

휴먼 헤드포즈 정보를 이용한 3차원 공간 내 응시점 추정

신채림 · 윤상석*

신라대학교

Estimation of a Gaze Point in 3D Coordinates using Human Head Pose

Chae-Rim Shin · Sang-Seok Yun*

Silla University

E-mail : cofla1324@naver.com / ssyun@silla.ac.kr

요 약

본 논문은 실내 공간에서 상호작용 로봇이 사용자의 시선이 응시하는 목표지점의 위치정보를 추정하는 방법을 제안한다. 저가의 웹캠으로부터 RGB 영상을 추출하고, 얼굴검출(Openface)모듈로부터 사용자의 헤드포즈 정보를 획득한 후 기하학적 연산을 적용하여 3차원 공간 내 사용자의 응시방향을 추정하게 된다. 추정된 응시방향과 테이블 상의 평면과의 상관관계를 통하여 최종적으로 사용자가 응시하는 목표 지점의 좌표를 추정하게 된다.

ABSTRACT

This paper proposes a method of estimating location of a target point at which an interactive robot gazes in an indoor space. RGB images are extracted from low-cost web-cams, user head pose is obtained from the face detection (Openface) module, and geometric configurations are applied to estimate the user's gaze direction in the 3D space. The coordinates of the target point at which the user stares are finally measured through the correlation between the estimated gaze direction and the plane on the table plane.

키워드

Head pose, Openface, Gaze, Tracking

I. 서 론

응시, 시선과 같은 사회적 신호를 이용하여 사용자의 현재 관심을 다른 사람과 공유하는 과정인 공동 주의는 아동 발달의 중요한 요소로 인식된다. 이러한 공동 주의가 부족한 자폐 스펙트럼 장애 아동은 다른 사람과의 의사소통과 상호작용에 어려움을 겪게 된다. 로봇은 이러한 어려움이 있는 아동의 상호 작용 기반 주의력을 높이는 보조자로서 검토되어지고 있다[1]. 이에, 자폐 스펙트럼 장애를 가진 아동은 공동 관심 및 사회적 상호작용의 향상을 위해 웹캠을 이용하여 사용자가 응시하는 방위 정보로부터 공간 내 목표 물체를 추정할 수 있는 연구들이 진행되어 왔다[2].

본 연구에서는 사용자가 바라보는 응시점을 헤

드포즈 인식과 기하학적 연산을 적용하여 추정하였으며, 최종적으로 응시하는 물체의 구별 가능한 거리의 정밀도를 검증하기 위한 프로세스를 제안한다.

II. 3차원 공간 내 응시지점 추정 프로세스

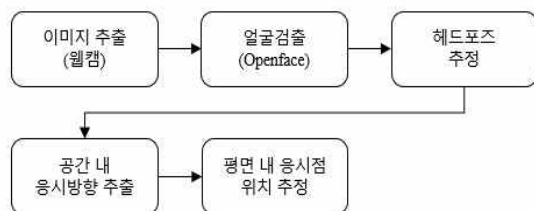


그림 1. 사용자 응시점 추정 시스템

* corresponding author

책상에 거치가 가능한 로봇의 몸통부위에 장착되는 웹캠의 RGB 영상을 통해 사용자의 이미지를 획득한 뒤, Openface 모듈로부터 사용자 얼굴의 특징점을 추출하여 사용자의 헤드 포즈를 연산한다. 추정된 헤드의 3차원 방위 정보에 기하학 연산을 적용하여 사용자가 응시하는 방향정보로부터 이 위치하는 평면 내의 응시점을 추정하게 된다.

웹캠으로부터 얻은 사용자의 헤드 위치 정보는 카메라 좌표계를 기준으로 나타나기 때문에 월드 좌표계를 기준으로 하는 사용자의 응시 목표 지점의 좌표를 획득하기 위해서 동차 행렬을 이용한 좌표계의 회전이 필요하다. 로봇플랫폼을 설계할 때 사용하는 치수정보를 바탕으로 로봇의 각 좌표계에 대한 행렬의 내적 곱을 계산하게 되며, 이를 통하여 사용자의 최종 응시점을 추정하게 된다.

전역 좌표계를 기준으로 하는 사용자 헤드 좌표의 연산과정은 수식 (1)과 같다. 여기서 W는 전역 좌표계, B는 Base_link, C는 Body_cam_link, H는 Head의 위치정보를 지칭한다. 각 동차행렬은 Translation과 Rotation으로 구성되어 있으며, 내적을 통해 구한 ${}^W H_H$ 의 변환행렬은 수식 (2)를 통해 구할 수 있다.

$${}^W H_H = {}^W H_B \cdot {}^B H_C \cdot {}^C H_H \quad (1)$$

$${}^W H_B = \begin{bmatrix} 1 & 0 & 0 & x_B \\ 0 & 1 & 0 & y_B \\ 0 & 0 & 1 & z_B \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$${}^B H_C = \begin{bmatrix} \cos\phi & \frac{-\sin(\theta+\phi)+\sin(\theta-\phi)}{2} & \frac{-\cos(\theta+\phi)+\cos(\theta-\phi)}{2} & x_C \\ \sin\phi & \frac{\cos(\theta+\phi)+\cos(\theta-\phi)}{2} & \frac{-\sin(\theta+\phi)-\sin(\theta-\phi)}{2} & y_C \\ 0 & \sin\theta & \cos\theta & z_C \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$${}^C H_H = \begin{bmatrix} \cos\beta & -\sin\beta & 0 & x_H \\ \frac{\sin(\alpha+\beta)-\sin(\alpha-\beta)}{2} & \frac{\cos(\alpha+\beta)+\cos(\alpha-\beta)}{2} & -\sin\alpha & y_H \\ \frac{-\cos(\alpha+\beta)+\cos(\alpha-\beta)}{2} & \frac{\sin(\alpha+\beta)+\sin(\alpha-\beta)}{2} & \cos\alpha & z_H \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

여기서 위치정보 x, y, z는 로봇에 설정한 좌표계 간의 Translation을 의미하며, θ , β 와 ϕ , α 는 각 B와 C를 기준으로 Z축과 X축에 대해 Rotation하는 방위 각도를 의미한다. 그리고, 위치정보 x_H, y_H, z_H 는 C를 기준으로 얻을 수 있는 사람의 헤드포즈의 3차원 위치 값을 의미한다.

$$P_x = x - \frac{z}{\tan(\theta)} \quad (3)$$

$$P_y = \sqrt{\left(\frac{z}{\tan(\theta)}\right)^2 + z^2 * \tan(\psi)} \quad (4)$$

따라서, 상기 좌표변환 공식을 이용하여 전역 좌표계로부터 사용자의 헤드 포즈의 값을 연산할 수 있다. 추출된 값을 삼각함수 공식을 이용하여 평면 공간에서 사용자 응시점 정보로 변환하게 된다. 응시점의 위치값은 수식 (3), (4)를 이용하여 추정하게 되는데, 각각 전역 좌표계를 기준으로 응시점의

평면 위치 정보(P_x, P_y)를 추출하게 된다. 수식(3)과 (4)에서 θ, ψ 는 각각 y축과 z축을 기준으로 회전하는 pitch와 yaw를 의미한다.

III. 실험 구성

응시점 정밀도 실험을 수행하기 위해 몸체에 웹캠을 장착한 탁상용 로봇 플랫폼을 책상 위의 평면에 위치시켰으며, 실험에 참가하는 사용자는 의자에 앉아서 책상 위 설정된 마크를 차례대로 응시하는 방법으로 실험을 진행하였다. 사용자가 응시해야하는 마킹된 9개의 좌표는 오차 추정을 위해 전역 좌표계를 기준으로 일정 거리만큼 떨어져 있도록 설정하였다(그림 2 참조).

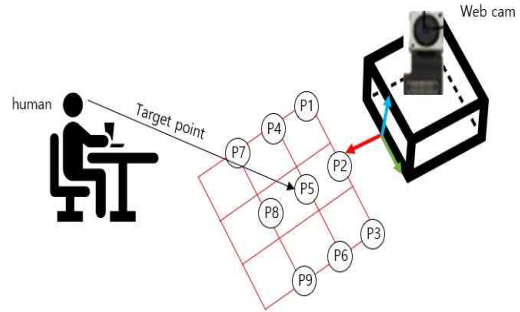


그림 2. 응시점 추정을 위한 기하학 구성도

마킹된 9개 좌표 중 좌표 P1은 전역 좌표계를 기준으로 (300mm, -150mm), 좌표 P2는 (300mm, 0mm), 좌표 P3는 (300mm, 150mm) 순으로 y축 방향으로 150mm의 간격을 두고 위치하며, 좌표 P4부터는 x축을 기준으로 100mm 간격을 두면서 순차적으로 위치를 설정하여 구성하였다.

목표 응시점의 정밀도를 측정하기 위한 실험은 20대인 3명의 참가자를 대상으로 진행하였으며, 참가자는 정해진 9개의 지점을 5초씩 순차적으로 응시하였고, 총 3회씩 반복하여 좌표당 9번의 위치를 측정하도록 하였다.

IV. 실험 결과

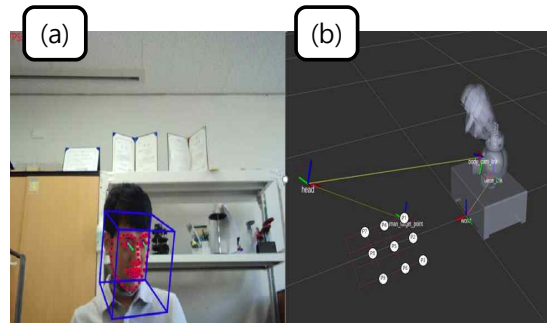


그림 3. 좌표변환 모델기반의 사용자의 응시점 추정 시뮬레이션

a) 헤드포즈 인식, b) 3D 내 응시점 추정

실제 9개 지점의 위치 좌표와 연산을 통해 획득한 추정 좌표의 오차를 측정하기 위해 사용자가 각 좌표를 응시하였을 때 사용자 시선의 위치, 그에 대한 평균과 표준편차를 계산하였다.

헤드포즈 인식을 통해 각 좌표별 사용자가 응시한 9개의 데이터의 응시점 추정은 그림 3과 같고, 각 응시점의 평균과 표준편차는 표 1에 표기한 바와 같다.

표 1. 목표 응시점에 대한 평균과 표준편차

마크	위치 (x, y)[mm]	평균±표준편차[mm]
P1	300, -150	358.61±48.52, -177.51±31.13
P2	300, 0	347.27±40.00, 30.00±37.75
P3	300, 150	330.31±75.79, 252.25±68.83
P4	400, -150	449.74±45.09, -135.20±36.14
P5	400, 0	410.68±43.59, 51.57±35.11
P6	400, 150	411.33±58.84, 252.02±122.20
P7	500, -150	503.72±41.73, -134.20±64.59
P8	500, 0	507.46±50.57, 28.08±53.90
P9	500, 150	480.99±44.53, 248.03±77.17

9개의 마크 위치에 따른 거리 오차에 대한 분석을 수행한 결과, 근거리 세 개의 좌표(P1, P2, P3)를 제외한 나머지 6개 좌표의 거리 오차는 평균 70.48mm 가 발생한 반면에, 3개의 좌표의 경우 평균 128.88mm 정도로 거리 오차가 상대적으로 크게 발생함을 확인하였다. 이는 사용자 헤드포즈 인식에서 3개 좌표의 yaw값이 나머지 좌표의 yaw값에 비해 평균 9.26도 높게 측정됨으로써 발생하는 오차이며, 동일 위치임에도 특정 참가자에 의해 yaw값이 높게 측정되는 현상에 의한 것으로 판단된다. 따라서, 사용자 별 인식되는 측정값이 가변적이고, 이는 근거리 물체의 응시점 오차의 증가로 응시점의 구분가능한 범위가 상대적으로 커지게 되는 취약점을 가지게 됨을 검증하였다.

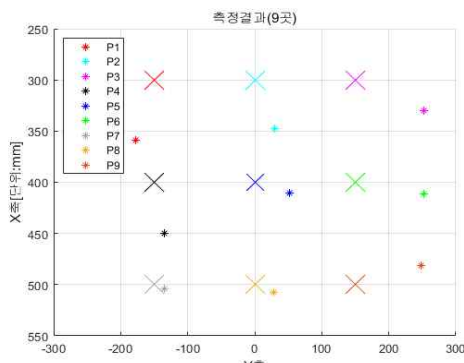


그림 4. 사용자 응시점 추정 평균 실험값 결과

응시점의 오차유발 요인은 Openface를 기반으로 얼굴검출정보를 이용하는 부분에서도 영향을 주는

것으로 분석되었다. 사용자의 응시점 추정을 위해 눈동자 응시정보를 3차원 좌표변환에 사용해야 하지만, 본 연구에서는 원거리에서도 응시정보의 추출 가능성을 높이기 위해 사용자 얼굴의 특징점을 매핑한 후 코의 위치를 기반으로 3차원 위치정보를 추출하도록 구성하였다. 따라서, 사용자의 눈동자와 코의 응시 시작지점이 상이한 부분과 응시방향의 수평 상태 유지여부가 응시점에 대한 편차를 가지는 것으로 판단된다. 이를 고려하여 실험을 진행한 결과 현재 프로세스에서는 마크된 좌표에 대한 위치가 약 70mm 이상인 경우 사용자가 어느 곳을 응시하고 있는지 정확하게 추정할 수 있음을 확인하였다.

V. 결 론

본 연구는 사용자가 지향하는 목표 응시점에 대해 얼굴인식 모듈과 기하학적 모델을 사용하여 추정 하였으며, 위치 정밀도의 실험결과를 토대로 사용자와 로봇이 마주보는 책상 위 평면상에 70mm의 거리오차를 가지는 물체의 위치를 구분할 수 있음을 확인하였다.

오차 발생의 원인으로 사용자의 눈동자 인식에 대한 고려, 그리고 웹캠 기반 사용자 얼굴 이미지 내 방위각 편차에 따른 응시점 오차의 연관관계가 존재함을 분석하였다.

향후에는 사용자 눈동자의 3차원 위치 정보와 방위 값을 이용하여 현재 발생하는 오차값을 개선시켜 사용자의 응시점의 구분 가능 거리를 줄일 수 있는 추가적인 연구를 진행할 계획이다.

Acknowledgement

본 연구는 2021년도 정부(과학기술정보통신부)의 지원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF- 2021R1F1A1063669).

References

[1] Evren Daglarh, Sare Funda Daglarh, Gülay Öke Günel, HaticeKöse, "Improving human-robot interaction based on joint attention", Applied intelligence, vol. 47 No. 1, 62-82, 2017, Jul.

[2] Da-Young Kim, Sang-Seok Yun, "Estimating user psychological state for joint attention using a robot in behavioral intervention", 한국정보통신학회논문지(J. Korea. Inst. Inf. Commun. Eng.) Vol. 23, No. 1: 399-406, 2019, Mar.

[3] Jae-Han Kim, "Design and Implementation of Real-time three dimensional Tracking system of gazing point", 대한전자공학회 학술대회, 2605-2608 (4 pages), 2003, 7