

영상 감시를 위한 군중 수 측정

송수한, 가기환, 이광국, 김희율
 한양대학교 전자통신컴퓨터공학부
 e-mail : shsong@vision.hanyang.ac.kr

Crowd Size Estimation for Video Surveillance

Su-Han Song, Ki-Hwan Ka, Gwang-Gook Lee, Whoi-Yul Kim
 Dept. of Electrical and Computer Engineering, Hanyang University

요 약

본 논문에서는 영상 감시 등의 응용을 위해 영상에서 자동으로 사람 수를 측정하는 군중 수 측정 방법을 제안한다. 제안한 방법에서는 전처리 과정으로 전경영상과 경계영상을 검출하여 객체의 픽셀 크기 히스토그램과 경계 방향 히스토그램을 특징으로 이용하고 카메라 투영행렬을 통해 픽셀 크기와 경계 방향에 대한 특징 정규화를 수행한다. 실제 사람 수와 얻어진 특징 히스토그램 간의 선형성은 사상 함수의 구성에 적용되며, 훈련 데이터를 통해 얻어진 사상 함수는 사람 수 측정에 이용되었다. 제안한 방법의 성능은 건물 내에서 촬영된 영상에 대한 실험 결과로 나타났으며 이 방법이 영상 감시 분야에 다양하게 적용될 수 있음이 확인되었다.

1. 서론

기차역 또는 광장과 같은 공공장소 내에서 군중의 수를 측정하는 영상 처리 기술은 위기 상황 감지와 같은 영상 감시에 이용될 수 있으므로 그 중요성이 날로 높아지고 있다. 이러한 공공장소 내 군중 수 측정을 위해 많은 방법들이 제안되었으며, 대표적인 방법으로는 픽셀 또는 텍스처와 같은 특징에 기반한 방법[1][2]과 인식에 기반한 방법[3][4]이 있다.

특징에 기반한 방법은 픽셀, 경계, 텍스처 등 영상의 특징에 기반하여 영상 내 군중의 숫자를 추정한다. 이러한 방법은 영상 내 많은 사람이 존재하여 가려짐이 많이 발생하여도 잘 대응할 수 있는 장점이 있다. 인식에 기반한 방법은 인식기에 의해 개개인의 사람을 인식하며, 정확한 숫자를 알 수 있고 추적을 통해 움직임까지 쉽게 알 수 있는 장점이 있으나, 많은 가려짐이 발생하는 영상에 대해서는 대응하기 어려운 문제가 있고 또 인식기의 성능에 시스템의 성능이 크게 좌우되는 문제가 있다.

한편, 일반적으로 투영된 영상에서는 카메라의 시점에 따라 동일한 형태의 사람이 서로 다르게 표현될 수 있는 문제가 있다[6]. 따라서 영상에서 나타나는 모든 사람에 대해 동일한 기준으로 둘 수 있는 정규화 방법이 필요하다.

이에 본 논문에서는 영상 내에서 추출된 특징에 기반한 방법을 카메라의 시점이나 위치에 독립적으로 대응하도록 정규화 시키는 방법을 제안하여 좀 더 정확한 군중 수 추정이 가능하도록 하였다. 제안한 방법은 그림 1과 같이 요약된다. 입력된 영상에 대해 전경 분리와 경계 검출과 같은 전처리가 수행되면, 객체의 픽셀 크기나 경계 방향을 얻을 수 있고 이 특징들은 카메라의 기하에 따라 정규화된다. 훈련 데이터를 통

해 구성된 사상 함수는 사람 수 추정에 적용되어 정확한 사람 수를 측정하도록 한다.

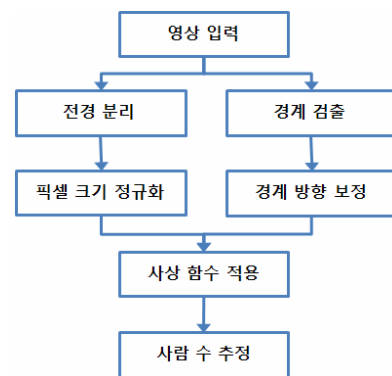


그림 1 제안한 방법의 흐름도

2. 특징 추출

영상에서 특징을 추출하기 위하여 전처리 과정으로 전경 영상(foreground image)을 획득해야 한다. 전경 영상 획득을 위해 우선 영상 내 모든 픽셀의 각 컬러 채널에 대해 배경 모델이 생성되었으며, 이 배경모델과 입력 영상의 차이가 전경 영상으로 이용되었다[5]. 움직이는 객체가 분할된 이후에는 군중 수 판단을 위한 특징 추출 과정이 이어진다. 군중 수 판단을 위한 특징으로는 전경 영역의 객체 크기와 각 경계 픽셀에서의 구배(gradient) 방향이 선택되었다. 그림 2는 추출 특징을 위한 영상의 예를 나타낸다. 그림 2(a)는 원영상에 관심 영역(ROI)의 마스크가 적용된 예이며, (b)는 분리된 전경 영역의 예, (c)는 전경 영역에 대한 경계 영상을 나타낸다. 경계 영상의 획득을 위해서 Canny 경계 검출자(edge detector)가 이용되었다.



(b) (c)

그림 2 특징 추출을 위한 영상의 예; (a) 원 영상의 관심영역, (b) 분리된 전경 영상, (c) 전경에 대한 경계 영상

이렇게 얻어진 전경 영상이나 경계 영상의 전체 픽셀 수는 군중 수 측정에 직접 이용될 수 있다[1]. 그러나 본 논문에서는 얻어진 특징을 히스토그램으로 구성하여 군중 수를 측정하였다[6]. 히스토그램을 기반으로 특징을 측정하게 되면, 객체 크기의 경우 잡음에 해당하는 전경을 제거할 수 있으며 그룹 형태로 몰려있는 사람들의 수 또한 측정할 수 있고, 경계 방향의 경우 수직 방향에 대한 픽셀 수에 따라 사람 수를 추정할 수 있다는 장점이 있다.

객체의 크기에 대한 히스토그램은 다음 수식에 나타난 것과 같이 정의 된다.

$$H_b(i) = \begin{cases} \sum_k S(k) & \text{if } B(i) \leq S(k) < B(i+1) \\ 0 & \text{elsewhere} \end{cases} \quad (1)$$

위 수식에서 B(i)는 객체의 크기를 나누는 기준이 되고, S(k)는 k 번째 객체에 해당하는 픽셀 수를 의미한다. S(k)가 i 번째 구간에 속할 경우 H_b(i)에 해당하는 픽셀 수만큼을 증가시킨다.

위의 수식에서 B(i)는 히스토그램의 i 번째 bin(bin)의 구간을 정의하는 객체의 크기를 나타내며, S(k)는 전경 영역 내의 객체 k 에 대한 픽셀 수를 나타낸다. 따라서, 히스토그램의 i 번째 bin의 값은 해당 bin의 구간에 포함되는 객체의 크기가 누적되어 계산된다. 본 연구에서는 픽셀 크기 히스토그램의 bin의 개수는 경험적으로 6 개를 이용하였으며, 최적의 bin 구간을 얻기 위하여 Lloyd-Max 양자화를 수행하였다.

경계의 방향 추출은 전경 영역 내의 각 경계 픽셀 위치에 대해 구배(Gradient) 연산자를 적용함으로써 얻어졌다. 얻어진 경계 방향에 대한 히스토그램은 0 에서 180 도에 이르는 8 방향으로 양자화하여 계산했다.

그림 3은 위에서 제시한 방법으로 그림 2에 대한

히스토그램을 구한 결과의 예이다.

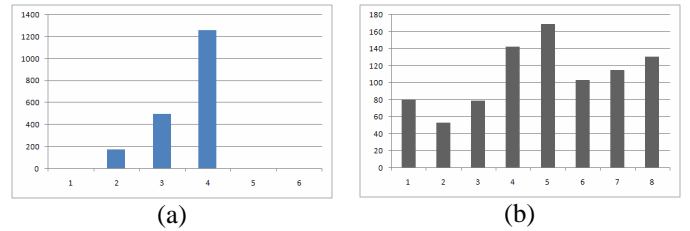


그림 3 추출된 특징에 대한 히스토그램; (a) 객체의 픽셀 크기에 대한 히스토그램, (b) 경계 방향에 대한 히스토그램

그림 3(a)는 그림 2에서 검출된 객체들의 픽셀 크기에 대한 히스토그램을 나타내며 각 객체들이 군집한 형태에 따라 다른 크기의 결과를 보인다. 그리고 그림 3(b)는 경계 방향에 대한 픽셀 수의 히스토그램으로 수직 방향에 해당하는 두 개의 bin이 지배적인 결과를 보인다.

3. 특징 정규화

2장에서 얻어진 이동 객체 크기와 경계의 방향에 대한 히스토그램 자체로 군중수의 측정을 직접 할 수도 있으나, 이럴 경우 카메라의 시점에 따라 변화할 수 있는 사람의 크기나 경계의 방향에 대응할 수 없다는 단점을 가질 수 있다. 그림 4는 언급한 상황에 대한 예로, 영상 내에서 유사한 형태의 객체가 각각 다른 크기와 방향으로 표현되어 있다. 이러한 특징의 변화에 독립적인 사람 수 추정을 위해서는 얻어진 특징들에 대한 픽셀 크기 정규화 및 경계 방향의 보정을 통한 정규화가 수행되어야 한다.

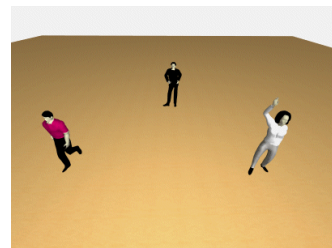


그림 4 객체의 특징이 위치에 따라 다르게 나타난 예

3.1 픽셀 크기 정규화

이에 본 논문에서는 픽셀 크기의 정규화를 위해 카메라 기하를 고려한 카메라 투영행렬(projection matrix)을 이용하여 영상의 각 픽셀에 대한 상대적 밀도를 추정하였다.

각 픽셀 위치에서 평균적인 사람 크기의 계산하기 위하여 사람을 3 차원 공간 상에서의 타원체로 모델링하였으며[4], 이 타원체가 2 차원 영상의 각 픽셀의 위치로 투영되었을 때의 면적을 계산하였다. 사람에게 해당하는 타원체는 장축의 길이가 170 cm, 단축의 길이가 장축의 0.3 배로 이루어져 있다고 가정하였다.

3×4 의 카메라 투영 행렬을 P 라 하고 4×4 의 타원체 행렬을 Q 라 할 때, 위의 타원체가 2 차원 영상

내로 사상(projection)된 타원 C 는 $C^{-1}=PQ^{-1}P^T$ 의 행렬에 의해 나타낼 수 있다. 각 픽셀의 위치에서 이러한 타원의 행렬이 계산되었으며, 그 타원의 면적이 해당 픽셀 위치에서 평균적인 사람의 크기로 이용되었다.

이렇게 얻어진 평균 사람의 크기를 이용하여 특징 정규화를 위한 상대적 밀도 맵이 계산되었다. 밀도 맵을 구하기 위하여 임의의 기준점을 설정하고 해당 위치에서의 사람을 모델링한 타원의 면적을 구하여 이를 A_r 로 정의하였다. 그리고, 관심 영역 내의 모든 픽셀들에 대해 사람을 모델링한 타원의 면적 $A(u, v)$ 를 계산하며, 이에 대한 기준점에서의 타원의 크기와의 비율 $A(u, v)/A_r$ 을 각 픽셀에 대한 상대적 밀도로 정의하였다. 이와 같이 관심 영역내의 모든 픽셀 위치에서 계산된 타원 크기의 비율이 밀도 맵이 되며, 이는 그림 5와 같이 표현될 수 있다. 그림에서 (a)는 원본 영상을 나타내며, (b)는 계산된 밀도 맵을 나타낸다.

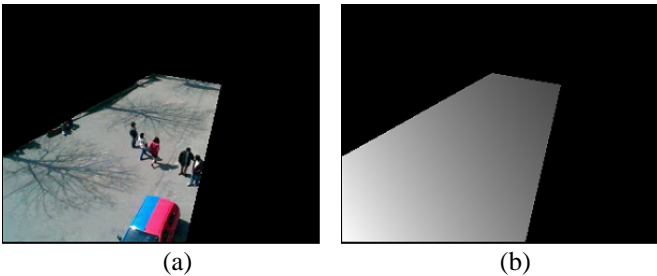


그림 5 관심영역에 대한 상대적 밀도 맵 적용의 예; (a) 원영상, (b) 상대적 밀도 맵

그림 5(b)에서 우측 상단의 한 픽셀은 좌측 하단의 한 픽셀보다 실제 공간에서 더 큰 공간을 차지하고 있는 것을 알 수 있으며, 각 위치에서 얻어진 픽셀에 대한 가중치를 곱해 줌으로써 모든 픽셀이 동일한 중요도를 가지도록 하였다.

3.2 경계 방향 보정

경계 방향의 보정 과정은 그림 6을 통해 설명될 수 있다. 그림 6(a)는 원 영상에서 추출된 각 이동 객체의 경계선 영상을 나타내며, 각 이동 객체의 중심을 지나는 실선은 해당 객체 위치에서의 수직 방향을 나타낸다. 이처럼 객체들의 수직 방향은 영상에서 카메라의 투영(projection)에 의해 모두 다른 방향을 가짐을 알 수 있다. 따라서, 수직 방향의 경계 픽셀이 방향 히스토그램 내에서 동일한 빈에 위치할 수 있도록 하기 위해서는 구배의 방향에 대한 보정이 수행되어야 한다. 그림 6(b)는 추출된 이동 객체를 해당 객체의 위치에서 수직 방향으로 회전시킨 예이다. 객체를 회전한 이후에 경계 방향 히스토그램을 계산함으로써 카메라 투영에 의한 방향 변화의 문제를 보정할 수 있다.

투영행렬식을 통해 소실점을 계산하여 그림 6(a)와 같이 각 객체의 중심을 지나는 실선으로 연결하면 각 객체에서의 수직방향을 알 수 있다. 이 실선과 영상에서 수직인 선과의 각도 θ 는 각 객체의 기울어짐을 나타내며, 각 객체에서 구해진 구배 방향을 보정하는

데에 이용된다. 보정이 이루어지면 그림 6(b)와 같이 객체의 위치에 독립적인 경계의 히스토그램을 구성할 수 있다.

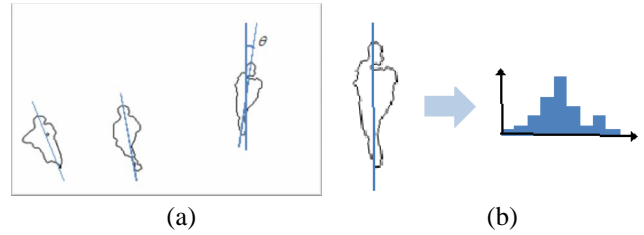


그림 6 경계 방향의 보정; (a) 객체 추출의 예, (b) 객체의 수직방향으로의 방향 보정한 예

3.3 정규화 적용

상기 과정이 모두 진행되고 나면, 얻어진 특징에 대한 정규화를 적용시킬 수 있다. 특징 정규화는 앞에서 설명한 상대적 밀도 맵과 경계 방향에 대한 보정의 결과로 구한다.

각 프레임에서 추출한 특징인 객체 크기와 경계의 방향을 나타내는 픽셀을 각각 $blob(u, v)$ 과 $or(u, v)$ 라 하고 상대적 밀도 맵을 $d(u, v)$ 라 할 때, 각 특징의 정규화는 다음의 수식들로 나타낸다.

$$blob'(u, v) = blob(u, v) / d(u, v) \quad (3)$$

$$or'(u, v) = or(u, v) / \sqrt{d(u, v)} \quad (4)$$

특징 정규화의 결과로 얻은 경계 방향 $or'(u, v)$ 는 경계 픽셀이 존재하는 부분에 대한 중요도를 반영한 결과로 경계 방향의 픽셀 크기는 상대적 밀도에 따라 선형적으로 증가시켰다. 그리고 객체의 크기에 대한 정규화 결과 $blob'(u, v)$ 는 각 픽셀이 중요도에 따라 면적 단위로 증가한다고 가정하여 정규화를 수행하였다.

4. 사상 함수

이렇게 얻어진 전경 영역의 이동객체 크기 히스토그램과 경계 방향에 대한 히스토그램은 공간 내 사람 수 추정을 위한 특징 벡터로 이용된다. 특징벡터와 공간 내 사람 숫자 사이의 선형적인 상관관계를 이용하면 사상함수를 구성할 수 있으며, 이는 다음의 수식을 통해 알 수 있다.

$$C = \sum_{i=1}^m H_b(i) \times W_b(i) + \sum_{j=1}^n H_e(j) \times W_e(j) \quad (5)$$

여기에서 C 는 추정된 군중의 수, H_b 와 H_e 는 각각 전경 영역의 객체 크기 히스토그램을 나타내며, W_b 와 W_e 는 각각의 히스토그램 빈에 대한 사상함수의 계수이다. 즉, 군중의 수는 위의 수식에 의해 각 히스토그램 빈의 값의 일차 함수로 근사화 되어 추정할 수 있게 된다.

이 사상함수의 계수 W_b 와 W_e 는 훈련과정을 통해

미리 계산 된다. 즉 N 장의 훈련 영상이 있을 때, 사람에게 의하여 훈련 영상 내의 군중 수가 측정이 된다. 또한 각각의 영상으로부터 히스토그램을 추출하여 다음과 같은 행렬식을 구성한다.

$$\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_N \end{bmatrix} = \begin{bmatrix} H_1(1) & H_1(2) & \dots & H_1(m+n) \\ H_2(1) & H_2(2) & \dots & H_2(m+n) \\ \vdots & \vdots & \ddots & \vdots \\ H_N(1) & H_N(2) & \dots & H_N(m+n) \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{m+n} \end{bmatrix} \quad (6)$$

위의 수식에서 N 은 훈련에 사용된 영상의 개수를 나타내고 m 은 객체의 크기 구간을 나눈 수를 나타내며 n 은 경계의 방향에 대한 구간을 나눈 수에 해당한다. H 는 히스토그램에 대한 행렬을 나타내고 S 는 각 훈련 영상에서 측정한 실제 사람의 수이며, W 는 훈련 과정을 통해 구하고자 하는 가중치 벡터가 된다. 사상 함수는 위의 수식과 같은 행렬식에 의해 얻어지고, 이는 각 영상 내에서 추출한 특징들과의 연산을 통해 선형적인 군중 수 추정을 하는 데 사용된다.

5. 실험 결과

본 논문에 대한 실험은 실내의 천장 높이에서 촬영된 영상에 대해 수행되었다. 군중 수 측정의 정확도 측정은 영상 내에 존재하는 실제 사람 숫자와 시스템이 추정한 사람 숫자의 비교를 통하여 이루어졌으며, 오차의 계산은 다음의 수식에 의해 이루어졌다. 수식에서 C_{GT}(i)는 i 번째 영상 내의 사람의 숫자, C(i)는 i 번째 영사에 대하여 시스템에 의해 추정된 사람의 숫자를 나타내며, 그 차이의 평균이 전체 평균 오차로 측정되었다.

$$error = \frac{\sum |C_{GT}(i) - C(i)|}{N} \quad (7)$$

정확도 측정 결과 평균 0.26 명의 오차가 발생하였다. 그림 7은 군중 수 추정의 결과를 시간에 따른 그래프로 나타낸 것이다. 그래프에서 흑색 막대는 실제 사람 수를 나타내며, 회색 막대는 제안한 방법에 의해 추정된 사람의 수를 나타낸다.

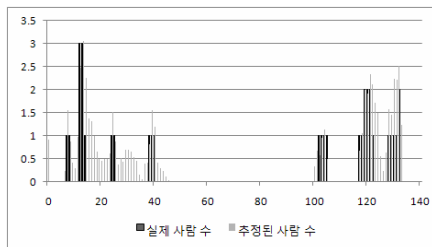


그림 7 제안한 방법의 실험 결과

그림 8은 잘못된 결과가 발생한 영상의 예를 나타낸다. 두 영상 모두에서 출입문 쪽의 밝기 변화에 의하여 전경 영역이 잘못 검출되었기 때문에 잘못된 군중 수 추정이 발생하였다. 이 같은 오류가 발생한 이

유는 전경 분리 시 영상에서 발생할 수 있는 그림자에 대한 제거가 충분히 이루어지지 못했기 때문이었다.



그림 8 실험에서 오차를 발생시킨 예

6. 결론

본 논문에서는 공공장소에서 영상의 특징에 기반한 군중 수 추정 방법을 제시하였고, 제안한 방법의 성능을 측정하기 위해 천장 높이에서 촬영된 영상을 실험에 이용하였다. 제안한 방법은 특징 정규화 방법을 통해 관찰 공간 내의 군중 수를 추정하고 영상 내로 투영되는 각 객체들의 수직 방향을 보정함으로써 공공 장소에 설치될 카메라의 시점과 위치에 독립적인 군중 수 추정이 가능하도록 하였다. 이렇게 정규화된 특징을 활용한 군중 수 추정 방법은 안정적인 결과를 산출하며, 영상 감시 분야에 다양하게 응용될 수 있을 것이다.

참고문헌

- [1] S. A. Velastin, J. H. Yin, A. C. Davies, M. A. Vicencio-Silva, R.E.Allsop, A. Penn, "Automated Measurement of Crowd Density and Motion using Image Processing," Seventh International Conference on Road Traffic Monitoring and Control, p. 127-32, 1994
- [2] R. Ma, L. Li, W. Huang and Q. Tian, "On Pixel Count Based Crowd Density Estimation for Visual Surveillance," IEEE Conference Cybernetics and Intelligent Systems, volume 1, pp. 170 - 173, 2004
- [3] X. Huang, L. Li, T. Sim, "Stereo-based Human Head Detection from Crowd Scene," International Conference on Image Processing, p. 1353-1356, 2004
- [4] T. Zhao, R. Nevatia, and F. Lv, "Segmentation and tracking of multiple humans in complex situations," Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 194-201, 2001.
- [5] S. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-Time Surveillance of People and Their Activities," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. 2000.
- [6] D. Kong, D. Gray, Hai Tao, "A Viewpoint Invariant Approach for Crowd Counting," in Proc. International Conference on Pattern Recognition, vol. 3, pp. 1187-1190, 2006