

광각 및 협각 카메라를 이용한 시선 위치 추적 시스템

정희원 박 강 령

Gaze Detection System by Wide and Narrow View Camera

Kang Ryoung Park *Regular Members*

요 약

시선 위치 추적이란 현재 사용자가 쳐다보고 있는 위치를 컴퓨터 시각 인식 방법을 이용하여 파악하는 연구이다. 일반적으로 사용자가 모니터 상의 한 위치를 쳐다보기 위해서는 얼굴 및 눈동자를 동시에 움직이는 경향이 있다. 기존의 시선 위치 추적 시스템은 사용자의 얼굴 전체를 취득할 수 있는 단 하나의 광각 카메라를 이용하여 사용자의 얼굴 및 눈동자 움직임을 추적하였다. 그러나 이러한 경우, 광각 카메라 내에 포함된 눈동자 영상의 해상도가 많이 떨어져서 사용자의 눈동자 움직임을 정확하게 추적하지 못하는 문제점이 있었다. 그러므로 이 논문에서는 얼굴 영상을 취득하기 위한 광각 카메라 및 눈 영역을 확대하여 취득하는 협각 카메라, 즉 2개의 카메라를 이용하여 시선 위치 추적 시스템을 구현하였다. 또한, 얼굴의 움직임 시 전체적인 위치가 변화될 눈동자의 움직임을 정확히 추적하기 위해, 협각 카메라에는 광각 카메라에서 추출된 얼굴 특징점의 위치를 기반으로 한 자동 초점 및 자동 상하/좌우 회전 기능이 포함되어 있다.

실험 결과, 얼굴 및 눈동자 움직임에 의한 모니터상의 시선 위치 정확도는 실험자가 눈동자는 고정으로 하고 얼굴만 움직여서 쳐다보는 경우에 약 3.1cm, 그리고 얼굴 및 눈동자를 같이 움직여서 쳐다보는 경우에 약 3.57cm의 최소 자승 에러성능을 나타냈다. 처리 속도도 Pentium-IV 1.8 GHz에서 약 30ms 이내의 처리 속도를 나타냈다.

Key words : Facial and Eye Gaze detection, auto focusing and auto pan/tilt

ABSTRACT

Gaze detection is to locate the position on a monitor screen where a user is looking by computer vision. Previous gaze detection system uses a wide view camera, which can capture the whole face of user. However, the image resolution is too low with such a camera and the fine movements of user's eye cannot be exactly detected. So, we implement the gaze detection system with a wide view camera and a narrow view camera. In order to detect the position of user's eye changed by facial movements, the narrow view camera has the functionalities of auto focusing and auto pan/tilt based on the detected 3D facial feature positions.

As experimental results, we can obtain the facial and eye gaze position on a monitor and the gaze position accuracy between the computed positions and the real ones is about 3.1 cm of RMS error in case of permitting facial movements and 3.57 cm in case of permitting facial and eye movement. The processing time is so short as to be implemented in real-time system (below 30 msec in Pentium -IV 1.8 GHz)

* 상명대학교 미디어학부(parkgr@smu.ac.kr)

논문번호 : 030217 - 0520, 접수일자 : 2003년 5월 20일

1. 서론

시선 위치 추적이란 현재 사용자가 응시하고 있는 위치를 파악하는 연구이다. 기존의 연구들은 주로 2차원 및 3차원 얼굴 움직임량을 추정하거나[1][15][20][21], 얼굴의 움직임 만에 의한 시선 위치를 파악하거나[2-8,16,17,19], 눈동자만의 움직임에 의한 사용자의 시선 위치를 파악하는 연구가 주종을 이루었다[9-14,18]. 그러나 얼굴 및 눈동자 움직임을 함께 고려하여 시선 위치를 파악하는 연구는 거의 이루어지지 않았다. Ohmura와 Ballard[4,5]등의 연구에서는 초기에 얼굴 특징점의 3차원 거리 정보를 알고 있어야 하며, 시선 위치를 파악하기 위해서는 많은 처리 시간(약 1분 이상)이 소요되는 단점이 있다. Gee[6]와 Heinzmann[7]들의 연구에서는 얼굴 좌표계에서의 시선 벡터의 방향을 계산하였을 뿐, 이로부터 모니터 상에 사용자 시선 위치 등을 구하지 않았다. 또한, 이들의 연구에서는 얼굴의 3차원 회전 및 이동이 동시에 발생하지 않는다고 가정했다. 이는 얼굴의 회전과 이동이 동시에 발생했을 때, 그들의 논문에서 사용하는 최소 자승 정합 알고리즘(least-square fitting algorithm)에서의 연산 복잡성과 처리 시간의 상승 등으로 3차원 움직임량을 정확하게 추정하기 어려웠기 때문이다. Rikert[8]의 연구에서는 학습 및 테스트 환경에서 얼굴 및 모니터 스크린사이의 거리가 변하지 않아야 한다는 가정이 있으며, 이러한 것은 실제 사용에 있어서 많은 불편함을 제공하게 된다. 기타, 다른 연구들[10][13][14][16][17]에서는 사용자로 하여금 구분점이 부착된 안경을 착용하게 함으로써 얼굴 특징점을 추적하는 연구를 수행하였는데, 이처럼 별도의 안경을 착용해야 하므로 사용자에게 불편함을 제공하는 결과를 낳게 되었다. 얼굴 움직임에 의해 모니터 상의 시선 위치를 파악했던 연구[2][3][19]에서는 눈동자의 움직임은 전혀 고려치 않고 단지 얼굴의 움직임 만에 의한 시선 위치 파악 연구를 수행하였다. 얼굴 및 눈동자의 움직임을 같이 고려하여 시선 위치를 파악한 박강령[22]의 연구에서는 한 개의 광각 카메라를 이용하여 전체 얼굴을 촬영하는 방식을 취하였다. 그러나

이러한 경우, 눈동자 영상의 해상도가 많이 떨어지며 그 결과 사용자의 정교한 눈동자 움직임을 추적하기 어려운 문제점이 있다. 그러므로 이 논문에서는 광각 카메라(얼굴의 움직임에 의한 시선 위치 추적용) 및 눈 영역을 확대하여 취득하는 협각 카메라(눈동자 움직임에 의한 시선 위치 추적용), 즉 2개의 카메라를 이용하여 시선 위치 추적 시스템을 구현하였다. 얼굴의 움직임 시 전체적인 위치가 변화될 눈동자의 움직임을 정확히 추적하기 위해 협각 카메라에는 추출된 얼굴 특징점의 위치를 기반으로 한 자동 초점 및 자동 상하/좌우 회전 기능이 포함되어 있다. 실험 결과, 보다 정확한 시선 위치 추출 성능을 얻을 수 있었으며, 전체 시스템의 가격 역시 사용자에게 부담스럽지 않을 정도로 저렴하게 시스템을 구현할 수 있었다.

2. 얼굴 특징점의 추적

모니터 상에 사용자가 응시하고 있는 위치를 파악하기 위해서는 입력 영상에서 얼굴 특징점(양 눈동자의 중심 및 눈의 구석, 콧구멍 및 입의 양 끝점)의 위치를 정확하게 추출할 수 있는 기술이 필수적으로 요구된다. 이 논문에서는 입력 영상으로부터 눈 영역을 먼저 추출한 후, 이를 기준으로 기타, 다른 얼굴 특징점들의 위치를 실시간으로 추적하는 방법을 사용하였다[22]. 일반적인 사무실 환경에서는 사용자의 얼굴 뒤에 복잡한 배경이 존재하며, 또한 외부 광등의 영향으로 얼굴 및 얼굴 특징점의 위치를 영상 처리 방법만으로는 실시간으로 추출하기 어렵다. 그러므로 이 논문에서는 그림 1과 같이 적외선 조명 기반 카메라 시스템을 개발하고, 이를 이용하여 입력 영상으로부터 얼굴 특징점을 실시간으로 추출하였다. 먼저, 입력 영상으로부터 눈 위치를 추출하기 위하여 본 논문에서는 적외선 조명을 켜고 눈의 각막(corneal)에서 발생하는 반사 위치(specular reflection)를 찾는 방법을 사용하였다[22]. 그림 1을 보다 자세히 설명하면 다음과 같다. IR-LED(InfraRed-light emitting diode) 조명(1)은 눈에서의 반사 위치를 생성하는 데 사용된다. 이 때 LED조명으로는 사람 눈으로 감지하기 어려운 880nm의 조명을 사용함으로써, 동작

중 사용자의 눈부심 현상이 없도록 하였다. 카메라 렌즈 앞에 부착된 HPF(high pass filter)(2)는 적외선 영역(800nm)이상의 적외선 조명만을 통과시키기 때문에, 카메라 센서를 통한 영상 취득 시 800nm이하의 외부광이 영향을 주지 못하게 된다. 즉, 입력된 영상 내에는 카메라에 설치되어 있는 적외선 조명이 주로 영향을 주기 때문에, 외부광에 의해 발생하는 영상내의 그림자 등의 영향은 적어지게 된다. 일반적인 카메라 CCD 센서의 경우, 가시광선보다는 적외선 영역으로 갈수록 영상의 감도가 낮아지는 특성을 나타낸다. 본 논문에서는 사용자의 눈에 눈부심을 주지 않기 위하여 880nm의 적외선을 사용했으며, 이러한 경우 영상에서의 감도가 많이 낮아 얼굴 특징점을 추출하기 위해 충분한 밝기를 내기 위해서는 여러 개(4개)의 적외선조명을 모아 사용해야만 했다.

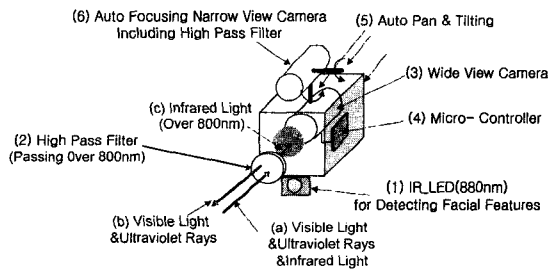


그림 1. 시선 위치 추적용 카메라
Fig. 1. The Gaze Detecting Camera

또한, 본 연구에서는 일반적인 인터레이스브 CCD(charge coupled device) 카메라와 카메라의 동작을 제어하기 위한 마이크로 컨트롤러(4)를 사용했다. 이를 이용하여 그림 2와 같이 CCD의 출력 신호 중 Even 과 Odd 필드의 VD(vertical drive)신호에 맞추어 조명의 On/Off를 조정하게 하였다. 그림 2를 보다 자세히 설명하면 다음과 같다. 초기에 사용자가 시선 위치 추적 시스템을 시작하려는 순간, PC쪽에 구현되어 있는 프로그램으로부터 카메라 마이크로컴으로 RS-232C 인터페이스를 거쳐 시작 신호가 전달된다(①). 이 다음 Frame부터, 카메라에 있는 마이크는 그림 2에 나타나 있듯이 CCD의 매 Even 및 Odd field의 시작 위치 VD 신호에 맞추어 적외선 조명을 계속 On/Off 시

키게 된다. 이처럼 적외선 조명을 On/Off시킨 영상이 입력되면, 이로부터 눈에 발생하는 적외선 반사 위치 및 이를 이용하여 얼굴 특징점의 위치를 추출하게 된다(②)[22]. 이 때 눈에 발생하는 반사위치(specular reflection)의 영상 그레이 값은 다른 반사 위치(뺨, 혹은 이마 등에 생기는 조명 반사)보다 상당히 높기 때문에, 차영상을 구하게 되면 눈에 발생하는 반사 위치를 쉽게 찾을 수 있게 된다[22].

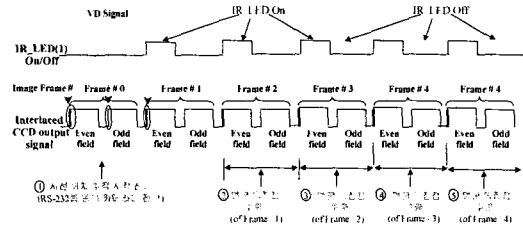


그림 2. 얼굴 특징점 추출을 위한 적외선 조명 조정
Fig. 2. The IR_LED controls for detecting facial features

그림 2. 얼굴 특징점 추출을 위한 적외선 조명 조정
Fig. 2. The IR_LED controls for detecting facial features

입력 영상으로부터 눈의 반사위치를 추출하게 되면, 이로부터 제한된 영역에서 정확한 눈 위치(눈동자의 중심 및 눈의 구석)를 추출하게 된다. 눈동자의 중심 위치를 추출하기 위해 이 논문에서는 원형 경계 검출 알고리즘(circular edge detection)을 사용하였다. 제한된 영역에서 눈동자 중심 위치를 추출하기 때문에, 정확한 위치를 추출하는 데에는 처리 시간이 거의 소요되지 않았다(pentium-II 550MHz PC에서 약 5~10 ms). 눈동자의 중심 위치를 추출한 후, 본 논문에서는 눈구석 모양 템플릿(eye corner shape template)을 이용하여 눈의 양 구석 위치를 추출하였다[22]. 본 논문에서는 그림 1에서와 같이 HPF(2)를 사용하므로 입력 영상 내외부광에 의한 영향을 최소화 할 수 있었으며, 그 결과 입력 영상이나 템플릿 등을 조명에 대해 별도로 정규화할 필요가 없었다. 이와 같은 눈구석 모양 템플릿과 함께 본 논문에서는 SVM(support vector machine)을 이용하여 정확한 눈의 구석 위치를 추출할 수 있었다[22]. 입력된 눈구석 영상은 30×30 픽셀 크기로 정규화하여 SVM의 입력으로 사용한다. 입력 데이

터로는 2000장의 연속된 영상을 취득하여(다양한 자세로 앉아 있는 20명의 사람×100장의 영상) 이로부터 8000개의 눈 모서리 샘플(4개의 눈 모서리 샘플/명×2000장의 영상)을 취득했으며, 이 외에 1000장의 영상을 추가로 획득하여 SVM 테스트를 위해 사용하였다. 실험 결과, 학습 데이터에 대한 분류 성능은 약 0.11%의 분류 에러(9/8000개)를 나타냈으며, 테스트 데이터 중에서는 약 0.2%(8/4000개)의 에러를 나타냈다. 또한, SVM을 사용했을 때의 분류 시간 역시 Pentium-II 550MHz 환경에서 13ms정도 소요되었다. 그러므로 실시간으로 얼굴 특징점 및 시선 위치를 파악하고자 하는 본 연구 분야에서는 충분히 사용 가능한 알고리즘임을 알 수 있었다. 이처럼 입력 영상에서 사용자의 눈 위치를 찾게 되면, 찾은 눈 위치에 대해 코 및 입의 상대적인 존재 가능 영역을 설정할 수 있게 된다. 이 논문에서는 설정된 존재 가능 영역 내에서, SVM을 이용하여 눈의 구석 위치를 찾았던 것과 유사한 방법으로 양 콧구멍 및 입의 끝점 위치를 찾았다. 얼굴 특징점 추출에 대한 실험 결과, 평균 1 픽셀(양 눈의 중심), 2 픽셀(양 눈의 구석 위치), 4 픽셀(양 콧구멍) 그리고 3 픽셀(입의 양 끝점)의 최소 자승 에러가 발생하였다. 본 논문에서의 입력 영상 크기는 640×480픽셀이며, 여기서 최소 자승 에러란 SVM으로 찾은 특징점의 위치와 사용자가 직접 눈으로 보고 찾은 위치 사이의 에러 값을 나타낸다. 본 논문에서는 총 3000장의 연속 영상(2000장의 학습 영상 및 1000장의 테스트 영상)을 이용하여 특징점 추출 성능을 테스트하였다

[22]. 입력 영상에서 얼굴 특징점의 위치가 추출되면, 그림 3과 같이 총 9개의 특징점(P1, P2, P3, P4, P5, P6, P7, Q1, Q2,)을 선정하게 되는데 이러한 특징점들은 얼굴의 3차원 움직임량(3차원 회전량 및 이동량) 및 눈동자 움직임량을 추정하기 위해 사용된다. 사용자가 모니터상의 한 지점을 응시하게 되면 그림 3의 (b)에서 나타나 있는 것처럼 9개의 특징점 위치가 (P'1, P'2, ~ P'7 Q'1, Q'2)로 변하게 된다. 여기서 (Q1, Q2, P1, P2, P3, P4) 와 (Q'1, Q'2, P'1, P'2, P'3, P'4)는 5장에서 언급할 눈동자 움직임에 의한 시선 위치 추출을 위하여서도 추가로 사용된다.

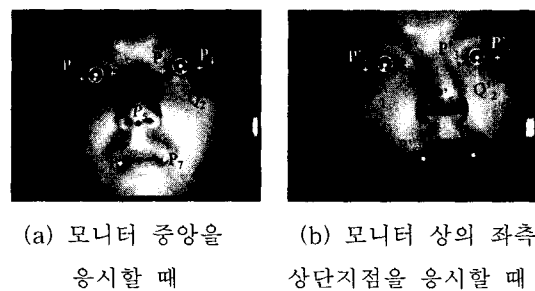


그림 3. 얼굴 및 눈동자 움직임에 의한 시선 위치 추출을 위해 사용되는 얼굴 특징점
Fig. 3. The feature points for detecting the gaze position by estimating 3D facial and eye movements

3. 초기 얼굴 특징점의 3차원 위치 추정

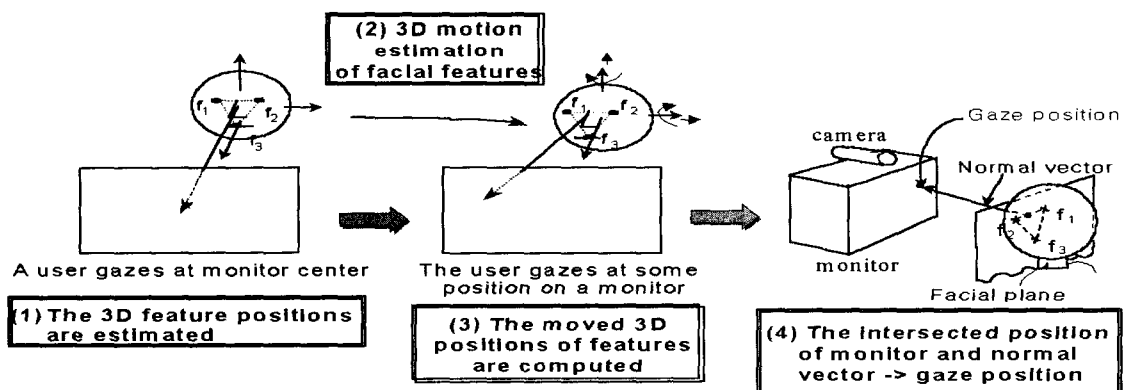


그림 4. 모니터 상의 시선 위치를 파악하기 위한 4단계
Fig. 4. 4 steps in order to compute a gaze position on a monitor

그림 4에 나타나 있는 것처럼, 얼굴 특징점을 추출한 후 모니터상의 시선 위치를 파악하기 위해서는 본 논문에서는 4단계의 과정을 거치게 된다[2][3][22]. 첫 번째 단계에서는 사용자가 모니터 상에 미리 지정된 3 위치(우측 상단, 중앙, 좌측 하단)를 쳐다봄으로써 얻어진 정보들을 바탕으로 초기 얼굴 특징점의 3차원 위치를 자동으로 구하게 된다. 두 번째 및 세 번째 단계에서 사용자가 모니터상의 한 지점을 쳐다보기 위해 얼굴을 움직이게 되는 경우(회전 및 이동), 3차원 움직임 추정(3D motion estimation) 및 아핀 변환(affine transformation)에 의해 변환된 얼굴 특징점의 위치를 구할 수 있게 된다. 마지막 단계에서는, 변환된 얼굴 특징점들의 위치로부터 하나의 얼굴 평면과 법선을 구할 수 있으며, 이 법선과 모니터 평면이 만나는 위치가 얼굴의 움직임에 의한 시선 위치가 된다. 세 번째 단계에서 얼굴 특징점의 변환된 3차원 위치가 계산되어 지면, 이들은 모니터 좌표계 및 첫 번째 단계에서 구한 카메라 매개변수 정보들을 이용하여 카메라 좌표계를 기준으로 계산되어 질 수 있다. 이와 같은 정보를 바탕으로, 본 논문에서는 정확한 눈동자의 위치를 취득하기 위해 협각 카메라(narrow view camera)를 자동 상하/좌우 회전 시킬 수 있게 된다(그림 1의 (5) 참조). 본 논문에서 사용하는 협각 카메라는 작은 화각(약 30 ~ 45 mm의 큰 초점거리)을 갖고 있으며, 이러한 특성 때문에 눈 영역을 확대하여 촬영할 수 있다. 또한 사용자가 모니터의 한 지점을 쳐다보기 위하여 얼굴을 심하게 회전하는 경우, 얼굴 영상을 취득하기 위한 광각 카메라(wide view camera)에서 한쪽 눈 영역이 아예 사라져 버리는 경우가 발생한다. 이러한 경우, 협각 카메라를 아무리 상하/좌우 회전한다고 하더라도, 광각 영상에서 사라진 한 쪽 눈 영역을 찾을 수 없게 된다. 그러므로 이러한 경우에 대비하기 위하여 본 논문에서는 사용자가 모니터 중앙을 쳐다보는 경우 양쪽 눈을 모두 추적하고, 이후 사용자가 얼굴을 심하게 회전하여 한쪽 눈을 추적할 수 없는 경우, 나머지 보이는 다른 쪽 눈만을 추적함으로써 눈동자 움직임에 의한 정확한 시선 위치를 파악하였다. 협각 카메라의 상하 및 좌우 회전을 위해 420 pps(pulse per second)를 갖는 2개의 스테핑 모터를 사용하였다. 또한, 일반적인 협

각 카메라의 경우 작은 범위의 DOF(depth of field)를 갖는 특성이 있으므로, 입력 영상은 사용자의 Z 방향 움직임에 따라 쉽게 초점이 흐려지는 특성을 나타낸다. 일반적으로 DOF는 카메라의 아이리스 크기가 작을수록, 피사체의 거리가 멀수록 커지는 특성을 나타낸다. 그러나 이 논문에서와 같이 사용자의 시선 위치를 파악하는 응용 분야에서는 모니터 앞에 약 50 ~ 70 cm의 다양한 Z거리 범위에서 사용자가 앉을 수 있으며, 카메라의 아이리스 크기를 작게 만드는 것은 궁극적으로 카메라 CCD 센서로 들어오는 빛의 양을 줄이게 되어 입력 영상이 어두워지는 결과를 낳게 된다. 그러므로 본 논문에서는 단초점 렌즈를 이용해서는 다양한 Z 거리를 갖는 사용자의 눈에 대한 초점을 맞추기 어렵기 때문에, 직경 10mm의 아이리스 크기를 갖는 가변 초점 렌즈 및 자동 초점용 모터(420 pps)를 이용하여 보다 선명한 눈 영상을 취득하였다(그림 1의 (6)참조). 이와 같은 자동 상하/좌우 회전 및 초점 기능은 그림 1에서 나타나 있는 카메라내의 마이크로 콘트롤러(그림 1의 (4)참조)에 의해 제어된다. 또한 협각 눈 영상의 초점을 맞추기 해서는 기본적으로 카메라를 기준으로 사용자 눈까지의 Z 거리를 알아야 한다. 본 논문에서는 그림 4의 단계 3에서 변환된 얼굴 특징점(특히 눈)의 3차원 위치 및 Z 거리를 구할 수 있으므로 이를 바탕으로 눈 영상의 자동 초점 작용을 하게 된다. 그러나 협각 카메라는 DOF(depth of field)가 굉장히 작기 때문에 이와 같은 Z 거리만 가지고는 정확한 자동 초점이 이루어지기 어렵다. 그러므로 본 논문에서는 영상내의 픽셀들 간의 변화 차이를 측정하는 단순한 초점 측정 알고리즘을 병행하여 사용하였다. 즉, 단계 3에서 계산된 얼굴 특징점(눈)의 3차원 위치 정보를 바탕으로 협각 카메라를 상하 및 좌우 회전을 하고 자동 초점을 수행함으로써 눈 영상을 취득한다. 이후, 취득된 영상은 PC로 전송된 후, 영상내의 픽셀들 간의 차이를 측정하는 초점 측정 알고리즘에 의해 영상의 초점 정도를 평가한다. 만일 영상의 초점 정도가 미리 정해놓은 임계치(본 논문에서는 초점 정도를 0에서 100까지의 범위를 갖는 연속적인 값으로 산정하였으며, 숫자가 커질수록 초점이 잘 맞는 상태로 설정하였다. 임계치는 90으로 사용하여, 초점 정도가 90 이상

이 되는 경우를 초점이 잘 맞은 상태로 간주하였다)를 넘지 못하는 경우, 카메라 마이크로 컨트롤러를 통해 자동 초점 렌즈에 움직임 명령을 전달하게 된다. 여기서, 초점이 맞지 않는 눈 영상이 취득된 경우, 초점 렌즈를 어느 방향으로 움직여야 하는 지(앞 혹은 뒤) 결정하기 어려운 문제가 있다. 이를 결정하기 위하여 본 논문에서는 기타 부가적인 정보를 바탕으로 하여 여러 가지 휴리스틱한 방법(눈 영상의 밝기 정보 혹은 피라미드 기반 전역 렌즈 탐색)등을 이용하여 정확하게 눈 초점 영상을 취득할 수 있었다. 초점이 맞은 눈 영상이 취득되면, 본 논문에서는 학습된 신경망(다층 퍼셉트론)을 이용하여 눈동자 움직임에 의한 모니터 상의 시선 위치를 파악하였다. 이후, 앞에서 구한 얼굴에 움직임에 의한 시선 위치와 눈동자 움직임에 의한 시선 위치는 산술적으로 더해짐으로써 최종적으로 모니터상의 시선 위치를 나타내게 된다. 이에 대한 자세한 내용은 5장에서 기술한다. 그림 4의 첫 번째 단계에 대한 자세한 방법은 [2]에서 참조할 수 있다. 첫 번째 단계에서 구한 초기 얼굴 특징점의 3차원 위치와 이 얼굴 특징점의 실제 위치(3D position tracker sensor로 측정) 사이의 최소 자승 에러(RMS error)는 약 1.15 cm(X축으로 0.64 cm, Y축으로 0.5cm, Z축으로 0.81cm)의 정확도를 나타냈으며, 실험에는 얼굴 특징점 추출을 위해 사용했던 20명분의 데이터를 이용하였다.

4. 얼굴의 3차원 움직임 및 이동량 추정

이 장에서 설명할 내용은 3장에서 소개한 그림 4의 2번째 단계에 속한다. 기존에 많은 3차원 움직임 추정 방법들이 연구되어 왔는데, 예를 들어 확장 칼만 필터(extended kalman filter)[1], 신경망[21] 그리고 아핀 투영 방법(affine projection method)[6][7] 등이 있다. 각 방법들은 많은 제한 요소와 문제점들을 가지고 있으며[22], 이러한 이유로 본 논문에서는 기존에 3차원 움직임 추정을 위해 가장 많이 사용되는 알고리즘인 확장 칼만 필터를 사용하였으며[1], 확장 칼만 필터와 아핀 변환을 이용하여 얼굴 특징점의 변화된 3차원 위치를 구하였다. 이에 대한 자세한 설명은 [1][20][22]에서 참조

할 수 있다. 본 논문에서 확장 칼만 필터의 예측 정확도는 3차원 위치 추적 장비(polhemus sensor)[22]와 비교되었다. 실험 결과, 확장 칼만 필터와 3차원 위치 추적 장비와의 정확도 차이는 이동량 및 회전량에서 각각 1.4cm 및 2.98°의 에러 성능을 나타냈다.

5. 모니터 상의 시선 위치 파악

5.1 얼굴의 움직임에 의한 시선 위치 파악

이 장에서 설명하는 내용은 그림 4의 3, 4번째 단계에 해당된다. 3장에서 설명한대로 모니터 좌표계를 기준으로 파악된 초기 얼굴 특징점의 3차원 위치(그림 3의 $P_1 \sim P_7$)는 얼굴 좌표계를 기준으로 변환되어 질 수 있다. 이후, 이 특징점의 위치는 사용자가 모니터 상의 한 지점을 쳐다 볼 때 발생하는 회전 $[R]$ 및 이동 $[T]$ (4장에서 설명한 확장 칼만 필터에 의해 추정되는) 행렬에 의해 변화되며, 이 행렬을 이용한 아핀 변환 계산에 의해 변화된 얼굴 특징점의 위치를 구할 수 있다. 이처럼 사용자가 모니터 상의 한 지점을 쳐다볼 때의 얼굴 특징점의 위치를 계산하면, 이를 모니터 좌표계를 기준으로 다시 변환하고, 이 특징점들이 형성하는 얼굴 평면 및 평면의 법선이 모니터와 만나는 위치가 사용자의 시선 위치가 된다[2].

5.2 눈동자의 움직임에 의한 시선 위치 파악

5장 1절에서 설명한 사용자의 시선 위치는 눈동자의 움직임은 고려하지 않고 얼굴의 움직임에 의해서만 모니터상의 시선 위치를 파악한 것이다. 그러나 대부분의 경우, 사용자가 모니터 상의 한 지점을 쳐다볼 때에는 얼굴 및 눈동자의 움직임이 동시에 발생한다. 그러므로 본 논문에서는 보다 정확한 시선 위치를 파악하기 위하여 사용자의 눈동자 움직임을 파악하는 연구를 수행하였다. 사용자의 눈동자 움직임은 그림 6과 같이 입력 영상에서 추출된 특징 정보를 기준으로 파악된다. 그림 6은 앞서 설명한 자동 상하/좌우 회전 및 초점 기능을 가진 협각 카메라에 의해 취득된 눈 영상이다. 이 그림에서 추출된 눈동자 영역 및 구석 위치가 표시되

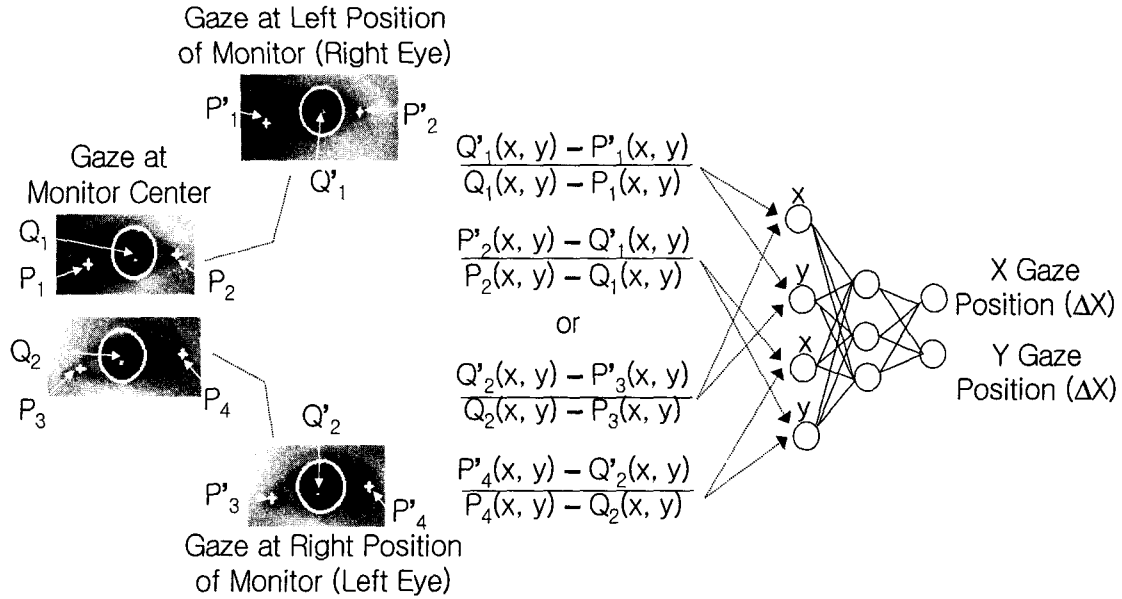


그림 5. 눈동자 움직임에 의한 시선 위치 파악을 위한 특징점 및 신경망 구조
 Fig. 5. The features and the neural network for detecting gaze position
 by eye movements

어 있다. 여기서 눈동자 및 눈구석 영역을 추출하기 위하여 원형 경계 검출(circular edge detection) 및 SVM기반 눈구석 템플릿 방법을 사용하였다. 이 방법은 2장에서 설명한 광각 카메라에서 눈 특징점을 추출하기 위해 사용했던 방식과 거의 유사하다. 전술한 바와 같이 사용자가 그의 얼굴을 심하게 회전하는 경우, 그의 한쪽 눈이 협각 카메라에서 보이지 않게 되는 경우가 종종 발생한다. 그러므로 본 논문에서는 그림 6의 (b)와 같이 사용자가 모니터 중앙을 쳐다 볼 때 양 눈을 모두 검출하고, 그림 6의 (a), (c)와 같이 사용자가 모니터의 한 지점을 쳐다보기 위해 얼굴을 심하게 회전하는 경우 보이게 되는 한쪽 눈(왼쪽 혹은 오른쪽 눈)만을 추적하게 된다. 일반적으로 사용자의 시선 위치에 따라 눈동자의 움직임 량 및 형태 정보는 변하게 된다. 특히 눈동자의 중심위치(Q_1, Q_2)와 왼쪽 혹은 오른쪽 눈의 구석 위치($P_1, P_2, P_3, P_4, P'_1, P'_2, P'_3, P'_4$)사이의 거리는 사용자의 시선 위치에 따라 변화하게 된다. 본 논문에서는 그림 5와 같이 눈동자의 움직임 량 및 시선 위치 사이의 관계를 학습하기 위하여 신경망(다층 퍼셉트론)을 사용하였다. 여기서 신경망을 위한 입력 값은 사용자가 모니터 중앙을 쳐다

볼 때 취득된 눈 영상에서의 눈 중심과 구석 사이의 거리로 정규화된다. 이와 같은 정규화가 필요한 이유는 본 논문에서는 줌 렌즈 없이 초점 렌즈만을 이용하여 눈 영상을 취득하고 있기 때문에, 사용자의 얼굴이 카메라에 접근할수록 눈 영상이 확대되는 결과를 낳게 되기 때문이다. 즉, 사용자가 카메라에 접근하여 앉은 경우가 모니터에서 멀리 떨어져서 앉은 경우보다 눈 영상의 크기가 커지게 되므로, 동일한 모니터상의 한 지점(예로 우측 상단)을 쳐다보더라도 눈동자의 움직임 량 자체가 커지게 된다. 이 논문에서 모니터 앞에 앉은 사용자의 Z거리는 50 ~ 70 cm 사이에 다양하기 때문에, 그림 5에서와 같은 Z 거리 정규화는 필요하다. 여기서 Z거리는 모니터 중앙에서 사용자 눈까지의 직선거리를 나타내는 것이며, 본 논문에서는 polhemus position tracker[23] 장비를 이용하여 측정하였다.

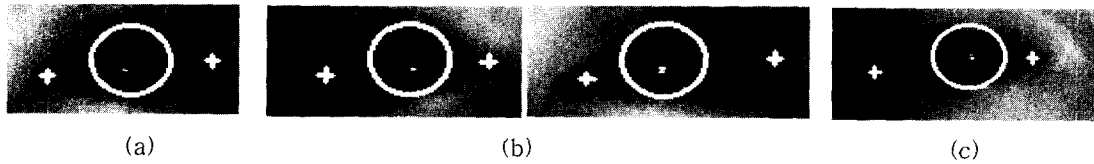


그림 6. 자동 상하/좌우 회전 및 초점 카메라를 이용한 눈 영상 취득

Fig. 6. The captured eye image with auto pan/tilt/focusing narrow view camera (a)Gazing at right position of monitor (b)Gazing at monitor center (c)Gazing at left position of monitor

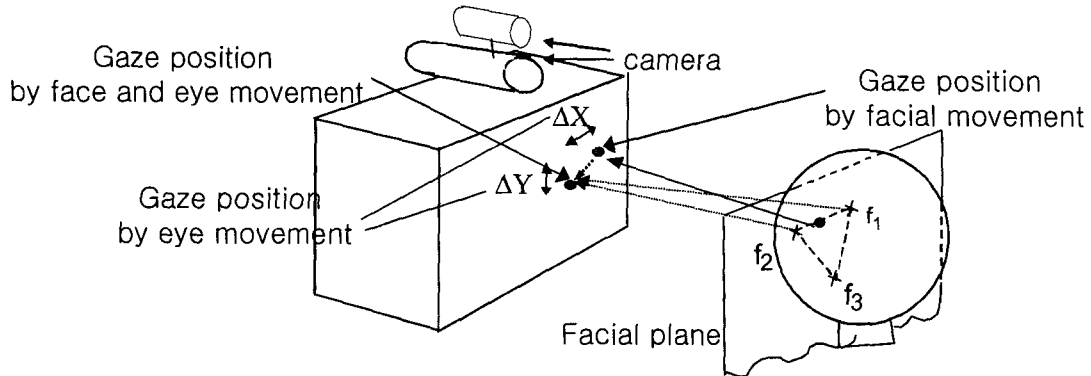


그림 7. 얼굴 및 눈동자 움직임에 의한 모니터상의 시선 위치 파악

Fig. 7. Detecting gaze position on a monitor by face and eye movements

이처럼 눈동자 움직임에 의한 시선 위치를 파악하게 되면, 5장 1절 및 2절의 결과를 산술적으로 종합하여 최종적으로 얼굴 및 눈동자 움직임에 의한 모니터 상의 시선위치를 그림 7과 같이 파악할 수 있게 된다. 즉, 5장 1절에서 설명한 대로 얼굴 움직임에 의한 시선위치를 파악한 후, 이를 기준으로 추가로 눈동자 움직임에 의한 시선 위치 변이량(ΔX , ΔY)을 파악함으로써 최종적인 모니터 상의 시선위치를 계산할 수 있게 되는 것이다.

6. 실험 결과 및 고찰

본 논문에서 제안하는 시선 위치 파악의 정확도는 표 1, 2와 같이 기존의 연구 결과 [2][3][19][22]와 비교하였다. 실험 데이터로는 총 10명의 사용자가 모니터 상의 23개 지점을 쳐다볼 때 취득된 데이터를 사용하였다. 여기서 시선 위치 에러란 사용자가 모니터 상에 실제 쳐다보는 위치와 계산되어 나온 위치사이의 최소 자승 에러를 나타낸다. 실험은 크게 두 가지로 나누어 표 1과 같이 사용자가 모니터상의

한 지점을 쳐다볼 때 얼굴의 움직임에 의해서만 응시하는 경우와 표 2와 같이 얼굴 및 눈동자 움직임을 같이 이용하여 자연스럽게 쳐다보는 경우로 구분하여 수행하였다. 먼저 얼굴 움직임에 의해서만 쳐다보는 경우의 정확도는 표 1과 같이 본 논문에서 제안한 방법이 기존의 많은 방법들에 비해 가장 우수함을 알 수 있다. 그리고 표 2와 같이 얼굴 및 눈동자 움직임이 같이 발생하는 실험 데이터에 대한 시선 위치 추출 정확도 역시 본 논문에서 제안하는 방식이 가장 우수함을 알 수 있다. 전반적으로는 얼굴의 움직임만이 존재하는 경우보다 얼굴 및 눈동자 움직임이 같이 존재하는 경우의 시선 위치 정확도가 다소 떨어짐을 알 수 있다.

방법	선형 보간법[19]	단일 신경망[19]	다중 신경망[19]	[2] 방법	[3] 방법	[22] 방법	본 논문의 방법
최소 자승 에러	5.1	4.23	4.48	5.35	5.21	3.40	3.10

표 1. 얼굴만의 움직임에 의한 시선 위치 추출 정확도

Table 1. Gaze error about test data including only facial movements (cm)

방법	선형 보간법[19]	단일 신경망[19]	다중 신경망[19]	[2] 방법	[3] 방법	[22] 방법	본 논문의 방법
최소 자승 에러	11.8	11.32	8.87	7.45	6.29	4.8	3.57

표 2. 얼굴 및 눈동자 움직임을 같이 고려한 시선 위치 추출 정확도

Table 2. Gaze error about test data including face and eye movements (cm)

두 번째 실험으로는 1280×1024 픽셀 해상도를 갖는 19인치 모니터에 수평 및 수직 150 픽셀 간격(2.8cm 간격)으로 반경 5 픽셀의 점들을 표시하고, 사용자가 이를 쳐다볼 때의 정확도를 측정하였다. 이를 통해 총 560개의 테스트 샘플(10명분×56개의 시선 위치)을 취득하게 되었으며, 이러한 실험 조건은 Rikert의 연구[8]에서와 동일하다. 실험 결과 실제 응시 위치와 본 논문에서 제안한 방법에 의한 응시 위치 사이의 최소 자승 에러는 약 3.61cm로 나타났으며, 이러한 실험 결과는 Rikert의 방법에서의 정확도(약 5.08 cm의 최소 자승 에러)보다 우수함을 알 수 있다. 또한 Rikert는 사용자와 모니터사이의 Z거리는 학습 및 실제 사용 시 고정되었다고 가정하고 있으나, 본 논문에서 그러한 불편한 제약을 사용하고 있지 않다. 이의 일환으로 본 논문에서는 사용자의 Z거리를 변화시켜가면서(55, 60, 65cm) 시선 위치 정확도를 측정하였다. 여기서 Z거리는 모니터 중앙에서 사용자 눈까지의 직선거리를 나타내는 것이며, 본 논문에서는 polhemus position tracker[23] 장비를 이용하여 측정하였다.

실험 결과, 최종 자승 에러는 Z거리가 55cm일 때 3.45cm, Z거리가 60cm일 때 3.57cm 그리고 Z거리가 65cm일 때 3.63cm의 결과를 나타냈다. 전술한 바와 같이 본 논문에서의 방법은 사용자의 Z거리 변화에 거의 영향을 받지 않는다는 것을 보여주고 있다. 또한 Rikert의 방법은 처리 시간이 많이 소요되는 문제점(alphastation 333MHz에서 약 1 분)이 있었으나 본 논문의

방법은 이에 비해 상당히 빠르고 실시간으로 구현 가능한 결과를 나타내고 있다(Pentium-IV 1.8GHz에서 약 30ms 이내).

7. 결론

시선 위치 추적이란 현재 사용자가 쳐다보고 있는 위치를 컴퓨터 시각 인식 방법을 이용하여 파악하는 연구이다. 기존의 시선 위치 추적 시스템은 사용자의 얼굴 전체를 취득할 수 있는 단 하나의 광각 카메라를 이용하여 사용자의 얼굴 및 눈동자 움직임을 추적하였다. 그러나 이러한 경우, 광각 카메라 내에 포함된 눈동자 영상의 해상도가 많이 떨어져서 사용자의 눈동자 움직임을 정확하게 추적하지 못하는 문제점이 있었다. 그러므로 이 논문에서는 얼굴 영상을 취득하기 위한 광각 카메라 및 눈 영역을 확대하여 취득하는 협각 카메라, 즉 2개의 카메라를 이용하여 시선 위치 추적 시스템을 구현하였다.

결과에 나타나 있는 것처럼, 시선 위치 추출 정확도는 약 3.57cm 이며, 사용자의 Z거리 움직임에 영향을 받지 않는 것으로 조사되었다. 또한 이러한 시선 위치 추출 에러는 추가적인 얼굴 혹은 눈동자 움직임에 의한 시선 위치 이동(마우스 드래깅과 같은 효과)으로 보정되어 질 수 있다. 그리고 본 연구에서는 이러한 시선 위치 추적 카메라를 약 30만원대 정도의 가격으로 개발하였으며 이로부터 일반 소비자들에게는 저렴한 가격으로 공급되어 질 수 있을 것

으로 기대한다.

향후, 보다 정확한 시선 위치 추출을 위해서는 눈동자 위치를 보다 정확히 추적할 수 있는 방법이 요구된다. 특히 본 논문에서는 눈동자 움직임을 추적하기 위한 특징점 추출을 위해 적외선 조명의 반사 위치를 찾는 방식을 채택하고 있으나(그림 1), 안경 사용자의 경우 안경면에서 적외선 조명에 의한 큰 반사 위치(specular reflection)가 발생하는 바람에 눈 특징점의 위치를 추출하지 못하는 경우가 종종 발생했다. 이러한 문제점을 해결하기 위하여 향후 연구에서는 카메라를 기준으로 양쪽 방향(왼쪽 및 오른쪽)에 조명을 배치시키고 서로 번갈아 키는 동작을 함으로써 안경에 의한 반사 문제를 해결하고자 한다.

Acknowledgement

본 연구는 한국과학재단 지정 생체인식연구센터의 지원을 받아 이루어졌습니다.

참 고 문 헌

- [1] A. Azarbayejani. "Visually Controlled Graphics". IEEE Trans. PAMI, Vol. 15, No. 6, pp. 602-605, 1993
- [2] K. R. Park et al., "Gaze Point Detection by Computing the 3D Positions and 3D Motions of Face", IEICE Trans. Inf. and Syst., Vol. E.83-D, No.4, pp.884-894, Apr 2000
- [3] K. R. Park et al., "Gaze Detection by Estimating the Depth and 3D Motions of Facial Features in Monocular Images", IEICE Trans. Fundamentals, Vol. E.82-A, No. 10, pp. 2274-2284, Oct 1999
- [4] K. OHMURA et al., "Pointing Operation Using Detection of Face Direction from a Single View". IEICE Trans. Inf.&Syst., Vol. J72-D-II, No.9, pp. 1441-1447, 1989
- [5] P. Ballard et al., "Controlling a Computer via Facial Aspect". IEEE Trans. on SMC, Vol. 25, No. 4, pp. 669-677, 1995
- [6] A. Gee et al., "Fast visual tracking by temporal consensus", Image and Vision Computing. Vol. 14, pp. 105-114, 1996
- [7] J. Heinzmann et al., "3D Facial Pose and Gaze Point Estimation using a Robust Real-Time Tracking Paradigm". Proceedings of ICAFG, pp. 142-147, 1998
- [8] T. Rikert et al., "Gaze Estimation using Morphable Models". Proc. of ICAFG, pp. 436-441, 1998
- [9] A.Ali-A-L et al., "Man-machine interface through eyeball direction of gaze". Proc. of the Southeastern Symposium on System Theory, pp. 478-82, 1997
- [10] A. TOMONO et al., "Eye Tracking Method Using an Image Pickup Apparatus". European Patent Specification-94101635, 1994
- [11] Seika-Tenkai-Tokushuu-Go, ATR Journal, 1996
- [12] Eyemark Recorder Model EMR-NC, NAC Image Technology Cooperation
- [13] Porrill-J et al., "Robust and optimal use of information in stereo vision". Nature. vol. 397, no. 6714, pp. 63-6, Jan 1999
- [14] Varchmin-AC et al., "Image based recognition of gaze direction using adaptive methods. Gesture and Sign Language in Human-Computer Interaction". Int. Gesture Workshop Proc. Berlin, Germany, pp. 245-57, 1998
- [15] J. Heinzmann et al., "Robust real-time face tracking and gesture recognition". Proc. of the IJCAI, Vol. 2, pp. 1525-1530, 1997
- [16] Matsumoto-Y, et al., "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement". Proc. the ICAFG. pp. 499-504, 2000
- [17] Newman-R et al., "Real-time stereo tracking for head pose and gaze

- estimation". Proceedings the 4th ICAFGR. pp. 122-8, 2000
- [18] Betke-M et al., Gaze detection via self-organizing gray-scale units. Proc. Int. Workshop on Recog., Analy., and Tracking of Faces and Gestures in Real-Time System. pp. 70-6, 1999
- [19] K. R. Park et al., "Intelligent Process Control via Gaze Detection Technology". EAAI, Vol. 13, No. 5, pp. 577-587, 2000
- [20] T. BROIDA et al., "Recursive 3-D Motion Estimation from a Monocular Image Sequence". IEEE Trans. Aerospace and Electronic Systems, Vol. 26, No. 4, pp. 639-656, 1990
- [21] T. Fukuhara et al., "3D-motion estimation of human head for model-based image coding". IEE Proc., Vol. 140, No. 1, pp. 26-35, 1993
- [22] K. R. Park et al., "Facial and Eye Gaze detection". LNCS, Vol.2525, pp. 368-376, 2002
- [23] <http://www.polhemus.com>

박강령(Kang Ryoung Park)

정회원



1994년 2월 연세대학교 전자공학과 (학사)

1996년 2월 연세대학교 대학원 전기·컴퓨터 공학과 (석사)

2000년 2월 연세대학교 대학원 전기·컴퓨터 공학과

(박사)

2000년 3월 ~ 2003년 2월 LG전자기술원 Digital Vision Group 홍채인식팀 선임연구원

2003년 3월 ~ 현재 상명대학교 미디어학부 전임강사

<주관심분야> 영상처리, 컴퓨터비전, 컴퓨터그래픽스