

웨이블릿 변환 기반의 실시간 얼굴 검출 및 추적 알고리즘

송해진⁰, 고병철, 변혜란
연세대학교 컴퓨터과학과
(haejins⁰, soccer1, hrbyun)⁰@cs.yonsei.ac.kr

Real-time Wavelet transform-based Face Detection and Tracking

Haejin Song⁰, Byoungchul Ko, Hyeran Byun
Dept. of Computer Science, Yonsei University

요 약

본 논문은 실시간 카메라 입력 환경에서의 새로운 얼굴 검출 및 추적 알고리즘을 제안한다. 복잡한 배경과 다양한 조명 조건에 관계 없이 얼굴을 검출하고 추적하기 위해 세 종류의 웨이블릿 변환된 형판을 사용하고 특히 다양한 조명 조건을 극복하기 위해 최소-최대 정규화(Min-Max Normalization)와 히스토그램 평활화를 혼합 적용하여 매우 밝거나, 매우 어두운 영상에서의 얼굴 오 검출 및 놓침을 줄이도록 하였다. 또한 세가지 크기의 얼굴 형판을 이용함으로써 입력 영상에 존재하는 다양한 크기의 얼굴도 검출할 수 있었으며, 효과적인 얼굴 추적 기법을 통해 다음 프레임에서의 얼굴 위치를 예측하여 그 지점에서의 탐색 영역에 형판 정합을 수행함으로써 수행 시간도 단축시킬 수 있었다. 실험을 위해 다양한 조명 조건에 따라 여섯 종류로 분류한 동영상 데이터에서 제안한 알고리즘은 약 96.8%의 뛰어난 얼굴 검출율을 보여주었다.

1. 서론

비디오에서의 자동적인 얼굴 검출과 얼굴 추적 기술은 화상 회의, 보안 통제 시스템, 내용기반 비디오 색인 등과 같은 많은 상업적인 분야에서 중요한 연구 과제가 되고 있다. 그 결과 지난 10여년간 얼굴 검출과 추적에 대한 많은 연구들이 진행되어 왔다.

그러나 비디오 상에 존재하는 다양한 얼굴의 크기와 그 위치, 카메라 각도에 의한 얼굴의 회전, 그리고 얼굴의 표정과 얼굴의 가리워짐 및 조명 조건 등의 영향으로 비디오 혹은 정지 영상에서 얼굴이 존재하는지 여부와 그 정확한 위치를 찾아내는 것은 결코 쉬운 일이 아니다[1].

일반적으로 얼굴 검출 알고리즘은 몇 개의 카테고리로 분류된다. 첫번째로, 지식기반 방법[2]은 인간의 기본 얼굴에 대한 연구자들의 지식으로부터 추론된 규칙들에 기반한 방법으로 인간의 얼굴에 대한 지식을 정확하게 정의된 규칙으로 적용하기 쉽지 않다는 단점이 있다. 두번째, 특징기반 방법[3]은 얼굴 검출을 위해 얼굴의 불변하는 특징들을 이용한다. 여러가지 특징들 중 피부 색은 얼굴의 이동, 회전, 크기 변화 등에 덜 민감한 특성을 가지기 때문에 최근 가장 많이 사용되고 있다. 그러나 피부 색 모델은 조명 변화에 대해 매우 비효율적이기 때문에 피부 색과 모양, 추적을 위해[4] 움직임 정보 등을 결합하는 방식이 시도되고 있다. 세번째로 형판(template) 기반 방법[5]은 얼굴에 대한 몇 가지 표준 패턴을 만들어 영상의 탐색 윈도우 안에서 영상과 비교해 보는 것으로 단순하다는 장점이 있지만 얼굴의 회전이나 크기 변화, 다양한 빛의 변화 및 잡음에 민감한 단점을 가지고 있다. 마지막으로 신경망(neural network) 기반 방법들은[6][7] 다양한 영상들로부터 얼굴 영역과 비 얼굴 영역을 획득하여 신경망을 통해 얼굴과 비 얼굴로 학습한 뒤 입력 영상으로부터

얼굴을 찾으려 하는 방법을 사용한다. 신경망을 이용한 방법들은 정면 및 측면 얼굴 검출은 비교적 만족할만하나 계산량이 많고 다양한 회전 얼굴 검출에서는 좋은 결과를 주지 못하고 있다[1,8].

본 연구에서는 실시간 카메라 입력 영상에서 형판 정합(template matching) 방법을 든 새로운 얼굴 검출 및 추적 기법을 제안한다. 형판 정합과 관련된 여러 가지 문제점들을 해결하기 위하여 본 연구에서는 웨이블릿 변환된 얼굴 영상들로부터 세 가지 형태의 평균 형판을 만들어 웨이블릿 변환된 입력 영상과 정합하는 방법을 사용하였다.

2. 얼굴 검출

2.1. 전처리

일반적으로 화상회의나 비디오 폰 관련 연구들에서는 정적인 카메라 움직임과 고정된 배경을 가정하고 있으나 실생활에서 얻어지는 동영상은 전경과 배경에서 다양한 색상 및 밝기 변화를 포함하고 있다. 이것은 얼굴의 오 검출 및 놓침(missing) 원인이 되기 때문에 정확한 얼굴 검출을 위해 조명 보정은 필수 요소가 된다. 본 논문에서는 실생활에서 획득한 동영상들을 빛의 위치와 입력 환경에 따라 6개의 카테고리로 분류하고 조명 보정을 위해 식(1)의 최대-최소 정규화(Min-Max Normalization)를 적용하였다. 최대-최소 정규화는 입력 영상의 밝기 범위를 새로운 범위로 매핑 시키는 선형 변형이다.

$$y' = \left(\frac{y - \min_1}{\max_1 - \min_1} \right) (\max_2 - \min_2) + \min_2 \quad (1)$$

식(1)에서 \min_1 과 \max_1 은 입력 영상의 최소, 최대 밝기 값이며,

min_2 와 max_2 는 새로운 범위의 최소, 최대 밝기 값이다. 최소-최대 정규화를 위해 실험적으로 밝기 성분의 상위 8.5%를 참조 백색(reference white)으로, 하위 8.5%를 참조 흑색(reference black)으로 정의하여 새로운 밝기 범위를 정하여 주었다. 그러나 입력 영상이 매우 밝거나 어두운 경우에는 최대-최소 정규화를 적용하기 전에 히스토그램 평활화를 적용하여 영상의 밝기를 조절하여 주었는데, 이러한 혼합된 조명 보정 방법에 의해 매우 밝은 영상이나 매우 어두운 영상, 다양한 조명 위치에 의해 왜곡된 영상에서의 얼굴 오 검출이나 얼굴 놓침 현상을 줄일 수 있었다.

2.2. 피부 색을 이용한 형판 정합 후보 영역 설정

많은 연구들이 피부 색 성분과 밝기 성분은 서로 독립적이라는 가정을 세우고 있지만, Jain의 연구[9]에서는 실제 피부 색은 비선형적으로 밝기 성분에 의존적이라고 제안하였다. 따라서 입력 영상의 YCbCr 컬러 모델을 Jain이 제안한 식에 의해 밝기 성분에 독립적인 컬러 모델로 변환한 뒤 피부 색을 이용하여 형판 정합 횟수를 줄이기 위한 형판 정합 후보 영역을 설정한다.

2.3. 형판 정합을 이용한 얼굴 검출

본 연구는 웨이블릿 변환된 예제 얼굴들의 평균 영상으로 얼굴 형판을 생성하였는데 63명으로부터 3장씩, 189개의 얼굴을 이용하였으며 그림 1-(1)(2)와 같이 위로는 눈썹으로부터 아래로 윗 입술까지만 포함하는 사각형 얼굴 영역을 대상으로 하였다. 다양한 크기의 얼굴 검출을 수행하기 위해 40x40, 60x60, 80x80의 세 종류의 크기로 형판을 생성하였는데, 이것은 입력 영상의 크기를 QCIF (176x144)로 설정하여 최소 40x40 이상부터 최대 80x80 이하의 얼굴을 가정하였기 때문이다. 입력 영상에 웨이블릿 변환을 수행하면 4가지 타입의 웨이블릿 계수를 얻을 수 있고(그림 1-(3)-(7)) 변환된 얼굴영역은 각각 10x10, 15x15, 20x20 으로 매우 낮은 해상도의 형판을 구할 수 있게 된다.

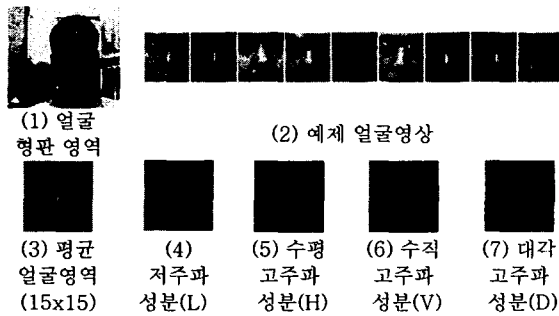


그림 1. 웨이블릿 계수를 이용한 형판 생성

얼굴 영역은 수평 고주파 성분(H)과 수직 고주파 성분(V)이 강하게 나타나는 특성을 가지기 때문에 수평, 수직 고주파 성분은 얼굴 영역 검출에 중요한 역할을 하는 반면 대각선 고주파 성분(D)은 두드러진 특성을 갖지 못하기 때문에 오직 저주파 성분(L)과 수평 고주파 성분(H), 수직 고주파 성분(V)의 계수만을 얼굴 형판으로 사용하였다. 따라서 각 얼굴 크기별 세 종류로 총 9개의 얼굴 형판이 얼굴 형판 정합에 이용된다.

입력 프레임에서 얼굴 영역을 검출하기 위해 웨이블릿 변환된 입력 프레임의 세 가지 성분 계수와 생성된 얼굴 영역 형판 사이의 거리를 측정하여 MAE(Mean Absolute Error)가 가장 작은 지점을 얼굴 영역으로 검출하게 되는데, 처음 세 프레임까지는 입력 영상 전체 영역 중 피부 색을 이용하여 형판 정합

의 후보 영역으로 선택된 지점들을, 이후 네 번째 프레임 부터는 이전 프레임들의 얼굴 위치를 이용해 추정된 다음 프레임에서의 얼굴 위치(m,n)에 대한 (m-3, n-3) 부터 (m+3, n+3) 까지의 탐색 영역 중 피부 색을 이용하여 형판 정합의 후보 영역으로 선택된 지점들을 대상으로 형판 정합을 수행하게 된다.

2.4. 회전된 얼굴 검출

일반적으로 정면 얼굴 영상들을 이용해서 만들어진 정면 형판만으로도 약 -10° ~ +10° 정도 회전된 얼굴은 검출이 가능하다[10]. 그러나 그 이상 회전된 얼굴의 경우엔 정면 형판만으로는 정확한 검출은 불가능하다. 따라서 정면 얼굴 검출 뿐 아니라 자연스럽게 회전된 얼굴을 검출하기 위해 웨이블릿 변환된 낮은 해상도의 입력 영상을 -20°, +20° 로 2번 회전시켜 정면 형판과 정합한 후 MAE가 가장 작은 회전 각도와 얼굴 크기를 조합하여 최종 얼굴 영역으로 검출하여 줌으로써 최대 -30° ~ +30° 까지 회전된 얼굴에 대해서도 정확하게 검출할 수 있도록 하였다.

3. 얼굴 추적

첫번째 프레임에서 찾아진 얼굴을 이어지는 프레임 들에서 추적하는 것은 얼굴 검출 시간을 줄이고 시스템으로 하여금 빠르게 움직이는 얼굴도 추적할 수 있도록 해 주어 실시간 비디오 응용 분야에서 필수 부분이 되고 있다. 대표적 얼굴 추적 알고리즘인 Xu의 알고리즘[11]은 다음 프레임에서의 얼굴 위치를 빠르고 정확하게 예측할 수는 있으나, 얼굴의 움직임은 속도가 일정하지 않으면 잘못된 예측 결과를 받게 된다. 따라서 본 논문에서는 Xu의 알고리즘을 수정하여 일정치 않은 속도로 움직이는 얼굴에 대해서도 다음 프레임에서의 정확한 얼굴 위치를 예측할 수 있는 방법을 제안한다. 시간 t 에서의 얼굴 위치를 $x(t) = [x(t), y(t)]$ 라고 가정하면 시간 (t+1)에서의 얼굴 위치는 이전 프레임 t-2, t-1 사이에서의 평균적인 얼굴 움직임 속도를 이용하여 다음의 식과 같이 예측할 수 있다.

$$v(t-1) = x(t-1) - x(t-2) \tag{2}$$

$$v(t) = x(t) - x(t-1) \tag{3}$$

$$\bar{x}(t+1) = x(t) + \left[\frac{v(t-1) + v(t)}{2} \right] \tag{4}$$

예를 들어, 1차원만 고려하여 t-2에서의 얼굴 위치가 2이고, t-1에서 3, 그리고 t에서는 4 이라면 Xu의 알고리즘의 경우 다음 프레임 t+1에서의 얼굴 위치는 $(\bar{x}(t+1) = 3x(t) - 3x(t-1) + x(t-2))$ 로 5이고 제안한 방법의 경우도 5로 같다. 그러나 움직이는 속도가 일정하지 않은 경우, 예를 들어 t-2에서 4이고, t-1에서는 2, t에서는 5인 경우 t+1에서의 얼굴 위치는 Xu의 알고리즘의 경우 13이 되지만 제안한 방법의 경우는 5가 된다. 따라서 예측된 얼굴 위치 5를 중심으로 탐색 영역을 설정하여 실제로 다음 프레임에서의 얼굴 영역을 검출하게 된다.

4. 실험 결과

본 알고리즘은 펜티엄 III 866 MHz PC에서 QCIF (176x144) 입력 영상을 대상으로 구현 및 실험 하였고 초당 6프레임 정도의 검출 속도를 보여주었다. 제안한 알고리즘의 성능을 실험하기 위하여 조명 조건 및 배경에 따라 6개의 카테고리를 정하여 예제 동영상을 만들었으며 만들어진 6개 카테고리의 동

영상은 홈페이지(<http://vip.yonsei.ac.kr/facetracking/>)에서 확인할 수 있다. 6개 카테고리의 동영상에 대해 실험한 결과 제안한 방법의 평균 얼굴 검출율은 약 96.8%로 측정되었고 조명 보정을 위한 전처리를 적용하지 않은 경우 시스템의 평균 얼굴 검출율은 약 58.7%로 측정 되었다. 전처리 방법의 성능 평가를 위해 히스토그램 평활화, 최소-최대 정규화, Contrast Stretching(cut-off 5%), Zero-Mean 정규화, Gamma Correction($\gamma = 1.5$) 방법을 각각 적용하여 실험하여 보았으나 그 결과는 각각 94.0%, 88.0%, 61.9%, 78.0%, 63.7%로, 제안한 최소-최대 정규화 방법의 뛰어난 성능을 확인할 수 있었다. 그림 2는 6개 카테고리 동영상에 대한 실험 결과를 보여준다.

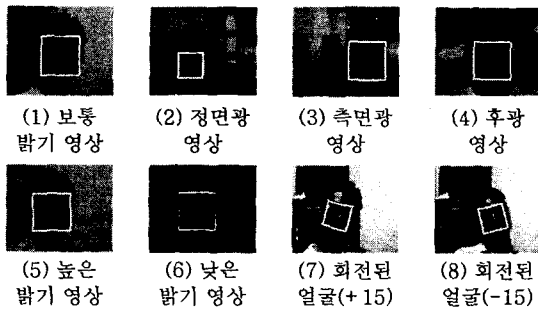


그림 2. 6개 동영상 카테고리 및 얼굴 검출 결과

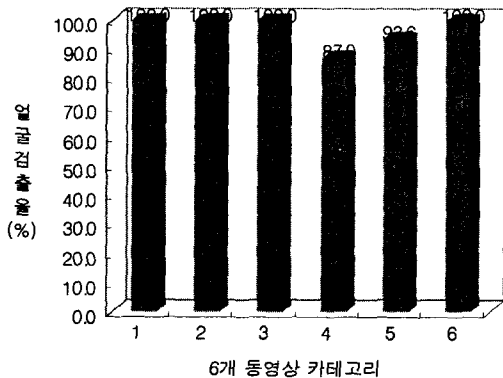


그림 3. 제안한 알고리즘의 얼굴 검출율(%)

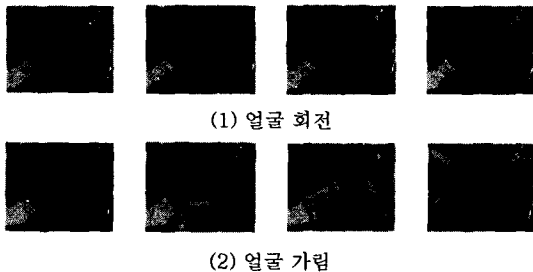


그림 4. 다양한 조건 하에서 실험 결과

본 논문이 제안한 알고리즘은 다양한 조명 조건에 대해 뛰어난 얼굴 검출 성능을 보여주었다. 최소-최대 정규화 방법은 다른 전처리 방법에 비해 뛰어난 조명 보정 효과를 확인시켜 주었으며, 제안한 얼굴 추적 기법 또한 빠르게 움직이는 얼굴에 대해서도 정확한 예측 결과를 보여주었다.

본 연구는 회전된 얼굴 검출만을 고려하였으나, 향후 측면 얼굴을 비롯하여 비교적 상하로의 회전이 큰 얼굴 검출에 대한 연구가 수행되어야 할 것이다. 또한 다양한 조명 조건을 갖고 비교적 작은 얼굴 영역에서의 눈과 같은 얼굴 요소 검출 연구를 수행하여 정확한 얼굴 영역 검출을 검증하고 얼굴 영역의 타원으로서의 검출도 가능해야 할 것이다.

6. 참고문헌

- [1] Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja, " Detecting Faces in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, Jan, 2002.
- [2] C. Kotropoulos, and I. Pitas, " Rule-based detection in frontal views," *International Conference on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2537-2540, 1997.
- [3] S.A. Sirohey, " Human face segmentation and identification," *Technical Report CS-TR-3176, University of Maryland*, 1993.
- [4] H. P. Graf, E. Cosatto, D. Gibbon, M.Kocheisen, and E. Petajan, " Mutimodal system for locating heads and faces," *The Second International Conference on Automatic Face and Gesture Recognition*, pp. 88-93, 1996.
- [5] V. Govindaraju, S.N. Srihari, and D. B. Sher, " A computational model for face location," *The third IEEE International conference on Computer Vision*, pp. 718-721, 1990.
- [6] H. A.Rowley, S. Baluja, and T. Kanade, " Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 22-38, Jan. 1998.
- [7] K. K. Sung and T. Poggio, " Example-based learning for view-based human face detection," *Technical Report A.I. Memo 1521, CBLIC paper 112, MIT*, Dec. 1994.
- [8] Kin C. Yow, and Roberto Cipolla, " Feature-Based Human Face Detection," *Second International Conference on Automatic Face and Gesture Recognition*, 1996.
- [9] Hsu, R.H., Mohamed, A.M. and Jain, A.K. " Face Detection in color images" , *IEEE Trans. on PAMI*, May, 2002, Vol. 24, pp.696-706.
- [10] Zhou, J, Lu, X. G., Zhang D., and Wu, C. " Orientation analysis for rotated human face detection" , *Image and Vision Computing*, 2002, (20), pp. 257-264.
- [11] Xu, G., and Sugimoto, T. " Rits Eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera" , *Int. Conf. on Pattern Recognition*, 1998, Vol. 2, pp. 1194-1197.

5. 결론 및 향후 연구 과제