm]

표 2. 좌, 우 얼굴 영상의 중심좌표
Table 2. Extracted center coordinates of the target face for four sample frames

프레임	[단위 : 픽셀]	
	좌측 얼굴 중심좌표	우측 얼굴 중심좌표
3	(272,88)	(183,95)
98	(209,80)	(106,89)
161	(161,72)	(68,83)

(160, 120)을 기준으로 x, y축 방향으로 이동된 표적 얼굴영상의 위치좌표 값을 나타낸 것이다.

표 2에서 획득된 3 프레임에 대한 위치좌표들은 조명의 밝기 및 명암도 차이에 따라 약간의 차이는 있으나 초당 30 프레임의 영상 처리 속도에도 적응적으로 위치정보를 검출함으로써 실시간적 처리의 가능성을 보여주고 있다. 또한, 표 2에서 획득된 위치좌표값은 각각의 프레임에서 입력된 표적영상의 실제 이동값인 (Δx, Δy)를 의미하며, 이 값은 식 (13)과 (14)를 통해 팬/틸트 제어각도로 변환되며, 변환된 제어값에 따라 카메라의 주시각이 제어됨으로써 이동 표적이 카메라 시야의 중심으로 이동하게 된다.

본 논문에서는 카메라의 광축이 표적영상의 주시점에 일치하도록 하는 교차식 카메라 설정법을 사용하였으므로 표적영상까지의 수직거리(D)는 식 (10)에 의해 구할 수 있으며, 팬/틸트 각도와 함께 구면좌표계를 형성하기 위한 각 카메라와 표적 영상까지의 광축 주시점 거리(r)는 식 (13)을 통해 산출할 수 있다. 또한, 표적영상의 실제 이동값인 (Δx, Δy)을 통해 제어될 팬/틸트 각도 θ과 φ는 각 시스템 엔코더를 사용하여 표 3과 같이 구할 수 있다.

표 3에서, 광축거리(r)과 팬/틸트 제어각도(θ, φ)는 표적 영상의 중심점에 대한 구면좌표계를 형성할 수 있으며, 이는 직교 좌표계로의 변환을 통해 표적물체의 현재의 직립 위치를 추정할 수 있는 정보로 제공된다.

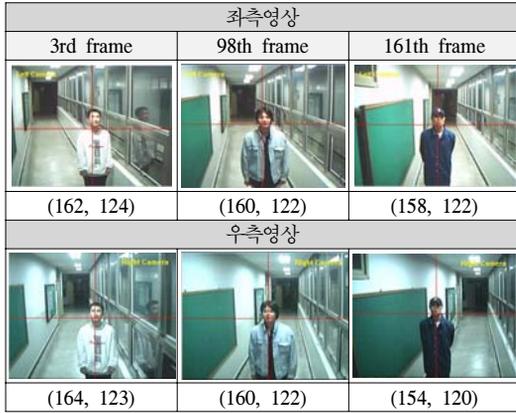


그림 8. 주시각이 제어된 스테레오 영상과 중심좌표
 Fig. 8. stereo image pairs and center coordinates just after being tracked with pan/tilt systems

그림 8은 실시간으로 입력되는 3 프레임의 스테레오 입력영상에 대해 중심좌표인 (160, 120)을 기준으로 표 3의 좌, 우 팬/틸트 각도를 통해 주시각이 제어된 이후의 좌, 우 영상을 나타낸 것이다. 그림 8에서 보면 320×240 해상도의 중심좌표인 (160, 120)을 기준으로 표적영상은 정확히 카메라 시야의 중앙에 위치하는 것을 볼 수 있으며, 실내조명과 밝기의 영향으로 ±8 픽셀의 오차를 보이고 있지만, 실험에 사용한 전체 프레임에 대한 표준 편차는 각각 1.82 및 1.11을 유지하며 얼굴 피부색이 서로 다른 사람에게도 적응적으로 주시각이 제어됨을 볼 수 있다. 또한, 입력되는 스테레오 영상 시퀀스에 대해 프레임 당 0.06초의 처리시간을 가지고 평균 2 프레임 이내에 추적이 이루어짐으로 제안된 감시추적 알고리즘의 실시간적 처리가 가능성이 분석되었다.

3.2 표적 영상의 3차원 위치좌표 검출 및 키 높이 추정 실험

표 4는 제안된 알고리즘을 통해 신출된 값과 실제 측정된 값과의 오차를 분석하기 위해, 표 3에서 검출된 구면 좌표계를 직교 좌표계로 변환한 표적

영상의 3차원 정보와 실험환경에서 측정한 실제좌표를 나타낸 것이다. 즉, 스테레오 카메라와 표적 영상까지의 수직 거리(D) 및 좌, 우 카메라의 광축 주시 거리(r)을 이용하여 구면 좌표계상의 (r, θ, ϕ)을 구성한 뒤, 식 (17), (18), (19)를 이용하여 직교 좌표계상의 (x, y, z)로 변환한 값이다. 표 4에서 보면 좌, 우측 영상 간에 존재하는 직교 좌표계의 오차는 평균 1.5cm 미만의 작은 오차가 존재함을 알 수 있는데, 이는 불규칙한 조명, 동일하지 않은 영상밝기 및 저가의 웹 카메라의 사용임을 고려할 때, 우수한 실험결과로 분석된다. 특히, z 값은 식 (17)과 같이 표적 영상의 팬/틸트 자체 높이인 46cm와 팬/틸트가 설치된 단상 높이인 179cm를 더한 전체높이($P_h + D_h$)인 225cm에서 z 값을 뺀 후, 일반적인 사람의 머리끝부터 코까지의 길이(f_h)인 약 15cm 만큼을 더해지게 되면 표 4의 표적 측정 좌표(x, y, M_h) 중 M_h 과 같이 대략적인 표적 영상의 키를 추정할 수 있다.

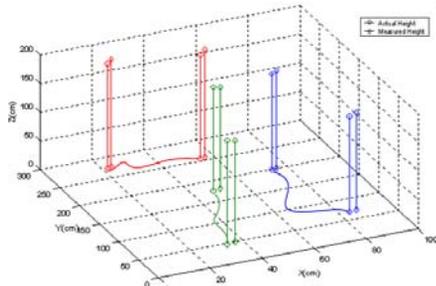
그림 9는 실험 참여 대상에 대한 추정된 신체 크기와 표준 편차로써, 그림 9 (a)는 우측 카메라에서 획득한 좌표값에 비해 실제 좌표와의 오차율이 상대적으로 작은 좌측 카메라에서 획득한 직교 좌표값을 이용하여 표적물체의 추정된 신체 크기를 3차원 공간상에 나타낸 것이다.

따라서, 검출된 직교 좌표계상의 (x, y) 점만을 추적하게 되면 표적 대상체의 이동 경로 및 용의자의 신체 크기를 쉽게 추정할 수 있음을 알 수 있다. 또한, 그림 9 (b)는 실험에 사용된 3사람에 대해 각각 6 프레임씩 샘플링하여 실제 신체 크기를 중심으로 이격된 편차를 나타낸 것이다. 즉, 실험에 참여한 3 사람의 신체 크기(167cm, 175cm, 185cm)를 기준으로 제안된 시스템의 추정된 결과오차를 나타낸 것이다. 또한, 그림 9 (b)과 같이 실험에 참여한 세 사람(3, 98, 161 프레임)에 대해 각각 6 프레임씩의 결과값을 이용한 표준편차는 1.63[SD], 0.4

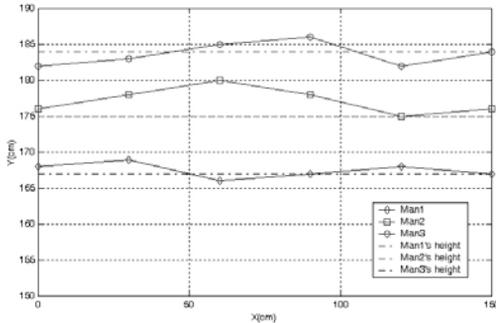
표 4. 표적 영상의 3차원 좌표검출
 Table 4. 3D coordinates extraction of target image

[단위 : cm]

프레임	좌측 카메라의 표적 직교좌표 (x_1, y_1, z_1)	좌, 우측 카메라간 오차	우측 카메라의 표적 직교좌표 (x_2, y_2, z_2)	추정된 신체 크기	실제 신체 크기
3	(74, 199, 48)	$ x =2, y =2, z =2$	(76, 201, 50)	169	167
98	(48, 176, 42)	$ x =3, y =2, z =4$	(51, 178, 38)	176	175
161	(22, 274, 25)	$ x =2, y =3, z =4$	(24, 277, 29)	186	184



(a) 실험에 참여한 세 사람의 추정된 신체 크기



(b) 실제 신체크기와 추정된 크기와의 표준편차

그림 9. 실험 참여 대상에 대한 추정된 신체 크기와 표준 편차
Fig. 9. Illustration of the height estimation and standard deviation

[SD] 및 1.07[SD]를 나타내었으며, 제안된 시스템에서 계산된 좌표값과 실제 실험대상으로 참여한 사람의 정확한 실제 위치와의 차이 역시 평균 1.2cm 미만의 오차와 전체적으로 1.18%의 오차를 보임으로써 보다 적응적인 감시추적 시스템의 실시간 구현 가능성을 제시하고 있다.

IV. 결론

본 논문에서는 팬/틸트가 탑재된 스테레오 카메라 기반의 새로운 사람 키 높이 추정기법을 제안하였다. 480 프레임의 테스트용 스테레오 영상을 사용한 실험 결과, 표적에 대한 실시간 위치정보 검출 및 적응적 추적 성능을 확인하였으며 특히, 팬/틸트 추적후 표적 중심좌표의 수평, 수직 편차가 1.82, 1.11로 유지되고, 표적의 추정된 신체 크기값 역시 계산치 및 측정치 사이의 표준편차가 평균 1.03의 작은 값으로 유지됨이 분석됨은 물론, 제안된 시스템에서 계산된 좌표값과 실제 실험대상으로 참여한 사람의 정확한 실제 위치와의 차이 역시 평균 1.2cm 미만의 오차와 전체적으로 1.18 %의 오차를 보임으로써 이를 통한 새로운 실시간 스테레오 표

적 감시추적 시스템의 구현 가능성을 제시하였다.

참고 문헌

- [1] P. Danielson, "Video surveillance for the rest of us: proliferation, privacy, and ethics education", 2002 International Symposium on Technology and Society, vol.1, no.1, pp.162-167, 2002.
- [2] Mita D. Desai, "Spotlight mode SAR stereo technique for height computation", IEEE Transactions on Image Processing, vol.6, no.10, pp.1400-1411, 1997.
- [3] J. S. Lee, J. H. Ko, E. S. Kim, "Real-time stereo object tracking system by using block matching algorithm and optical binary phase extraction joint transform correlator", Optics Communication, vol.191, pp.191-202, 2001.
- [4] J. S. Lee, C. W. Seo, E. S. Kim, "Implementation of opto-digital stereo object tracking system", Optics Communications, vol. 200, pp.73-85, 2001.
- [5] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space", IEEE Region Ten Conference(TENCON' 2000), vol.2, no.1, pp. 421-424, 2000.
- [6] D. Murray and A. Basu, "Motion tracking with an active camera" IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.16, no. 5, pp.449-459, 1994.
- [7] 고정환, 김은수, "PID기반의 팬/틸트 제어기를 통한 적응적 스테레오 물체추적 시스템의 구현", 한국 통신학회 논문지, vol.27, no.12A, pp.1235-1244, 2002.
- [8] R. Hartley and A. Zisserman, Multiple view geometry, Cambridge University Press, 2000.
- [9] K. H. Bae, J. S. Koo, E. S. Kim, "A new stereo object tracking system using disparity motion vector", Optics Communications, vol. 221, no 13, pp.23-35, 2003.
- [10] Smith, G.D.J., Bors and A.G., "Height estimation from vector fields of surface normals", 2002 14th International Conference on Digital Signal Processing, vol.2, pp.1031-

1034, 2002.

- [11] 박창주, 고정환, 김은수, “스테레오 영상 보정 알고리즘에 기반한 새로운 중간시점 영상합성 기법”, 한국통신학회 논문지, vol.29, no.5C, pp.632-641, 2004.

김 은 수 (Eun-Soo Kim)

종신회원

광운대학교 차세대 3D 디스플레이 연구센터

(<http://3DRC.org>)

한국통신학회지 Vol.29, No. 4c 참조

고 정 환 (Jung-hwan Ko)

정회원



1999년 2월 광운대학교 제어계
측공학과 학사

2001년 8월 광운대학교 전자공
학과 석사

2005년 2월 광운대학교 전자공
학과 박사

2005년 3월~현재 광운대학교

차세대 3D 디스플레이 연구센터 연구교수

<관심분야> 3D 로봇비전, 스테레오 카메라, 영상 통
신