

# 실시간 영상에서 객체 추출 및 추적에 관한 연구

장정화  
이즘테크노

e-mail:jangjunghwa@naver.com

## A Study of Object Extraction and Trace at Real Time Images

Jung-Hwa Jang  
Ismtechno Co., Ltd

### 요 약

본 논문에서는 실시간 영상에서 적응적 배경영상을 이용하여 객체를 추출하고 추적하는 방법을 제안한다. 입력되는 영상에서 배경영역의 잡음을 제거하고 조명에 강인한 객체 추출을 위하여 객체영역이 아닌 배경영역 부분을 실시간으로 갱신함으로써 적응적 배경영상을 생성한다. 그리고 배경영상과 카메라로부터 입력되는 입력영상과의 차를 이용하여 객체를 추출한다. 추출된 객체의 내부점을 이용하여 최소사각영역을 설정하고, 이를 통해 객체를 추적한다.

### 1. 서론

객체 추적은 카메라로부터 입력된 영상에서 움직임을 보이는 객체를 인식하고, 그 움직임을 추정하여 추적하는 것이다. 객체를 추적하는 방법은 보안, 의료, 군사, 교통, 제어분야 등 여러 분야에 응용될 수 있어 그동안 많은 연구와 개발이 이루어지고 있다[1]. 하지만, 기존의 실시간 객체 추적 시스템의 구현에는 많은 어려움이 존재한다. 우선 강인한 실시간 객체 추적 시스템을 구현하기 위하여 고가의 장비를 요구한다. 움직임을 감지하기 위한 센서를 장착한 카메라나 pan/tilt로 움직이는 카메라 등의 요구이다. 또한 객체 추출이나 추적에 많은 연산을 필요로 하는 알고리즘 등을 적용하여 객체 추출에는 뛰어난 성능을 보이지만 실시간 객체 추적에는 부적합하다.

기존의 객체 추출 및 인식 기법에서의 객체 추적과 얼굴 추출에 우수한 성능을 보이는 방법은 상대적으로 처리 속도가 떨어짐을 볼 수 있다. 반대로 처리 속도가 빨라 많은 프레임을 처리할 수 있는 방법은 객체 추출과 얼굴 추출에 뛰어난 성능을 보이지 못하고 있다[2].

본 논문에서는 보안 및 감시 시스템 분야에서 적용되어 질 수 있는 방법을 제안하기 위하여, 저가형

PC카메라의 움직임이 고정되어 있고 외부 환경이 아닌 실내 환경이며 배경영상의 변화가 거의 없다는 특수 환경으로 제약 조건을 가진다.

제안하는 알고리즘은 시간의 경과에 따라 배경영역에서 잡음생성을 줄이고자 객체영역 이외의 영역을 배경영역으로 갱신하여 항상 최신의 배경영상을 유지하도록 하였다. 이와 같이 생성된 적응적 배경영상과 설치된 PC카메라로부터 실시간으로 입력되는 입력영상의 차를 이용하여 객체의 크기와 위치를 탐지함으로써 객체를 추출한다. 이때 연산의 양을 줄이기 위하여 그물식 탐색방법을 이용한다. 그리고 추출된 객체의 내부점들에 MBR(Minimum Bounding Rectangle)을 설정하여 객체를 실시간으로 추적한다. 또한 RGB를 정규화하고, 정규화된 RGB 색상정보를 이용하여 MBR내의 객체로부터 얼굴영역을 추출한 후 사용자의 모바일 단말기로 그 결과를 전송한다.

### 2. 기존연구

#### (1) Low-Level Analysis

저수준 해석은 외곽선이나 모서리에 해당하는 에지(edges) 정보, 명암으로 객체를 구분할 수 있는 그

레이(gray) 정보, RGB 색상값을 이용하는 컬러(color) 정보, 비디오 시퀀스를 사용하여 움직이는 객체들의 위치를 파악할 수 있는 움직임(motion) 등이 객체에서 객체를 인식하고 추적하는데 많이 이용되고 있다[3].

(2) Feature Analysis

저수준 해석으로부터 생성되는 특징들은 애매하기 십상이다. 예를 들어서 컬러 모델을 사용하여 객체 영역들의 위치를 결정하는 것에서 유사한 컬러의 배경 객체들 또한 추출될 수 있다. 이것은 더 높은 수준의 특징해석에 의해서 해결될 수 있는 고전적인 다대일 정합문제이다. 많은 객체추출 기법들에서 객체형태의 지식은 그것들의 애매한 상태에서부터 다양한 특징들을 특성화하고 이어서 입증하기 위해서 채택되었다[2,4].

(3) Active Shape Models

능동 형상 모델들은 특징들의 실제적인 물리적 및 더 높은 수준의 외관을 묘사한다. 일단 특징에 대해 가까운 근접 내에서 방출되면 능동 형상 모델은 국소 이미지 특징들(모서리들, 밝기)과 상호작용하고 점차적으로 특징의 형상을 얻기 위해서 변형한다[5].

(4) Linear Subspace Methods

1980년에 사람 얼굴들을 효과적으로 표현하기 위해서 PCA를 사용하는 기법들을 개발하였다[6]. 다른 얼굴 이미지들의 조합이 주어지면 고유벡터들에 의하여 표현되는, 얼굴들의 분포의 주성분들을 찾는다.

3. 제안 기법

(1) 적응적 배경 영상 생성

PC카메라로부터 입력되어진 초기 설정된 배경영상과 객체의 움직임이 감지되지 않은 상태에서 어느 정도 시간이 지난 후의 배경영상은 사람의 감지 능력으로는 두 그림이 똑같아 보이지만, 시스템에서는 둘을 다르게 인식한다. 배경영상과 입력영상의 차이가 균집을 이루고 있으면 배경영상임에도 불구하고 새로운 객체로 인식하는 오류를 범할 수 있다.

전체적인 배경영상의 갱신은 많은 연산량을 필요로 하기 때문에 실시간 객체 추적 성능을 저하시킨다. 제안하는 시스템에서는 객체영역인  $n * m$  의 영역을 제외한 나머지 영역을 배경영상으로 갱신하면서 객체

추적을 동시에 이루어질 수 있도록 하였다.

(2) 객체의 내부점 추출

식 1~4에 따라 RGB채널 각각의 차이가 임계값  $\beta$ 보다 크다면, 객체 내부점 후보군으로 등록하게 된다. 제안하고자 하는 시스템은 보안 및 감시 시스템에 적용되어지기 때문에 사람이라는 객체 이외의 빛의 간섭 등으로 인한 배경에서의 작은 잡음 제거를 위해, 식 5를 이용하게 된다. 또한 객체의 일부분이지만 효율적인 잡음 제거를 위하여 어느 정도의 미세한 객체의 일부분은 무시를 한다. 식 1~4를 만족하는 객체 내부점 후보군에서 연속된 픽셀이  $\gamma$ 이하라면 잡음으로 판단하여 제거하고, 원하는 객체 내부점  $Obj(x_i, y_i)$ 를 추출하게 된다.

\*  $Obj(x_i, y_i)$  : 객체 내부점

\*  $Obj\_left(top, right, bottom)$  : MBR의 최대/최소 좌표

$$Obj\_left = \min [Obj(x_i)] - \delta \tag{1}$$

$$Obj\_top = \min [Obj(y_i)] - \delta \tag{2}$$

$$Obj\_right = \max [Obj(x_i)] + \delta \tag{3}$$

$$Obj\_bottom = \max [Obj(y_i)] + \delta \tag{4}$$

$$MBR = [Objleft, Objtop, Objright, Objbottom] \tag{5}$$

(3) MBR을 이용한 객체 추적

제안하는 알고리즘에서는 완전 차영상이 아닌 그물식 탐색 방법을 이용하기 때문에 일반적인  $n*n$  block median filter window를 사용하는 것이 아니라, horizontal median filter window를 이용한다. 또한 컬러 영상에 적용되는 median filtering을 위한 절차를 따른다.

식 6과 같이 각 화소 샘플에 대하여 각각의 RGB 채널에 대한 차들을 합한다. 이중 가장 작은 값을 가진  $Distance_i$ 는 필터의 출력인  $x_i$ 에 대응된다.

\*  $Distance_i$  : 각 컬러 요소들에 대한 차

\*  $N$  : 필터 윈도우에 표현되는 샘플수

\*  $i$  : 처리되는 화소

\*  $j$  : 다른 화소 샘플

\*  $x_{med} = x_i$  (단,  $i$ 는  $Distance_i$  중 최소값에 대응하는 값)

$$\sum_{i=1}^N |x_{med} - x_i| \leq \sum_{i=1}^N |y - x_i| \tag{6}$$

$$Distance_i = \sum_{j=1}^N (|red_i - red_j| + |green_i - green_j| + |blue_i - blue_j|) \quad (7)$$

객체 내부점  $Obj(x_i, y_i)$ 의  $x, y$ 좌표 중 식 7을 통하여 각각의 최대, 최소 좌표를 구하여 객체를 포함하는 최소한의 사각 영역을 설정해 준다.

(4) 얼굴 영역 추출

HSI 색상계의 변환에서 가장 복잡한 부분은 색상(hue)의 계산 부분이다. 색상은 RGB의 혼합 비율을 빨강을 기준으로 하는 일반전각으로 표현한다. 얼굴 영역 추출은 식 8~10에 따라 HSI 좌표계로 변환된 영상의 색상정보에서 피부영역으로 알려진 범위 값에서 추출된 부분을 균집화하여 선택한다.

$$H = \cos^{-1} \left[ \frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \right] \quad (8)$$

$$S = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \quad (9)$$

$$I = \frac{1}{3}(R+G+B) \quad (10)$$

색상을 통한 추출 방법은 수행속도가 빠르다는 장점이 있지만 조명을 통해 영상의 변화가 많이 일어나며 명도에 대한 정규화 과정이 필요하다. 이에 HSI 색상좌표계는 색상과 채도 값만을 사용함으로 인해 조명의 영향을 감소시킬 수 있는 장점이 있다. HSI 모델 실험에서 피부색의 범위를 다음과 같이 사용하였다.

$$H_{\min} = 0^\circ, \quad H_{\max} = 50^\circ$$

$$S_{\min} = 0.23, \quad S_{\max} = 0.68$$

정규화된 RGB를 이용한 방법은 정규화한 값을 사용하여 HSI와 유사한 효과를 거뒀다. 이때  $r^2 + g^2 + b^2 = 1$ 이라는 성질을 이용하여  $r', b'$ 의 두 채널 정보만을 사용하여 색을 구분하는 기준으로 사용할 수 있다. RGB 좌표계에서 색상의 정규화 식은 식 (11)과 같다.

$$(r', g', b) = \left( \frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b} \right) \quad (11)$$

이 방법은 R, B의 값만을 이용하여 얼굴 영역 추출에 이용할 수 있다. 정규화된 RGB 모델 실험에서 피부

색의 범위를 다음과 같이 사용하였다.

$$r'_{\min} = 0.3550, \quad r'_{\max} = 0.4300$$

$$b'_{\min} = 0.2500, \quad b'_{\max} = 0.3050$$

이 방식은 어두운 영역에서 후보 추출에 장점이 있으나 유사 색상영역에서는 구별이 어렵다. 이에 본 논문에서는 두 가지 방법을 혼합한 hybrid 방법론을 이용하였다.

4. 실험평가

표 1은 제안한 시스템과 기존 방법들과의 성능 평가이다. FPS는 처리속도 평가기준으로 초당 처리되는 프레임의 수를 나타낸다. 추출율은 하나의 프레임에서 19,200(160×120 해상도) 픽셀 중 실제 객체영역의 픽셀 수에 대한 추출된 객체영역의 픽셀 수를 나타낸다. 또한 추적율은 이전 프레임의 추출된 객체 영역 픽셀의 다음 프레임에서 추적된 성능을 나타낸다. 성능 평가에 사용된 영상은 160×120 해상도의 24 bit 컬러영상을 사용하였다. 표 1에서 보이듯이 초당 처리하는 프레임 수와 객체 추출율, 추적율의 복합적인 면에서 기존의 방식들에 비해 향상됨을 알 수 있다.

Table 1. Performance Comparison for System

Method	FPS	Detection rate	Tracking rate
Low-level Analysis	15~18	91.1%	90.2%
Feature Analysis	11~16	84.3%	82.5%
Active Shape Models	12~15	97.5%	93.2%
Linear Subspace Methods	5~9	98.0%	89.2%
Neural Networks	6~9	96.3%	91.4%
Statistical Approaches	5~8	92.6%	95.7%
Proposed System	17~23	97.6%	98.3%

5. 결론

본 논문에서는 PC 카메라로부터 입력되는 영상으로부터 배경영상의 실시간 갱신을 통해 객체의 위치를 탐지하여 객체를 추적하고, 객체의 내부점을 이용한 MBR에서의 얼굴영역을 추출한 후 모바일

단말기로 모니터링 또는 검색할 수 있는 기법을 제안하였다. 실험은 고정 PC 카메라와 배경영상의 변화가 거의 없다는 제한된 환경 조건에서 실시간으로 배경영상의 갱신과 객체의 추적이 안정적임을 보여주었다. 이는 객체의 추출 및 추적 알고리즘이 빠르게 수행되어 객체 인식과 결합하여 객체의 움직임 정보와 인식을 통한 보안 및 감시 시스템 등 응용 분야에 적용될 수 있다는 것을 기대할 수 있도록 하여 준다. 하지만 입력영상과 배경영상 각각의 R, G, B 값의 차를 이용, 객체의 일부분임을 인식할 때, 잡음과 조명의 영향으로 인하여 실험에 실패한 경우도 발생하여 이에 대한 보완이 필요하다.

### 참고문헌

- [1] E. Hjelm and B.K. Low, "Face Detection: A Survey," Computer Vision and Image Understanding, Vol. 83 No. 3, pp. 236-274, 2001.
- [2] T.V. Pham, M. Worring, and A.W.M. Smeulders, "Face detection by aggregated Bayesian network classifiers," Technical Report 2001-04, Intelligent Sensory Information Systems Group, University of Amsterdam, 2001.
- [3] Augi, T. et al., "Contour Tracking and Synthesis in Image Sequences," SPIE '95, pp. 834-845, 1995.
- [4] K. Okada and C. von der Malsburg, "Analysis and synthesis of human faces with pose variations by a parametric piecewise linear subspace method," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 761-768, 2001.
- [5] M. Rogers and J. Graham, "Robust active shape model search," Proceedings of the 7th European Conference on Computer Vision, Vol. 4, pp. 517-530, 2002.
- [6] Kawato S. and Ohya H., "Automatic Skin-Color Distribution Extraction for Face Detection and Tracking," in Proceedings of ICSP, pp. 1415-1418, 2000.