Pan-Tilt-Zoom-Camera에서 AAM과 CAMSHIFT를 이용한 얼굴 검출 및 추적

배정완°, 최권택, 변혜란
연세대학교 컴퓨터과학과
{baejw°, choikt, hrbyun}@cs.yonsei.ac.kr

Face Detecting and Tracking using Active Appearance Models and CAMSHIFT with a Pan-Tilt-Zoom-Camera

Jeongwan Bae°, KwunTaeg Choi, Hyeran Byun
Yonsei University

요 약

감시 시스템에서 많이 사용되는 팬틸트줌(Pan-Tilt-Zoom) 카메라로 객체 검출과 추적을 할 때 카메라를 섬세하게 제어하는 것이 중요하다. 본 논문은 팬틸트줌 카메라를 이용하여 얼굴을 검출 및 추적하는 감시 시스템 구성과 카메라 제어 방법을 제안한다. 얼굴 검출을 위해서 P. Viola가 제안한 Haar-like feature를 이용한 빠른 객체 검출방법을 이용하고 얼굴 추적을 위해서 CAMSHIFT와 AAM을 이용하여 얼굴추적과 얼굴 특징 정보 추출이 가능한 감시 시스템 구현을 하였다.

Keywords: 팬틸트줌 카메라, 얼굴 추적, AAM, CAMSHIFT, 얼굴 검출

1. 서 론

감시 카메라를 이용하여 얼굴을 검출하고 추적하는 연구는 최근에 활발히 연구되고 있다. 감시 카메라의 종류를 4가지로 나눠볼 수 있는데 고정 카메라(Fixed Camera), 팬틸트 카메라(Pan-Tilt Camera), 팬틸트줌 카메라(Pan-Tilt-Zoom Camera), 그리고 전방위 카메라(Omniview Camera)이고 본 논문은 먼 거리에 있는 사람의 얼굴을 추적하고 얼굴의 특징을 추출하는데 적합한 팬틸트줌 카메라를 사용한다.

감시 시스템 중에 각각 다른 역할을 하는 두 대 이상의 카메라들(고정 카메라, 팬틸트줌 카메라)을 이용하는 감시 시스템이 있다. 이런 감시 시스템에서는 사람추적을 위해 고정 카메라를 이용하여 배경모델링을 하여 전경을 찾아내는 방법[1, 2]을 많이 사용한다. 그리고 팬틸트줌 카메라를 이용하여 사람이나 다른 객체들을 인식하거나 자세인식, 표정인식 등을 할 수 있다.

얼굴 검출은 특징 기반 방법[4], 신경망을 이용한 방법[5], 색상 기반 방법[6] 등이 많이 연구되고 있다. 감시 시스템에서 실시간 수행이 가능한지는 중요한 요소이므로 실시간 얼굴 검출이 가능하고 100%에 가까운 검출률을 보일 수 있는 P. Viola가 제안한 Haar-like feature를 이용한 빠른 객체 검출 방법[7]을 이용하였다.

얼굴 추적 역시 실시간 수행이 되어야 하므로 실시간이 가능한 추적 방법인 CAMSHIFT[8]와 condensation[9] 그리고 adative Kalman filtering등이 이용될 수 있다. 하지만 이러한 방법들의 공통된 단점은 추적하는 객체가 카메라에서 너무 멀리 있으면 실패하게 된다는 점이다. 그러므로 추적하고자 하는 얼굴이 갑자기 작아지거나 커지지 않도록 줌 조절(Zoom Control[3])이 요구된다. 본 논문은 고정 카메라에서 배경 모델링을 통해 사람의 위치 정보와 크기 정보를 얻고 팬틸트줌 카메라에서 팬과 틸트와 줌 조절을 통해 얼굴 검출을 수행하고 CAMSHIFT와 AAM(Active Appearance Models)[10]을 이용하여 얼굴 추적을 수행하였다. 2장과 3장은 본 논문에서 제안하는 시스템 구성과 카메라 제어방법과 함께 적용된 얼굴 검출과 추적 기술에 관하여 설명한다. 4장은 제한된 실내공간에서 자유롭게 움직이는 사람들의 얼굴을 추적하는 실험을 한 결과이다.

2. 시스템 구성

넓은 실내 공간에서 여러 사람들이 자유롭게 움직이고 있는 상황에서 팬틸트줌 카메라만 이용하여 실시간 얼굴 검출은 쉽지 않은 문제이다. 예를 들어 팬틸트줌 카메라에서 받은 영상에서 얼굴의 크기가 10x10정도로 작다면 얼굴 검출이 용이하지 않다. 검출에 성공했더라도 추적이 가능한 얼굴 크기가 되도록 줌인(ZoomIn)한 후 추적을 하다가 팬틸트줌 카메라가 얼굴의 움직이는 속도를 못 따라가서 추적에 실패하게 되면 다시 추적하기가 힘들고 한 사람의 얼굴을 추적할 때 나머지 사람들은 추적할 수 없다. 감시 영역 전체를 항상 볼 수 있는 또 다른 고정 카메라가 필요하다. 또 다른 고정 카메라를 사용하여 사람들의 위치와 크기를 알 수 있다면 얼굴 후보 영역을 선정할 수 있어 얼굴 후보 영역으로 얼굴 검출이 가능하게 줌인하여 얼굴 검출을 할 수 있다. 사람의 위치와 크기를 알아 내는

방법은 간단하게 배경모델링을 사용하여 새로운 객체가 나타나면 찾을 수 있다[1, 2].

전체 시스템 구성은 고정 카메라에서 사람추적을 항상 하고 팬틸트줌 카메라에서 얼굴 검출이 이뤄지면 바로 다음 프레임(frame)부터 얼굴 추적을 수행하다가 추적에 실패하는 경우 고정된 카메라에서 그 얼굴에 해당하는 사람을 계속 추적하고 있었기 때문에 다시 그 사람의 얼굴 후보 영역에서 얼굴 검출을 하고 추적을 다시 실행한다.

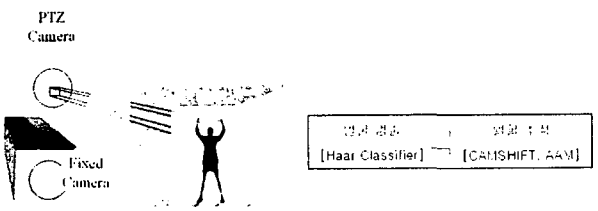

그림 1. 감시 시스템 구성

구체적인 카메라 제어는 우선 얼굴 검출을 시도할 때 얼굴후보영역을 줌인한 후 얼굴 검출을 한다. 검출이 안 되면 팬(Pan)과 틸트(Tilt) 조절을 하여 다른 얼굴 후보 영역을 살펴본다. 다음으로 얼굴 검출이 이뤄지면 얼굴 추적을 하기 위해 추적하기에 충분한 얼굴 색 정보와 특징 정보를 가짐과 동시에 실내 공간에서 감안 할 수 있는 어느 정도 빠른 움직임을 카메라가 따라갈 수 있을 정도의 영상에서 얼굴의 비율을 결정하여야 한다. 320x240영상에서 실험한 결과 얼굴 폭이 영상 너비의 1/4이 항상 되도록 줌 조절하며 얼굴을 추적할 때 위 두 조건을 만족한다.

2.1. 얼굴 검출

얼굴검출에서 복잡한 주위 환경, 다양한 조명 조건, 여러 가지 자세 문제가 있어 다양한 포즈와 조명 그리고 얼굴의 일부가 가려진 상황까지 고려하기 힘들다. 이런 얼굴 검출의 어려움과 얼굴 검출 후 얼굴 추적(정면 얼굴 AAM을 초기값으로 준다)을 고려하여 볼 때 정면얼굴에 국한되어 얼굴검출을 하는 것이 높은 성공률을 보장하며 추적을 쉽게 한다. 정면 얼굴을 검출하기 위해 P. Viola가 Haar-like feature를 이용한 제안한 빠른 객체 검출 방법[7]을 이용하였다.

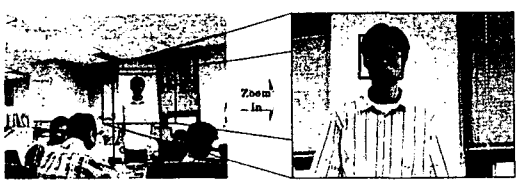

그림 2. 전체영상에서 얼굴 후보 영역으로 줌인 후 얼굴 검출

3. 얼굴 추적(CAMSHIFT, AAM)에서 팬틸트줌 카메라 제어

얼굴 검출이 되면 검출된 얼굴 영역으로 CAMSHIFT[8]를 수행한다. 얼굴이 있는 사각 영역의 HSV color histogram을 구하고 색상 확률분포를 구하여 그 정보로 얼굴 영역을 찾아서 추적을 한다. CAMSHIFT는 Mean Shift와 달리 커널의 크기, 형태의 변환에 영향을 받지 않는다는 장점이 있으나 얼굴 영역이 매우 작아지면 추적이 힘들다. 그러므로 영상에서 일정한 얼굴 크기를 유지하는 것이 중요하다. CAMSHIFT에서 얻은 얼굴 폭을 이용하여 그림 4

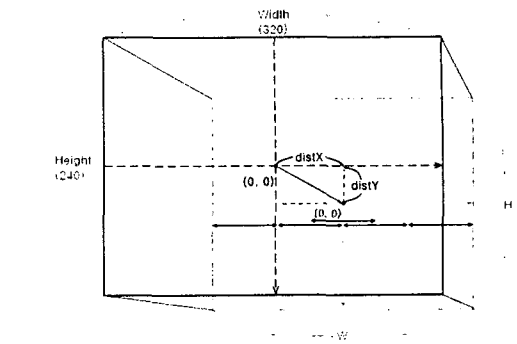

그림 3. 얼굴 추적할 때 얼굴이 항상 영상 중심에 위치하고 일정한 크기가 되도록 팬틸트줌 조절

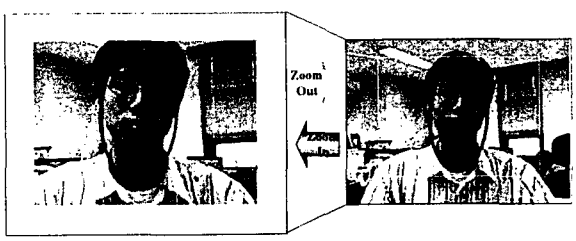

그림 4. CAMSHIFT를 이용한 얼굴 추적 (일정한 얼굴 크기를 유지하기위한 줌 조절)

처럼 줌 조절을 한다. 그림3에서 팬틸트줌 조절 연산을 위하여 아래 두 식을 사용한다. 아래 식 (1)은 얼굴의 중심을 영상의 중심으로 이동하기 위한 팬과 틸트하는 과정이다. pointX와 pointY는 얼굴의 중심이 좌표 값이고 Width와 Height는 영상의 폭과 높이이고 distX와 distY는 팬과 틸트할 거리이다. 식 (2)는 줌을 할 영역의 폭(W)을 계산하는 것이다. face_width는 얼굴의 폭이고 Ratio는 얼굴이 영상에서 차지할 비율이다. Ratio가 4이면 320X240영상에서 얼굴의 폭이 80픽셀이 된다.

$$\begin{aligned} distX &= pointX - \frac{Width}{2} \\ distY &= pointY - \frac{Height}{2} \end{aligned} \qquad (1)$$

$$W = \frac{face_width}{Width} \times Ratio \qquad (2)$$

CAMSHIFT로 얼굴 포즈 정보를 알 수 있어 이를 이용하여 AAM[10]의 초기값을 준다. 포즈 정보가 정면일 경우, 정면 얼굴 AAM을 이용하여 추적하고 포즈 정보가 왼쪽 포즈이거나 오른쪽 포즈일 경우, 좌 우 측면 각각에 대한 AAM을 이용하여 추적한다. 그러나 본 논문에서는 검출된 얼굴에 CAMSHIFT를 적용하여 나온 포즈 정보는 정면이므로 추적을 시작하는 첫 프레임에서는 정

면 평균 AAM을 초기값으로 주고 추적하고 다음 프레임부터는 전 프레임에서 추적된 AAM 결과를 초기값으로 이용하여 추적한다. 그림 5는 이와 같이 AAM을 이용하여 얼굴을 추적 한 결과이다.

(a) (b) (c)

(a) 첫 프레임(CAMSHIFT) (b) 첫 프레임(AAM 초기화)
(c) 20번째 프레임(AAM추적)

그림 5. CAMSHIFT 정보를 이용한 AAM 초기화와 얼굴 추적

4. 실험 및 결과

본 실험은 Pentium4, CPU1.8Ghz의 성능을 가지는 PC에서 SONY SNC-RZ30 팬틸트줌 카메라를 이용하여 320x240 영상이 초당 15프레임 속도로 복잡한 실내 환경에서 실험했다. 고정된 카메라가 사람추적을 하며 만약 얼굴추적에 실패했을 경우 고정된 카메라에서 사람의 위치, 크기정보를 다시 받아서 추적을 재시도 한다. 본 연구에서는 추적 대상이 뛰다든지 갑작스런 빠른 움직임을 보여 얼굴 추적에 실패하는 경우는 고려하지 않고 천천히 움직이거나 조금 빨리 걸을 때 나타나는 얼굴 움직임 속도에 대해서만 실험하였다.

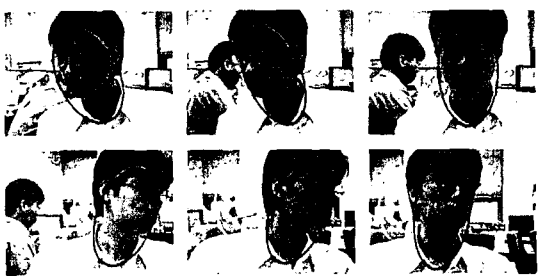

그림 6. 평상시 보다 조금 빠르게 움직이는 얼굴 추적

추적은 우선 CAMSHIFT로 이뤄지고 그 결과 AAM은 초기 얼굴 위치를 알게 된다. 그러나 추적을 시작하는 두 번째 프레임부터 AAM은 CAMSHIFT로부터 독립적으로 추적을 하므로 매 프레임마다 AAM이 어느 정도 정확히 얼굴을 추적하며 얼굴의 특징 추출을 하는지 알아야 추적을 실패하지 않고 있다는 것을 알 수 있다. 매 프레임에서 유사도값을 이용하여 이를 판단할 수 있다. 여기서 말하는 유사도값은 AAM이 얼굴을 찾은 후 실제 얼굴 텍스쳐 정보와 현재 AAM을 통해 찾은 얼굴 텍스쳐 정보간에 차를 표현하는 수치이고 4이하의 값이 나오면 얼굴 추적에 성공했다고 판단할 수 있다. 표1은 얼굴의 움직임이 느린 경우와 빠른 경우로 구분하여 300프레임의 추적영상에 대하여 실험한 결과이다. 여기서 평균유사도 0.29와 0.34는 각 프레임에서 유사도를 측정하여 프레임수로 나누어 평균을 낸 값으로 얼굴 추적을 잘 했다는 것을 보여 준다. 이는 팬틸트줌 연산이 얼굴 움직임을 보정하였기 때문에 조금 빠른 움직임에서도 실제로 영상 프레임 간에 얼굴의 움직임이 작아서 AAM이 얼굴 추적을 실패하는 확률을 낮췄다.

표1. 움직이는 얼굴에서 AAM 추적 정확성 실험 결과

AAM	test set 1 (얼굴 움직임이 느린 동영상)	test set2 (얼굴 움직임이 빠른 동영상)
평균 유사도	0.29	0.34

비록 사람이 뛰어 다니는 것처럼 매우 빠른 움직임에 관하여 실험하지는 못했지만 복잡한 실내 환경에서 얼굴의 일부가 가려지는 일이 없을 때 CAMSHIFT와 AAM, 둘 다 추적이 안 되는 경우가 거의 없었다.

AAM은 얼굴을 추적하면서 얼굴의 특징을 추출하는 역할을 하고 얼굴의 텍스쳐 정보나 포즈 정보와 같은 추출된 얼굴의 특징 정보를 이용하여 얼굴 인식을 할 수 있다.

참고 문헌

[1] L. Davis, I. Haritaoglu, and D. Harwood, "W4: real-time surveillance of people and their activities," IEEE Trans. PAMI, vol. 22, no. 8, pp. 809-830, 2000.

[2] T. Funahasahi, M. Tominaga, T. Fujiwara, H. Koshimizu, "Hierarchical face tracking by using PTZ camera," Proc. Int. Conf. IEEE(FGR'04), pp. 427 - 432, 2004.

[3] X. Clady, F. Collange, F. Jurie, and P. Martinet, "Object tracking with a pan-tilt-zoom camera: application to car driving assistance," Proc. Int. Conf. Robotics, Automation, Vol.2, pp. 1653-1658, 2001.

[4] Y. Yagi, "Facial feature extraction from frontal face image," Signal Processing Procedings, 2000. WCCC-ICSP 2000. 5th Int. Conf., Vol. 2, pp. 1225-1232, 2000.

[5] H.A. Rowley, S. Baluja, t. Kanade, "Neural network-based face detection," IEEE Trans. PAMI, Vol. 20, pp. 23-38, 1998.

[6] S.Z. Li, L. Zhu, Z.Q. Zhang, and H.J. Zhang, "Statistical Learning of Multi-view Face Detection," In Proc. 7th European Conference on Computer Vision Copernhagen, Denmark. 2002.

[7] P. Viola and M. J. Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features," IEEE CVPR, Vol. 1, No. 2, pp. 511-518, 2001.

[8] G. Bradski, "Computer vision face tracking for use in a perceptual user interface," Intel Tech. Journal, Q2, 1998.

[9] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," Int. Journal of Computer Vision, vol. 29, no. 1, pp. 5-28, 1998.

[10] T. Cootes, G. Edwards, and C. Taylor. "Active appearance models," In Proc. European Conference on Computer Vision, pages 484–498, 1998.