

기하학적 정보를 이용한 영상 시퀀스에서 높이 추정에 관한 연구

김상훈, 김종수, 윤용인, 김진태*, 최종수
중앙대학교 첨단영상전문대학원 영상공학과
*한서대학교 컴퓨터 정보학과

Height Measurement using Geometric informations from image sequences

Sang-Hoon Kim, Jong-Su Kim, Yong-In Yoon, Jin-Tae Kim*, Jong-Soo Choi
Dept. of Image Engineering,
Advanced Graduate School of Image, Multimedia and Film, Chung-Ang Univ.
*Dept. of Computer Information Science, HanSeo Univerity

요약

본 논문에서는 보안 시스템에서 사람 인식을 위한 중요한 단서 중의 하나인 사람의 키를 측정하는 알고리즘으로, 이미지 시퀀스에서 사람의 영역을 추출하고 기하학적 정보를 이용해 추출된 사람의 키를 측정하고자 한다. 이를 위해 단일 이동 물체 사람을 대상으로 하여 시퀀스 이미지에서 사람의 움직임 정보를 추출하고, 추출된 영역에서 수직 히스토그램 투영을 하여 사람의 중심선을 찾아 머리와 발의 좌표점을 추출한다. 추출된 좌표점들은 소실점과 소실선의 기하학적 해석과 미리 입력한 다른 물체의 기준 높이를 가지고 실세계에서의 사람의 키를 측정하게 된다.

I. 서론

비디오 카메라 보안 시스템의 이용은 개인과 공공의 목적으로 증가 추세에 있다. 이 시스템은 날로 증가하고 있는 범죄와 사건을 해결하는 실마리를 제공해 줄 수 있다. 기존의 논문들에서 모니터링, 조절, 추출, 인식, 확인과 같은 특별한 목적을 위해서 어떻게 CCTV를 장치하는가에 방향을 제시하고 있다. 이런 제시는 특별한 목적을 위한 운영 시스템에서 시스템을 운영하는 기사들에게 도움을 주고 있다. 일반적으로 사건이 비디오에 포착되었을 때, 사건의 이해를 위해서 또는 개인의 신상을 파악하기 위해서 테이프를 돌려본다. 많은 경우에 증거나 정보목적의 자료로 비디오의 화질이나 내용이 충분하다. 그러나, 대부분이 영상 향상이나 판독하는데 부적합 경우이다. 테이프의 질이 개인을 식별하는데 충분하지 못한 경우로서, 개

인이 카메라로부터 멀리 떨어져 있는 경우 또는 위장을 했을 때 발생할 수 있다. 이러한 현상을 발생시키는 다른 요소로는 질이 안 좋은 녹화장비, 조명 효과 등을 들 수 있겠다. 이런 경우 개인을 식별하기 위해 다른 방법에 의존해야 하는데 그 중에 중요한 하나의 요소로 사람의 키가 되겠다.

이 논문에서는 최소한의 기하학적 정보를 이용해 이미지로부터 사람의 영역을 추출하고 이로부터 한장의 영상뿐 아니라 이미지 시퀀스로부터 사람의 높이를 계산하는 알고리즘에 대해 소개한다.

그림 1은 본 논문의 전체적인 알고리즘 구조를 나타낸 것이다.

II. 시퀀스 영상에서 영역 추출.

1. 시퀀스 영상에서 영역 추출.
고정된 카메라를 이용하여 시퀀스 영상에서 이동물

체의 존재 여부를 감지하기 위해서 본 연구에서는 기준 프레임을 사용하였다. 카메라를 통해서 처음 입

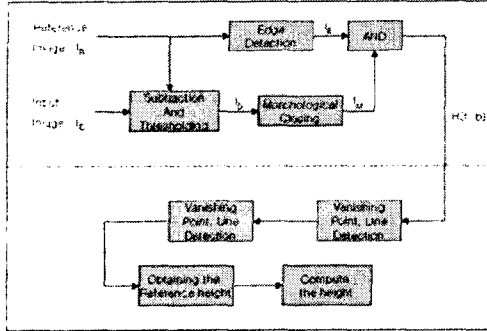


그림 1. 전체 알고리즘 구조.

력된 영상을 기준프레임으로 설정하여, 현재 입력된 영상에서 기준 프레임을 빼게 되면 움직임 영역을 추출할 수 있게 된다. 하지만 기준 프레임을 이용할 경우, 조명의 변화에 따라 매우 민감하게 반응하므로 본 연구에서는 조명의 변화가 작은 시퀀스 영상을 대상으로 하였고 단일 물체의 움직임을 대상으로 하여 움직임 영역을 추출하였다.

현재 입력 영상에서 기준 프레임의 영상을 빼게 되면, 움직임이 존재하는 영역에서는 화소값의 변화가 발생하게 되고 따라서 이들의 변화율은 움직임의 정도를 나타내게 된다. 움직임이 있는 영역에서는 큰 에너지를 갖는 영상 I_d 를 얻을 수 있다. 따라서 다음과 같은 수식으로 나타낼 수 있게 된다. 또한 식(1)은 문턱치 th 를 이용하여 식(2)와 같이 근사화 시킬 수 있다.^[1]

$$\frac{\partial I(x, y, kt)}{\partial t} \approx \frac{I(x, y, t) - I(x, y, t - \delta kt)}{\delta t} \quad (1)$$

$$I_d(x, y) = u\{(|I(x, y, t) - I(x, y, t - \delta kt)| - th)\} \quad (2)$$

여기서, δt 는 영상 취득 간격이며, $I(x, y, t)$ 와 $I(x, y, t - \delta kt)$ 는 각각 시간 t 와 $t - \delta kt$ 에서 취득한 영상의 밝기 함수(intensity function)이다. $u[\]$ 는 단위계단 함수를 의미하며, 따라서 문턱값 th 에 의한 이진화 과정이 이루어진다. 기준 영상과 현재 입력 영상의 차영상에서는 주로 움직임 영역에서 에너지가 집중되지만, 잡음에 의하여 그 밖의 영역에서도 미세한 이동성분들이 생기게 된다. 이런 잡음 성분들은 영역 추출의 정확성을 떨어지게 한다. 본 연구에서는 이런 잡음 성분들을 제거하기 위해 문턱치 th 를 이용하였고, th 는 다음과 같이 나타낼 수 있다.

$$th = \frac{1}{A} \sum_x \sum_y (I(x, y, t) - I(x, y, t - \delta kt)) - w \quad (3)$$

즉, 문턱치 th 는 차성분 영역에서 평균을 취하고,

가중치 w 를 더하였다. 여기서 A 는 $I(x, y, t) - I(x, y, t - \delta kt)$ 가 0이 아닌 픽셀의 개수이고, 가중치 w 는 실험적인 결과로 +10을 사용하였다. 또한 더욱 정확한 경계 영역을 추출하기 위해 수리형태학적 닫힘 연산을 수행하였다.^[2] 수리형태학적 닫힘 연산은 다음과 같이 나타낼 수 있다.

$$I_m = I_d \cdot K = (I_d \oplus K) \ominus K \quad (4)$$

여기서 \oplus 와 \ominus 는 각각 팽창, 침식 연산을 의미하며, \cdot 는 닫힘 연산을 의미하는데, 다음과 같이 구현된다.

$$I_d \oplus K = \{C \in E^N | C = a + b \text{ for some } a \in I_d \text{ and } b \in K\} \quad (5)$$

$$I_d \ominus K = \{x \in E^N | a + b \in I_d \text{ for every } b \in K\} \quad (6)$$

여기서 E^N 은 a 와 b 로 이루어진 공간이며, N 은 2로서 2차원 영상 평면을 나타낸다. K 는 구조화 요소(Structuring element)로서 3×3 크기의 흑화소 마스크를 이용하였다. 위의 수리형태학적 닫힘 연산을 사용하여 미세한 분리된 영역들을 연결하여 물체의 경계를 더욱 자세히 나타낼 수 있다.

정확한 사람 높이와 소실점과 소실선의 측정을 위해 본 연구에서는 Canny 연산자를 사용하였다.^[3] Canny 연산자는 다음과 같은 세가지 조건을 만족해야 하기 때문에 다른 에지 검출기 보다 더욱 상세한 외곽 영역을 검출할 수 있다.

- (1) 좋은 검출
- (2) 좋은 국부화
- (3) 하나의 에지에 대응하는 결과

Canny 연산자에서 경계 검출 연산자의 표준편차 크기(σ)의 결정은 주로 영상내의 추출하고자 하는 물체 형태내에 존재하는 경계강도나 복잡도에 의존한다. 작은 σ 값을 적용하는 경우 상세한 경계들이 추출되며, σ 값이 커질수록 물체의 외곽 성분에 충실한 경계점들만을 추출하게 된다.^[4] 본 연구에서는 내부에 존재하는 경계보다는 외곽경계가 중요한 정보이므로 외곽경계 이외의 미세한 경계 성분들을 억제할 수 있도록 σ 를 1.2로 설정하여 경계영상 I_e 를 추출하였다.

다음으로, 구해진 I_d 와 I_e 를 식(7)과 같이 논리곱 연산을 수행하게 되면 움직임 영역을 추출할 수 있다.

$$I_{me}(x, y) = I_m \odot I_e \quad (7)$$

여기서 \odot 는 논리곱 연산을 나타낸다.

이렇게 구해진 I_{me} 에 최외곽 점들을 포함하는 영역을 설정하게 되면 움직임 영역을 표시할 수 있게 된다.

2. 움직임 영역에서 키 높이 좌표 점 추출.

추출된 움직임 영역에서 사람 중심의 키 높이의 좌표점을 추출하기 위해 I_{me} 에 대하여 수직 방향의 히스

토그램 투영을 실시한다. 수직 방향의 히스토그램 투영에서 히스토그램 최대값을 나타내는 위치에서 영역의 최상위 좌표점과 최하위 좌표점이 사람의 키 높이의 좌표점으로 결정한다.

III. 추출된 사람 영역의 높이 측정

본 논문에서 이용될 카메라 모델은 central projection 이다. 우리는 기준평면에 대한 소실선과 수직성분의 소실점, 그리고 영상에서 기준 높이가 주어지면 지면으로부터 어떠한 점의 높이도 계산할 수 있다. 알고자 하는 사람의 높이는 차영상을 통해 영역을 추출하여 계산하였다

이 장에서는 기준 평면에 대해서 점들의 거리를 계산하는 알고리즘에 대해 설명한다.

1. 소실점과 소실선(Vanishing points lines)

소실점과 소실선은 매우 강력한 기하학적 단서이다. 이것들은 선들의 방향성과 평면들의 방위에 대한 많은 정보를 전달한다. 이러한 소실선과 소실점은 카메라와 장면 시점과의 명백한 기하학적 지식없이 이미지로부터 직접적으로 계산할 수 있다. 때로 이러한 소실점과 소실선은 이미지 밖에 놓일 수 있으나 계산에는 아무런 영향을 미치지 않는다.^{[5][6][7]}

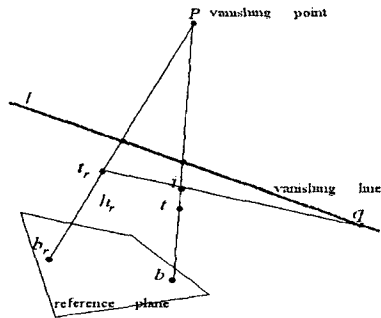


그림 2. 사영 기하학적 해석.

2. 목적물의 직선 에지들(Object's straight edges).

소실점과 소실선을 계산하기 위한 첫 번째 과정은 정확한 목적물의 직선 에지를 찾는 것이다. 직선 선분은 Canny edge detector를 이용해서 이미지에서 찾을 수 있다. 그러나, 실세계에서 물리적 에지들이 사영된 선들은 occlusions에 의해서 깨어져 보일 수 있다. 선택된 에지들을 병합하기 위해 수직적 회귀(orthogonal regression)에 기반한 단순 병합 알고리즘(A simple merging algorithm)을 통해 해결할 수 있다.

3. 소실점의 계산(Computing vanishing points)

실세계에서의 평행한 선들은 각각 소실점에서 교차한다. 그러므로, 소실점은 적어도 그러한 2개 이상의 선들로 정의되어진다. 그러나 2개 이상의 선들을 이용할 수 있는 경우, 소실점을 찾기위해 Maximum Likelihood Estimate(MLE) 알고리즘을 적용할 수 있다. 따라서 소실점을 계산하기 위해서 실세계에서 서로 평행한 선들의 벡터 곱을 통해서 계산할 수 있다. 사영공간에서 두 직선

$$l_1 = [a_1, b_1, c_1]^T \quad (8)$$

$$l_2 = [a_2, b_2, c_2]^T \quad (9)$$

의 교점은 $l_1 \wedge l_2 = l_1 \times l_2$ 이다.^[8]

4. 소실선의 계산(Computing vanishing lines)

평행한 이미지들의 선들은 서로 각각 그리고 평면에 대해서 소실선에 위치한 점들에서 교차한다. 그러므로 그러한 선들의 2개의 집합은 평면 소실선을 정의하는데 충분하다. 사영공간에서 두 점을 지나는 직선은 벡터곱

$$n_1 \vee n_2 = n_1 \times n_2 = \begin{vmatrix} 0 & z_1 & -y_1 \\ -z_1 & 0 & x_1 \\ y_1 & -x_1 & 0 \end{vmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \quad (10)$$

로 표시된다.^[8] 그러므로 평면에 대한 소실선을 찾기 위해, 평면에 대한 두 소실점을 벡터 곱으로 나타낼 수 있다.

계산된 소실점과 소실선을 이용해서 영상에서 사람의 실제 키에 해당되는 점 i를 계산한다.

$$i = (p \times b) \times (t_r \times (l \times (b_r \times b))) \quad (11)$$

이로부터 동일선상의 네 점 p, i, t, b를 구성할 수 있다. 그리고 p점은 수직 방향으로 기준 평면, 즉 지면으로부터 무한 거리의 위치에 놓이게 된다는 것을 알게 된다. 그리고, b와 t로 표시되는 알고자 하는 사람의 높이는 앞에서 추출한 사람의 영역으로부터 알 수 있다. 이를 통해 우리는 b와 t 사이 즉, 추출된 사람의 높이 h를 구할 수 있다. 그것은 다음의 2x2 line-to-line homography에 의해서 계산되어진다.^[5]

$$\begin{pmatrix} x_w \\ 1 \end{pmatrix} = H \begin{pmatrix} x_i \\ 1 \end{pmatrix} \quad (12)$$

여기서, x_i 는 이미지 위치, x_w 는 실세계 위치를 나타내며 H는 2x2 homography matrix로써 다음과 같이 나타낼 수 있다.

$$H = \begin{pmatrix} h_r(d(p, b) - d(i, b)) & 0 \\ -d(i, b) & d(p, b)d(i, b) \end{pmatrix} \quad (13)$$

여기서, $d(a, b)$ 는 a와 b 사이의 거리를 나타낸다.

이로부터 추출된 영역의 높이는 2-벡터 S의 구성 요소인 s_1 그리고 s_2 에 의해서 높이를 계산할 수 있

다. [5][8] 이때, $s = H \begin{pmatrix} d(t, b) \\ 1 \end{pmatrix}$, $h = \frac{s_1}{s_2}$ 이다.

단계인 평행한 선들을 찾는 과정과 카메라 왜곡에서도 생길 수 있다.

IV. 실험 및 결과

아래의 그림 1은 시퀀스 영상에서 움직임의 영역과 사람 키의 높이 좌표점을 추출한 영상이다. 머리와 발의 좌표점은 수직 방향의 투영 히스토그램을 통해 최대값을 가지는 라인에 위치한 최대y값과 최소y값의 좌표로 추출된다.

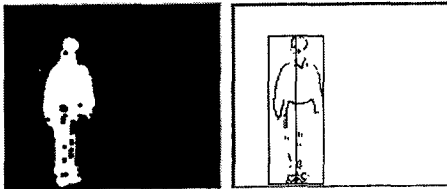


그림 3. 움직임 영역 추출과 높이 점 추출.

다음의 그림4-5는 평행한 라인들을 이용하여 소실점과 소실선을 구하고, 위의 움직임 영역에서 추출한 t, b의 좌표값을 가지고 사람의 키를 계산하는 과정을 나타낸다.



그림4. 소실점과 소실선을 구하기 위한 평행한 선.

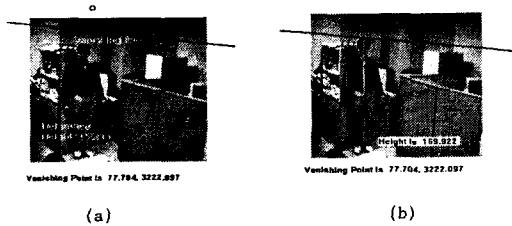


그림 5. (a) 평행한 선들을 이용하여 소실점(좌표값) 과 소실선(파란선). 초록선은 높이를 구하기 위한 기준 높이
(b) 소실점, 소실선을 이용한 사람 키의 측정.

그림 5의 (b)에서는 측정된 사람의 키를 보여 주고 있다. 여기서 측정된 키는 169.922cm 이지만, 실제 키는 170cm이다. 이와 같은 오차는 조명이나 그림자의 영향으로 정확한 머리, 발의 좌표점을 추출하지 못한 것에 기인한다. 또한 정확한 소실점을 구하기 위한 전

V. 결론 및 향후 과제

본 논문에서는 입력되는 영상으로부터 사람의 영역을 추출하고 기하학적 정보를 이용해 추출된 사람의 키를 측정하였다. 본 논문에서는 단지 단일 물체, 즉 한 명의 사람이 있을 경우에 대해서만 높이를 측정하였지만, 다중 물체가 존재하는 경우에 대해서도 측정이 가능하도록 보완할 예정이다. 그 밖의 향후 연구로는 사람을 정확히 추출하기 위해서 조명 효과에 강건한 알고리즘을 구현하는 것이 있겠다.

[참고문헌]

- [1] 이규원, 김영호, 이재구, 박규태 “무인감시장치 구현을 위한 단일 이동물체 추적 알고리즘”, 1995, 11. 한국전자공학회, 제 32권[1].
- [2] J. Serra, “Image Analysis and Mathematical Morphology”, Vol.1, Academic Press, New York, 1982.
- [3] John Canny, “A Computation Approach to Edge Detection”, IEEE Trans.on PAMI, Vol.8, No.6, pp.679-698, Nov.1986.[2]
- [4] Kashi Rao. “Extracting Salient Contours for Target Recognition : Algorithm and Performance Evaluation”, Optical Engineering, Vol.32, No.11, pp.2690-2697, Nov.1993[4]
- [5] A. Criminisi, A. Zisserman, L. Van Gool, S. Bramble, and D. Compton. “A new approach to obtain height measurements from video”, In proc. of SPIE, Boston, Massachusetts, USA, volume 3576, 1-6 November 1998.
- [6] D. Leibowitz, A Criminisi, and A. Zisserman. “Creating architectural models from images”, In Proc. EuroGraphics, Sep 1999.
- [7] O. Faugeras. “Three-Dimensional Computer vision : a Geometric Viewpoint”, MIT Press, 1993
- [8] R. Hartly, A. Zisserman “Multiple view geometry in computer vision” Oxford university press, 2000