

논문 2002-11-1-04

비디오에 의한 감시 및 관측(VSAM) 시스템을 위한 사람의 계수기법

도용태*

A People Counting Technique for Video Surveillance and Monitoring(VSAM) Systems

Yongtae Do*

요 약

비디오 센서에 의한 감시 및 관측(VSAM)에 있어 사람은 중요하면서도 해석하기 어려운 목표이다. 본 논문에서는 연속 영상에서 자동으로 사람을 추적하거나 행위를 해석하는 작업에 선행하는 절차로서, 그 수를 계측하는 방법에 대해 다룬다. 정적인 비디오 카메라의 영상에서 사람만을 추출해낸 이진 영상을 얻고, 최상위 화소만을 연결한 선에서 지역최소점을 근거로 사람군을 나눈다. 이와 같이 나누어진 각 영역의 특성은 군내에서의 사람들의 상대적 위치에 따라 변화하게 되며, 이를 구분하기 위하여 완전 가림, 부분적 가림, 완전 분리된 개인, 오분리의 4가지 상태를 정의하였다. 연속적인 영상에서 나누어진 영역들의 상태가 변화하여가는 과정을 분석하여 관측 지역내 사람의 수를 추정하였으며, 이의 유용성을 실제 실험을 통하여 확인하였다.

Abstract

People are important targets for video surveillance and monitoring(VSAM) but difficult to be analyzed. In this paper, a technique to count people in image sequences is dealt as a prerequisite procedure for automatic tracking and behaviour analysis. A group of people is divided at local minima of the line connecting the highest pixels on the binary image of the people extracted from the image taken by a stationary video camera. As the properties of the divided regions vary according to the relative positions of the people in a group, different states are assigned for the completely occluded, partially occluded, completely separated individual, and wrongly divided regions. By analyzing the transition of the states of divided regions, the number of people on the site monitored is estimated. The technique is checked in real experimental situations.

1. 서 론

비디오에 의한 감시 및 관측(video surveillance and monitoring: VSAM)은 군사, 보안 및 교통 제어 등의 목적으로 근래에 와서 그 중요성이 부각되고 있으며 연구도 활발하다⁽¹⁾. 사람은 VSAM에 있어 차량과 더불어 가장 중요한

목표이지만, 그 연구는 상대적으로 부족하며 결과 또한 만족스럽지 못한 형편이다. 특히 사람은 차량에 비해 훨씬 큰 자유도(degree of freedom)를 가지고 있어서 형태의 변화가 다양하며 그 이동 행태가 도로만을 따라가는 차량과는 달리 예측하기 힘들므로 이를 추적하고 해석하는 일은 상대적으로 어렵다. 이러한 관점에서 목표에 대한 풍부한 정보와 이의 시간에 대한 변화를 알 수 있게 해주는 비디오 영상은 기존의 정지 영상에 비해 확실한 장점을 가진다.

본 논문에서는 감시 및 관측을 위한 센서로 사용된 비디오 카메라의 연속 영상에서 사람의 숫

* 대구대학교 정보통신공학부 (School of Computer and Communication Engineering, Taegu University)

<접수일자 : 2000년 12월 4일>

자를 계속하는 방법을 다룬다. 이 문제는 후에 자세히 설명되는 바와 같이 영상에서 사람군 (group of people)을 각 개인의 영역으로 나누는 작업을 요구한다. 이같은 문제는 두 가지 관점에서 그 중요성을 찾아 볼 수 있다. 첫째로, 피관측 사람군이 몇 명으로 구성되어 있는지 파악하는 일은 다른 VSAM 작업을 위한 기초가 될 수 있다. 얼마나 많은 사람이 있는지를 알지 못하는 상황에서는 각 사람들이 제대로 추적될 수도 없고, 또 그들의 행위가 올바르게 분석될 수도 없다. 즉, 관측 지역내에 있는 사람군에서 각 자를 구분해내는 일은 그들을 해석하는 출발점이다. 둘째로, 이러한 문제를 무시한 기존의 사람 영상 분석기법들은 여러 가지 현실적 제약을 가지고 있다. 많은 경우 단수의 사람을 전제로 하고 있고, 다수가 해석될 경우에는 해상도가 높아 각 사람들의 주요한 몸 부분들이 분명히 확인될 수 있는 경우가 전제된다. 그러나 이러한 조건들은 하나의 카메라로 넓은 영역을 원거리에서 관측하는 경우에는 적용되기 곤란하다. 더욱이 배경으로부터 추출된 사람의 이진 영상은 자주 불완전하고, 특히 배경과 사람의 색깔이 부분적으로 유사할 때 이 문제는 두드러진다. 이를 극복하기 위해서 수리형태학적(morphological) 기법이 사용되기도 하였으나⁽²⁾, 계산량이 많아지며⁽³⁾ 문제를 해결하지 못하는 경우도 자주 발생한다.

기존의 비디오에 의한 사람 영상의 처리는 크게 인식, 추적, 행위해석, 그리고 형태분석으로 나눌 수 있다^(1,4). 우선 사람을 인식하는 작업을 위해 Lipton 등⁽⁵⁾은 사람이 차량에 비해 훨씬 복잡한 형태를 가진다는 점을 이용하여 목표물 영역의 면적과 둘레의 길이로부터 사람과 차량을 분리하는 척도를 제안한 바 있다. Pfinder⁽⁶⁾는 방안에 있는 한 사람의 손과 머리 그리고 몸통을 찾고 추적하며, 이로부터 행위를 해석하는 프로그램이다. 또 Ghost⁽⁷⁾는 사람의 외부윤곽 (silhouette)에 근거하여 다양한 자세 형태에서 몸의 각 부분을 인식하기 위해 작성된 프로그램이다. 하지만 본 논문에서 다루고 있는 바와 같이 비디오에서 사람을 계수하는 문제에 대한 연구는 드물며, 교통량의 분석을 위한 Heikkila 등의 연구결과⁽³⁾를 찾을 수 있는 정도였다. 그들은 Quasi-topological 코드의 사용을 제안하였지만, 해상도가 높아 사람의 형태가 뚜렷이 드러

날 때만 유효하다는 문제가 있으며, 비디오에 의해 제공되는 연속적인 영상들을 충분히 활용하는 방법도 제시하지 못하고 있다.

본 논문에서는 비디오 영상을 이루는 각 프레임에서 최소한의 절차에 의해 사람을 계수하는 방법과 연속 영상내에서 그 결과가 어떻게 변화하는지 살펴봄으로써 군내 사람들의 상대적 위치의 작용에 대해 이해하는 기법을 제안한다. 우선 다음 장에서 전체 시스템의 구성에 대해 대략적으로 소개한 후, III장에서 각 영상 프레임에서 사람군을 각 개인영역으로 나누는 기법을 기술하며, 이어 IV장에서 연속 영상에서 나누어진 영역들의 변화를 추적하고 이를 기반으로 계수하는 기법을 제시한다. V장에서는 제안된 기법을 실제 적용한 예를 보이고, 마지막으로 VI장에서 결론이 주어진다.

II. 영상의 처리구조

비디오 카메라로부터 얻어진 사람 영상의 처리는 배경으로부터 사람을 추출하는 작업에서 시작된다. 기본적으로 이는 정해진 관측 영역내에서 이동하는 물체를 찾고, 이가 사람인지 판단하는 절차로 구성될 수 있다. 영상내에서 이동물체를 찾기 위한 기존의 접근법은 다음의 세 가지로 구분될 수 있을 것이다. 우선 몇 프레임간의 차분 영상을 얻는 방법⁽⁸⁾이 있는데, 이는 동적인 환경에서 적응적으로 사용될 수 있으나 실제 물체의 영역을 추출해내는 것이 간단치 않다. 이에 비해 배경(background)영상을 기억한 후 이를 이용하는 방법^(6,7)은 이동 물체의 추출이 간단하고 직접적이지만, 환경이 심하게 동적일 때는 적용하기 곤란하다. 마지막으로 optical flow를 이용하는 방식⁽⁹⁾은 카메라가 움직이는 경우에도 사용할 수 있지만, 일반적으로 실시간 처리하기에는 계산량이 과다하다. 본 논문에서는 [1,2]에서 제안된 적응적 배경영상 기억법을 사용하여 이동물체를 추출한 이진영상을 얻은 후 이를 분석하였다. 즉, 비디오 영상내 각 화소값들의 시간에 대한 평균치와 표준편차를 연속 프레임에서 구하면서, 이의 변화가 지정한 기준치 이상이 되면 전경 (foreground) 영역에 포함시키고, 일단 일어난 변화가 일정 시간이상 지속되는 경우에는 다시 배경 영역을 갱신(update)하는 데 사용하는 방

식이다.

영상내에서 전경 영역이 검출되면, 우선 이 사람이 사람인지 아닌지를 판단하여야 한다. 이러한 목적으로 신경망을 사용하거나, 형태에 관한 척도를 사용하는 방법이 제안된 바 있다⁽¹⁾. 그러나 어느 방법도 완벽할 수는 없으므로, 여기서는 목적상 사람만 존재하는 영상을 직접 확인한 후 사용함으로써 이 절차에서 생기는 오류가 본 논문에서 다루는 계수의 과정에 영향을 미치지 않도록 하였다. 즉, 전경 영역이 사람이라는 것은 본 연구에서 기본 전제로 사용되었다.

관측되는 사람군에 대한 비디오의 연속 영상으로부터 자동으로 몇 명의 사람이 있는가를 추정하기 위해서 두 가지 접근법이 고려되었다. 첫 번째는 연속 영상중에서 임의의 기준에 의해 가장 신뢰할 만하다고 추정되는 한 장의 영상을 선택하고 이것에 의존하여 결과를 추정하는 것이고, 두 번째는 구하고자 하는 정보와 관련이 있는 어떤 특징이나 파라메타가 연속된 영상에서 어떻게 변화해 나가는가를 살피고 이로부터 최종적인 결과를 추정하는 방식이다.

전자의 방식을 사용할 경우에는 가장 좋은 프레임을 선택하기 위한 기준, 즉 각 프레임으로부터 얻어지는 정보의 질이나 양, 혹은 그로부터 얻어지는 결론에 대한 확실성의 척도(measure of certainty)가 먼저 적절히 정의되어야 한다. 즉, 관측된 사람군의 연속 영상들 중에서 이들을 대표하는 한 영상을 선택하기 위한 근거가 정의되어야 하는 것이다. 우리는 이를 위해 각 영상 프레임마다 전경영역을 영역에 속하는 각 개인별로 수직으로 나눈 후, 이 나눔을 계수할 때, 나눔이 얼마나 신뢰할 수 있는 것인지를 영역내 색의 균질도(color homogeneity)나 밝기 균질도(gray level homogeneity)를 척도로 정의하여 사용함으로써 판단하였다. 이는 한 사람의 색상이나 밝기는 균질할 것이라는 가정에 의한 것인데, 한 사람이 아래 위 다른 색상이나 밝기의 옷을 입고 있을 수 있으므로 횡적으로만 그 분산을 측정하였다. 아래는 밝기 균질도를 사용할 경우 나누어진 영역의 평균분산을 얻기 위한 식이며, 색 균질도를 사용할 경우에는 R,G,B 신호에 대해 동일한 식을 사용할 수 있다.

$$var = \frac{1}{N_d} \sum_{k=1}^{N_d} \left[\frac{1}{N_k} \sum_{i=k_{min}i}^{k_{max}i} \sum_{j=k_{min}j}^{k_{max}j} \{ \mu_{ki} - I(i,j) \}^2 \right] \quad (1)$$

단, N_d 는 영상 프레임내 나누어진 영역의 수,

N_k 는 k번째 영역내 화소의 수,

k_{mini} 와 k_{maxi} 는 각각 k번째 영역내 최소행과 최대행의 번호,

k_{minj} 와 k_{maxj} 는 각각 k번째 영역내 최소열과 최대열의 번호,

μ_{ki} 는 k번째 영역내 i행의 밝기값 평균,

$I(i,j)$ 는 전경영역내 (i,j)좌표 화소의 밝기값

한편 후자의 접근법을 사용하는 한 방식으로, 본 논문에서는 각 프레임에서 전자의 방식에서와 같이 전경 영역의 사람군 영상을 일단 나누고, 이를 계수하며 그 결과가 연속 영상속에서 어떻게 변화해 나가는지 살피고 이로부터 관측된 사람군에 포함된 사람의 수를 추정하고자 하였다. 예를 들어 각 영상 프레임에서 나누어진 영역의 갯수가 늘거나 혹은 줄어들면, 어떤 사람이 가려지거나 가려졌다가 다시 출현하는 상황이 일어났다고 이해하는 방법이다.

V장에서 볼 수 있듯이 상기의 두 접근법을 각각 적용하였을 때 후자의 방식이 더욱 좋은 결과를 얻을 수 있었다. 이는 전자의 접근법을 사용하는 경우에는 어떤 사람이 다른 사람을 거의 가리게 될 때 확실성의 척도가 오히려 높아지는 문제가 발견되었고, 연속 영상에서 시간의 변화에 따라 목표의 상대적 위치가 변화해 나가는 것을 이해하는 방법이 부재하는 것이 문제가 될 수 있다고 사료되었다. 반면 후자의 경우, 비교적 적용이 간단하고 그 결과도 신뢰성이 높았으며 비디오 센서의 이점을 충실히 활용할 수 있었다. 따라서 본 논문에서는 이 방법을 사용하는 절차와 적용의 결과를 중심으로 기술한다.

후자의 방법을 실현하기 위한 구체적인 사항들은 다음의 두 절에서 서술될 것이며, 여기서 그 절차의 전체적인 구조를 서술한다. 우선 사람군을 전경 영역으로 하는 이진 영상이 주어지면, 여기서 각 열(column)별로 최고점들을 이은 선(이후 'draping line'이라고 부르겠다)을 구한다. 이 선에서 지역 최소(local minima)를 구하고, 이 점들에서 균을 수직으로 나눈다. 다음에 나누어진 균들의 구조적 특성, 예를 들어 폭, 높이, 중심좌표 등을 구하게 되고, 이들 특성으로부터 각 나누어진 영역들의 상태(state)를 주로 상호간의 가려짐을 중심으로 추정하게 된다. 이러한 절차를 연속된 영상 프레임들에 대해 수

행하면서 그 결과가 어떻게 변화하는지 살펴본 후, 사람의 수를 추정하게 된다. 이러한 절차는 그림 1과 같이 요약될 수 있고, 여기서 각 블럭은 중요한 본 논문의 구성요소로서 이후 상술될 것이다.

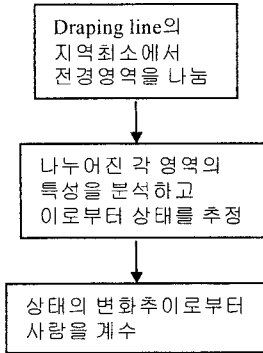


그림 1. 사람 계수의 절차.
Fig. 1. Steps of people counting.

III. 이진화된 전경영역의 Draping Line을 이용한 나눔

영상에서 사람을 계수한다는 것은 사람의 영역을 각 개인별로 나누는 것을 필요로 한다. 이러한 작업은 각 개인을 추적하고, 그 형태와 행위를 분석하고, 기록하기 위한 기본적인 작업이 된다. 이를 위해서는 가장 먼저 영상내에서 사람군을 전경영역으로 추출하여야 하며, 이는 결국 배경 영역과 전경 영역을 구분하는 이진화의 작업이다. 본 논문에서는 전절에서 서술된 바와 같이 [1,2]에서 제안된 적응적 배경 영상 기억법을 다음과 같이 사용하였다.

연속 영상중 n번째 프레임내 화소 $I_n(i, j)$ 에 대해서 시간에 대한 평균 밝기 $\overline{I_n(i, j)}$ 와 표준편차 $\overline{\sigma_n}$ 를 아래와 같이 갱신한다.

$$\overline{I_{n+1}(i, j)} = aI_{n+1}(i, j) + (1-a)\overline{I_n(i, j)} \quad (2.a)$$

$$\overline{\sigma_{n+1}(i, j)} = a|I_{n+1}(i, j) - \overline{I_{n+1}(i, j)}| + (1-a)\overline{\sigma_n(i, j)} \quad (2.b)$$

여기서 a 는 프레임 주파수를 고려하여 결정되는 파라미터이며, 만약 어떤 화소의 값이 그 평균으로부터 표준편차의 2배이상 차이가 나면 전경영

역으로 고려한다. 이처럼 얻어진 이진 영상은 전경 영역 내부에 많은 구멍(holes)이 존재하는 등 불완전하므로 2번의 불림(dilation)과 한번의 침식(erosion)을 수행하여 최종의 이진 영상을 얻었다.

일단 사람군의 영상이 배경으로부터 분리되어 지면 이를 각 개인별 영역으로 다시 나누는 작업을 수행한다. 본 논문에서는 이를 위해 전경 영역의 draping line을 얻고 이의 지역최소에서 수직으로 영역을 나누었는데 다음과 같은 가정과 관측을 사용하였다.

(i) 영상에서 사람들의 자세는 대체로 서있는 형태이다. 이 가정은 카메라가 적절히 설치되고 사람들이 보통의 자세로 걷거나 뛰는 상황에서 받아들일 수 있다. 이 가정을 사용하면 사람군의 영역을 나누는 작업이 수직으로만 제한되므로 문제가 간단해진다.

(ii) 사람의 상측부분(머리와 몸통)은 하측부분(다리)에 비해 연속 영상에서 변화가 적다. 이 관측은 사람 영역의 상측부에 기반한 처리가 보다 신뢰성 높은 결과를 얻게 함을 의미한다.

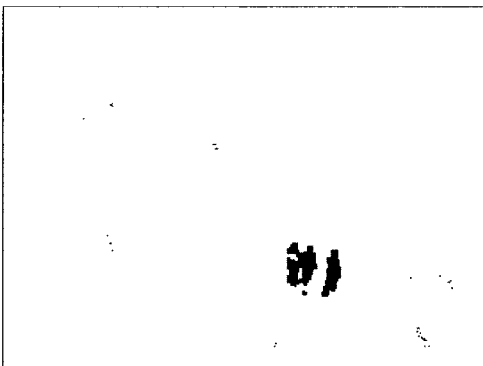
(iii) 인근에 있는 두 사람 사이에는 draping line 상에 하나의 지역최소점이 존재한다. 이러한 가정은 둘 중 한 명이 다른 사람을 거의 완전하게 가리는 경우나 사람이 손을 올리는 경우와 같이 특수한 경우를 제외하고는 받아들일 수 있다. 필요하다면 draping line 상의 잡음에 의해 지역 최소가 나타나는 경우에 대비해서 선상의 이웃 점들과 평활화(smoothing)를 행할 수 있다.

그림 2는 상기와 같은 방식으로 영역을 나누는 예를 보여준다. 그림(a)는 입력된 한 영상 프레임이고, 이로부터 전경 영역을 추출한 것이 그림(b)이다. 그림에서 사람군 부분만을 얻은 것이 그림(c)이고, 이의 draping line은 그림(d)이다. Draping line의 지역최소에서 영역을 수직으로 나누게 되는 데 이의 결과가 그림(e)에 주어졌다. 일단 3개로 나누어진 영역은 다음 단계에서의 해석을 위해 각 나누어진 영역을 둘러싸는 사각형(surrounding rectangles)에 의해 표시된다. 사각형의 좌측과 우측 경계는 수직 나눔의 결과를 이용하고 상측 경계는 수직으로 나뉜 영역내 최상점(즉, 전경 영역에 속하며 수직으로 나누어진 영역내의 화소 $I(i, j)$ 들 중 가장 작은 i 를 가지는 화소)에 의해 결정되지만, 하

측 경계는 영역내 최하점보다는 영역에 속하는 열(column)들의 하점(즉, 영역내 각 j에 대해 가장 높은 i를 가지는 화소)들 중 중간값(median)에 의해 결정되었으며, 이는 그림(f)에서 확인할 수 있다. 이처럼 영역내에서의 최하점으로 사각형의 하측 경계를 결정짓지 않는 이유는 상기의 사항(ii)에서 관측된 대로 현실적으로 나타나는 하측부의 프레임간 불안정의 문제를 극복하기 위한 것이다. 실험의 결과 경계선을 분석하는 방법^(3,4)이나 전경 영역의 이진영상에서 수직의 투영(projection)⁽¹⁰⁾을 사용하는 방법에 비해 본 논문에서 사용된 기법은 전경영역의 불완전한 추출이나 사람이 이동할 때 상측부를 약간 기울이는 각도 등에 덜 민감하였다.



(a)



(b)

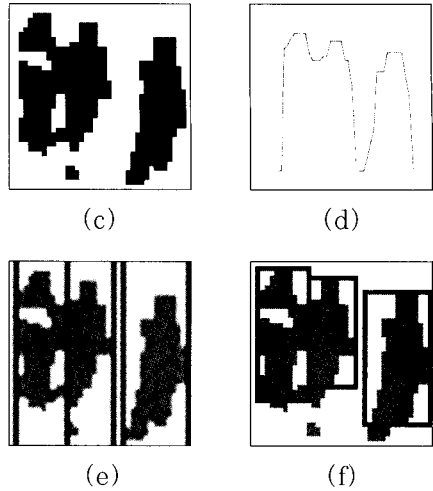


그림 2. Draping에 의한 사람군 영상의 나눔: (a) 원 영상, (b) 전경영역의 추출, (c) 사람군 부분 추출, (d) Draping line, (e) 수직의 분할, (f) 수평의 분할로 영역분할의 완성.

Fig. 2. Dividing an image of people group by draping: (a) An original image, (b) Extraction of the foreground region, (c) Extraction of the image patch of people group, (d) Draping line, (e) Vertical division, (f) Completed by horizontal division.

IV. 나누어진 영역의 상태 지정, 변환 그리고 계수

비디오 카메라의 관측영역에 잡힌 사람들은 이동함에 따라 상대적 위치가 달라지게 되므로, 동일한 사람군이라고 할지라도 연속 영상에서 영역의 나눔은 일정하지 못하게 된다. 특히 사람이 다른 사람에 의해 심하게 가려지는 경우 이를 한 영상 프레임에서 알아내기란 거의 불가능할 수 있다. 따라서 연속된 영상에서 각 프레임의 영역 나눔의 결과가 어떻게 변화해 가는지를 해석할 필요가 생긴다.

각 영상 프레임에서 III장에서 서술된 방식으로 나누어진 사람군 영역들이 연속되는 영상에서 어떻게 변화해 가는지를 효율적으로 살피기 위해서 각 영역의 상태를 지정하였다. 각 영역의 상태는 나누어진 영역이 사람인지 하는 것과 그 영역이 다른 영역에 의해 얼마나 가려졌는지에 대한 잠정적인 믿음을 표현하기 위한 수단으로 사용되었다. 즉, draping line의 지역최소에 의해

1) 이 때 화소 행렬(pixel array)의 원점은 영상 프레임의 좌측상단 모서리에 설정되었고, i와 j는 각각 종축과 횡축의 좌표를 의미한다.

나뉘어진 각 영역은 적어도 다음의 네가지 상태 중 하나에 속한다고 가정하였고, 그 예를 그림 3에 보였다.

(i) CO(completely occluded) : 적어도 두 명이상의 사람이 같은 영역에 있지만 draping line에 의한 나눔은 이를 구분하지 못하였다. 이런 상황은 주로 그들 사이의 가려짐의 정도가 커서 두 인접한 사람사이에는 draping line상의 지역 최소가 존재한다는 가정이 만족되지 못하는 경우에 발생하며, 특히 목표물이 원거리에 있어서 해상도가 높지 못한 상황에서 일어날 가능성이 크다.

(ii) PO(partially occluded) : 한 사람이 한 영역안에 있지만 인접 영역과 연결되어 있다. 이런 경우는 인접한 사람끼리 부분적으로 가려짐이나 연결됨이 있지만 draping line상에는 지역최소가 존재하여 적절히 나누어진 경우이다.

(iii) IP(individual person) : 한 사람이 한 영역안에 있으며 인접 영역과 연결됨이 없다. 즉, 한 사람이 영상내에서 다른 사람과 겹쳐짐이 없이 완전히 독립되어 있는 경우이다.

(iv) WD(wrongly divided) : 나누어진 영역은 사람의 소지품이거나 다른 영역에 포함된 사람의 한 부분일 뿐이다. 이런 상황은 사람이 가방등 소지품을 들고 가거나 팔등 말단부의 일부가 draping line에서 지역최소를 유발시켜 별개의 영역으로 나누어진 경우이다.

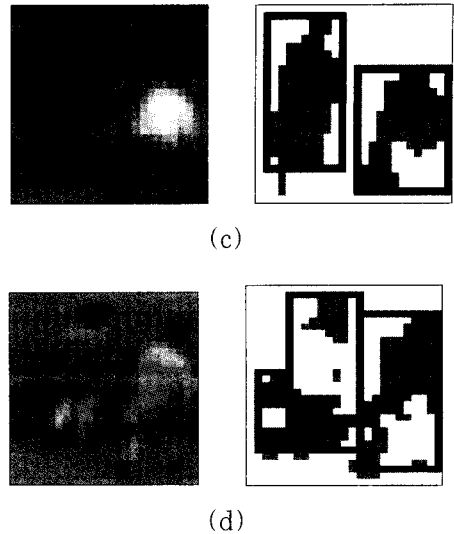
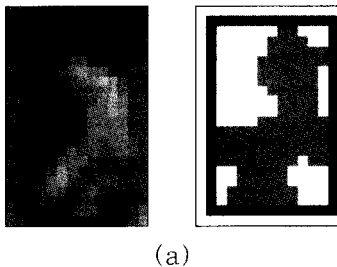


그림 3. 나누어진 영역의 4가지 상태에 : (a) CO, (b) PO, (c) IP, (d) WD.

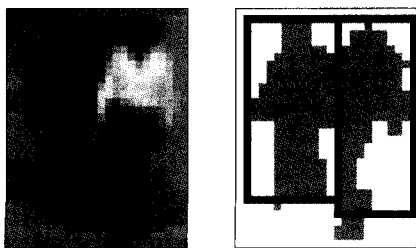
Fig. 3. Examples of the 4 states of divided regions : (a) CO, (b) PO, (c) IP, (d) WD.

계수를 목적으로 주어진 영상 프레임에서 나누어진 영역의 상태를 확인하고, 연속된 영상에서 상태를 변화시키는 절차는 그림 4에 나타낸 바와 같다. 우선 주어진 영상 프레임에서 앞 절에서 설명한 방식과 같이 draping line의 지역최소를 이용하여 사람군을 독립된 영역으로 나눈다. 나누어진 영역중 WD가 아니라고 판정된 영역들에 대해서 몇가지 특성을 분석하게 되고, 이를 C_rect 변수에 저장한 후 이에 기초하여 각 영역의 상태를 판정하게 된다. 만약 현재 처리하는 프레임이 관측장면에 관한 연속 영상중 최초의 프레임이 아니라면 이전 프레임의 해석중에 얻어진 변수 P_rect를 현재 프레임내 영역을 해석하는 데 참고하게 되며, 그렇지 못한 경우에는 현재 프레임내 사각형의 특성 변수 C_rect만에 의해 판정하게 된다. P_rect와 C_rect에 저장되는 특성은 다음과 같다.

- coor-영역을 구분하는 사각형 좌측상단과 우측하단의 행과 열 좌표
- connect- 사각형의 좌표에 기준하여 이웃 사각형들과의 연결여부를 나타낸다.
- state-판정된 상태(state). 판정 전에는 비어(null) 있다.
- pat-현재 사각형이 비롯된 이전 프레임에서의 사각형, 즉 부모 사각형을 지시한다.



(a)



(b)

- num-사각형내 사람의 숫자. CO가 아닌 경우에는 1이다.
- flag-사각형의 합침(merger)이나 나눔(division)이 일어난 경우에는 이를 표시한다.

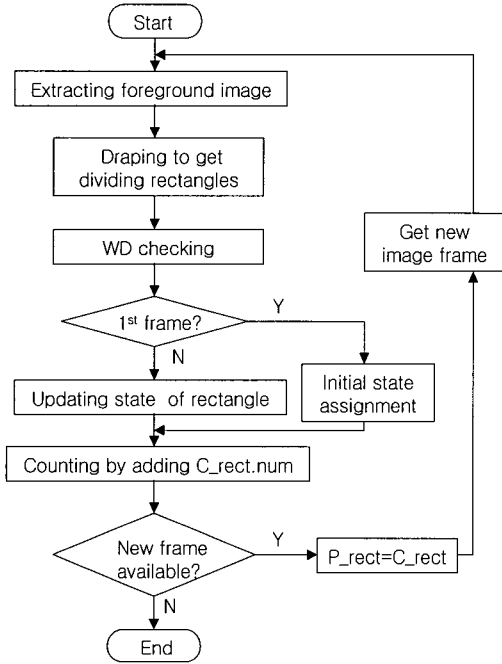


그림 4. 상태확인에 의한 계수의 절차.
Fig. 4. Counting steps by state recognition.

나누어진 영역중 WD를 판정하고 프레임내 사각형의 리스트로부터 이를 제외하는 작업은 다음과 같은 절차에 의한다. 즉, 영상내 N개 사각형의 좌표, $C_rect[i].coord$, $i=1, \dots, N$, 로부터 사각형의 폭과 높이, 그리고 수직 중심을 구한 후 이들 중 하나라도 미리 정해진 임계치보다 작으면 WD로 판정한다. 즉, 폭이 일정 너비에 미치지 못하거나, 좌 또는 우편의 사각형에 비해 너무 낮게 위치하는 작은 사각형은 잘못 분리된 것이라고 간주하는 것이다. 이는 영상내 사람의 크기를 고려하고, WD가 주로 몸의 움직임이 활발한 부분인 다리부분이나 손에 들려 있는 물체에 의해 나타난다는 사실을 관측한 결과이다. 그림 5는 WD를 판정하기 위한 프로그램의 흐름도이다. 여기서 th_W , th_H , th_C 는 각각 사각형의 폭, 높이, 수직중심의 이웃 사각형에 대한 상대적 위치에 대한 임계치로 판정에 사용되는 값이다.

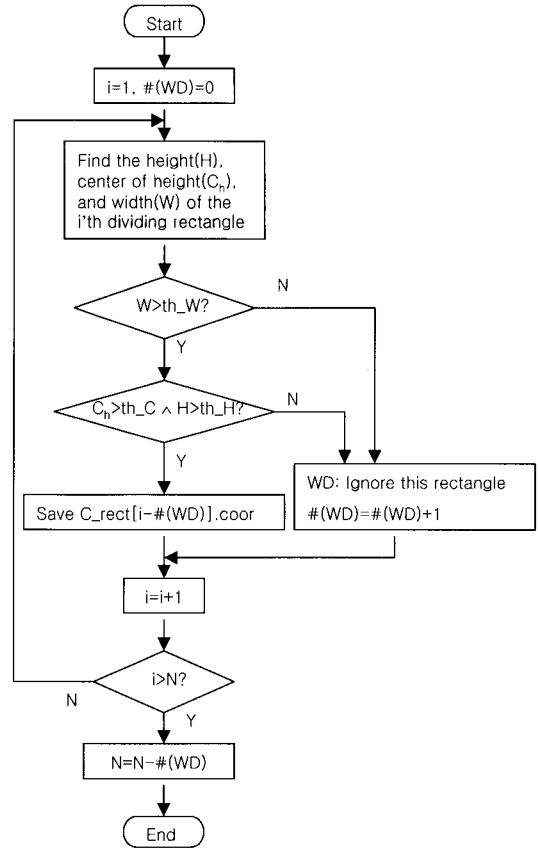


그림 5. WD 판정 루틴의 흐름도.
Fig. 5. A flow chart for WD checking routine.

만약 현재의 영상 프레임이 관측되고 있는 목표에 관한 최초의 프레임인 경우, 최초 상태지정을 위한 절차가 실행된다. 최초 상태지정(initial state assignment)을 위한 루틴의 출력은 단지 PO 또는 IP인데, 단순히 $C_rect.connect$ 만을 살펴서 주변 사각형과 연결되어 있는 경우 PO로 그렇지 않은 경우에는 IP로 지정한다.

현재의 프레임이 관측장면에 대한 최초의 프레임이 아닌 경우에는 직전 프레임을 해석한 결과를 참조하게 된다. 이 때, 세가지 경우를 생각할 수 있는데, 그 첫째는 이전 프레임과 현재의 프레임에서 사각형의 개수에 변화가 없는 경우이고, 둘째는 개수가 감소된 경우, 셋째는 개수가 증가된 경우이다. 먼저 개수가 변화하지 않는 경우에는 직전 프레임에서 얻어진 결과를 기초로 사각형이 직전 프레임에서 CO였던 경우와 아니었던 경우로 나눈 후, 후자의 경우에는 좌우 연

결성을 참조하여 IP 또는 PO로 판정한다. 그림 6은 이러한 절차를 보여준다.

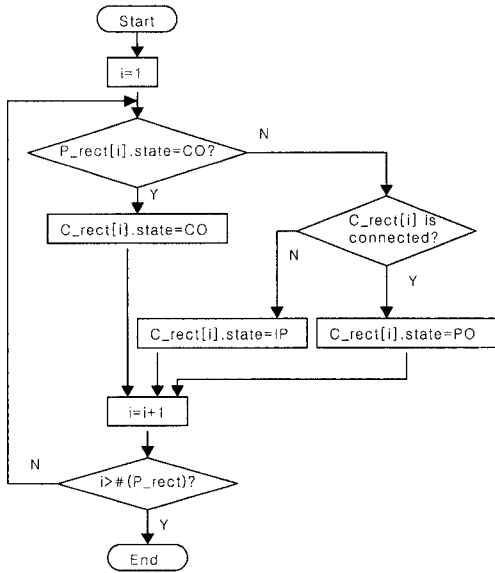
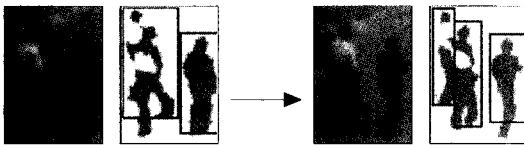
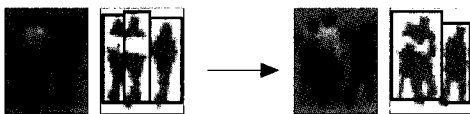


그림 6. 이전 프레임과 현재 프레임에서 나누어진 영역의 수가 같은 경우의 상태확인.

Fig. 6. State recognition when the numbers of regions divided on previous and current frames are the same.



(a)



(b)

그림 7. 연속 영상에서 상태 변환의 예 : (a) CO → PO, (b) PO → CO.

Fig. 7. Examples of state transition : (a) CO → PO, (b) PO → CO.

한편 그림 7에서 예를 보인 바와 같이 이전 프레임과 비교하여 영역 나눔의 개수변화가 있는 경우에는 이를 확인하고 상태를 갱신해야한다. 우선 숫자가 적어진 경우는 직전 프레임에서 PO

또는 CO 상태에 있던 이웃하는 두 영역이 현재 프레임에서는 합쳐져 CO 상태로 됨을 의미한다. 따라서 합쳐져서 새로 생긴 사각형을 찾고 내부에 포함된 사람의 수는 이를 잉태한 두 부모 사각형내 사람 수를 합친 값으로 하는 작업이 필요하다. 이를 위해서는 합쳐진 두 사각형과 결과된 새 사각형을 직전 프레임과 현재 프레임에서 찾아낼 필요가 있다. 본 논문에서는 두 프레임 사이에서 사각형들 사이의 폭을 비교하였다. 즉, 다음 식에서와 같이 직전 프레임의 i 번째 사각형과 $i+1$ 번째 사각형 폭, $PW(i)$ 와 $PW(i+1)$ 의 합과 현재 프레임의 i 번째 사각형의 폭 $CW(i)$ 의 차이가 어떤 임계치 이하이면 합침이 일어났다고 간주하는 것이다.

$$fn_mrg(i) = \sqrt{[PW(i) + PW(i+1) - CW(i)]^2} < th_M \quad (3)$$

단, $1 \leq i \leq \#(C_rect)$

임계치 th_M 은 작은 값으로 초기화한 후 $(\#(P_rect) - \#(C_rect))$ 개의 합침이 일어날 때까지 증가시킨다. 그림 8은 합침이 일어난 경우의 프로그램 루틴을 보여준다.

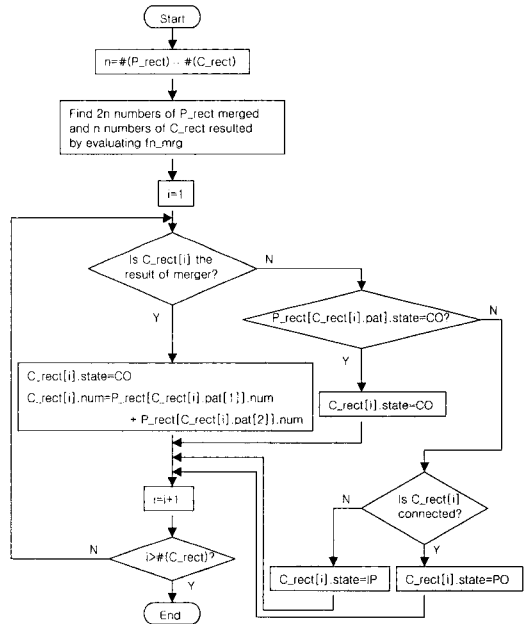


그림 8. 나누어진 영역의 수가 감소된 경우의 처리 순서.

Fig. 8. Processing steps when the number of divided regions decreases.

만약 사각형의 개수가 전의 프레임의 경우와 비교하여 많아졌다면, 이는 이전 프레임의 사각형이 현재 프레임에서 분리되었음을 의미하며, 다음의 식을 사용하여 분리된 사각형과 분리를 통하여 결과된 사각형들을 찾는다.

$$fn_div(i) = \sqrt{[PW(i) - CW(i) - CW(i+1)]^2} < th_D \quad (4)$$

단, $1 \leq i \leq \#(P_rect)$

이 때 임계치 th_D 는 $(\#(C_rect) - \#(P_rect))$ 개의 나눔이 얻어질 때까지 초기의 작은 값에서 차츰 증가된다. 그림 9는 이러한 경우의 처리 단계를 보여준다.

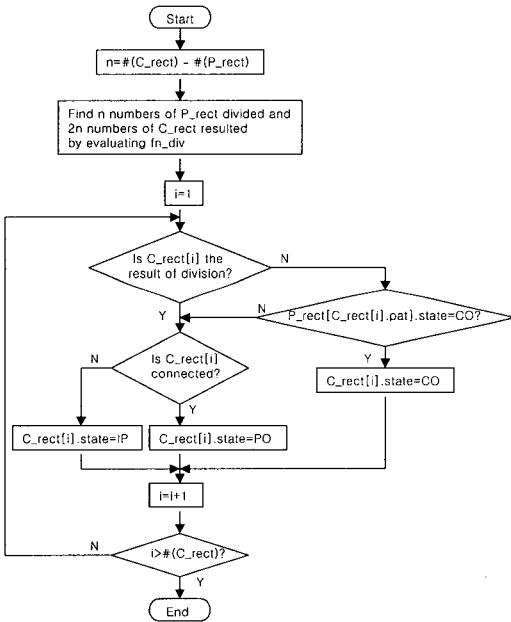


그림 9. 나누어진 영역의 수가 증가된 경우의 처리순서.

Fig. 9. Processing steps when the number of divided regions increases.

V. 실험

제안된 계수 방법을 실제 실험을 통하여 확인하였다. 두 다른 장소의 2층과 6층 건물 옥상에 설치된 비디오 카메라를 사용하여 2명에서 5명

까지 출현하는 사람군의 비디오 영상을 얻고 분석하였다. 제안된 계수 기법은 Matlab에 의해 프로그램된 후 200MHz Pentium PC에서 약 0.5[frame/sec]의 속도로 실행되었다. 이 때 각 영상 프레임의 크기는 320×240 이다.

표 1은 실험의 결과를 보여주며, 모든 경우에 제안된 방식은 정확한 계수를 하였다. 특히 5명이 출현하는 경우, 상호간의 가려짐이 심하여 정확하게 숫자가 구분되는 영상 프레임은 2/49에 불과하지만 제안된 기법은 5명이 영상내에 있다는 것을 알아낼 수 있었다. 이 경우의 구체적 예는 그림 10에 보였다. 실험에서 WD의 판정을 위해 사용한 사각형 폭과 높이의 임계치는 각각 $th_W = 6[\text{pixel}]$ 과 $th_H = 19[\text{pixel}]$ 이었고, 중심의 수직 높이는 이웃 사각형의 40%이상, 즉 $th_C = 0.4 \times (\text{이웃 사각형의 수직 중심 평균})$, 을 임계치로 하였다.

한편 동일한 영상에 신뢰도의 척도로 식(1)을 적용한 경우 그림 11의 그래프를 얻을 수 있었다. 그림에서 볼 수 있듯이 밝기 균질도의 분산이 가장 작은 경우는 17번째 프레임에서 발생했으며, 이 때 영상내 나누어진 영역의 계수는 4로 부정확했다. 이는 II장에서 서술된 바와 같이 심하게 가려진 개체들이 상호간에 부분적인 가림이 있는 개체들에 비해 오히려 균질하게 판정되는 이유로 인한 것이다.

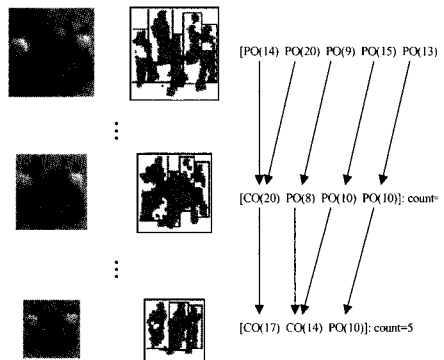


그림 10. 연속 영상에서 상태 변환을 관측함으로써 사람을 계수하는 예. (n)에서 n은 사각형의 폭.

Fig. 10. People counting example by monitoring the state transition in an image sequence. The n of (n) is the width of dividing rectangle.

표 1. 실험에 사용된 다양한 비디오 영상의 특성과 나눔의 결과.

Table 1. The property of video image sequences and division results.

비디오에 출현하는 사람의 실숫자		2	3	4	5
처리된 영상 프레임의 수		38	45	32	49
각 영상 프레임에서 draping line에 의해 나누어진 영역의 수	과다 나눔	8	0	1	0
	정확한 나눔	30	32	19	2
	과소 나눔	0	13	12	47
한 사람의 영상 해상도	최대 크기	16×33	29×51	21×42	22×47
	최소 크기	7×19	7×22	7×20	6×19

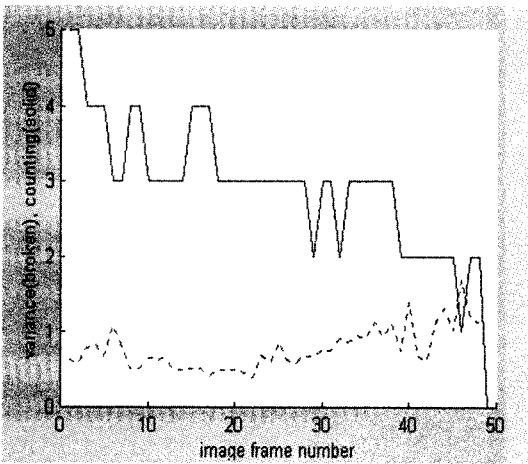


그림 11. 각 프레임당 나누어진 영역의 밝기 분산과 계수의 결과.

Fig. 11. The brightness variance of divided regions and counted result for each frame.

VI. 고찰

사람은 차량과 더불어 VSAM 시스템의 가장 중요한 목표이지만 그 연구의 성과는 아직 상대적으로 크게 부족하다. 현재 이 분야의 연구는 주로 사람의 추적과 행위 해석등에 집중되는 경향이 있지만, 이에 선행되어야 할 것은 몇 명의 사람이 영상내에 있는지, 그들 각자의 영역은 어떻게 구분되어질 수 있는지 하는 문제이다. 본 논문에서는 이러한 점을 고려하여 비디오의 연속 영상에서 사람을 계수하기 위한 기법이 제안되었다.

각 영상 프레임에서 사람군의 영상이 추출되고, 이를 draping line을 이용하여 수직으로 나누었다. 이같은 접근법은 기존의 기법에 비해 단

순하면서도 전경 영역의 불완전한 추출이나 사람의 자세에 비교적 덜 민감하다는 장점을 확인할 수 있었다. 나누어진 각 영역에는 그 가려짐의 정도에 따라 상태가 할당되고 이의 변화를 추적함으로써 다양한 상황하에서도 바른 계수가 가능했다.

본 논문의 영역은 사람의 계수 문제에 한정되어 있지만, 이를 추적이나 행위 해석과 같은 다른 작업과 연결시키는 일은 시스템의 완성도를 위해 필요한 작업이다. 따라서 사람에 관한 VSAM의 기존 연구결과와 본 논문의 결과를 연결시키는 작업은 향후의 연구과제가 될 것이다.

참고 문헌

- [1] T. Kanade *et al.*, "Advances in cooperative multi-sensor video surveillance," Proc. DARPA Image Understanding Workshop, pp.3-24, 1998.
- [2] H. Fujiyoshi and A. J. Lipton, "Real-time human motion analysis by image skeletonization," Proc. IEEE WACV98, 1998.
- [3] J. Heikkila and O.Silven, "A real-time system for monitoring cyclists and pedestrians," Proc. 2nd IEEE Int. Workshop on Visual Surveillance, pp.74-81, 1999.
- [4] I. Haritaoglu, "A real time system for detection and tracking of people and recognizing their activities," Ph.D dissertation, University of Maryland, College Park, 1999.

- [5] A. J. Lipton, *et al.*, "Moving target classification and tracking from real-time video," Proc. DARPA Image Understanding Workshop, 1998.
- [6] C. Wren, *et al.*, "Pfinder: real-time tracking of the human body," IEEE Trans. PAMI, vol.19, no.7, pp.780-785, 1997.
- [7] I. Haritaoglu, *et al.*, "Ghost: A human body part labeling system using silhouettes," Proc. Image Understanding Workshop, pp.229-235, 1998
- [8] C. Anderson *et al.*, "Change detection and tracking using pyramid transformation techniques," Proc. SPIE: Intelligent Robots and Computer Vision, vol.579, pp.72-78, 1985.
- [9] J. Barron *et al.*, "Performance of optical flow techniques," Int. J. Computer Vision, vol.12, no.1, pp.42-77, 1994.
- [10] R. Haralick and L. Shapiro, Computer and Robot Vision, Addison-Wesley Pub., Reading, MA, 1992.

著 者 紹 介

도 용 태

경북대학교 전자공학과(공학사)

서강대학교 전자공학과(공학석사)

Univ. of Hull(영) 전자공학과(공학박사)

CMU(미) Robotics Institute (객원부교수)

Univ. of Wisconsin(미), ECE (명예연구원)

현재 대구대학교 정보통신공학부 (부교수)

주관심 분야 : 자동화용 센서시스템 및 로봇시각