

역전파 신경망을 이용한 동영상에서의 얼굴 검출 및 트래킹

지승환, 김용주, 김정환, 박민용

연세대학교 전자공학과

E-mail : mignpark@bubble.yonsei.ac.kr

Face Detection and Tracking in Sequential Images using Backpropagation

Abstract

In this paper, we propose the new face detection and tracking algorithm in sequential images, which have complex background. In order to apply face detection algorithm efficiently, we convert the conventional RGB coordinates into CIE coordinates and make the input images insensitive to luminance. And human face shapes and colors are learned using neural network's backpropagation. For variable face size, we make mosaic size of input images vary and get the face location with various size through neural network. Besides, in sequential images, we suggest face motion tracking algorithm through image subtraction processing and thresholding. At this time, for accurate face tracking, we use the face location of previous image. Finally, we verify the real-time applicability of the proposed algorithm by the simple simulation.

1. 서론

영상 전화나 서비스 로봇, 얼굴 인식 시스템 등 휴먼 인터페이스 기술 분야에서 사람 얼굴의 특징점 검출은 가장 중요하고도 많이 다루어지고 있는 부분이다. 특히 장애자에게 음식을 제공하는 로봇이나[3], 비밀 유지를 위한 얼굴의 자동 인식 시스템[1,2] 등 휴먼 기반의 컴퓨터 비전 시스템은 수동이든 자동이든 간에 사람의 입이나 눈 등 얼굴의 특징점을 검출하지 않으면 안된다. 이러한 얼굴의 특징점 추출은 얼굴 영역의 탐색 과정 없이 직접 특징점을 추출하는 경우도 있으나 대개 얼굴의 전체 영역의 검출 과정이 선행되어야만 한다.

최근 몇년 동안 영상에서 얼굴 영역을 검출하는 연구가 활발히 진행되었다. 특히 KL 변환을 이용한 고유 얼굴(Eigenface)을 이용하는 방법이나[1] 변형 가능한 템플

리트를 사용한 방법[4,8] 등에서 좋은 성능을 나타내었다. 또 퍼지나 뉴럴 네트워크와 같은 여러 학습 알고리즘에서도 좋은 향상을 나타내었다[2]. 또 영상 처리 이론이나 제어 이론에서 사용되어지고 있는 여러 알고리즘들도 얼굴 탐색을 위해 다양한 방법의 접근이 시도되고 있다.

그러나 복잡한 실제 배경에서 얼굴을 검출하기는 그리 쉬운 과정은 아니다. 대개 배경에 제한을 가하거나 목표 얼굴에 제한을 가하여 얼굴을 검출한다. 또 복잡한 이론과 과도한 계산량으로 실시간으로 처리하기에는 다소 무리가 있다. 또 얼굴에 비추어지는 조명에 따라서 매우 민감한 결과를 나타낸다.

본 논문에서는 정지 영상에서 얼굴을 검출한 다음, 연속적으로 변화하는 실시간 영상에서 자동적으로 얼굴을 트래킹하는 알고리즘을 제안한다. 실시간 처리를 목적으로

로 간단하고도 빠른 뉴런 네트워크의 역전파 학습 알고리즘을 사용하여 얼굴을 찾아내는 방법을 제안한다. 제 2장에서는 동영상 처리를 위한 비전 시스템에 대해서 간단히 설명하고, 3장에서는 노이즈와 빛의 양에 무감한 입력 영상을 위한 전처리 과정에 대해서 설명하고, 제 4장에서는 전처리 과정을 거친 입력 영상에 뉴런 네트워크 알고리즘을 적용, 정지 영상에서 얼굴을 검출하는 기법에 대해 설명한다. 5장에서는 연속된 동영상에서 얼굴을 트래킹하는 방법에 대해서 기술하고, 마지막으로 6장에서 결론을 내리기로 한다.

2. 비전 시스템

논문 실험을 위한 비전 시스템은 320×240 컬러 영상을 제공해주는 컬러 CCD 카메라와 카메라로부터 출력된 영상을 획득할 수 있는 영상 캡처 카드, 그리고 영상의 출력 및 얼굴 검출, 트래킹을 위한 소프트웨어 프로그램으로 되어 있다. 실험의 목적은 CCD 카메라로부터 얻은 정지 영상으로부터 얼굴 영역을 검출하고 그 다음 연속적으로 이어지는 동영상에서 그 얼굴 영역을 계속적으로 트래킹하려는 것이다.

3. 전처리 과정

3-1 입력 영상의 모자이크 패턴 처리

본래의 비전 시스템으로부터 얻어지는 컬러 영상은 320×240이므로 검출을 찾기위해 탐색해야 할 영역이 너무 많다. 따라서 이 영상을 다운사이징(down-sizing)할 필요가 있다. 이 다운사이징 과정은 입력 정보의 많은 상소를 가지오지만 그만큼 탐색 공간도 줄어드는 효과를 가져올 뿐만 아니라 카메라로부터의 영상 획득사 수반하지 모르는 노이즈에 매우 강한 장점을 나타낸다. 이 과정은 5×5 크기의 kernel을 이용하여 그 평균값을 모자이크 패턴의 값으로 대체하여 이루어진다. 이 경우 경우 입력은 64×68로 다운스케일링되어 탐색 영역을 줄일 수가 있다. 그림 1, 2는 각각 320×240 입력 영상과 모자이크화된 64×68의 모자이크 영상을 보여준다.



그림 1. 입력 영상



그림 2. 모자이크 영상

3-2 CIE 좌표계로의 변환

RGB 좌표계로 되어 있는 입력 영상을 다음과 같은 CIE 좌표계로 변환한다[5].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5943 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

이 CIE 좌표계는 영상의 Chromaticity와 luminance의 두 성분으로 나누어질 수가 있다. 이 중 luminance 성분은 조명에 매우 민감하기 때문에 이 성분을 제거하면 조명의 영향에 둔감한 결과를 얻을 수가 있다. 이 과정은 다음과 같은 과정을 거치면 된다.

$$D = X + Y + Z$$

$$x = \frac{X}{D} \quad y = \frac{Y}{D} \quad z = \frac{Z}{D}$$

여기서 $x + y + z = 1$ 이므로 x, y 두 값만 가지고 표현할 수 있다. 모자이크화된 각 픽셀의 RGB 값은 위와 같은 CIE Chromaticity 값 x, y 로 변환되고 따라서 각 픽셀의 처리는 이 두값으로 처리하면 된다.



그림 3. 정규화된 CIE 좌표로 변환된 입력 영상

4. 정지 영상에서의 얼굴 영역 검출

4-1 뉴럴 학습 데이터의 생성

3절에서 제시된 전처리 과정을 거친 컬러 영상 데이터에서 8×12 크기의 배경을 제외한 얼굴 안쪽 부위만을 추출해 뉴럴의 입력으로 사용한다. 입력 영역의 각 셀은 전절에서 설명한 x, y 두값을 가지므로 총 입력은 192개가 된다. 또 출력은 FACE/NONFACE의 2출력이 된다. 학습 알고리즘은 뉴럴 네트워크에서 가장 많이 사용되어지고 있는 역전파(backpropagation) 알고리즘을 사용한다. 비전 시스템의 CCD 카메라와 영상 캡춰 보드의 해상도가 그리 높지 않아 영상의 질이 낮고 따라서 배경과 얼굴 역시 선명하지 않은 경우에는 얼굴의 정확한 기하학적 구조를 이용하는 기존 몇몇 알고리즘은 적용하기 힘든데 반해 뉴럴 네트워크는 보다 융통성있는 결과를 가져다 준다. 그림 4.은 모자이크 영상으로부터 얻어진 8×12 뉴럴 학습 데이터 영역이다. 이 영역의 RGB 좌표를 CIE로 바꾼 값이 뉴럴의 입력 데이터가 된다. 실험은 20대 남자 14명 선택하여, 각 10 개씩의 입력 영상을 받아 총 140개의 데이터를 학습시켰다.

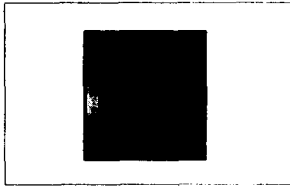


그림 4. 추출된 학습 데이터(8×12)

4-2 정지 영상에서의 얼굴 검출

실제 입력 영상이 들어오면 3절에서 제시된 모자이크 패턴 처리한 후 8×12 셀의 뉴럴 네트워크 입력을 차례로 내려받혀 입력한다. 각 입력에 대한 출력값은 FACE/NONFACE인데 그 값을 8×12 윈도우의 중앙에 기록한다. 뉴럴 처리가 끝나면 출력이 FACE로 된 부분이 밀집된 부분을 검출해 얼굴 위치로 정한다. 그림 5는 뉴럴 네트워크에 의한 결과 영상을 나타낸다. 그림에서 어두운 부분이 NONFACE로 출력된 부분이고, 밝은 색 부분이 FACE로 출력된 부분이다. 이 경우 가운데 밝은 영역을 검출로 검출한다. 다양한 사이즈의 얼굴을 검출하기 위해

서는 입력 영상을 다양한 셀 크기로 모자이크 영상화해 각각을 뉴럴의 입력 영상으로 처리하면 작은 얼굴에 대해서도 얼굴 위치를 찾아낼 수가 있다.



그림 5. 뉴럴에 의한 결과 영상

5. 동영상에서의 얼굴 영역 트래킹

정지 영상에서 탐색된 얼굴 영역을 기반으로 하여 동영상에서의 얼굴의 트래킹을 진행한다. 동영상에서의 얼굴 트래킹의 가장 중요한 요소는 영상 내의 움직임을 감지하는 일이다. 정지 영상에서 탐색된 얼굴 영역을 기반으로 이후 연속된 입력 영상들에 대해서는 시간의 따른 변화를 감지할 수 있는 cell by cell 이미지 감산(Image Substraction) 처리와 thresholding을 통해 영상 내의 움직임이 있는 부분을 검출해낼 수 있다. 이때 전 영상에 기억해둔 얼굴의 위치가 다음 영상의 얼굴 위치를 찾는 데 중요한 정보가 된다. 전 영상을 기억해야 하는 단점을 가지고 있지만 일일이 뉴럴 네트워크로 찾는 것보다는 훨씬 수월한 결과를 가져온다.

6. 결론

본 논문에서는 복잡한 환경을 가진 동영상에서 사람의 얼굴을 검출, 트래킹하는 방법을 제안하였다. 기존 RGB 좌표를 CIE 좌표로 변환하여 처리하여 기존 방법의 큰 문제점 중의 하나인 조명의 영향을 크게 줄일 수 있었고, 다양한 사이즈의 모자이크 패턴 처리로 뉴럴 네트워크를 복잡한 환경에서 여러가지 크기의 정면 얼굴을 검출할 수가 있었다. 이 과정은 독립적으로 행해질 수 있으므로 병렬 분산 처리가 가능하여 보다 빠른 시간을 요구하는 실시간 응용 프로그램에 적용할 수 있다. 기동인 얼굴이나 일 모습 또한 뉴럴에 의한 학습으로 검출할 수 있는 가능성도 제시하였다. 이러한 얼굴 검출 및 트래킹 시스템

템은 뉴럴에 의한 빠른 컴퓨터 비전 시스템을 이용한 로봇 등 많은 응용 분야에서 사용되어질 수 있다. 탐색된 얼굴 위치에 형상 모델(Wire Frame)을 취함으로써 얼굴의 분석 및 합성 분야에 유용하게 쓰일 수 있을 뿐만 아니라, 실시간 처리 가능한 얼굴 인식 등 여러 응용 분야에 중요한 전처리 과정으로 사용되어질 수 있을 것이다. 추후 과제로는 보다 빠른 알고리즘으로의 개선과 얼굴 인식 등 실제 시스템으로의 적용 등이 필요하다고 생각된다.

Analysis of Images and Patterns, pp432-439, 1995.

- [8] A. Yuille, D. Cohen, and P. Hallinan, "Feature Extraction from Faces using Deformable Template", in Proc. IEEE conf. on Computer Vision and Patt. Recog., 1989, pp104-109

8. 참고 문헌

- [1] M. Turk and A. Pentland "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3: pp71-86, 1991.
- [2] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi and Andrew D. Back, "Face Recognition: A Convolution Neural Network Approach", *IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition*.
- [3] Kazuhiko Kawamura, Sugato Bagchi, Moenes Iskarous and Magued Bishay "Intelligent Robotic Systems in Service of the Disabled", *IEEE Transactions on Rehabilitation Engineering Vol.3 No.1*, 1995.
- [4] 지승환, 김은태, 김용주, 박민용, 최장석, "변형가능한 템플릿과 퍼지타원 모델을 사용한 얼굴의 3차원 형상모델 정합", *제어계측·자동화·로보틱스 연구회 합동 학술 발표회 논문집*, pp380-383, 1996.
- [5] M.A.Sid-Ahmed, "Image Processing: Theory, Algorithms & Architectures", *MaGraw-Hill International Editions*, pp277-375
- [6] Henry A. Rowley, Schumeet Baluja, and T. Kanade, "Human Face Detection in Visual Scenes", *Technical Report CMU-CS-95-158, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, July 1995*.
- [7] Kah-Kay Sung and T. Poggio, "Learning Human Face Detection in Cluttered Scenes", In *Computer*