

사람의 움직임 추적을 위한 다중 카메라 기반의 지면 위 발의 대응

Multiple Camera-Based Correspondence of Ground Foot for Human Motion Tracking

서 동욱, 채 현욱, 조강현*
(Dong-Wook Seo, Hyun-Uk Chae, and Kang-Hyun Jo)

Abstract : In this paper, we describe correspondence among multiple images taken by multiple cameras. The correspondence among multiple views is an interesting problem which often appears in the application like visual surveillance or gesture recognition system. We use the principal axis and the ground plane homography to estimate foot of human. The principal axis belongs to the subtracted silhouette-based region of human using subtraction of the predetermined multiple background models with current image which includes moving person. For the calculation of the ground plane homography, we use landmarks on the ground plane in 3D space. Thus the ground plane homography means the relation of two common points in different views. In the normal human being, the foot of human has an exactly same position in the 3D space and we represent it to the intersection in this paper. The intersection occurs when the principal axis in an image crosses to the transformed ground plane from other image. However the positions of the intersection are different depend on camera views. Therefore we construct the correspondence that means the relationship between the intersection in current image and the transformed intersection from other image by homography. Those correspondences should confirm within a short distance measuring in the top viewed plane. Thus, we track a person by these corresponding points on the ground plane. Experimental result shows the accuracy of the proposed algorithm has almost 90% of detecting person for tracking based on correspondence of intersections.

Keywords: multi-camera calibration, correspondence among multiple cameras, multiple background models, homography, Kalman filter

I. 서론

최근 다중 카메라를 이용하여 구조화된 공간에서 인체의 감시 및 제스처 인식 시스템에 대한 연구가 활발히 진행 중이다. 이러한 시스템에서는 객체의 강인한 검출 및 추적이 중요하다. 다중 카메라를 이용하게 되면 단일 카메라가 갖고 있는 제한된 시야를 확장하거나 가림 현상에 대한 문제 해결이 가능하다. 또한 다중 시점을 통해 얻은 정보를 이용하여 객체의 위치, 방향, 형태, 이동 경로 등과 같은 정보를 얻을 수 있다. 이와 같이 여러 가지 장점을 이용할 수 있는 다중 카메라 시스템에서는 다중 시점 사이의 대응 문제, 카메라 교정, 다중 카메라의 위치 설정, 카메라간의 스위칭, 그리고 데이터 혼합과 같은 문제들을 가지고 있다. 이 논문에서는 여러 가지 문제들 중에서 다중 시점 사이의 대응 문제에 대하여 기술한다. 다중 카메라 사이의 대응은 같은 시간에서 다중 시점에서 보이는 객체의 일치를 말한다. 다시 말하면 각 시점에서 보이는 객체의 정보를 혼합하여 하나의 객체로 재구성하는 것이다.

다중 카메라 사이에서 객체의 대응 문제를 해결하기 위해 영역 또는 특징점을 기반으로 하는 방법들을 적용할 수 있다. 영역기반의 방법은 다중 시점의 대응을 위한 특징으로 특정 영역을 이용하는 것이다. 여러 가지 특징 중 컬러는 일반적

으로 교차되는 시점에서 객체의 일치 문제를 해결하는데 많이 사용된다. Mittal and Davis [1]는 교차되는 시점에서의 사람의 일치를 위해 가우시안 컬러 모델(Gaussian color models)을 이용하였다. 또한 Orwell et al. [2]과 Krumm et al. [3]는 다중 시점에서 사람의 일치를 위해 컬러 히스토그램(color histogram)을 이용하였다. 그러나 컬러를 이용하는 방법은 교차된 시점에서 대응 문제를 해결하는데 완벽한 방법이 아니다. 만약 여러 사람이 비슷한 컬러의 옷을 입고 있거나 한 사람이 앞면에는 검은색 뒷면에는 흰색의 옷을 입었을 경우는 한 사람이 아닌 두 사람이라고 인식할 확률이 높다.

특징점을 기반으로 하는 방법은 두 가지 경우로 나누어 생각할 수 있다. 첫 번째는 특징점의 3차원 정보를 이용하는 것이다. Utsumi et al. [4]는 객체의 중심점을 구하고 이것의 월드좌표계에서의 3차원 위치를 예측하여 일치 문제를 해결하였다. 또한 Kelly et al. [5]는 다중 카메라의 교정을 한 다음 3차원 환경 모델을 만들었다. 그리고 공간상에 존재하는 객체의 비슷한 3차원 위치를 구하여 일치 문제를 해결하였다. 그러나 항상 특징점의 3차원 위치가 일치하는 것이 아니기 때문에 이것을 보완하기 위해 2차원 정보를 추가하여 해결하였다. 두 번째는 3차원 정보와 2차원 정보를 함께 사용하는 것이다. Khan et al. [6,7]는 교차된 시점에서 사람의 일치를 위해 사람의 발의 2차원 위치를 호모그래피(homography)를 이용하여 바닥 면(ground plane)에 정의하였다. 그러나 이것도 항상 발의 위치를 구할 수 있는 것이 아니기 때문에 완벽한 방법이 아니다.

이 논문에서는 교차되는 시점에서 객체의 일치 문제를 해

* 책임저자(Corresponding Author)

논문접수 : 2007. 10. 12., 채택확정 : 2008. 5. 8.

서동욱, 채현욱, 조강현 : 울산대학교 전기전자정보시스템공학과

(dwseo@islab.ulsan.ac.kr/hwchae@islab.ulsan.ac.kr/jkh2008@islab.ulsan.ac.kr)

※ 본 연구는 2단계 두뇌한국(BK)21, NARC, 산업자원부, 그리고 한국산업기술재단의 지역혁신 인력양성 사업의 지원에 따라 연구되었음.

결하기 위해 주축과 발의 위치를 이용한다. 주축과 발의 위치를 구하기 위해 먼저 움직이는 사람의 정확한 검출이 필요하다. 움직이는 사람을 정확하게 검출하기 위해 다중 배경 모델을 이용한다. 연속되는 영상 안에서 픽셀의 밝기 변화는 다양하게 일어난다. 다양하게 변화하는 픽셀의 밝기를 단일 배경 모델을 이용하여 특징 짓기가 쉽지 않기 때문에 다중 배경 모델을 이용한다.

교차되는 시점에서 움직이는 사람을 일치시키기 위해 주축을 이용한다. 3차원 공간상에서 있는 사람은 각 시점에서 보이는 모습은 달라도 월드좌표계에서의 위치는 동일하다. 즉, 3차원 공간상에 있는 사람의 주축과 2차원 영상의 주축은 동일하게 된다. 만약 두 영상간의 호모그래피를 이용하게 되면 두 주축의 교점을 얻을 수 있고 이는 결국 3차원 공간상에서 주축과 바닥 면이 만나는 교점과 일치하게 된다. 각 시점에서의 검출된 영역의 주축은 주축 모멘트(principal axis moment)를 이용하여 얻는다.

이 논문은 다음과 같이 구성된다. 2장에서는 다중 카메라 교정 및 다중 배경모델에 관해 서술한다. 3장에서는 움직이는 사람 및 주축 검출에 관해 서술한다. 4장에서 다중 시점의 대응 및 사람의 움직임 추적에 관해 서술한다. 5장에서는 제안된 방법의 검증을 위해 실험 및 결과를 보여준다. 마지막으로 6장에서는 결론을 맺는다.

II. 다중 카메라 교정 및 다중 배경 모델

1. 다중 카메라 교정

카메라 교정은 카메라의 내부변수와 외부변수를 구하는 일련의 과정이다. 내부변수는 카메라의 기하학적 특성, 광학적 특성을 나타내는 것이고 외부변수는 월드좌표계에서의 카메라의 3차원 위치와 방향을 나타내는 것이다. 다중 카메라의 교정은 다섯 대의 카메라를 다섯 쌍의 스테레오 카메라로 묶어 교정한다.

일반적인 카메라 교정은 형태를 미리 알고 있는 패턴을 놓고 그것을 관측함으로써 교정을 한다. 핀홀 카메라 모델(pinhole camera model)의 사영방정식은 다음과 같다.

$$\begin{aligned} \tilde{m} &= P\tilde{M} = A[R|T]\tilde{M} \\ \tilde{m} &= [u \ v \ 1], \tilde{M} = [X \ Y \ Z \ 1] \end{aligned} \tag{1}$$

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} p_1^T & p_{14} \\ p_2^T & p_{24} \\ p_3^T & p_{34} \end{bmatrix}$$

여기서 \tilde{m} 은 영상의 2차원 좌표이고 \tilde{M} 은 영상의 2차원 좌표에 대응되는 패턴의 3차원 좌표이다. P 는 3×4의 사영행렬(perspective projection matrix)을 나타낸다. A 는 카메라의 내부변수이고 R , T 는 카메라의 외부변수를 나타내는 회전행렬(rotation matrix)과 병진벡터(translation vector)이다. 즉, 카메라 교정은 A , R , T 을 구하는 과정을 말한다.

스테레오 카메라의 교정은 단일 카메라 교정을 통해 각 카메라의 사영행렬, 회전행렬, 그리고 병진벡터를 구한 후 그 결과를 이용하여 두 카메라 간의 회전행렬과 병진벡터를 구할 수 있다. 두 카메라의 좌표계와 교정 패턴의 좌표계와의

관계는 다음 식을 만족한다.

$$\begin{aligned} M_{C1} &= R_1 M_w + T_1 \\ M_{C2} &= R_2 M_w + T_2 \end{aligned} \tag{2}$$

여기서 M_{C1} , M_{C2} 는 두 대 카메라 1, 2의 좌표계이고 M_w 는 교정 패턴의 좌표계이다. 단일 카메라 교정법을 이용하여 각 카메라의 사영행렬 P_1 , P_2 을 구하게 되고 구해진 사영행렬을 이용하여 각 카메라의 R_1 , R_2 , T_1 , T_2 을 구하게 된다. 두 카메라간의 관계는 다음 관계를 통해 구해진다.

$$\begin{aligned} M_{C1} &= R_1 R_2^T M_{C2} - R_1 R_2^T T_2 + T_1 \\ &= R M_{C2} + T \end{aligned} \tag{3}$$

$$R = R_1 R_2^T, \quad T = -R_1 R_2^T T_2 + T_1$$

여기서 R , T 는 두 카메라간의 회전과 병진을 나타낸다. 다중 카메라 교정은 다섯 쌍의 스테레오 카메라에 대해 개별적으로 교정을 하게 된다.

2. 다중 배경 모델

움직이는 객체의 검출을 위해 다중 배경 모델을 사용한다. 연속되는 영상 안에서는 픽셀의 밝기 값에 많은 변화가 일어난다. 그림 1은 동일한 공간에서 100 프레임 동안의 한 픽셀의 밝기 값의 변화를 보인다.

그림 1(a)는 100 프레임 동안 픽셀의 밝기 값의 변화가 일정하게 일어나는 것을 알 수 있다. 따라서, 이러한 경우는 하나의 배경 모델을 정의한다. (b)는 20에서 30 프레임 사이에서, 픽셀의 밝기값이 일시적으로 크게 변화한 것을 알 수 있다. 이는 20에서 30 프레임 동안에 움직이는 객체가 지나감을 의미한다. (c)에서는 65 프레임에서 픽셀의 밝기 값의 변화가 크게 나타나고 있다. 이는 65 프레임까지는 하나의 배경 모델을 가지고 65에서 90 프레임까지는 또 다른 하나의 배경 모델을 가지게 된다는 것을 의미한다. 또한 90에서 100 프레임에서 일정 시간 동안 밝기의 변화가 있는데 이는 움직이는 물체가 일정 시간 동안 머문다는 것을 알 수 있다. 이처럼 동일한 공간에서도 연속되는 영상에서는 픽셀의 밝기 값의 변화가 다양하게 일어나는 것을 알 수가 있다. 따라서 단일 배경 모델을 이용하여 특징을 표현하기가 어렵기 때문

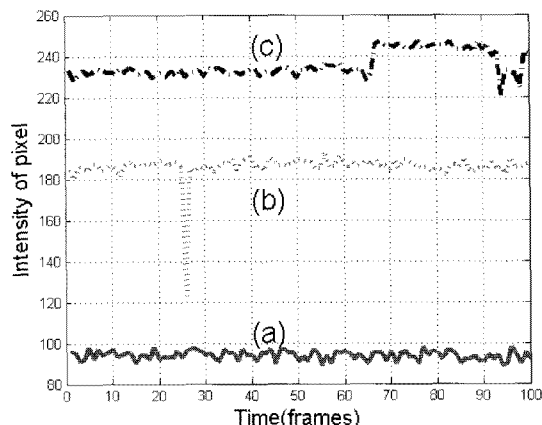


그림 1. 100 프레임 동안의 픽셀 변화.
Fig. 1. Intensity history plot of a pixel in 100 frames.

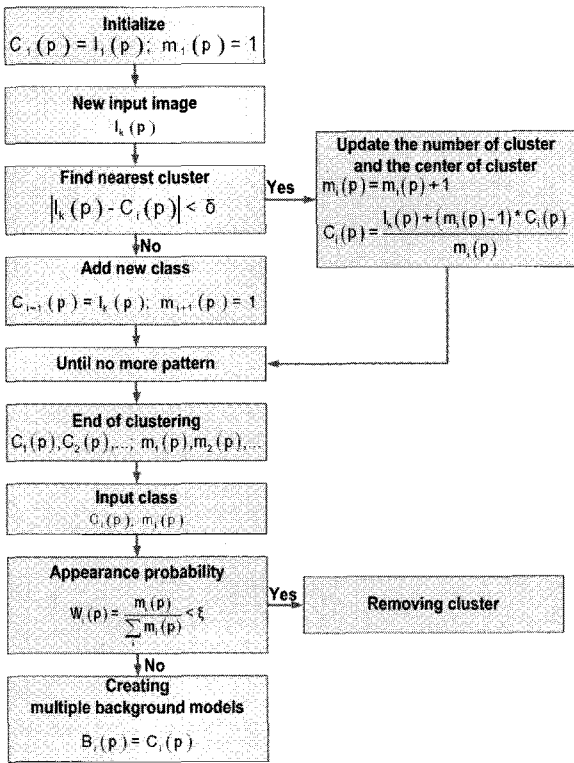


그림 2. 다중 배경 모델의 생성 과정.
Fig. 2. The processing of multiple background models.

에 다중 배경 모델을 이용한다. 다중 배경 모델의 생성 과정은 크게 세 단계로 나눌 수가 있다. 첫 번째 단계는 픽셀의 밝기 값의 온라인 클러스터링 (online clustering) 과정이다. 두 번째 단계는 나타날 확률이 작은 가중값을 가진 클래스를 제거하는 과정이다. 세 번째 단계는 제거하고 남은 클래스를 이용하여 다중 배경 모델을 만드는 과정이다[8,9]. 그림 2는 다중 배경 모델 생성 과정을 보인다.

첫 번째 단계는 픽셀의 밝기 값을 온라인 클러스터링 기반으로 하여 픽셀을 분류하는 과정이다. 연속되는 영상에서 N 개의 프레임을 선택한다.

$$I_k(p), \quad k=1,2,3,\dots,N \quad (4)$$

여기서 $I_k(p)$ 는 k번째 프레임의 픽셀 p의 밝기 값을 나타낸다. $C_i(p)$ 와 $m_i(p)$ 는 i번째 클래스의 평균값과 넘버를 나타낸다. 그리고 δ 는 클래스를 나누기 위한 문턱값(threshold value)이다. 첫 번째로 들어오는 픽셀의 밝기 값이 초기 클래스의 평균값이 되고 초기 클래스의 넘버는 1이 된다. 두 번째 단계는 나타나는 빈도수가 낮은 클래스를 제거하는 과정이다. 각 클래스의 가중값은 다음 식을 이용하여 구한다.

$$W_i(p) = \frac{m_i(p)}{\sum_{i=1}^{N(p)} m_i(p)}, \quad i=1,2,3,\dots,N(p) \quad (5)$$

여기서 $W_i(p)$ 는 i번째 클래스의 가중값이고 $N(p)$ 는 클래스의 전체 개수이다. 문턱값 보다 작은 가중값을 제거한다.

마지막으로 다중 배경 모델을 생성한다.

$$B_i(p) = C_i(p), \quad (i=1,2,3,\dots,n(p) \quad n(p) \leq N(p)) \quad (6)$$

여기서 $B_i(p)$ 는 i번째 클래스의 배경 모델이다. 빈도수가 낮은 클래스가 제거된 후의 개수를 $n(p)$ 라 하고 다중 배경 모델을 생성 후 새로운 가중값을 계산한다.

$$W(p)_i = \frac{m_i(p)}{\sum_{i=1}^{n(p)} m_i(p)}, \quad i=1,2,3,\dots,n(p) \quad (7)$$

III. 인체의 영역 및 주축 검출

1. 인체의 영역 검출

이 논문에서는 다중 카메라를 이용하여 구조화된 공간에서 움직이는 객체는 사람만 존재한다고 가정한다. 인체의 영역 검출은 앞 장에서 정의한 다중 배경 모델을 이용한다. 그림 3은 인체의 영역 검출 과정을 보여준다. 여기서 일정 시간 동안에 생성된 다중 배경 모델은 MBM_i라고 한다. i는 배경 모델의 개수를 말한다. 현재와 과거 영상을 I_k 와 I_{k-1} 로 정의한다. 먼저 두 가지의 다른 영상을 획득한다. 하나는 다중 배경 모델 MBM_i과 현재 영상 I_k 의 차연산의 결과 영상 I_{Diff_k} 이고 또 다른 하나는 현재 영상 I_k 와 과거 영상 I_{k-1} 의 차연산의 결과 영상 $I_{Diff_{k-1}}$ 이다.

$$I_{Diff_k}(x,y) = |I_k(x,y) - MBM_i(x,y)| \quad i=1,2,3,\dots \quad (8)$$

$$I_{Diff_{k-1}}(x,y) = |I_k(x,y) - I_{k-1}(x,y)|$$

여기서 얻은 결과 영상에서의 각 픽셀은 움직이는 객체일 확률이 높은 값들이다. 두 쌍의 문턱값 σ_b 와 σ_k 을 이용하여 문턱값보다 높은 값은 남기고 낮은 값은 제거한다. 픽셀의

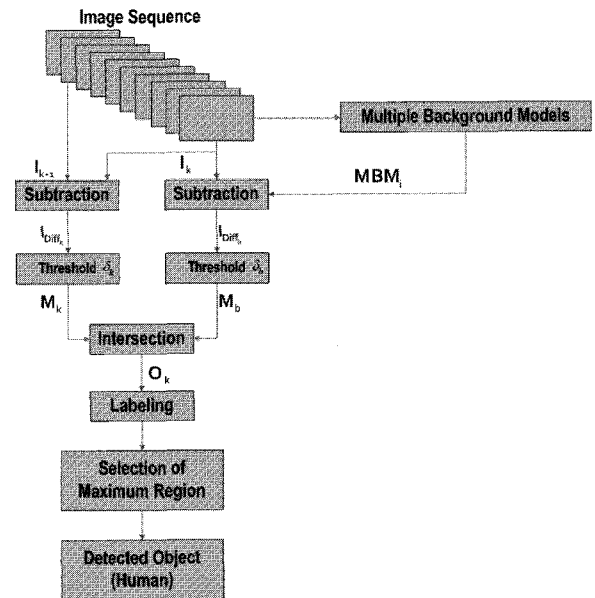


그림 3. 인체의 영역 검출 과정.
Fig. 3. The process of human region detection.

밝기가 '1'인 것이 객체이고 '0'인 것은 배경이 된다. 다중 카메라에 의해 구조화된 공간에서 움직이는 객체는 사람만 존재한다고 가정하였다. 즉, 움직이는 사람을 제외한 모든 객체들은 고정된 물체로 생각하기 때문에 움직이는 객체의 영역의 크기는 다른 고정된 객체들의 영역의 크기보다는 크게 될 것이다. 따라서 움직이는 객체의 영역은 차연산 결과에서 여러 영역들 중에서 가장 큰 영역이 움직이는 객체가 될 확률이 높다. 그래서 문턱값은 여러 영역들 중에서 가장 큰 영역만을 검출할 수 있는 값을 반복적인 실험을 통해 선택한다.

$$M_b(x, y) = \begin{cases} 1 & I_{Diff_b}(x, y) > \sigma_b \\ 0 & otherwise \end{cases} \quad (9)$$

$$M_k(x, y) = \begin{cases} 1 & I_{Diff_k}(x, y) > \sigma_k \\ 0 & otherwise \end{cases} \quad (10)$$

마지막으로 두 개의 결과 영상 M_b 와 M_k 를 이용하여 움직이는 객체를 얻는다.

$$O_k(x, y) = \begin{cases} 1 & M_k(x, y) \cap M_b(x, y) \\ 0 & otherwise \end{cases} \quad (11)$$

여기서 O_k 는 움직이는 객체를 검출한 결과 영상이다. 검출된 결과 영역들을 레이블링(labeling)하여 가장 큰 영역을 선택한다. 선택된 영역이 움직이는 사람이 된다.

2. 주축 검출

검출된 객체의 주축은 주축 모멘트를 이용하여 얻는다. 디지털 영상 $f(x, y)$ 에 대한 pq차 모멘트(pq-th moments)는 다음과 같다.

$$M_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (12)$$

여기서 x, y 는 영상내의 임의의 한 점의 좌표, $f(x, y)$ 는 픽셀의 밝기 값을 나타낸다.

주축을 구하기 이전에 검출된 영역의 무게 중심을 계산한다.

$$\bar{x} = \frac{M_{10}}{M_{00}}, \bar{y} = \frac{M_{01}}{M_{00}} \quad (13)$$

여기서 \bar{x} 와 \bar{y} 가 영역의 무게 중심이다. M_{10}, M_{01}, M_{00} 는 p와 q에 의해 결정된다.

영역의 무게중심에 의해 주축 방향은 다음과 같이 결정된다.

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{2I_{xy}}{I_{yy} - I_{xx}} \right) \quad (14)$$

여기서 (14)가 주축 모멘트이다. 여기서 θ 는 영역에 의해 나타나는 주축과 기준 좌표계가 이루는 각이다. 그리고 I_{xx}, I_{yy}, I_{xy} 는 영역 내의 각 점에 대하여 각각 x축 방향과 y축 방향으로의 관성 모멘트, 그리고 x와 y 두 축의 방향에 대한 상승 모멘트이다[10].

IV. 다중 시점의 사이의 대응

1. 호모그래피

호모그래피는 2차원 공간 간의 변환으로 정의된다. 한 카메라에서 보이는 바닥 면 위의 임의의 한 점과 다른 카메라에서 보이는 같은 점에 대한 관계를 설명한다. 호모그래피 행렬(homography matrix) H는 3×3으로 정의한다.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (15)$$

두 시점에서의 바닥 면 위에서 대응되는 점들을 (x_i, y_i) 와 (x'_i, y'_i) 이라 한다. (15)를 사용하여 두 점 사이의 대응관계는 다음과 같이 정의된다.

$$\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (16)$$

이 논문에서 행렬 H 는 바닥 면에 부착된 랜드마크(landmark)를 이용하여 구한다.

2. 다중 시점의 기하학적 관계 및 대응

그림 4는 다중 시점의 기하학적 관계를 보이고 있다. L'_S 는 시점 i에서 사람 S의 주축이다. 영상 평면 i에서 영상 평면 j로 호모그래피 행렬을 H 라 한다. 영상 평면 j에서 영상 평면 i로 역방향의 변화는 H^{-1} 가 된다. H^{-1} 에 의해 영상 평면 j에서 주축 L'_k 는 영상 평면 i에서의 L''_{KS} 가 된다. L''_{KS} 와 주축 L'_S 와 교점은 Q''_{KS} 이다. 즉, 이 교점은 3차원 공간상에서 주축 L과 바닥 면과의 교점 X가 된다. 즉, 교점 Q''_{KS} 는 바닥 면의 점이라 할 수 있다. 왜냐하면 영상 평면 i에서 주축 L'_S 가 바닥 면에 투영되면 g'_S 가 되고 영상 평면 j에서 주축 L'_k 가 바닥 면에 투영되면 g'_k 가 된다. 3차원 공간상에서 두 직선의 교점은 X가 된다. 영상 평면 j에서 L'_k, Q''_{SK}, L''_{SK} 는 동일한 방법으로 정의된다.

두 영상에서 두 점 Q''_{SK} 와 Q''_{KS} 는 대응된 점이라 가정한다. $Q''_{KS}(x_1, y_1, 1), X'_S(x'_1, y'_1, 1), Q''_{SK}(x_2, y_2, 1),$ 그리고 $X'_K(x'_2, y'_2, 1)$ 라 하면 두 점 사이의 관계는 다음과 같이 정의

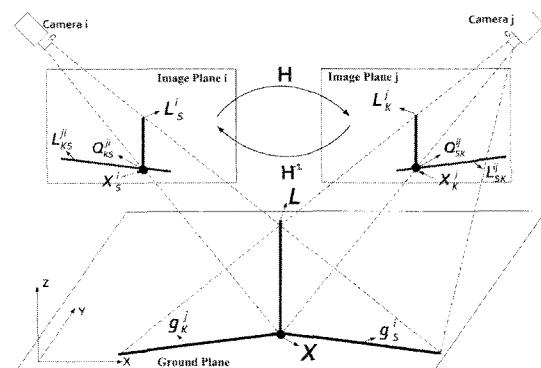


그림 4. 다중 시점 사이의 기하학적 관계.
Fig. 4. The geometrical relationship between multiple views.

된다.

$$\begin{aligned} X_S^i &= \mathbf{H}^{-1} Q_{SK}^i \\ X_K^j &= \mathbf{H} Q_{KS}^j \end{aligned} \quad (17)$$

여기서 두 점의 쌍 (X_S^i, Q_{KS}^i) , (X_K^j, Q_{SK}^j) 을 이용하여 3차원 위치를 구하면 3차원 공간상의 교점 X 와 동일한 점이 된다. 두 점 사이의 기하학적 거리는 다음과 같이 정의한다.

$$\begin{aligned} d_1(Q_{KS}^i, X_S^i) &= \sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2} < \varepsilon_1 \\ d_2(Q_{SK}^j, X_K^j) &= \sqrt{(x_2 - x_2')^2 + (y_2 - y_2')^2} < \varepsilon_2 \end{aligned} \quad (18)$$

여기서 d_1 과 d_2 가 문턱값보다 작다면 두 교점 Q_{SK}^i 와 Q_{KS}^j 는 대응되는 점이 된다. 지금까지 두 대의 카메라에 대한 대응에 관해 서술하였다. 나머지 카메라에 대해서도 동일한 방법을 적용한다.

3. 사람의 움직임 추적

다중 시점 사이의 대응되는 점들을 이용하여 각 시점에서 사람의 움직임을 추적한다. 이 논문에서는 다중 카메라에 의해 구조화된 공간에서는 움직이는 객체는 사람만 존재한다고 가정하였다. 또한 만약 한 시점에서 교점이 다른 물체에 의해 가려진다면 다른 시점들에서 매칭되는 점들을 이용하여 예측한다. 사람의 움직임의 추적은 칼만 필터(Kalman filter)[11]를 이용한다.

V. 실험 및 결과

이번 장에서는 제안된 알고리즘의 검증을 위해 실험과정 및 결과를 보여준다. 실험을 위해 5대의 카메라를 그림 5와 같이 설치하였다. 카메라의 번호는 C1~C5로 부여하였다.

카메라는 동일 기종의 5대의 카메라를 사용하였다. 실험을 위한 영상은 실내에서만 촬영된 영상을 이용하였다.

1. 다중 카메라 교정

다중 카메라 교정은 다섯 대의 카메라를 다섯 쌍의 스테레오 카메라로 묶어 교정을 한다. 각 쌍에 대한 스테레오 카메라 교정 결과는 표 1과 같다. 다중 카메라 교정 오차의 범위는 X축 방향에서는 0.03mm ~ 9.90mm이고 Y축 방향에서는 0.04mm ~ 8.42mm이다. 그리고 Z축 방향에서의 오차 범위는 0.07mm ~ 11.46mm가 된다.

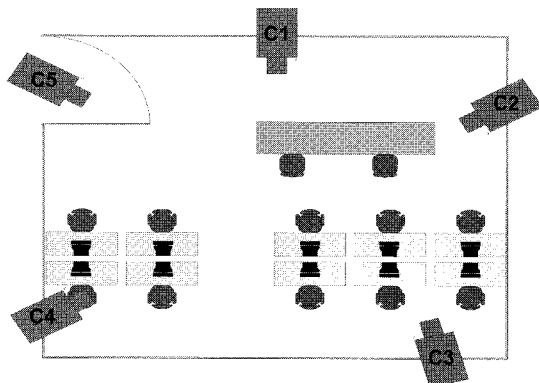


그림 5. 다중 카메라 배치도.
Fig. 5. The configuration of multiple cameras.

표 1. 다중 카메라 교정 오차.

Table 1. Multi-camera calibration error.

Stereo pairs	Error type	Multi-camera calibration error [mm]		
		X	Y	Z
C1, C2	Min	0.03	0.04	0.28
	Mean	1.89	2.92	2.67
	Max	5.10	8.31	5.15
C2, C3	Min	0.21	0.24	0.29
	Mean	2.20	1.95	2.68
	Max	9.66	5.10	7.31
C3, C4	Min	0.12	0.18	0.49
	Mean	4.54	2.65	4.30
	Max	8.06	8.42	11.46
C4, C5	Min	0.89	0.10	0.07
	Mean	5.81	2.45	2.61
	Max	9.90	7.06	7.65
C5, C1	Min	0.33	0.76	0.12
	Mean	2.03	2.76	1.40
	Max	4.63	6.72	3.00

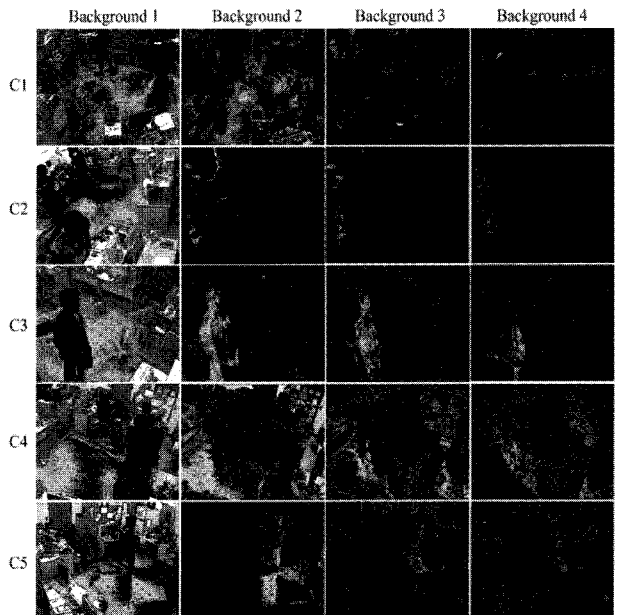


그림 6. 다중 배경 모델.
Fig. 6. The multiple background models.

2. 다중 배경 모델

그림 6은 100 프레임 동안에 생성된 다중 배경 모델을 보여준다. 각 영상마다 네 장의 배경 모델이 생성된 것을 볼 수 있다. 맨 위부터 행이 각 C1 ~ C5를 나타내고 각 열이 배경 모델이다. 다중 배경 모델을 만들기 위한 문턱값은 실험을 통해 결정하였다.

3. 사람 및 주축 검출

그림 7은 다중 배경 모델을 이용하여 움직이는 사람 및 주축 검출의 결과를 보여준다. 각 카메라에 대해서 149, 726, 그리고 1325 프레임에서의 사람 및 주축의 검출 결과이다. 녹



그림 7. 사람 및 주축 검출 결과.
Fig. 7. The results of detected person and principal axis.

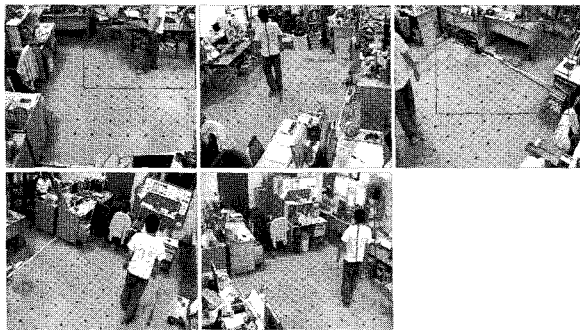


그림 8. 잘못 검출된 인체의 영역과 주축의 예.
Fig. 8. Example of incorrectly detected person and principal axis.

표 2. 인체의 영역 검출결과.
Table 2. The result of detected human region.

	전체	검출된 객체	잘못 검출된 객체
C1	1586	1348/1586	111/1348
		84.99 %	8.23 %
C2	1586	1523/1586	25/1523
		96.03 %	1.64 %
C3	1586	1361/1586	62/1361
		85.81 %	2.35 %
C4	1586	1445/1586	18/1445
		91.11 %	1.25 %
C5	1586	1491/1586	12/1491
		94.01 %	0.8 %
Total	7930	7168/7930	228/7168
		90.39 %	3.18 %

색 바운딩 박스의 부분이 검출된 사람의 영역이고 파란색 선이 검출된 주축이다.

그림 8은 잘못 검출된 사람 및 주축의 결과를 보여준다. 영역이 검출되지 않은 것은 다중 배경 모델을 이용하여 움직이는 객체를 검출할 때 영역이 가장 큰 것을 움직이는 객체로 검출하게 되는데 이 조건을 만족하지 못하거나 배경으로 인식된 경우이다. 또한 움직이는 객체가 아닌 다른 부분이 더 큰 영역으로 검출되었기 때문이다. 이를 해결하기 위해서는 배경 모델의 업데이트를 하거나 컬러와 같은 또 다른 특징을 이용하면 된다. 그러나 이 논문에서는 고려하지 않고 대응되는 교점을 구할 때의 조건을 이용하여 잘못된 검출결과를 배제하게 된다.

표 2는 사람의 검출결과를 보인다. 실험을 위해서 1586장의 영상을 사용하였고 90% 이상의 검출 결과를 얻은 것을 알 수 있다. 검출되지 않은 부분은 움직이는 사람의 영역이 아닌 다른 영역의 크기가 더 큰 영역으로 검출되었기 때문이다. 이는 이 논문에서는 배경 모델의 업데이트를 하거나 컬러와 같은 다른 특징을 고려하지 않았기 때문이다. 검출된 결과 중에 사람이 아닌 다른 객체가 검출된 경우는 3.18%로 나왔다. 이는 대응되는 교점을 구할 때의 조건을 이용하여 배제하게 된다.

4. 다중 시점 사이의 대응

그림 9는 다중 시점 사이의 대응 결과 보여준다. 147, 945 프레임에서 스테레오 카메라 쌍에 대한 대응 결과를 보인다.

빨간색 바운딩 박스는 사람의 영역을 나타내는 것이고 파란색 선은 주축을 나타낸다. 그리고 흰색 선은 각 주축에 대한 호모그래피에 의한 변화된 것이다. 빨간색 점이 두 주축에 의한 교점이다. 두 시점 사이에서 기하학적 오차가 가장 작은 것이 대응점이다. 그림 10은 같은 시간대에서 다섯 대의 카메라들에 대한 대응 결과를 보여준다. 101, 220, 그리고 1330 프레임에서의 대응 결과이다. 대응된 교점은 빨간색 점이다. 다섯 대에서 동시에 대응된 결과는 비슷한 위치를 보이고 있다.

5. 사람의 움직임 추적

각 시점에서 대응관계를 찾은 다음에 그 결과를 이용하여 사람의 움직임을 추적한다. 그림 11은 카메라 1과 카메라 2에서의 사람의 움직임 경로를 보인다.

그림 12는 칼만 필터(Kalman Filter)를 이용하여 시점 1에서



그림 9. 다중 시점 사이의 대응 결과.
Fig. 9. The results of correspondence between multi-views.

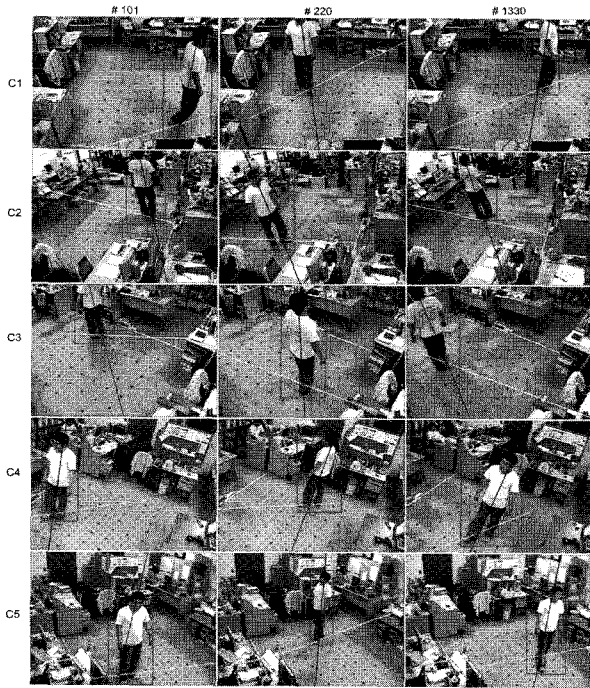
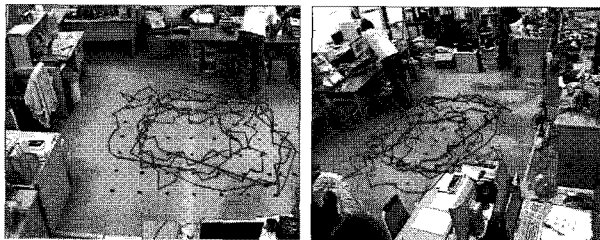


그림 10. 다섯 대의 카메라에서 대응 결과.
Fig. 10. The results of correspondence between five cameras.



(a)시점 1에서 움직임의 경로 (b)시점 2에서 움직임의 경로

그림 11. 각 시점에서 움직임의 경로.
Fig. 11. Trajectories in each view.

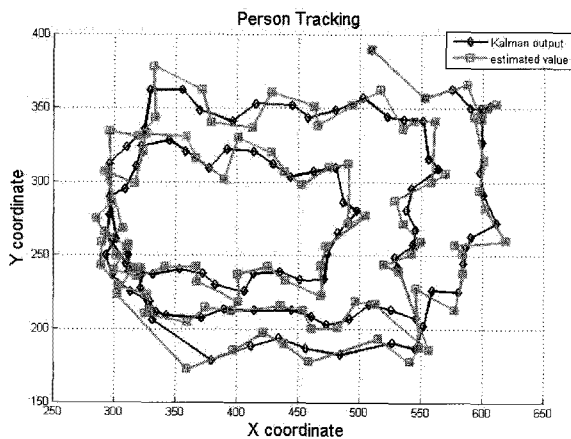


그림 12. 3차원 공간상에서 사람의 움직임 추적.
Fig. 12. Tracking a person in 3D space.

사람의 움직임을 추적한 결과이다. 초록색 선은 예측된 결과이고 파란색 선이 이를 바탕으로 한 칼만 필터의 결과이다.

VI. 결론

본 논문에서는 다중 카메라를 이용하여 구조화된 공간에서 다중 시점 사이의 대응 문제에 대하여 서술하였다. 다중 시점 사이의 대응 문제를 해결하기 위해 움직이는 객체의 주축과 호모그래피를 이용하였다. 이 논문에서는 움직이는 객체는 사람만 존재한다고 가정하였다. 움직이는 객체를 정확하게 얻기 위해 다중 배경 모델을 이용하였다. 단일 배경 모델을 이용해서는 한 영상에서 다양하게 변화하는 상황을 특징지을 수 없기 때문에 다중 배경 모델을 이용하였다.

3차원 공간상에 존재하는 사람의 주축과 바닥 면의 교점을 이용하였다. 교차되는 시점에서 주축과 바닥 면의 교점은 동일한 점이다. 이 관계를 이용하여 다중 시점 사이의 대응을 해결하였다. 이 교점은 한 시점에서 보이는 주축과 다른 시점에서 호모그래피에 의해 변화된 주축과의 교점과 동일하다.

두 영상에 존재하는 교점과 호모그래피에 의해 변화된 점 사이의 기하학적 오차가 작은 두 교점이 대응되는 점이 되었다. 만족되는 점을 이용하여 사람의 움직임을 추적하였고 3차원 공간상의 위치도 구하였다. 움직이는 사람과 주축이 정확하게 검출되면 두 점 사이의 기하학적 오차가 거의 영에 가까운 값을 가지는 대응점을 얻을 수 있다. 그러나 영역이 부정확하고 주축이 부정확하게 검출되면 잘못된 대응관계를 얻을 수 있다. 이를 해결하기 위해 움직이는 객체에 대한 모델을 설정하거나 컬러와 같은 또 다른 특징을 이용하면 더 정확한 영역을 검출할 것이다. 또한 이 논문에서는 움직이는 사람이 한 명만 있는 경우에 대해서만 생각하였는데 여러 명이 존재하는 상황에서 이들을 추적하고 행동을 해석하는 것이 향후 과제로 남아있다.

참고문헌

- [1] A. Mittal and L. S. Davis, "M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene using Region-Based Stereo," *Proc. European Conf. Computer Vision*, pp. 18-36, May 2002.
- [2] J. Orwell, P. Remagnino, and G. A. Jones, "Multiple Camera Color Tracking," *Proc. IEEE Int'l Workshop Visual Surveillance*, pp. 14-24, 1999.
- [3] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer, "Multi-Camera Multi-Person Tracking for EasyKiving," *Proc. IEEE Int'l Workshop Visual Surveillance*, pp. 3-10, July 2000.
- [4] A. Utsumi, H. Mori, J. Ohya, and M. Yachida, "Multiple Human Tacking using Multiple Cameras," *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 498-503, Apr. 1998.
- [5] P. Kelly, A. Katkee, D. Kuramura, S. Moezzi, S. Chatterjee, and R. Jain, "An Architecture for Multiple Perspective Interactive Video," *Proc. ACM Multimedia*, pp. 201-212, Nov. 1995.
- [6] S. Khan and M. Shah, "Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1355-1360, Oct. 2003.
- [7] S. Khan, O. Javed and M. Shah, "Tracking in Uncalibrated Cameras with Overlapping Field of View," *Proc. IEEE Int'l*

Workshop Performance Evaluation of Tracking and Surveillance, pp. 84-91, Dec. 2001.

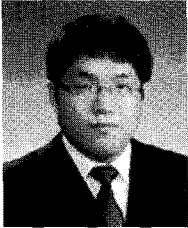
[8] M. Xiao, C. Han, and X. Kang, "A Background Reconstruction for Dynamic Scenes," *Int'l Conf. on Information Fusion*, pp 1-7, 2006.

[9] T. Kim and K.-H. Jo, "Robust Background Acquisition from Dynamic Scene caused by a Moving Camera," *ICIA 2007*,

Jul. 2007.

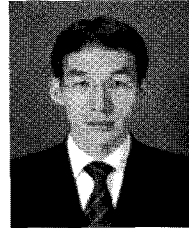
[10] Y.-H. Yoon, "The Korean Sign Language Recognition System using the Hidden Markov Models and Moment Invariants," University of Ulsan, 2003.

[11] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *SIGGRAPH 2001*, Aug. 2001.



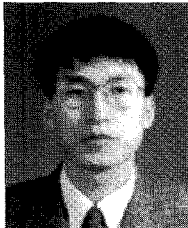
서동욱

2006년 울산대학교 전기전자정보시스템 공학부 졸업. 2008년 동대학원 전기전자 정보시스템공학과 졸업(공학석사). 관심 분야는 이동로봇, 컴퓨터비전, 지능시스템제어.



채현욱

2007년 울산대학교 전기전자정보시스템 공학부 졸업. 2007년~현재 동 대학원 전기전자정보시스템공학과 석사과정 재학 중. 관심분야는 얼굴검출 및 인식, 컴퓨터비전, 지능시스템제어.



조강현

1989년 부산대학교 정밀기계공학과 졸업. 1993년 일본 오사카대학 대학원 전자제어기계공학과 졸업(공학석사). 1997년 일본 오사카대학 대학원 전자제어기계공학과 졸업(공학박사). 1997년~1998년 ETRI 시스템공학연구소 포스트닥터 연구원. 1998년~현재 울산대학교 전기전자정보시스템공학부 부교수. 2005년 7월~2006년 7월 일본 큐슈대학 객원 교수 및 KIST 지능로봇연구센터 객원 연구원. 관심분야는 휴먼인터랙션(HCI), 컴퓨터비전, 지능형로봇 및 시스템제어.