

동영상에서 물체의 움직임 검출 방법

이연철*, 김은이, 황상원, 김항준
경북대학교 컴퓨터공학과 인공지능연구소

Motion Detection Method For Nonstationary Video Sequences

Yeon Chul Lee*, Eun Yi Kim, Sang Won Hwang, Hang Joon Kim
Dept. of Computer Engineering, Kyungpook National University
yclee@ailab.knu.ac.kr

요약

본 논문에서는 카메라의 움직임이 있는 비디오 영상에서 물체의 움직임을 검출하기 위한 방법을 제안한다. 입력 영상에서 카메라의 움직임이 있을 경우, 이를 제거하기 위해 카메라의 움직임을 추정, 보상한다. 보상된 영상에서 적응적 임계값을 사용하여 물체의 움직임을 검출한다. 제안된 방법은 카메라의 움직임이 있는 비디오 영상뿐만 아니라, 시간에 따라 조명이 변하거나 잡음이 포함된 비디오 영상에 대해 좋은 결과를 얻을 수 있다.

I. 서론

비디오 영상에서 물체의 움직임을 검출하는 것은 이미지 처리의 가장 흥미 있는 분야 중 하나이다. 이는 물체 추적, 침입자 감시 시스템, 자동차 감시 시스템에서부터 영상 분할과 압축까지 다양한 분야에 사용되고 있다 [1,2,3].

비디오 영상에서 물체의 움직임을 검출한다는 것은 움직임이 있는 물체와 움직임이 없는 배경을 서로 구별시켜 주는 것이다. 이를 위한 많은 연구가 이루어지고 있으며, 지금까지 연구된 방법들 중에 가장 단순하고 널리 사용하는 방법은 고정된 임계값을 적용하여 임계값보다 낮은 밝기 차이를 가진 부분은 변화가 없는 배경으로 구별하고, 임계값보다 큰 밝기 차이를 가진 부분은 변화가 있는 물체로써 구별하는 것이다. 이런 임계값의 선택은 비디오 영상 그 자체 뿐 아니라 시간에 따라 변화하는 조명에 상당히 의존적이다.

또한, 위의 방법은 모든 프레임들에게 고정된

임계값을 적용하기 때문에, 시간에 따라 조명이 변하거나 비디오 영상에 잡음이 있을 때 부적절하다. 즉, 물체의 움직임을 검출하기 위한 임계값의 선택은 입력 영상에 따라 유동적으로 선택되어야 한다. 이러한 임계값의 선택은 카메라의 움직임이 없는 환경에서만 이루어지기 때문에, 카메라의 움직임이 있는 환경에서는 카메라의 움직임 제거가 선행되어야 한다.

이런 문제점들을 해결하기 위해서, 본 논문에서는 정적인 환경뿐 아니라 동적인 환경에서도 적응적 임계값을 사용하여 물체의 움직임을 검출하는 방법을 제안한다. 두 장의 입력 영상에서 카메라의 움직임이 있을 경우, 이를 제거하기 위해 어핀 모션 모델(Affine Motion Model)을 이용한다 [4,5]. 보상된 비디오 영상에서 물체의 움직임을 검출하기 위한 임계값의 선택은, 연속된 두 프레임의 밝기 차이는 세 개의 가우시안 분포(Gaussian Distribution)의 혼합으로 구성되어 있다고 가정하며 [6], 이를 이용해서 가우시안 모델과 데이터 사이에서의 절대적인 차의 합이 최소가 되는 값을 임계값으로 결정한다. 이를 통해 얻어진 임계값을 이용해 이진화 영상인 CDM(Change Detection Mask)을 얻어내고, 후처리 과정을 통해 물체의 움직임을 검출한다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 카메라의 움직임을 제거 후 보상하는 방법을 서술하고, III장에서는 적응적 임계값을 계산하고, 후처리를 하는 방법을 서술하며, 실험 결과는 IV장에서 보이고, 마지막 V장에서 결론을 맺는다.

II. 카메라 움직임 추정 및 보상

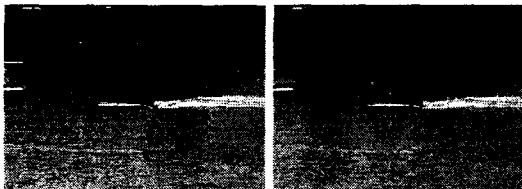
비디오 영상 중에는 카메라의 움직임이 있을 수 있으며, 이는 카메라의 줌-인, 줌-아웃, 회전, 이동이나 이들의 혼합으로 구성되어 있다. 본 논문에서는 카메라의 움직임을 어파인 모션 모델을 이용해서 표현한다 [4]. 어파인 모션 파라미터들은 LSM(Least Square Method)을 사용해 추정한다. 이때, 최소화해야 할 에러 함수는 식 (1)과 같다.

$$E(a) = \sum_{i=1}^N \{ [U_X(x_i, y_i) - \tilde{U}_X(x_i, y_i)]^2 + [U_Y(x_i, y_i) - \tilde{U}_Y(x_i, y_i)]^2 \} \quad (1)$$

$$\tilde{U}_X(x_i, y_i) = a_1x + a_2y + a_3$$

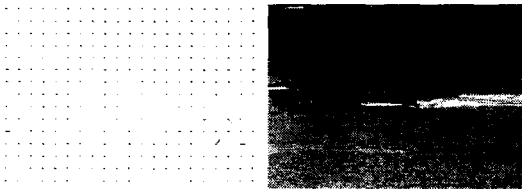
$$\tilde{U}_Y(x_i, y_i) = a_4x + a_5y + a_6 \quad (2)$$

여기서, $U_X(x_i, y_i)$, $U_Y(x_i, y_i)$ 는 픽셀 (x_i, y_i) 에서의 로컬 모션 벡터(Local Motion Vector)들이고, $\tilde{U}_X(x_i, y_i)$, $\tilde{U}_Y(x_i, y_i)$ 는 글로벌 모션 벡터(Global Motion Vector)이며, (a_1, \dots, a_6) 는 어파인 모션 파라미터들(Affine Motion Parameters)이다. 로컬 모션 벡터들은 Full Search Block-Matching 알고리즘을 이용하여 얻는다.



(가) 4번째 프레임

(나) 5번째 프레임



(다) 모션 벡터

(라) 카메라 움직임이 제거된 영상

그림 1. "Coast Guard" 의 카메라 움직임 추정 및 보상

그림 (1)은 카메라의 움직임(이동)이 있는 "Coast Guard" 영상에서 카메라의 움직임을 제거한 영상을 보여주고 있다. 그림 (1)의 (가)와 (나)

는 "Coast Guard" 영상에서의 4, 5번째 프레임이고, (다)는 5번째 프레임에서의 로컬 모션 벡터이며, (라)는 어파인 모션 모델을 통해 카메라의 움직임을 제거한 후 보상된 5번째 프레임이다.

카메라의 움직임을 추정 및 보상을 통해, 비디오 영상에서 물체의 움직임 검출은 카메라 움직임의 유무에 관계없이 연속된 두 장의 프레임간의 밝기 차이로부터 얻어질 수 있게 된다.

III. 적응적 임계치 및 후처리

비디오 영상에서 물체의 움직임 검출은 입력 영상들의 연속된 두 프레임간의 밝기 차이로부터 얻어진다. 연속된 두 프레임에서 물체의 움직임을 검출하기 위한 임계값의 계산은 두 프레임에서 얻어진 밝기 차의 히스토그램이 세 개의 가우시안 분포를 따른다는 가정에서 계산되어진다 [6]. 세 개의 가우시안 확률 분포는 식 (3)과 같이 정의된다.

$$f(i) = \frac{\varphi_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{i-\mu_1}{\sigma_1}\right)^2} + \frac{\varphi_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{i^2}{2\sigma_2^2}} + \frac{\varphi_3}{\sqrt{2\pi}\sigma_3} e^{-\frac{1}{2}\left(\frac{i-\mu_3}{\sigma_3}\right)^2} \quad (3)$$

왼쪽과 오른쪽의 항은 물체의 움직임 때문에 발생하는 밝기 차이의 변화들을 나타내며, 가운데 항은 움직임이 없거나 잡음이 존재하여 생기는 밝기 차이의 변화들을 나타낸다.

임계값의 결정은 식 (4)에 정의된 적합 기준 함수를 계산함으로써 결정되어진다. 이 기준 함수를 최소화하는 t 가 임계값이 된다.

$$H(t) = \sum_{i=\min(d)}^{\max(d)} |P(i) - f(i)| \quad (4)$$

여기서, d 는 두 프레임간 밝기 차의 범위이고, $P(i)$ 는 i 번째 값의 확률이며, $f(i)$ 는 세 개의 가우시안 확률 분포 중에 i 번째 값이 속하는 가우시안 확률 분포 값이다. 또한, $f(i)$ 는 식 (5)로 근사될 수 있다.

식 (4)에서 계산된 임계값은 두 장의 연속적인 프레임의 밝기 차에 대한 하나의 CDM를 발생시킨다. CDM은 움직임 있는 영역과 움직임이 없는 영역으로 구별 시켜주는 이진화 영상이며, 이는 식 (6)와 같이 정의된다.

$$f(i) \approx \begin{cases} \frac{\varphi_1}{\sqrt{2\pi\sigma_1}} e^{-\frac{1}{2}\left(\frac{i-\mu_1}{\sigma_1}\right)^2}, & \min(d) \leq i < -t \\ \frac{\varphi_2}{\sqrt{2\pi\sigma_2}} e^{-\frac{i^2}{2\sigma_2^2}}, & -t \leq i \leq t \\ \frac{\varphi_3}{\sqrt{2\pi\sigma_3}} e^{-\frac{1}{2}\left(\frac{i-\mu_3}{\sigma_3}\right)^2}, & t < i \leq \max(d) \end{cases} \quad (5)$$

$$CDM(i, j) = \begin{cases} 1, & \text{if } |D(i, j)| > t \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

여기서, $D(i, j)$ 는 현재 프레임과 이전 프레임에서 픽셀 (i, j) 의 밝기 차이 값이며, t 는 임계값이다.

그림 2는 적응적 임계값을 사용하여 CDM을 생성하는 것을 보여 주고 있다. (가)와 (나)는 "Table Tennis"의 4, 5번째 프레임이고, (다)는 적응적 임계값을 사용해 얻어진 5번째 프레임의 CDM이다.

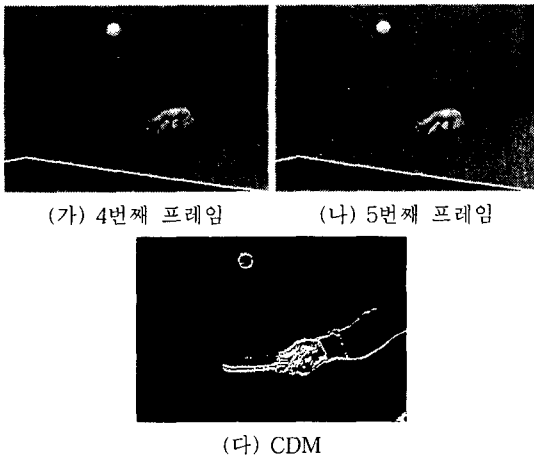


그림 2. "Table Tennis"의 움직임이 있는 물체의 검출

그러나, 그림 (다)에서처럼, CDM은 잡음을 포함하고 있으며, 이는 후처리 과정을 통해 해결한다.

CDM은 물체의 움직임이 있는 부분의 내부가 아닌 외부의 경계만이 변한 것으로 나타난다. 이런 문제점들은 인접 픽셀의 연결성 성분(Connected Component) 조사에 의해 쉽게 해결

될 수 있다. 움직임이 있는 부분은 서로 연결되어 있고, 잡음은 연결성이 없는 조금만 영역을 형성한다.

따라서, 이런 잡음은 연결성 성분의 크기에 의해 쉽게 제거될 수 있다. 하나의 연결성 성분이 임계값보다 크다면, 움직임이 있는 부분에 속한다고 가정하고, 그렇지 않으며 잡음으로 간주하고 제거한다. 그림 3은 그림 2의 (다)에 연결성 성분 조사를 이용하여 잡음을 제거한 CDM을 보여주고 있다.

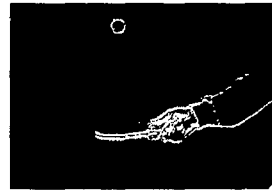


그림 3. 후처리 과정을 통해 잡음이 제거된 영상

IV. 실험 결과

본 논문에서 제안한 방법을 정적인 환경의 비디오 영상들과 동적인 환경의 비디오 영상들에서 실험을 해 보았다. CDM에서 잡음을 제거하기 위하여 연결성 요소의 최소, 최대 크기를 경험적으로 얻어진 값들을 입력으로 주었다.

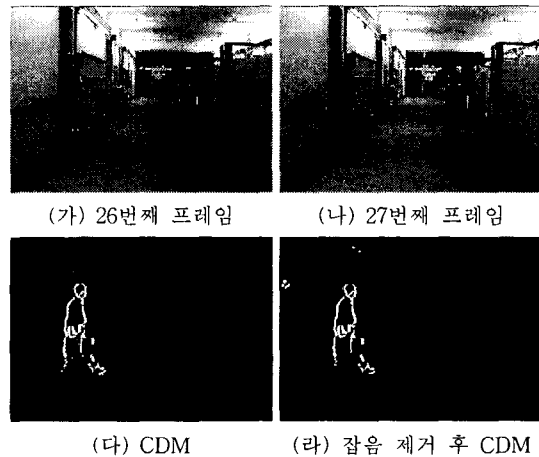


그림 4. 카메라의 움직임이 없는

"Hall monitor" 영상에서의 움직임 검출

그림 4는 카메라의 움직임이 없는 "Hall monitor" 영상에서의 실험 결과를 보여주고 있다. 이 영상은 매 프레임마다 조명의 변화가 일어나

는 것으로 알려져 있다. (가)와 (나)는 26, 27번째 프레임이고, (다)는 27번째 프레임의 CDM이다. (라)는 연결성 성분을 이용해 얻어진 CDM이다.

그림 5는 카메라의 움직임(줌-아웃)이 있는 "Table Tennis"에서의 실험 결과를 보여주고 있다. (가)와 (나)는 31, 32번째 프레임이고, (다)는 카메라의 움직임을 제거하지 않고 적응적 임계값을 적용한 영상이고, (라)는 카메라의 움직임을 제거한 후 적응적 임계값을 적용한 영상이다. (마)는 (라)영상에서 연결성 성분을 이용해 후처리를 한 CDM이다.

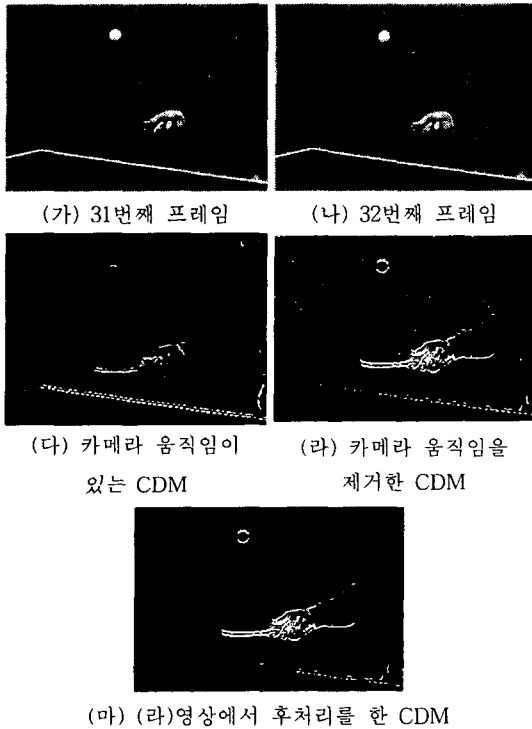


그림 5. 카메라의 움직임이 있는 "Table Tennis" 영상에서의 움직임 검출

V. 결론

본 논문에서는 카메라의 움직임이 있는 비디오 영상에서 물체의 움직임을 검출하기 위한 방법을 제안하였다. 제안된 방법은 카메라의 움직임이 있는 비디오 영상뿐만 아니라 조명의 변화가 있거나 잡음이 있을 경우에도 적용될 수 있음을 보였다. 연속적인 두 프레임들에 카메라 움직임이 있을 경우 이파인 모션 모델을 이용해 카메라의 움직임을 추정, 보상한 후 두 프레임 간의 밝기 차

이에 대한 히스토그램을 이용하여 임계값을 계산하였고, 이렇게 얻어진 임계값을 사용하여 물체의 움직임을 검출하였다. 또한, 이를 통해 얻어진 영상에 잡음을 없애기 위한 후처리를 수행함으로써 더 정확한 움직임을 검출할 수 있었다.

하지만, 제안된 방법은 빠른 움직임을 포함하는 동영상에 적용하는데는 제한이 있다. 향후 연구에서는 이러한 문제를 해결할 뿐만 아니라, 다양한 동영상에 대해 실험해 볼 것이다.

참고문헌

- [1] D. Le Gall, "MPEG: A video compression standard for multimedia applications", *IEEE Trans. on Image Processing*, vol. 6, no. 1, pp. 189-202, 1997.
- [2] L. H. Chen, S. Chang, "A video tracking system with adaptive predictors", *Pattern Recognition*, vol. 25, no. 10, pp. 1171-1180, 1992.
- [3] Kui Zhang, Bober M., Mittler J., "Image sequence coding using multiple-level segmentation and affine motion estimation" *IEEE Journal on Selected Areas in Communications*, Vol. 15 9, pp. 1704 -1713, Dec. 1997.
- [4] M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, Member, IEEE, C. Ahn, Y. Ho, "A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no. 8, pp. 1216-1226, Dec. 1999
- [5] C. Fuh, P. Maragos, "Affine Models for Image Matching and Motion Detection", *Acoustics, Speech, and Signal Processing (ICASSP '1991)*, vol. 4, pp. 2409 -2412, 1991
- [6] N. Habili, A. Moini, N. Burgess, "Automatic Thresholding for change Detection in Digital Video", in *Proc. SPIE Visual Communications and Image Processing*, vol 4067, pp. 133-142, 2000.