지역적영역의 상대적 유사도를 이용한 실시간 얼굴추적

이재혁*, 신동화*, 김현정**, 원일용*
*호서전문학교 사이버해킹보안과
**건국대학교 컴퓨터공학과
e-mail : gmrxhlwk@naver.com, sin_d_h@daum.net, nygirl@konkuk.ac.kr, clccclcc@shoseo.ac.kr

Real-time Face Tracking using the Relative Similarity of Local Area

JeaHyuk Lee*, DongWha Shin*, HyunJung Kim**, ILYong Weon*
* Cyber Hacking Security Seoul Hoseo Technical College
** Dept, of Computer Science and Engineering Konkuk University

요 약

객체의 인식과 추적은 컴퓨터 비전 및 영상처리 분야에서 연구가 활발히 진행되고 있다. 특히 얼굴을 인식하고 추적하는 기술은 많은 분야에서 응용될 수 있다. 기존에 연구되어 온 기준 프레임과 관찰 프레임 사이의 차를 이용하여 객체를 인식하고 추적하는 방식은 관찰 대상이 다수인 경우 동일성을 확보하기에는 어려움이 많다. 따라서 본 논문에서는 각각의 프레임에서 빠르게 얼굴 영역을 인식하고, 독립적으로 인지된 얼굴들의 동일성을 연결하는 방법을 제시한다. 제안된 방법의 유용성은 실험으로 검증하였으며, 어느 정도 의미 있는 결과를 관찰할 수 있었다.

1. 서론

동영상에서 객체를 인식하고 추적하는 방법에 대한 연구는 꾸준하게 진행되어 왔다[1,2]. 이러한 연구 중 사람의 얼굴을 대상으로 인식하고 추적하는 연구는 생체인증 시스템, 출입보안 시스템, 보안 감시 시스템, 군사 시스템 등 많은 분야에서 응용되고 있다[1].

객체를 인식하고 추적하는 기존 연구들은 기준 프레임과 관찰 프레임과의 차 영상을 통해 객체의 위치를 인지하고 또 영상의 방향적 특성을 이용하여 다음 영상에서 앞 객체의 연속성을 연결하는 방법들을 사용한다[2]. 이러한 방법은 제한된 환경에서는 어느 정도 성과를 내기도 하지만, 추적하는 대상이 같은 유형의 다수인 경우 어려움이 있다[3].

서로 다른 프레임에서 동일인을 연속적으로 인식하는 방법에는 전통적 얼굴 인식 시스템 방법이 있다[4]. 전통적인 방법은 인식하고자 하는 대상을 미리 학습하고, 각각의 프레임에서 인식되는 얼굴을 판단하는 방법이다. 하지만 이러한 방법을 실시간 동영상에서 적용하면 처리 속도의 문제가 있고, 또 불특정 다수를 대상으로 하는 경우에는 학습이 불가능하다는 제약이 발생한다[5].

본 논문에서는 특정 범위의 장소에서 획득한 동영상들을 대상으로 시작 프레임부터 연속적으로 이어지는 프레임에서 각각 사람의 얼굴을 인지하고, 인지한 동일인을 연속적으로 추적하는 방법이다. 따라서 본 논문에서 제안하는 방법은 각각의 프레임에서 빠르게 얼굴 영역을 인식하기 위해 Haar-like Adaboost 를 사용하고, 각각의 프레임에서 독립적으로 인지된 얼굴들의 동일성을 연결하기 위해 인식된 얼굴들 사이의 SAD(Sum of Absolute Difference)를 계산하는 방법이다. 제안된 방법의 유용성은 실험으로 검증했다.

본 논문은 다음과 같이 구성되어 있다. 2 장에서는 관련 연구를 언급하고, 3 장에서는 제안하는 방법을 설명했다. 4 장에서는 실험 및 분석, 5 장에서는 결론 및 향후 과제를 언급했다.

2. 관련 연구

2.1 Haar-like Adaboost 를 이용한 얼굴인식

Haar-like Feature 는 현재 영상에서 얼굴 인식을 수행할 때 가장 많이 사용되고 있으며, 사각형의 집합으로 구성된 특징(feature) 이미지의 집합으로써 이미지 영역 누적 테이블을 이용해 매우 빠른 시간에 특징(Feature)을 구할 수 있다는 점에서 널리 사용되고 있다. 하지만 기초적인 간단한 사각형들로만 구성된 feature 만으로는 복잡한 이미지를 인식하기 어려우므로 단순한 Haar-like Feature 의 집합을 각각의 가중치와 함께 합산하여 복잡한 이미지에 대한 인식기를 구성한다[6].

Fig 1. Haar-like Feature Prototype Set

Harr-like 와 함께 사용되는 Adaboost 는 얼굴인식을 위한 Haar-like Feature 를 이용하기 위해 가장 널리 쓰이고 있는 알고리즘이다. 분류 규칙을 차례대로 생성하고, 이전의 분류 규칙을 적용하여 얻은 관측값에서 표본 데이터의 분포를 재조정한다. 단계를 반복하여 나온 여러 개의 약한 분류기(week classifier)와 가중치 값들의 조합으로 강한 분류기(strong classifier)를 생성하는 알고리즘이다[7].

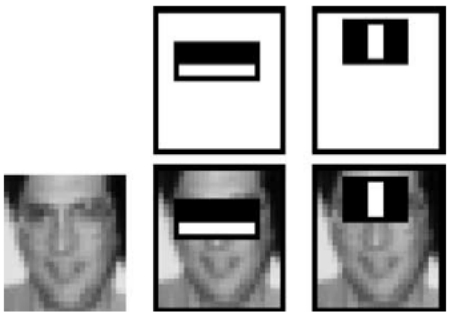

Fig 2. Week Classifier Example [11]

실제로 여러 분류기와 학습기를 조합하여 하나의 메타 학습기를 만드는 기법으로 하나의 학습방법에 따라 발생하는 약점을 극복할 수 있다. 하지만 여러 학습 및 분류기를 사용함으로써 시간적인 오버헤드가 크기 때문에, Adaboost 는 단순하지만 빠른 소위 약한 분류기(weak classifier)를 조합하여 하나의 분류기(final classifier)를 생성할 때 주로 사용된다[8, 9].

2.2 SAD 정합 알고리즘

절대값 차이의 합(Sum of Absolute Difference 이하 SAD) 연산은 이미지 사이의 유사성을 측정하기 위해 사용되는 알고리즘이다. 비디오 압축 등의 분야에서 널리 사용되고 있으며, 대표적인 스트림 데이터 처리 방법이다.

$$SAD(x, y, \partial) = \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} | SI(x, y) - RI((x + \partial), y) |$$

H = 윈도의 열 길이
W= 윈도의 행 길이
∂ = 윈도의 차이 값
SI = Standard Image
TI = Target Image

최적의 정합점을 계산하기 위하여 윈도의 차이 값 ∂가 0 일 때부터 최대계산 값 Δ 까지 변화시키면서 (Δ+1) 개의 SAD(x, y, ∂) 값 중 최솟값을 갖는 ∂을 최적 정합점(Best matching point)으로 선택한다. Δ 가 지나치게 크면 유사하게 생긴 물체가 범위 내에 존재할 경우 계산에 오류가 발생할 가능성이 증가 된다. 또한 계산해야 하는 값의 수가 증가하므로 계산량이 증가하게 된다. 반면, Δ 가 적게 설정하면 기준 시야에 접근하는 물체를 인식하지 못하는 오류가 발생하게 된다.

따라서 Δ 는 계산량 및 인식할 장면의 특성을 고려하여 적절히 줄여주되, 지나치게 작아져서 오류가 생기지 않도록 해야 한다[10].

3. 시스템 설계

제안하는 시스템은 크게 2 단계로 구성되어 있다. 각각의 프레임에서 얼굴 영역을 인지하는 단계와 인지된 프레임 사이의 얼굴의 동일성을 연결하는 단계로 나누어진다. 얼굴 영역 인식을 위해 Haar-like Adaboost 방법을 사용한다. 정확도를 높이기 위하여 얼굴 후보 영역에 대한 색채 비율을 이용하여 정확도를 높인다.

각각의 프레임에서 인식된 얼굴들에 대하여 연속성을 연결하기 위해 우리는 각각의 프레임의 얼굴 영역들의 유사도를 계산하여 사용한다. 정확도를 높이기 위하여 전 단계에 인식된 얼굴들의 위치를 기준으로 특정 범위 내에서 유사도를 계산한다.
두 얼굴의 유사도를 계산 하는 식은 아래와 같다.

(1) Calculation of Similarity

$$Similarity\ (face1, face2) = \frac{1}{SAD(face1, face2)}$$

앞 프레임에서 인지된 얼굴이 다음 프레임에서의 연속 얼굴로 인지하는 방식은 각각의 얼굴에 대한 기준 얼굴과의 상대적 유사도를 계산하여 가장 높은 얼굴을 기준 얼굴의 연속된 얼굴로 인식한다. 상대적 유사도를 계산하는 식을 아래와 같다.

(2) Calculation of RS

$$RS = \frac{Sim(source\ face, target\ face(i))}{\sum Sim(source\ face, target\ face(i))}$$

Sim = Similarity
RS=Relative similarity

Fig 3 는 인지된 얼굴에 대한 다음 프레임에서의 연속 얼굴로 인지하는 방식을 알고리즘으로 간략하게 표현했다.

1. Obtain a subset f_n of the human face regions from an camera input frame.
2. Label f_n from step 1.
3. obtain another subset B_n of the face regions from next input frame.
4. Measure individual similarity values between f_n and B_n.
5. Label B_n with the highest similarity values to f_n ,as the same face.

Fig 3. Face Tracking Algorithm

4. 실험 및 결과

4.1 실험환경 및 데이터

제안한 시스템은 C++을 사용하여 Windows 에서 구현되었으며, 이미지 처리를 위해 Open CV 2.45 를 사용했다.

실험을 위해 데이터는 고정된 장소를 지정하고 카메라는 CAM(640 * 480 /초당 15 프레임)을 사용했다. 영상은 Table 1 에서와 같이 예측 가능한 움직임을 보이는 영상 27 개를 사용했으며, 10~20 초 분량으로 구성되어 있다. 각 영상에는 다양한 인원의 얼굴들이 포함되어 있다.

Table 1. Sample of Video

분류	1명	2명	3명	4명	합계
영상 수	13	8	3	3	27

4.2 결과 및 분석

Fig 4 는 3 명의 얼굴이 포함된 영상에서 얼굴 영역이 인식된 프레임을 표시하였다. face 1 을 중심으로 다음 프레임에서 인지된 얼굴과의 RS (Relative similarity)값을 계산하여 얼굴의 동일성을 추적하는 실험 결과의 일부를 표시하였다.

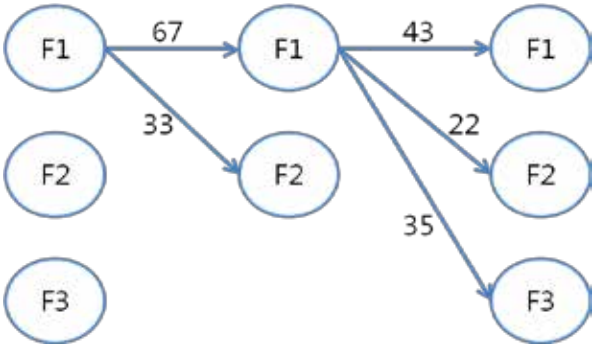

Fig 4. Relative Similarity of face 1

얼굴 영역을 인지한 프레임과 다음 프레임별 RS 값을 계산하여 얼굴의 연속성을 연결할 때, Threshold 값에 따라 실제 성능을 달라질 수 있다.

Fig 5 은 Threshold 값에 따른 실험결과를 그래프로 표시한 것이다.

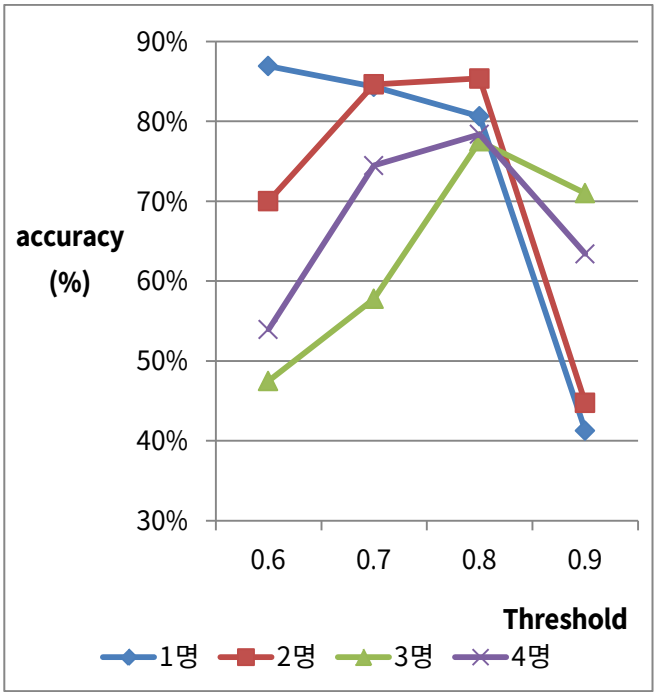

Fig 5. Calculation the Accuracy

실험 결과, Threshold 값이 증가할수록 정확도 역시 일반적으로 증가하였다. 그러나 어느 기준값 이상으

로 넘어서면 정확도가 줄어드는 것을 볼 수가 있었다. 이러한 이유는 프레임에서 인지된 얼굴에 대한 다음 프레임에서의 연속된 얼굴로 인지하기까지 많은 프레임이 소모되기 때문으로 추정된다.

이 실험에서 얻을 수 있는 Threshold 값의 최적 값은 0.8 이다. 동영상에 포함된 사람의 인원수가 많아질수록 정확도는 조금씩 감소했지만, Threshold 값에서 큰 차이는 보이지 않았다. 기준값에서 전체적인 에러율은 19% 정도인데, 앞선 프레임에서는 인식되었던 얼굴이 이어지는 프레임에서는 인지되지 않는 경우가 있었다. 이 경우는 오류가 자주 발생하였다. 또한 두 프레임에서 측정하여 비교하는 영역의 크기가 일정 값 이하로 감소하면 에러율이 급격하게 증가되고, Threshold 값이 0.9 이상일 때, 부정오류(False Negative)가 급격히 증가하는 현상이 관찰되었다. 이는 SAD 가 단순하게 영역의 값만을 1:1 로 고려하기 때문으로 추정된다. 따라서 영상이 일정 범위 이상 확대되거나 축소되면 SAD 는 정확도 계산에 어려움이 있는 것으로 예상된다.

5. 결론 및 향후 과제

본 연구에서 우리는 다수의 얼굴이 포함된 동영상의 각 프레임의 얼굴을 인식하고, 프레임별 얼굴의 동일성을 연결하여 추적하는 방법을 제안하였다.
얼굴 영역 인식은 Haar-like Adaboost 을 이용하였으며, 얼굴의 동일성을 연결하기 위해서 프레임별 얼굴 영역의 유사도를 상대적으로 계산해 결정하였다.

실험을 통해 제안한 시스템의 유용성을 검증하였는데, 어느 정도 의미 있는 결과를 관찰할 수 있었다.

추후 연구과제는 아래와 같다. 첫째, 프레임별 인식된 얼굴 영역의 유사도를 계산하는 방식을 개선하여 동일 얼굴의 추적 정확도를 높이는 것이 필요하다. 둘째, 제안한 시스템을 실시간 환경에서 사용하기 위해 계산을 빠르게 수행할 수 있는 연구도 필요하다

참고문헌

[1] F.Zhou, H.B.Duh and M. Billinghurst, “Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR,”IEEE International Sym-posium on Mixed and Augmented Reality,pp. 193-202, 2008

[2] A. Bakowski, G.A. Jones, “Video Sureillance Tracking using Color region Adjacency Graphs”, Image Processing and its Applications, Conference Publication No.465, 1999, pp.794-798

[3] Seung-Young Lee, Tae-Rim Jung, Chang-Wu Hur, Kwang-Ryol Ryu, A Study on the Position Tracking of Moving Image for Surveillance System, KIMICS Integrated Conference ,Fall 2006, pp. 205-208

[4] Peter J. B. Hancock, A. Mike Burton, and Vicki Bruce. “Face Processing: Human perception and principal components analysis” Memory and Cogrition. Volume: 24, Issue: 1, pp.26-40, 1996

[5] A.Pentland, B.Moghaddam, T.Starner, and M. Turk, “View-Based and Modular Eigenfaces for face Recognition”, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 84-91, 1994

[6] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. CVPR, pp.511–518, 2001

[7] Freund, Yoav, Robert Schapire, and N. Abe. "A shortintroduction to boosting", Journal-Japanese Society For Artificial Intelligence, pp.771-780, 1999

[8] Robert. E. Schapire and Yoram Singer, “ Improved boosting algorithms using confidencerated predictions,” Machine Learning, Vol. 37, no.3, pp. 297-336, 1999

[9] Y. Freund, R. E. Schapire, “A Short Introduction to Boosting”, Journal of Japanese Society for Artificial Intelligence, Vol. 14, no. 5, pp. 771-780, 1999

[10] Jongsu Yi, SeungGu Yang, JunSeong Kim, Development of a SAD Correlater for Real-time Stereo Vision, IEEK Vol.45. No.1. 2008. 1, pp.55-61

[11] Viola and MJ Jhones, 2001. 2. 1, “Robust real-time object detection", Compaq Cambridge research Laboratory