



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월14일  
(11) 등록번호 10-2420924  
(24) 등록일자 2022년07월11일

(51) 국제특허분류(Int. Cl.)  
G06K 9/00 (2022.01) G06N 3/08 (2006.01)  
(52) CPC특허분류  
G06V 40/18 (2022.01)  
G06N 3/08 (2013.01)  
(21) 출원번호 10-2020-0132465  
(22) 출원일자 2020년10월14일  
심사청구일자 2020년10월14일  
(65) 공개번호 10-2021-0155317  
(43) 공개일자 2021년12월22일  
(30) 우선권주장  
1020200072251 2020년06월15일 대한민국(KR)  
(56) 선행기술조사문헌  
KR1020190100982 A\*  
K. Wang et al., 'Deep eye fixation map learning for calibration-free eye gaze tracking,' In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Application, 2016.03., pp.47-55 (2016.03.14.) 1부.\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
세종대학교산학협력단  
서울특별시 광진구 능동로 209 (군자동, 세종대학교)  
(72) 발명자  
김용국  
경기도 성남시 분당구 동판교로 156, 911동 503호(삼평동, 봇들마을9단지금호어울림아파트)  
(74) 대리인  
송인호, 최관락

전체 청구항 수 : 총 6 항

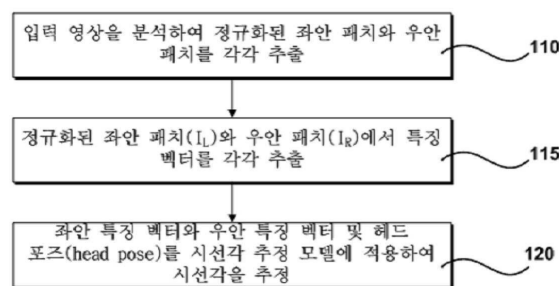
심사관 : 황승희

(54) 발명의 명칭 딥러닝 기반 3D 시선 예측 방법 및 그 장치

(57) 요약

딥러닝 기반 3D 시선 예측 방법 및 그 장치가 개시된다. 딥러닝 기반 3D 시선 예측 방법은 가시광 입력 영상에서 눈 영역을 각각 검출한 후 정규화하여 우안 패치 및 좌안 패치를 각각 추출하는 단계; 상기 우안 패치 및 상기 좌안 패치를 우안 특징 추출 모듈과 좌안 특징 추출 모듈에 적용하여 우안 특징 벡터와 좌안 특징 벡터를 각각 추출하는 단계; 및 상기 우안 특징 벡터, 상기 좌안 특징 벡터 및 헤드 포즈를 학습된 시선각 추정 모델에 적용하여 시선각을 추정하는 단계를 포함한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	1711090675
과제번호	2017-0-00731-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발(R&D)
연구과제명	딥러닝 기반 시청자 주의 및 감정상태 판별을 이용한 맞춤형 광고 플랫폼 연구 개발
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31
공지예외적용	: 있음

---

## 명세서

### 청구범위

#### 청구항 1

가시광 입력 영상에서 눈 영역을 각각 검출한 후 정규화하여 우안 패치 및 좌안 패치를 각각 추출하는 단계;

상기 우안 패치 및 상기 좌안 패치를 우안 특징 추출 모듈과 좌안 특징 추출 모듈에 적용하여 우안 특징 벡터와 좌안 특징 벡터를 각각 추출하는 단계; 및

상기 우안 특징 벡터, 상기 좌안 특징 벡터 및 헤드 포즈를 학습된 시선각 추정 모델에 적용하여 시선각을 추정하는 단계를 포함하되,

상기 우안 특징 추출 모듈과 상기 좌안 특징 추출 모듈은 상기 우안 패치 및 상기 좌안 패치에 대해 5개의 컨볼루션 레이어를 통과한 결과값과 공간 주의 모듈을 적용한 결과값인 공간 가중치 매트릭스를 시그모이드 연산한 최종 가중 활성화맵을 각각 특징 벡터로 생성하되,

상기 우안 패치 및 상기 좌안 패치는 기준점(x)에 상응하여 헤드 롤(head roll)이 제거된 상태에서 중심(p)에서 일정 사이즈로 잘린 정규화된 패치 영상이며,

상기 시선각은 요(yaw) 및 피치(pitch)인 것을 특징으로 하는 시선 예측 방법.

#### 청구항 2

제1 항에 있어서,

상기 시선각 추정 모델은,

3개의 완전 연결층과 선형 회귀 모듈을 포함하되,

상기 우안 특징 벡터와 상기 좌안 특징 벡터 및 상기 헤드 포즈는 상기 3개의 완전 연결층을 통해 결합된 후 학습된 상기 선형 회귀 모듈에 적용되어 시선각이 추정되는 것을 특징으로 하는 시선 예측 방법.

#### 청구항 3

제2 항에 있어서,

상기 선형 회귀 모듈은 예측된 시선각과 실제 시선각을 이용하여 계산되는 손실 함수가 최소가 되도록 학습되는 것을 특징으로 하는 시선 예측 방법.

#### 청구항 4

삭제

#### 청구항 5

삭제

#### 청구항 6

제1 항 내지 제3 항 중 어느 하나의 항에 따른 방법을 수행하기 위한 프로그램 코드를 기록한 컴퓨터로 판독 가능한 기록매체 제품.

**청구항 7**

가시광 입력 영상에서 눈 영역을 각각 검출한 후 정규화하여 우안 패치 및 좌안 패치를 각각 추출하는 전처리부;

상기 우안 패치를 분석하여 우안 특징 벡터를 추출하는 우안 특징 추출 모듈과 상기 좌안 패치를 분석하여 좌안 특징 벡터를 추출하는 좌안 특징 추출 모듈을 포함하는 특징 추출부; 및

상기 우안 특징 벡터, 상기 좌안 특징 벡터 및 헤드 포즈를 모델에 적용함으로써 선형 회귀 분석을 통해 시선각을 추정하는 시선각 추정부를 포함하되,

상기 우안 특징 추출 모듈과 상기 좌안 특징 추출 모듈은 상기 우안 패치 및 상기 좌안 패치에 대해 5개의 컨볼루션 레이어를 통과한 결과값과 공간 주의 모듈을 적용한 결과값인 공간 가중치 매트릭스를 시그모이드 연산한 최종 가중 활성화맵을 각각 특징 벡터로 생성하되,

상기 전처리부는 입력 영상에서 얼굴 영역을 검출하고, 상기 검출된 얼굴 영역에서 양쪽 눈 영역을 각각 검출한 후 기준점(x)에 상응하여 헤드 롤(head roll)을 제거한 상태에서 중심(p)에서 일정 사이즈로 잘라 정규화하여 상기 우안 패치 및 상기 좌안 패치를 각각 추출하되,

상기 시선각은 요(yaw) 및 피치(pitch)인 것을 특징으로 하는 시선 예측 장치.

**청구항 8**

제7 항에 있어서,

상기 모델은,

3개의 완전 연결층과 선형 회귀 모듈을 포함하되,

상기 우안 특징 벡터와 상기 좌안 특징 벡터 및 상기 헤드 포즈는 상기 3개의 완전 연결층을 통해 결합된 후 학습된 상기 선형 회귀 모듈에 적용되어 시선각이 추정되는 것을 특징으로 하는 시선 예측 장치.

**청구항 9**

삭제

**청구항 10**

삭제

**발명의 설명**

**기술 분야**

[0001] 본 발명은 딥러닝 기반 3D 시선 예측 방법 및 그 장치에 관한 것이다.

**배경 기술**

[0003] 시각 및 인지 처리 측면에서 눈의 움직임과 시선 추정은 매우 중요하다. 특히, 안구 운동은 인간의 시선 집중, 감정 분석 및 행동장애 식별 연구를 위해 폭넓게 연구되어 왔다.

[0004] 시선 추정은 인간-컴퓨터 상호 작용, 심리학, 장애 연구, 내비게이션, 운전자 행동 감지, 로봇 수술 및 마케팅 연구 등과 같이 광범위한 어플리케이션을 가지고 있기 때문에 컴퓨터 비전 영역에서 연구되어 있다.

[0005] 시선 예측을 위한 이전 모델 및 특징 기반 방법들은 조명 조건, 카메라 보정 방법, 및 개별 머리 포즈 변화에 따라 제한을 가지고 있다. 컴퓨터 비전 연구자들은 최근 대규모 시선 데이터 세트의 가용성으로 인해 일반적으로 컨볼루션 신경망(CNN)을 사용하여 제어되지 않는 환경에서 사람의 시선을 추정하는 외모 기반 방법을 탐색했다.

[0006] 딥러닝 접근 방식이 자연 환경에서 인간의 시선을 추정하는데 놀라운 성공을 거두었음에도 불구하고 현재 접근 방법은 약3.6도의 정확도를 달성하는데 그치고 있어 실시간 어플리케이션에 적용하기는 어려운 실정이다.

**발명의 내용**

**해결하려는 과제**

[0008] 본 발명은 웹캠 영상만으로 관찰자의 시선을 정확하게 예측할 수 있는 딥러닝 기반 3D 시선 예측 방법 및 그 장치를 제공하기 위한 것이다.

[0009] 또한, 본 발명은 다양한 자연 얼굴 영상만을 이용해 모델을 학습시키며, 실행 시간이 길지 않고 성능이 우수한 딥러닝 기반 3D 시선 예측 방법 및 그 장치를 제공하기 위한 것이다.

**과제의 해결 수단**

[0011] 본 발명의 일 측면에 따르면, 웹캠 영상만으로 관찰자의 시선을 정확하게 예측할 수 있는 딥러닝 기반 3D 시선 예측 방법이 제공된다.

[0012] 본 발명의 일 실시예에 따르면, 가시광 입력 영상에서 눈 영역을 각각 검출한 후 정규화하여 우안 패치 및 좌안 패치를 각각 추출하는 단계; 상기 우안 패치 및 상기 좌안 패치를 우안 특징 추출 모듈과 좌안 특징 추출 모듈에 적용하여 우안 특징 벡터와 좌안 특징 벡터를 각각 추출하는 단계; 및 상기 우안 특징 벡터, 상기 좌안 특징 벡터 및 헤드 포즈를 학습된 시선각 추정 모델에 적용하여 시선각을 추정하는 단계를 포함하는 3D 시선 예측 방법이 제공될 수 있다.

[0013] 상기 시선각 추정 모델은, 3개의 완전 연결층과 선형 회귀 모듈을 포함하되, 상기 우안 특징 벡터와 상기 좌안 특징 벡터 및 상기 헤드 포즈는 상기 3개의 완전 연결층을 통해 결합된 후 학습된 상기 선형 회귀 모듈에 적용되어 시선각이 추정될 수 있다.

[0014] 상기 선형 회귀 모듈은 예측된 시선각과 실제 시선각을 이용하여 계산되는 손실 함수가 최소가 되도록 학습될 수 있다.

[0015] 상기 시선각은 요(yaw) 및 피치(pitch)이다.

[0016] 상기 우안 패치 및 상기 좌안 패치는 기준점(x)에 상응하여 헤드 롤(head roll)이 제거된 상태에서 중심(p)에서 일정 사이즈로 잘린 정규화된 패치 영상일 수 있다.

[0018] 본 발명의 다른 측면에 따르면, 웹캠 영상만으로 관찰자의 시선을 정확하게 예측할 수 있는 장치가 제공된다.

[0019] 본 발명의 일 실시예에 따르면, 가시광 입력 영상에서 눈 영역을 각각 검출한 후 정규화하여 우안 패치 및 좌안 패치를 각각 추출하는 전처리부; 상기 우안 패치를 분석하여 우안 특징 벡터를 추출하는 우안 특징 추출 모듈과 상기 좌안 패치를 분석하여 좌안 특징 벡터를 추출하는 좌안 특징 추출 모듈을 포함하는 특징 추출부; 및 상기 우안 특징 벡터, 상기 좌안 특징 벡터 및 헤드 포즈를 모델에 적용함으로써 선형 회귀 분석을 통해 시선각을 추정하는 시선각 추정부를 포함하는 3D 시선 예측 장치가 제공될 수 있다.

[0020] 상기 모델은, 3개의 완전 연결층과 선형 회귀 모듈을 포함하되, 상기 우안 특징 벡터와 상기 좌안 특징 벡터 및 상기 헤드 포즈는 상기 3개의 완전 연결층을 통해 결합된 후 학습된 상기 선형 회귀 모듈에 적용되어 시선각이 추정될 수 있다.

[0021] 상기 전처리부는, 상기 입력 영상에서 얼굴 영역을 검출하고, 상기 검출된 얼굴 영역에서 양쪽 눈 영역을 각각 검출한 후 기준점(x)에 상응하여 헤드 롤(head roll)을 제거한 상태에서 중심(p)에서 일정 사이즈로 잘라 정규화하여 상기 우안 패치 및 상기 좌안 패치를 각각 추출할 수 있다.

**발명의 효과**

[0023] 본 발명의 일 실시예에 따른 딥러닝 기반 3D 시선 예측 방법 및 그 장치를 제공함으로써 가시광 영상을 이용하여 정확하게 관찰자의 시선을 예측할 수 있는 이점이 있다.

[0024] 또한, 본 발명은 적외선 영상 이용 없이 다양한 가시광 얼굴 영상만을 이용하여 모델을 학습시켜 실행 시간이 길지 않으면서도 성능이 우수한 시선 예측이 가능한 이점이 있다.

**도면의 간단한 설명**

- [0026] 도 1은 본 발명의 일 실시예에 따른 딥러닝 기반 3D 시선 예측 방법을 나타낸 순서도.
- 도 2는 본 발명의 일 실시예에 따른 좌안 패치와 우안 패치 추출을 설명하기 위해 도시한 도면.
- 도 3은 본 발명의 일 실시예에 따른 특징 추출 모듈의 세부 구조를 도시한 도면.
- 도 4는 본 발명의 일 실시예에 따른 시선각 추정부의 상세 구성을 도시한 도면.
- 도 5는 본 발명의 일 실시예에 따른 MPIIGaze 데이터 셋 및 EYEDIAP 데이터 셋으로부터 예제 이미지를 예시한 도면.
- 도 6은 본 발명의 일 실시예에 따른 데이터 셋을 증강 방법을 설명하기 위해 도시한 도면.
- 도 7은 본 발명의 일 실시예에 따른 3D 시선 예측 장치의 구성을 도시한 도면.
- 도 8은 본 발명의 일 실시예에 따른 추정된 시선각과 실제 시선각을 비교한 도면.

**발명을 실시하기 위한 구체적인 내용**

- [0027] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0028] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.
- [0030] 도 1은 본 발명의 일 실시예에 따른 딥러닝 기반 3D 시선 예측 방법을 나타낸 순서도이고, 도 2는 본 발명의 일 실시예에 따른 좌안 패치와 우안 패치 추출을 설명하기 위해 도시한 도면이며, 도 3은 본 발명의 일 실시예에 따른 특징 추출 모듈의 세부 구조를 도시한 도면이고, 도 4는 본 발명의 일 실시예에 따른 시선각 추정부의 상세 구성을 도시한 도면이며, 도 5는 본 발명의 일 실시예에 따른 MPIIGaze 데이터 셋 및 EYEDIAP 데이터 셋으로부터 예제 이미지를 예시한 도면이고, 도 6은 본 발명의 일 실시예에 따른 데이터 셋을 증강 방법을 설명하기 위해 도시한 도면이다.
- [0031] 단계 110에서 시선 예측 장치(100)는 입력 영상을 분석하여 정규화된 좌안 패치와 우안 패치를 각각 추출한다.
- [0032] 입력 영상인 가시광 영상은 조명 조건, 카메라 조건에 따라 해상도가 상이하다. 따라서, 본 발명의 일 실시예에서는 입력 영상에서 좌안 영역과 우안 영역을 각각 검출한 후 정규화한 후 좌안 패치와 우안 패치를 각각 추출할 수 있다.
- [0033] 시선 예측 장치(100)는 가시광 영상인 입력 영상에서 얼굴 영역을 검출하고, 검출된 얼굴 영역에서 양쪽 눈 영역을 각각 검출할 수 있다. 이어, 시선 예측 장치(100)는 검출된 양쪽 눈 영역을 정규화한 후 양쪽 눈 패치를 각각 추출한다.
- [0034] 예를 들어, 시선 예측 장치(100)는 입력 영상에서 머리 롤(head roll)을 제거한 정규화된 양쪽 눈 패치를 각각 추출할 수 있다.
- [0035] 본 발명의 일 실시예에 따르면, 입력 영상은 웹캠(web cam)에서 촬영되는 가시광 영상일 수 있다.
- [0036] 시선 예측 장치(100)는 입력 영상에서 카메라 파라미터와 무관하게 외관 변화(appearance variation)를 극복하여 시선을 정확하게 예측하기 위해 양쪽 눈 패치를 추출함에 있어 머리 롤(head roll)을 제거한 정규화된 양쪽 눈 패치를 추출할 수 있다. 이에 대해 보다 상세히 설명하기로 한다.
- [0037] 입력 영상을 I라 칭하기로 한다. 입력 영상 I와 기준점(x)가 주어지는 경우, 목표는 컨버전 행렬 M(conversion matrix)을 계산하는 것이다. 여기서, 컨버전 행렬 M은 수학적 1을 이용하여 계산될 수 있다.

수학식 1

$$M = SR$$

[0038]

[0039] 회전 행렬 R을 사용하면 머리 좌표계와 카메라의 x축은 평행하다. 따라서, 스케일 행렬(S)는 가상 카메라가 고정된 거리  $d_s$ 로부터 기준점(x)를 보도록 정의될 수 있다. 여기서, 스케일 행렬(S)는 수학식 2와 같이 나타낼 수 있다.

수학식 2

$$S = \text{diag}(1, 1, d_x / \|P\|)$$

[0040]

[0041] 여기서,  $\text{diag}()$ 는 대각 행렬을 나타내고,  $\|P\|$ 는 기준점으로부터의 이동(translation)을 나타내는 행렬이다.

[0042] 시선 예측 장치(100)는 입력 영상 I에 대해 변형 행렬 W(transformation matrix)를 사용하여 투영 왜곡(perspective warping)을 통해 정규화할 수 있다. 이를 수학식으로 나타내면 수학식 3과 같다.

수학식 3

$$W = C_s M C_n^{-1}$$

[0043]

[0044] 여기서,  $C_s$ 는 정규화된 카메라의 투영 행렬을 나타내고,  $C_n$ 는 실제 카메라 행렬을 나타낸다.

[0045] 결과적으로 시선 예측 장치(100)는 가시광 입력 영상에서 양쪽 눈 영역을 각각 검출한 후 기준점(x)에 대한 머리 롤(head roll)이 제거된 상태에서 중심(p)에서 일정 사이즈(W x H)를 가지는 잘린 패치로 양쪽 눈 패치를 각각 추출할 수 있다.

[0046] 도 2에는 양쪽 눈 패치 추출에 대한 일 예가 도시되어 있다. 시선 예측 장치(100)는 웹캠 영상에서 얼굴 영역을 검출하고, 검출된 얼굴 영역에서 양쪽 눈 영역을 검출한 후 기준점(x)에 대해 머리 롤(head roll)이 제거된 정규화된 눈 패치( $I_L$ ,  $I_R$ )를 각각 추출할 수 있다.

[0047] 단계 115에서 시선 예측 장치(100)는 정규화된 좌안 패치( $I_L$ )와 우안 패치( $I_R$ )에서 특징 벡터를 각각 추출한다.

[0048] 정규화된 좌안 패치와 우안 패치는 각각의 특징 추출부로 입력되며, 각각의 특징 추출부에 의해 특징 벡터가 각각 추출될 수 있다. 편의상 이하에서는 좌안 특징 벡터와 우안 특징 벡터라 칭하기로 한다.

[0049] 좌안 패치와 우안 패치는 좌안 특징 추출 모듈과 우안 특징 추출 모듈로 각각 입력된다. 여기서, 좌안 특징 추출 모듈과 우안 특징 추출 모듈은 서로 가중치를 공유하지 않는다. 즉, 좌안 특징 추출 모듈과 우안 특징 추출 모듈은 서로 다른 가중치를 가지도록 학습될 수 있음은 당연하며, 독립적으로 학습/동작될 수 있다.

[0050] 좌안 특징 추출 모듈과 우안 특징 추출 모듈의 세부 구성은 동일하다. 좌안 특징 추출 모듈과 우안 특징 추출 모듈의 구조는 도 3에 도시된 바와 같다.

[0051] 특징 추출 모듈은 특징 추출을 위해 5개의 컨볼루션 레이어를 포함하며, 공간 주의 모듈의 출력값인 공간 가중

치 매트릭스가 5개의 컨볼루션 레이어를 통과한 결과값(활성맵)과 시그모이드 연산되어 최종 가중 활성화맵이 특징 벡터로 각각 생성될 수 있다.

[0052] 공간 주의 모듈은 필터 크기가 1 x 1인 3개의 컨볼루션 레이어와 정제된 선형 유닛(ReLU)로 구성된다. 따라서, 특징 추출 모듈은 5개의 컨볼루션 레이어를 통과한 결과값(활성화맵)과 공간 주의 모듈을 적용한 결과값인 공간 가중치 매트릭스를 이용하여 최종 가중 활성화 맵이 생성될 수 있다. 이를 수학식으로 나타내면 수학식 4와 같다.

**수학식 4**

$$V = W \odot U$$

[0053] 여기서, W는 공간 주의 모듈의 출력 결과인 공간 가중치 매트릭스를 나타내고, U는 5개의 컨볼루션 레이어를 통과한 활성화 맵을 나타낸다.

[0054] 가중 활성화 맵은 최종 최대 풀링 계층을 통과하며 차원이 축소될 수 있다. 공간 가중치 매트릭스와 최종 가중 활성화 맵의 각 요소별 곱셈을 통해 최대 풀링 계층을 통과함으로써 특징 벡터의 차원(크기)가 감소될 수 있다.

[0055] 단계 120에서 시선 예측 장치(100)는 좌안 특징 벡터와 우안 특징 벡터 및 헤드 포즈(head pose)를 시선각 추정 모델에 적용하여 시선각을 추정한다.

[0056] 여기서, 추정된 시선각은 요(yaw)와 피치(pitch)일 수 있다. 여기서, 요(yaw)와 피치(pitch)는 눈동자 움직임에 따른 요(yaw)와 피치(pitch)일 수 있다.

[0057] 결과적으로 본 발명의 일 실시예에 따르면, 3D 시선 예측 장치(100)는 2D 가시광 입력 영상을 이용하여 눈동자 움직임을 검출하여 3D 시선각을 추정할 수 있다.

[0058] 시선각 추정 모델에 대해 도 4를 참조하여 보다 상세히 설명하기로 한다.

[0059] 시선각 추정 모델은 3개의 완전 연결 층과 선형 회귀 모듈을 포함한다.

[0060] 이때, 완전 연결 층의 후단에는 BN층(Batch-normalization layer)과 ReLU층이 각각 위치될 수 있다.

[0061] 본 발명의 일 실시예에 따르면, 시선각 추정 모델에 포함된 3개의 완전 연결층의 크기는 각기 상이할 수 있다. 즉, 3개의 완전 연결층의 크기는 각각 512, 256 및 2일 수 있다. 완전 연결층의 사이즈가 반드시 이로 제한되는 것은 아니며 구현 방법에 따라 완전 연결층의 개수 및 사이즈는 상이해질 수도 있음은 당연하다.

[0062] 좌안 특징 벡터와 우안 특징 벡터를 결합하기 전에, 드롭 아웃층(p=0.5)를 512 크기의 완전 연결층에 연결한 후 BN 층과 ReLU층을 연결하였다.

[0063] 이후, 256 크기의 완전 연결 층이 추가되며, 256 크기의 완전 연결층은 BN층과 ReLU 층과 연결되고, 2 크기의 완전 연결 층이 추가될 수 있다. 또한, 본 발명의 일 실시예에 같이 BN층을 완전 연결층과 연결하여 사용하는 경우 성능이 개선되는 것을 알 수 있다.

[0064] 최종 층에 헤드 포즈가 입력될 수 있다.

[0065] 가장 후단 레이어를 통해 헤드 포즈 벡터가 입력될 수 있다.

[0066] 이로 인해, 선형 회귀 모듈은 좌안 특징 벡터, 우안 특징 벡터 및 헤드 포즈를 이용하여 시선각 요(yaw) 및 피치(pitch)를 계산할 수 있다.

[0067] 선형 회귀 모듈은 훈련 데이터 셋을 이용하여 사전 학습되어 있는 것을 가정하기로 한다. 선형 회귀 모듈은 예측된 시선각과 실제 시선각도 사이의 유클리드 거리를 추정하여 손실 함수를 계산할 수 있다. 이때, 선형 회귀 모듈은 손실 함수가 최소가 되도록 학습될 수 있다.

[0068] 예를 들어, 손실 함수는 수학식 5와 같이 정의될 수 있다.



수학식 5

$$\mathcal{L}_{ed} = \sum_{i=1}^N \|\hat{g}_i - g_i\|^2$$

[0070]

[0071] 여기서, N은 전체 이미지 개수를 나타내고,  $\hat{g}_i$  는 i번째 영상의 예측된 시선각을 나타내며,  $g_i$  는 i번째 영상의 실제 시선각을 나타낸다.

[0072] 본 발명의 일 실시예에 따르면, 시선각 추정 모델은 시선 추적에 있어 세계적으로 표준으로 이용되는 MIPGaze와 EYEDIAP 데이터 셋을 이용하여 사전에 학습될 수 있다. 도 5에는 MIPGaze와 EYEDIAP 데이터 셋에 포함된 영상들의 일 예가 도시되어 있다. MIPGaze와 EYEDIAP 데이터 셋에 각각 영상들이 포함되어 있으나, 이들만을 이용하여 모델을 학습하는 경우 정확도가 상대적으로 낮아질 수 있다. 따라서, 본 발명의 일 실시예에서는 MIPGaze와 EYEDIAP 데이터 셋 이외에도, 해당 데이터 셋의 영상에 가우시안 블러, 감마 변환, 노이즈 추가하여 데이터 셋을 늘려서 학습에 이용할 수 있다(도 6 참조).

[0073] 데이터 셋에 포함된 영상에 감마 보정된 영상을 추가함으로써 서로 다른 조명 조건(예를 들어, 어두운 조명, 밝은 조명)에 강인하게 적응하도록 모델을 훈련시킬 수 있다.

[0074] 또한, 카메라의 흐림 조건에서의 모델을 더욱 강력하게 훈련하기 위해 OpenCV 가우시안 블러를 이용하여 커널 사이즈를 7 x 7, 3 x 3 등으로 하는 보정된 영상을 생성하여 추가적인 훈련이 가능하도록 하였다. 또한, 가우시안 솔트 페퍼(salt pepper)와 같은 다양한 노이즈를 눈 패치에 추가한 영상을 이용하여 모델을 훈련할 수 있다.

[0075] 본 발명의 일 실시예에서는 이와 같이 훈련 데이터 셋을 생성함에 있어, 가우시안 블러, 감마 보정, 노이즈 추가 등을 통해 데이터를 증강하여 다양한 조명 조건, 카메라 조건에 대해서도 학습이 가능하도록 하여 모델의 정확도를 높일 수 있는 이점이 있다.

[0077] 도 7은 본 발명의 일 실시예에 따른 3D 시선 예측 장치의 구성을 도시한 도면이고, 도 8은 본 발명의 일 실시예에 따른 추정된 시선각과 실제 시선각을 비교한 도면이다.

[0078] 도 7을 참조하면, 본 발명의 일 실시예에 따른 3D 시선 예측 장치(100)는 전처리부(710), 특징 추출부(715), 시선각 추정부(720), 학습부(725), 메모리(730) 및 프로세서(735)를 포함하여 구성된다.

[0079] 전처리부(710)는 가시광 영상(입력 영상)에서 얼굴 영역을 검출하고, 검출된 얼굴 영역에서 양쪽 눈 영역(좌안 영역과 우안 영역)을 각각 검출한 후 정규화하여 좌안 패치와 우안 패치를 각각 추출하기 위한 수단이다.

[0080] 이에 대해서는 도 1을 참조하여 설명한 바와 동일하므로 중복되는 설명은 생략하기로 한다.

[0081] 특징 추출부(715)는 좌안 특징 추출 모듈과 우안 특징 추출 모듈을 포함한다. 좌안 특징 추출 모듈은 좌안 패치를 이용하여 좌안 특징 벡터를 추출하기 위한 수단이다. 또한, 우안 특징 추출 모듈은 우안 패치를 이용하여 우안 특징 벡터를 추출하기 위한 수단이다.

[0082] 특징 추출부(715)에 포함된 좌안 특징 추출 모듈과 우안 특징 추출 모듈의 세부 구성은 도 3에 도시된 바와 같다. 이에 대해서는 전술한 바와 동일하므로 중복되는 설명은 생략하기로 한다.

[0083] 시선각 추정부(720)는 선형 회귀 모듈을 포함한다. 선형 회귀 모듈 전단에는 3개의 완전 연결층이 위치되며, 3개의 완전 연결층을 통해 좌안 특징 벡터와 우안 특징 벡터가 결합될 수 있다. 또한, 최종 레이어(층)에는 헤드 포즈가 입력될 수 있다.

[0084] 따라서, 선형 회귀 모듈은 좌안 특징 벡터, 우안 특징 벡터 및 헤드 포즈를 포함하는 벡터값이 입력되며, 선형 회귀 모듈은 좌안 특징 벡터, 우안 특징 벡터 및 헤드 포즈를 이용하여 3D 시선각을 추정할 수 있다. 3D 시선각은 요(yaw) 및 피치(pitch)일 수 있다.

[0085] 학습부(725)는 훈련 데이터 셋을 이용하여 시선각 추정부(720)를 학습하기 위한 수단이다.

[0086] 학습부(725)는 손실 함수가 최소가 되도록 선형 회귀 모듈을 학습할 수 있다. 손실 함수에 대해서는 이미 전술

한 바와 동일하므로 이에 대한 추가 설명은 생략하기로 한다.

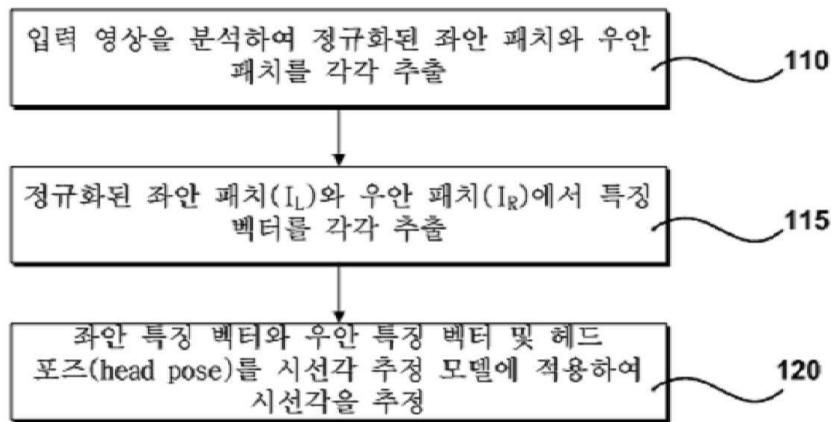
- [0087] 훈련 데이터 셋은 이미 전술한 바와 같이, MIPGaze와 EYEDIAP 데이터 셋을 이용하되, MIPGaze와 EYEDIAP을 이용하여 가우시안 블러, 감마 변환, 노이즈 추가 등을 통해 데이터를 추가하여 학습에 이용할 수 있다.
- [0088] 메모리(730)는 본 발명의 일 실시예에 따른 딥러닝 기반 3D 시선 예측 방법을 수행하기 위해 필요한 다양한 명령어들(프로그램 코드들)을 저장하기 위한 수단이다.
- [0089] 프로세서(735)는 본 발명의 일 실시예에 따른 3D 시선 예측 장치(100)의 내부 구성 요소들(예를 들어, 전처리부(710), 특징 추출부(715), 시선각 추정부(720), 메모리(730) 등)를 제어하기 위한 수단이다.
- [0090] 도 8에는 3D 시선각을 예측한 결과가 예시되어 있다. 도 8에서 그린은 예측된 시선각을 나타내고, 붉은색은 실제 시선각을 나타낸다. 도 8에서 보여지는 바와 같이, 가시광 영상에서 정확하게 3D 시선각이 예측되는 것을 알 수 있다.
- [0092] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0093] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0094] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

**부호의 설명**

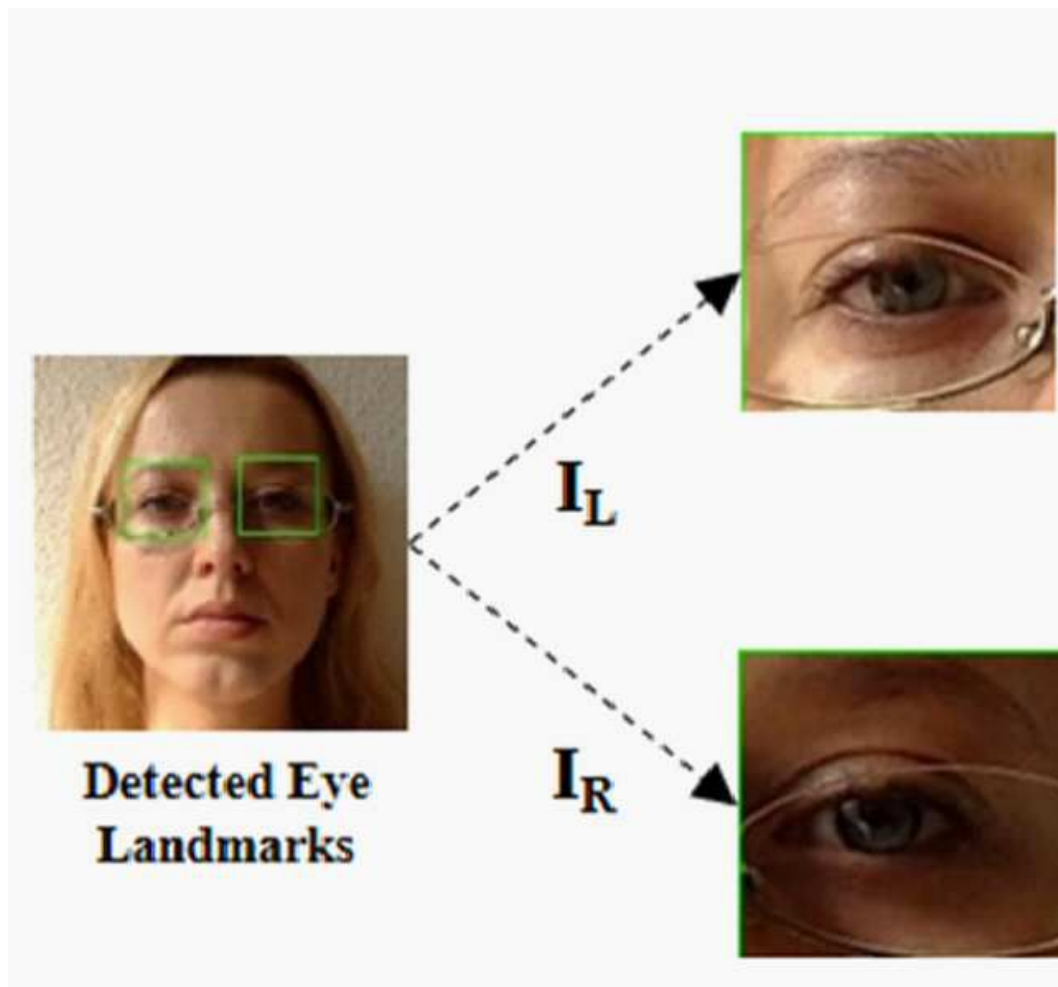
- [0096] 100: 시선 예측 장치
- 710: 전처리부
- 715: 특징 추출부
- 720: 시선각 추정부
- 725: 학습부
- 730: 메모리
- 735: 프로세서

도면

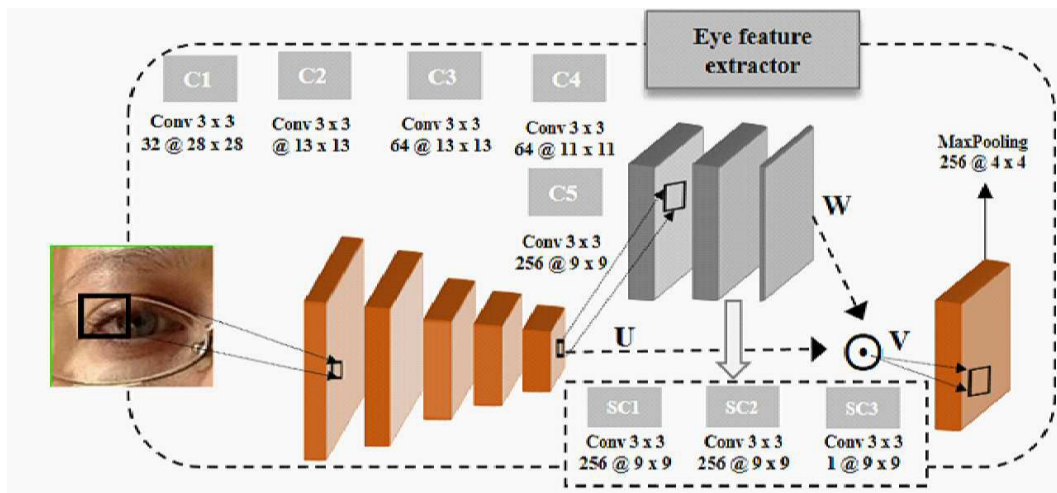
도면1



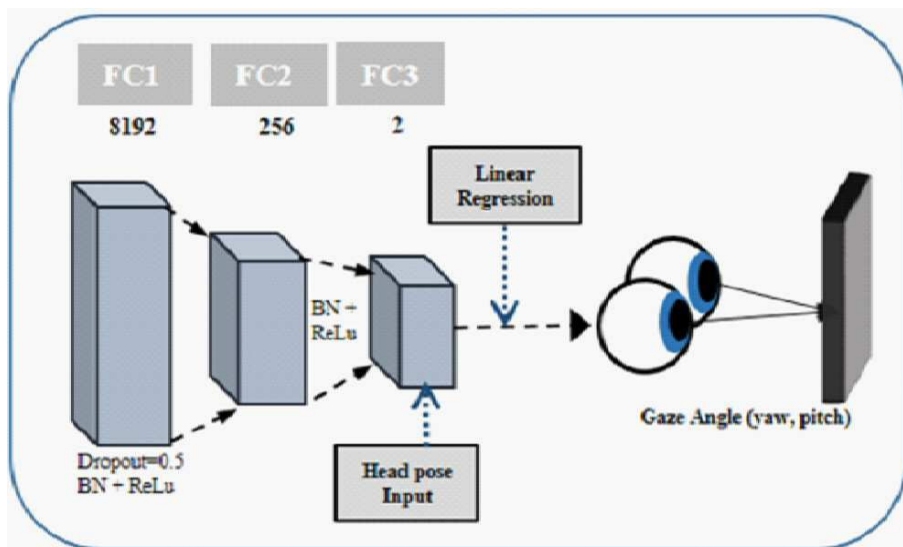
도면2



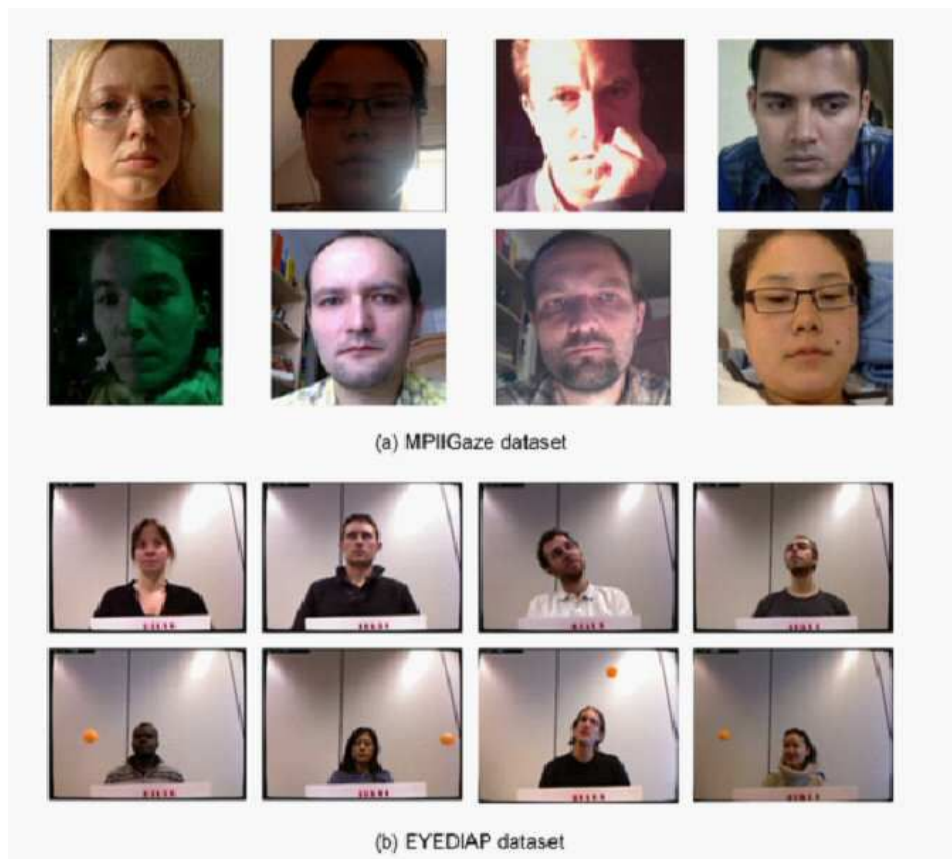
도면3



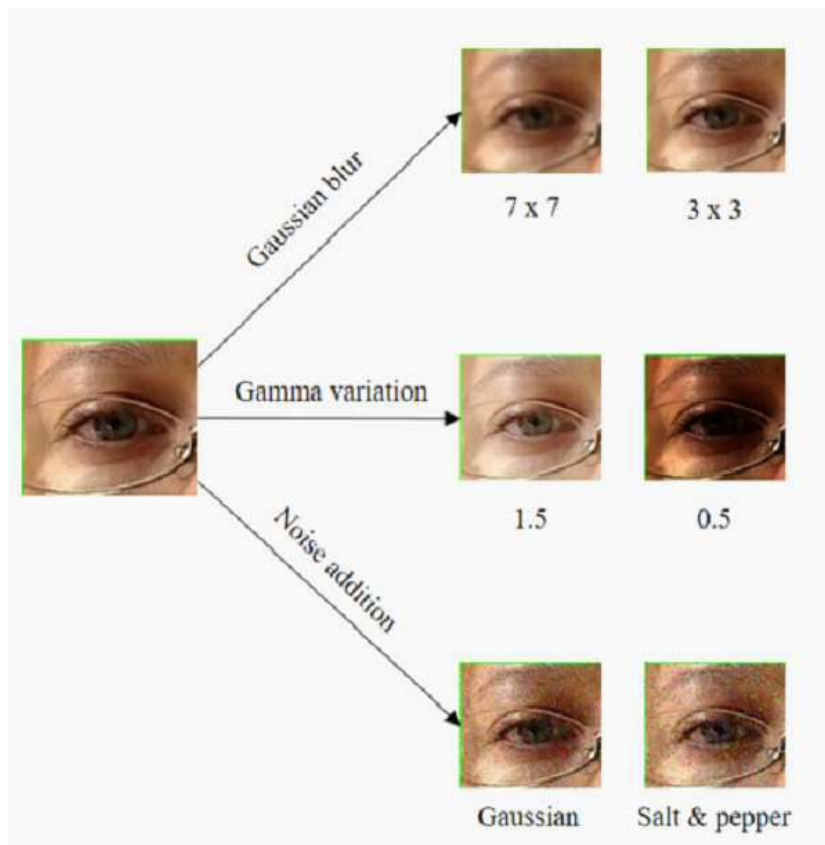
도면4



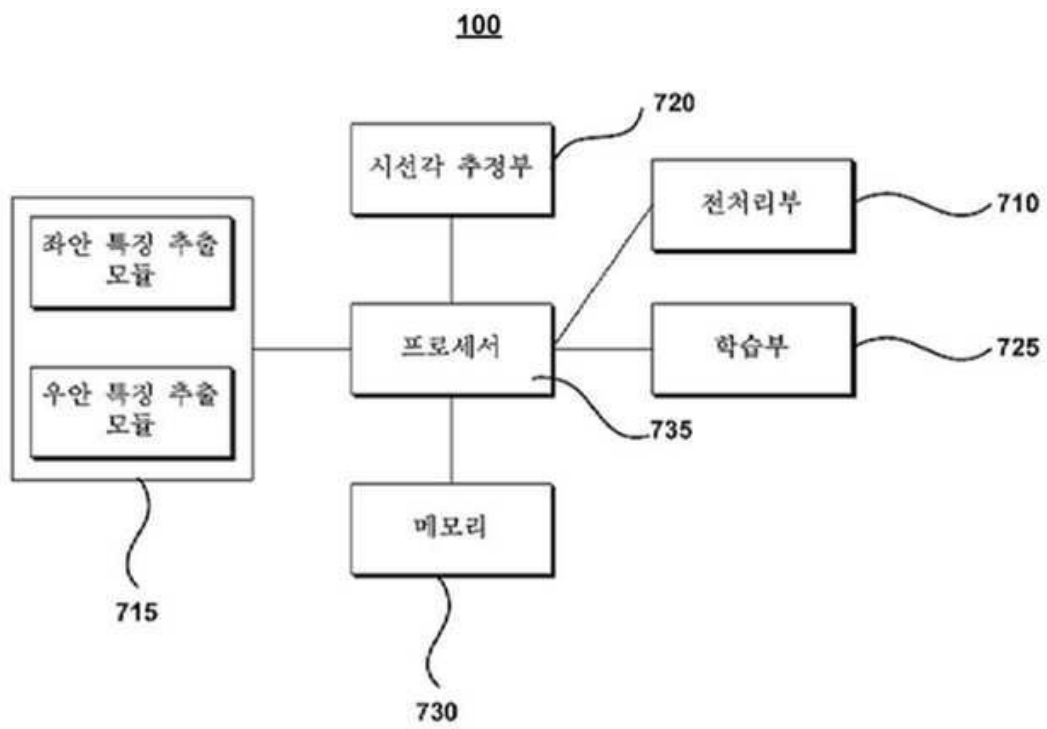
도면5



도면6



도면7



도면8

