(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0176749
(43) 공개일자 2024년12월24일

(51) 국제특허분류(Int. Cl.)
G08G 1/00 (2006.01) *G06N 20/20* (2019.01)
G06N 3/08 (2023.01) *G06Q 10/04* (2023.01)
G08G 1/01 (2006.01) *G08G 1/123* (2006.01)
G08G 1/16 (2006.01)
(52) CPC특허분류
G08G 1/20 (2013.01)
G06N 20/20 (2021.08)
(21) 출원번호 10-2023-0107732
(22) 출원일자 2023년08월17일
심사청구일자 2024년06월05일
(30) 우선권주장
1020230077567 2023년06월16일 대한민국(KR)

(71) 출원인
포티투닷 주식회사
서울특별시 강남구 남부순환로 2621, 9층(도곡동, 에스앤아이강남빌딩)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
김하영
경기도 수원시 영통구 대학3로 55 (이의동, 광교역참누리포레스트)
연규환
경기도 안양시 만안구 안양천서로 289 (안양동, 주공뜨란채)
(뒷면에 계속)
(74) 대리인
특허법인 무한

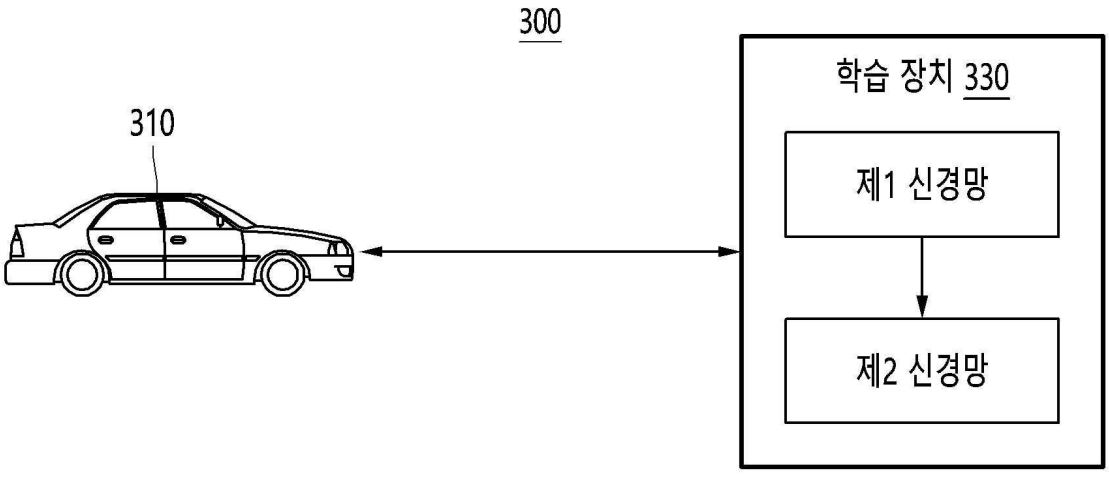
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 **차량의 주행 경로를 결정하기 위한 학습 방법 및 이를 수행하는 장치**

(57) 요 약

차량의 주행 경로를 결정하기 위한 학습 방법 및 이를 수행하는 장치가 개시된다. 일 실시예에 따른 방법은, 복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 복수의 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여, 제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성하는 동작과, 상기 제1 정보 및 상기 제1 어텐션 계수에 기초하여, 제2 신경망을 통해 제2 어텐션 계수를 생성하는 동작과, 상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다.

대 표 도 - 도3



(52) CPC특허분류
G06N 3/08 (2023.01)
G06Q 10/047 (2023.01)
G08G 1/0125 (2013.01)
G08G 1/0141 (2013.01)
G08G 1/123 (2013.01)
G08G 1/166 (2013.01)
(72) 발명자
정성균
경기도 성남시 분당구 판교역로 100 (백현동, 백현마을6단지아파트)
김진규
서울특별시 광진구 아차산로 599, 107동 801호 (광장동, 광장동신동아 파밀리에)
문석하
서울특별시 동대문구 왕산로 52, 102동 9층 1020호 (용두동,신설동역대우아이빌)

명 세 서

청구범위

청구항 1

차량의 주행 경로(driving route)를 결정하기 위한 학습 방법에 있어서,

복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 복수의 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여, 제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성하는 동작;

상기 제1 정보 및 상기 제1 어텐션 계수에 기초하여, 제2 신경망을 통해 제2 어텐션 계수를 생성하는 동작; 및

상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 상기 제2 신경망을 학습시키는 동작

을 포함하고,

상기 어텐션 계수는,

상기 주변 차량들이 상기 복수의 차량들 중 목표 차량의 주행 경로에 영향을 미치는 정도와 관련하여, 상기 주변 차량들 각각에 대해 설정된 가중치인, 학습 방법.

청구항 2

제1항에 있어서,

상기 제1 신경망은,

상기 목표 차량, 및 상기 주변 차량들 간의 상호작용을 예측하기 위한 것인, 학습 방법.

청구항 3

제1항에 있어서,

상기 제2 신경망은,

상기 목표 차량의 주행 경로를 예측하기 위한 것인, 학습 방법.

청구항 4

제1항에 있어서,

상기 제1 어텐션 계수를 생성하는 동작은,

상기 제1 정보를 인코딩(encoding)하여, 상기 제1 정보에 대응하는 제1 잠재 벡터(latent vector)를 생성하는 동작;

상기 제2 정보를 인코딩하여, 상기 제2 정보에 대응하는 제2 잠재 벡터를 생성하는 동작;

상기 제1 잠재 벡터, 및 제2 잠재 벡터를 멀티 임베딩(multi-embedding)하여, 임베딩 벡터(embedding vector)를 생성하는 동작; 및

상기 임베딩 벡터를 키(key) 및 밸류(value)로 사용하고, 상기 제1 잠재 벡터를 쿼리(query)로 사용하는 크로스 어텐션 레이어(cross attention layer)를 이용하여 상기 제1 어텐션 계수를 생성하는 동작

을 포함하는, 학습 방법.

청구항 5

제1항에 있어서,

상기 제2 어텐션 계수를 생성하는 동작은,

상기 제1 정보를 인코딩하여, 상기 제1 정보에 대응하는 제1 잠재 벡터를 생성하는 동작;

상기 제1 잠재 벡터에 기초하여, 상기 목표 차량이 주행 가능한 경로의 차선에 관한 정보를 추출하는 동작;

상기 차선에 관한 정보에 상기 목표 차량의 주행 가능성에 관한 신뢰도 값(confidence value)을 할당하는 동작; 및

상기 신뢰도 값이 할당된 정보에 기초하여, 제1 잠재 벡터를 업데이트하는 동작

을 포함하는, 학습 방법.

청구항 6

제5항에 있어서,

상기 제2 어텐션 계수를 생성하는 동작은,

상기 제1 잠재 벡터를 쿼리 및 밸류로 사용하고, 업데이트된 제1 잠재 벡터를 키로 사용하는 크로스 어텐션 레이어를 이용하여 상기 제2 어텐션 계수를 생성하는 동작

을 포함하는, 학습 방법.

청구항 7

제1항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 제1 신경망으로부터 지식 증류 방식으로(knowledge distillation) 상기 제2 신경망을 학습시키는 동작

을 포함하는, 학습 방법.

청구항 8

제7항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 증류 손실(distillation loss)을 계산하는 동작

을 포함하는, 학습 방법.

청구항 9

제8항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 증류 손실에 기초하여, 상기 제1 어텐션 계수와 상기 제2 어텐션 계수 간의 차이가 최소가 되도록 종단간

방식으로(end-to-end) 상기 제2 신경망을 학습시키는 동작

을 포함하는, 학습 방법.

청구항 10

차량의 주행 경로를 결정하기 위한 학습 장치에 있어서,

인스트럭션들을 포함하는 메모리; 및

상기 메모리와 전기적으로 연결되고, 상기 인스트럭션들을 실행하기 위한 프로세서를 포함하고,

상기 프로세서에 의해 상기 인스트럭션들이 실행될 때, 상기 프로세서는 복수의 동작들을 수행하고,

상기 복수의 동작들은,

복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여, 제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성하는 동작;

상기 제1 정보 및 상기 제1 어텐션 계수에 기초하여, 제2 신경망을 통해 제2 어텐션 계수를 생성하는 동작; 및

상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 상기 제2 신경망을 학습시키는 동작

을 포함하고,

상기 어텐션 계수는,

상기 주변 차량들이 상기 차량들 중 목표 차량의 주행 경로에 영향을 미치는 정도와 관련하여, 상기 주변 차량들 각각에 대해 설정된 가중치인, 학습 장치.

청구항 11

제10항에 있어서,

상기 제1 신경망은,

상기 목표 차량, 및 상기 주변 차량들 간의 상호작용을 예측하기 위한 것인, 학습 장치.

청구항 12

제10항에 있어서,

상기 제2 신경망은,

상기 목표 차량의 주행 경로를 예측하기 위한 것인, 학습 장치.

청구항 13

제10항에 있어서,

상기 제1 어텐션 계수를 생성하는 동작은,

상기 제1 정보를 인코딩(encoding)하여, 상기 제1 정보에 대응하는 제1 잠재 벡터(latent vector)를 생성하는 동작;

상기 제2 정보를 인코딩하여, 상기 제2 정보에 대응하는 제2 잠재 벡터를 생성하는 동작;

상기 제1 잠재 벡터, 및 제2 잠재 벡터를 멀티 임베딩(multi-embedding)하여, 임베딩 벡터(embedding vector)

를 생성하는 동작; 및

상기 임베딩 벡터를 키(key) 및 밸류(value)로 사용하고, 상기 제1 잠재 벡터를 쿼리(query)로 사용하는 크로스 어텐션 레이어(cross attention layer)를 이용하여 상기 제1 어텐션 계수를 생성하는 동작

을 포함하는, 학습 장치.

청구항 14

제10항에 있어서,

상기 제2 어텐션 계수를 생성하는 동작은,

상기 제1 정보를 인코딩하여, 상기 제1 정보에 대응하는 제1 잠재 벡터를 생성하는 동작;

상기 제1 잠재 벡터에 기초하여, 상기 목표 차량이 주행 가능한 경로의 차선에 관한 정보를 추출하는 동작;

상기 차선에 관한 정보에 상기 목표 차량의 주행 가능성에 관한 신뢰도 값(confidence value)을 할당하는 동작; 및

상기 신뢰도 값이 할당된 정보에 기초하여, 제1 잠재 벡터를 업데이트하는 동작

을 포함하는, 학습 장치.

청구항 15

제14항에 있어서,

상기 제2 어텐션 계수를 생성하는 동작은,

상기 제1 잠재 벡터를 쿼리 및 밸류로 사용하고, 업데이트된 제1 잠재 벡터를 키로 사용하는 크로스 어텐션 레이어를 이용하여 상기 제2 어텐션 계수를 생성하는 동작

을 포함하는, 학습 장치.

청구항 16

제10항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 제1 신경망으로부터 지식 증류 방식으로(knowledge distillation) 상기 제2 신경망을 학습시키는 동작

을 포함하는, 학습 장치.

청구항 17

제16항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 증류 손실(distillation loss)을 계산하는 동작

을 포함하는, 학습 장치.

청구항 18

제17항에 있어서,

상기 제2 신경망을 학습시키는 동작은,

상기 증류 손실에 기초하여, 상기 제1 어텐션 계수와 상기 제2 어텐션 계수 간의 차이가 최소가 되도록 종단간 방식으로(end-to-end) 상기 제2 신경망을 학습시키는 동작

을 포함하는, 학습 장치.

발명의 설명

기 술 분 야

[0001] 아래 개시는 차량의 주행 경로를 결정하기 위한 학습 방법 및 이를 수행하는 장치에 관한 것이다.

배 경 기 술

[0003] 자율주행은 최근 몇 년 동안 상당한 진전을 보였지만, 주변 교통 에이전트의 움직임을 정확하게 예측하는 것은 안전을 보증하는 측면에서 과제들이 남아있다. 종래 기술들은 차선, 에이전트 정보 등 다양한 정보가 포함된 대규모 데이터를 활용한 행동 예측에 집중해 왔다. 하지만 이러한 종래 기술들은 과거 정보만을 사용하기 때문에 에이전트 간의 상호작용을 예측하는 데 불확실성이 존재하며, 이로 인해 에이전트 간의 충돌이나 잘못된 경로 예측이 발생할 수 있다.

[0004] 지식 증류 기법(knowledge distillation)은 높은 수준의 지식을 가진 모델(예: 티처 네트워크(teacher network))가 그 지식을 가지지 않은 모델(예: 학생 네트워크(student network))에게 그 지식을 전달하는 과정을 의미한다. 보다 구체적으로 지식 증류는 사전에 학습된 티처 네트워크로부터 생성된 데이터를 이용하여 학생 네트워크를 학습시키는 방법을 의미한다.

[0005] 위에서 설명한 배경기술은 발명자가 본원의 개시 내용을 도출하는 과정에서 보유하거나 습득한 것으로서, 반드시 본 출원 전에 일반 공중에 공개된 공지기술이라고 할 수는 없다.

발명의 내용

해결하려는 과제

[0007] 차량의 주행 경로를 예측함에 있어 불확실성을 제거하기 위해서, 차량의 미래 정보를 이용하여 차량의 주행 경로를 예측하는 기술이 요구된다.

[0008] 일 실시예는 복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 복수의 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보로부터 차량의 주행 경로를 예측하기 위한 신경망(neural network)을 학습시키는 기술을 제공할 수 있다.

[0009] 일 실시예는 차량의 주행 경로를 예측하기 위한 신경망(neural network)을 지식 증류 기법을 통해 학습시키는 기술을 제공할 수 있다.

[0010] 다만, 기술적 과제는 상술한 기술적 과제들로 한정되는 것은 아니며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0012] 일 실시예에 따른 차량의 주행 경로(driving route)를 결정하기 위한 학습 방법은, 복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 복수의 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여, 제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성하는 동작과, 상기 제1 정보 및 상기 제1 어텐션 계수에 기초하여, 제2 신경망을 통해 제2 어텐션 계수를 생성하는 동작과, 상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다. 상기 어텐션 계수는, 상기 주변 차량들이 상기 복수의 차량들 중 목표 차량의 주행 경로에 영향을 미치는 정도와 관

련하여, 상기 주변 차량들 각각에 대해 설정된 가중치일 수 있다.

[0013] 상기 제1 신경망은, 상기 목표 차량, 및 상기 주변 차량들 간의 상호작용을 예측하기 위한 것일 수 있다.

[0014] 상기 제2 신경망은, 상기 목표 차량의 주행 경로를 예측하기 위한 것일 수 있다.

[0015] 상기 제1 어텐션 계수를 생성하는 동작은, 상기 제1 정보를 인코딩(encoding)하여, 상기 제1 정보에 대응하는 제1 잠재 벡터(latent vector)를 생성하는 동작과, 상기 제2 정보를 인코딩하여, 상기 제2 정보에 대응하는 제2 잠재 벡터를 생성하는 동작과, 상기 제1 잠재 벡터, 및 제2 잠재 벡터를 멀티 임베딩(multi-embedding)하여, 임베딩 벡터(embedding vector)를 생성하는 동작과, 상기 임베딩 벡터를 키(key) 및 밸류(value)로 사용하고, 상기 제1 잠재 벡터를 쿼리(query)로 사용하는 크로스 어텐션 레이어(cross attention layer)를 이용하여 상기 제1 어텐션 계수를 생성하는 동작을 포함할 수 있다.

[0016] 상기 제2 어텐션 계수를 생성하는 동작은, 상기 제1 정보를 인코딩하여, 상기 제1 정보에 대응하는 제1 잠재 벡터를 생성하는 동작과, 상기 제1 잠재 벡터에 기초하여, 상기 목표 차량이 주행 가능한 경로의 차선에 관한 정보를 추출하는 동작과, 상기 차선에 관한 정보에 신뢰도 값(confidence value)을 할당하여 제1 잠재 벡터를 업데이트하는 동작을 포함할 수 있다.

[0017] 상기 제2 어텐션 계수를 생성하는 동작은, 상기 제1 잠재 벡터를 쿼리 및 밸류로 사용하고, 업데이트된 제1 잠재 벡터를 쿼리로 사용하는 크로스 어텐션 레이어를 이용하여 상기 제2 어텐션 계수를 생성하는 동작을 포함할 수 있다.

[0018] 상기 제2 신경망을 학습시키는 동작은, 상기 제1 신경망으로부터 지식 증류 방식으로(knowledge distillation) 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다.

[0019] 상기 제2 신경망을 학습시키는 동작은, 상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 증류 손실(distillation loss)을 계산하는 동작을 포함할 수 있다.

[0020] 상기 제2 신경망을 학습시키는 동작은, 상기 증류 손실에 기초하여, 상기 제1 어텐션 계수와 상기 제2 어텐션 계수 간의 차이가 최소가 되도록 종단간 방식으로(end-to-end) 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다.

[0021] 일 실시예에 따른 차량의 주행 경로를 결정하기 위한 학습 장치는, 인스트럭션들을 포함하는 메모리와, 상기 메모리와 전기적으로 연결되고, 상기 인스트럭션들을 실행하기 위한 프로세서를 포함하고, 상기 프로세서에 의해 상기 인스트럭션들이 실행될 때, 상기 프로세서는 복수의 동작들을 수행하고, 상기 복수의 동작들은, 복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 상기 차량들 중 주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여, 제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성하는 동작과, 상기 제1 정보 및 상기 제1 어텐션 계수에 기초하여, 제2 신경망을 통해 제2 어텐션 계수를 생성하는 동작과, 상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다. 상기 어텐션 계수는, 상기 주변 차량들이 상기 차량들 중 목표 차량의 주행 경로에 영향을 미치는 정도와 관련하여, 상기 주변 차량들 각각에 대해 설정된 가중치일 수 있다.

[0022] 상기 제1 신경망은, 상기 목표 차량, 및 상기 주변 차량들 간의 상호작용을 예측하기 위한 것일 수 있다.

[0023] 상기 제2 신경망은, 상기 목표 차량의 주행 경로를 예측하기 위한 것일 수 있다.

[0024] 상기 제1 어텐션 계수를 생성하는 동작은, 상기 제1 정보를 인코딩(encoding)하여, 상기 제1 정보에 대응하는 제1 잠재 벡터(latent vector)를 생성하는 동작과, 상기 제2 정보를 인코딩하여, 상기 제2 정보에 대응하는 제2 잠재 벡터를 생성하는 동작과, 상기 제1 잠재 벡터, 및 제2 잠재 벡터를 멀티 임베딩(multi-embedding)하여, 임베딩 벡터(embedding vector)를 생성하는 동작과, 상기 임베딩 벡터를 키(key) 및 밸류(value)로 사용하고, 상기 제1 잠재 벡터를 쿼리(query)로 사용하는 크로스 어텐션 레이어(cross attention layer)를 이용하여 상기 제1 어텐션 계수를 생성하는 동작을 포함할 수 있다.

[0025] 상기 제2 어텐션 계수를 생성하는 동작은, 상기 제1 정보를 인코딩하여, 상기 제1 정보에 대응하는 제1 잠재 벡터를 생성하는 동작과, 상기 제1 잠재 벡터에 기초하여, 상기 목표 차량이 주행 가능한 경로의 차선에 관한 정보를 추출하는 동작과, 상기 차선에 관한 정보에 신뢰도 값(confidence value)을 할당하여 제1 잠재 벡터를 업데이트하는 동작을 포함할 수 있다.

[0026] 상기 제2 어텐션 계수를 생성하는 동작은, 상기 제1 잠재 벡터를 쿼리 및 밸류로 사용하고, 업데이트된 제1 잠

재 벡터를 쿼리로 사용하는 크로스 어텐션 레이어를 이용하여 상기 제2 어텐션 계수를 생성하는 동작을 포함할 수 있다.

[0027] 상기 제2 신경망을 학습시키는 동작은, 상기 제1 신경망으로부터 지식 증류 방식으로(knowledge distillation) 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다.

[0028] 상기 제2 신경망을 학습시키는 동작은, 상기 제1 어텐션 계수, 및 상기 제2 어텐션 계수에 기초하여, 증류 손실(distillation loss)을 계산하는 동작을 포함할 수 있다.

[0029] 상기 제2 신경망을 학습시키는 동작은, 상기 증류 손실에 기초하여, 상기 제1 어텐션 계수와 상기 제2 어텐션 계수 간의 차이가 최소가 되도록 종단간 방식으로(end-to-end) 상기 제2 신경망을 학습시키는 동작을 포함할 수 있다.

도면의 간단한 설명

[0031] 도 1은 일 실시예에 따른 자율 주행 방식을 설명하기 위한 도면이다.

도 2는 일 실시예에 따른 자율 주행 장치에 포함되는 하드웨어를 도시한 블록도이다.

도 3은 일 실시예에 따른 차량의 주행 경로를 결정하기 위한 학습 시스템의 일 예를 나타낸다.

도 4는 도 3에 도시된 학습 장치의 일 예를 나타낸다.

도 5는 도 4에 도시된 미래 인식 어텐션 모듈의 일 예를 나타낸다.

도 6은 도 4에 도시된 LAM 기반의 자기 어텐션 모듈의 일 예를 나타낸다.

도 7은 도 6에 도시된 LAM 레이어의 동작을 설명하기 위한 도면이다.

도 8은 일 실시예에 따른 차량의 주행 경로 예측에 있어 종래 기술과의 비교를 설명하기 위한 일 도면이다.

도 9는 일 실시예에 따른 차량의 주행 경로 예측에 있어 종래 기술과의 비교를 설명하기 위한 다른 도면이다.

도 10은 일 실시예에 따른 차량의 주행 경로를 결정하기 위한 학습 방법의 흐름도의 일 예를 나타낸다.

도 11은 일 실시예에 따른 장치의 일 예를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0032] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0033] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0034] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

[0035] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0036] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0037] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

[0039] 도 1은 일 실시예에 따른 자율 주행 방식을 설명하기 위한 도면이다.

[0040] 도 1을 참조하면, 자율 주행 장치(예: 도 2의 자율 주행 장치(40))는, 차량에 장착되어 자율 주행 차량(10)을 구현할 수 있다. 자율 주행 차량(10)은 운전자의 조작 없이 스스로 주행할 수 있는 자동차일 수 있다.

[0041] 자율 주행 차량(10)에 장착되는 자율 주행 장치(40)는, 주변의 상황 정보를 수집하기 위한 다양한 센서들(예: 도 2의 센서부(41))을 포함할 수 있다.

[0042] 자율 주행 장치(40)는 자율 주행 차량(10)의 전면에 장착된 이미지 센서 및/또는 이벤트 센서를 통해, 전방에서 운행 중인 선행 차량(20)의 움직임을 감지할 수 있다. 자율 주행 장치(40)는 자율 주행 차량(10)의 전면은 물론, 옆 차로에서 운행중인 다른 주행 차량(30)과, 자율 주행 차량(10) 주변의 보행자 등을 감지하기 위한 센서들을 더 포함할 수 있다.

[0043] 자율 주행 차량(10) 주변의 상황 정보를 수집하기 위한 센서들 중 적어도 하나는, 도 1에 도시한 바와 같이 소정의 화각(FoV: field of view)을 가질 수 있다. 자율 주행 차량(10)의 전면에 장착된 센서가 도 1에 도시한 바와 같은 화각(FoV)을 갖는 경우에, 센서의 중앙에서 검출되는 정보가 상대적으로 높은 중요도를 가질 수 있다. 이는 센서의 중앙에서 검출되는 정보에, 선행 차량(20)의 움직임에 대응하는 정보가 대부분 포함되어 있기 때문일 수 있다.

[0044] 자율 주행 장치(40)는 자율 주행 차량(10)의 센서들이 수집한 정보를 실시간으로 처리하여 자율 주행 차량(10)의 움직임을 제어하는 한편, 센서들이 수 집한 정보 중에 적어도 일부는 메모리 장치(예: 도 2의 메모리 시스템(47))에 저장할 수 있다.

[0046] 도 2는 일 실시예에 따른 자율 주행 장치에 포함되는 하드웨어를 도시한 블록도이다.

[0047] 도 2를 참조하면, 자율 주행 장치(40)는 센서부(41), 프로세서(46), 메모리 시스템(47), 및 차체 제어 모듈(48) 등을 포함할 수 있다.

[0048] 센서부(41)는 복수의 센서들(42-45)을 포함할 수 있다. 복수의 센서들(42-45)은 이미지 센서, 이벤트 센서, 조도 센서, GPS 장치, 가속도 센서 등을 포함할 수 있다. 센서들(42-45)이 수집한 데이터는 프로세서(46)로 전달될 수 있다.

[0049] 프로세서(46)는 센서들(42-45)이 수집한 데이터를 메모리 시스템(47)에 저장할 수 있다. 프로세서(46)는 센서들(42-45)이 수집한 데이터에 기초하여 차체 제어 모듈(48)을 제어하여 차량의 움직임을 결정할 수 있다.

[0050] 메모리 시스템(47)은 둘 이상의 메모리 장치들과, 메모리 장치들을 제어하기 위한 시스템 컨트롤러를 포함할 수 있다. 메모리 장치들 각각은 하나의 반도체 칩으로 제공될 수 있다. 메모리 시스템(47)의 시스템 컨트롤러 외에, 메모리 시스템(47)에 포함되는 메모리 장치들 각각은 메모리 컨트롤러를 포함할 수 있다. 메모리 컨트롤러는 신경망과 같은 인공지능(AI: artificial intelligence) 연산 회로를 포함할 수 있다. 메모리 컨트롤러는 센서들(42-45) 또는 프로세서(46)로부터 수신한 데이터에 소정의 가중치를 부여하여 연산 데이터를 생성하고, 연산 데이터를 메모리 칩에 저장할 수 있다.

[0051] 차체 제어 모듈(48)은 프로세서(46)로부터 명령을 수신하여 차량의 움직임을 제어할 수 있다.

[0053] 도 3은 일 실시예에 따른 차량의 주행 경로를 결정하기 위한 학습 시스템의 일 예를 나타낸다.

[0054] 도 3을 참조하면, 학습 시스템(300)은 차량(310) 및 학습 장치(330)을 포함할 수 있다. 차량(310)은 사람 및/또는 사람을 운반하는 운송 수단을 의미하며, 예를 들어, 자동차, 기차, 선박, 보트, 항공기, 킥 보드 및/또는 자전거와 같은 운송 수단을 포함할 수 있다. 차량(310)은 학습 장치(330)와 통신을 위한 통신 모듈(도시되지 않음)을 내장하거나 외부 장치를 통해 학습 장치(330)와 통신을 할 수 있다. 학습 장치(330)는 차량(310)의 내

부에 설치 되어있거나 차량(310)의 외부에 있을 수 있다. 또한, 학습 장치(330)에 의해 학습된 제2 신경망(neural netowrk)(예: 도 4의 430)을 포함하는 장치(미도시)가 차량(310)의 내부에 설치될 수 있다. 차량(310)은 도 1의 자율 주행 차량(10), 또는 도 2의 자율 주행 장치(40)가 포함된 차량일 수 있다. 차량(310)은 목표 차량(예: 도 8의 805) 및 주변 차량(예: 도 8의 815)을 포함할 수 있다. 목표 차량(805)은 주행 경로를 결정하고자 하는 차량일 수 있다. 주변 차량(815)는 목표 차량(805)의 주변에 있는 차량일 수 있다.

[0055] 학습 장치(330)는 제1 신경망(예: 도 4의 410), 또는 제2 신경망(430)을 포함할 수 있다. 제1 신경망(410)은 목표 차량(805) 및 주변 차량들(예: 도 8의 815) 간의 상호 작용을 예측하기 위한 것일 수 있다. 목표 차량(805) 및 주변 차량들(815) 간의 상호 작용은 목표 차량(805)의 주행 경로에 주변 차량들(815)이 미치는 영향을 의미할 수 있다. 목표 차량(805) 및 주변 차량들(815) 간의 상호 작용은 목표 차량(805)의 위치 및 주변 차량들(815)의 위치와 관련된 것일 수 있다.

[0056] 제2 신경망은(430)은 목표 차량(805)의 주행 경로를 예측하기 위한 것일 수 있다. 제2 신경망(430)은 목표 차량(805)의 주행 경로를 예측함에 있어, 차량들(주변 차량들(815) 및 목표 차량(805))의 과거 주행 경로로부터 차량들(주변 차량들(815) 및 목표 차량(805))의 미래 위치를 예측할 수 있다.

[0057] 제1 신경망(410) 및 제2 신경망(430)은 심층 신경망(Deep Neural Network)을 포함할 수 있다. 제1 신경망(410) 및 제2 신경망(430)은 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), 퍼셉트론(perceptron), 다층 퍼셉트론(multilayer perceptron, MLP), FF(Feed Forward), RBF(Radial Basis Network), DFF(Deep Feed Forward), LSTM(Long Short Term Memory), GRU(Gated Recurrent Unit), AE(Auto Encoder), VAE(Variational Auto Encoder), DAE(Denoising Auto Encoder), SAE(Sparse Auto Encoder), MC(Markov Chain), HN(Hopfield Network), BM(Boltzmann Machine), RBM(Restricted Boltzmann Machine), DBN(Depp Belief Network), DCN(Deep Convolutional Network), DN(Deconvolutional Network), DCIGN(Deep Convolutional Inverse Graphics Network), GAN(Generative Adversarial Network), LSM(Liquid State Machine), ELM(Extreme Learning Machine), ESN(Echo State Network), DRN(Deep Residual Network), DNC(Differentiable Neural Computer), NTM(Neural Turning Machine), CN(Capsule Network), KN(Kohonen Network) 및 AN(Attention Network)를 포함할 수 있다

[0058] 학습 장치(330)는 복수의 차량들(310)의 과거 주행 경로에 관한 제1 정보(예: 도 4의 405-1), 및 복수의 차량들(310) 중 주변 차량들(815)의 미래 주행 경로에 관한 제2 정보(예: 도 4의 405-3)에 기초하여, 제1 신경망(410)을 통해 제1 어텐션 계수(attention coefficient)(예: 도 4의 455)를 생성할 수 있다.

[0059] 제1 정보(405-1), 및 제2 정보(405-3)는 차량(310)으로부터 실시간으로 전송되는 것이 아닌, 이전 시점에 저장된 정보일 수 있다. 예를 들어, 과거 주행 경로는 실시간으로 현재 시점을 기준으로 과거 시점의 주행 경로를 의미하는 것이 아니고, 이전 시점 중 특정 시점을 기준으로 과거 시점의 주행 경로일 수 있다. 또한, 미래 주행 경로는 실시간으로 현재 시점을 기준으로 미래 시점의 주행 경로를 의미하는 것이 아니고, 이전 시점 중 특정 시점을 기준으로 미래 시점의 주행 경로일 수 있다.

[0060] 또한, 제1 정보(405-1)는 복수의 차량들(310)의 과거 위치에 관한 정보를 포함할 수 있다. 제2 정보(405-3)는 주변 차량들(815)의 미래 위치에 관한 정보를 포함할 수 있다.

[0061] 어텐션 계수(455)는 목표 차량(805)과 주변 차량들(815)간의 상호 작용과 관련된 것일 수 있다. 어텐션 계수(455)는 주변 차량들(815)이 목표 차량(805)의 주행 경로에 영향을 미치는 정도와 관련하여, 주변 차량들(815) 각각에 대해 설정된 가중치를 의미할 수 있다. 예를 들어, 목표 차량(805)이 주행 시 주변 차량들(815)의 위치(또는 주행 경로)에 따라, 목표 차량(805)의 주행 경로가 달라질 수 있다. 이때, 어텐션 계수(455)는 주변 차량들(815) 중 목표 차량(805)의 주행 경로에 영향을 미치는 정도가 클수록 주변 차량들(815)에 높은 가중치가 부여된 것일 수 있다.

[0062] 학습 장치(330)가 제2 신경망(430)을 학습시키는 동작에 대해서는, 도 4를 참조하여 이하에서 상세히 설명하도록 한다.

[0064] 도 4는 도 3에 도시된 학습 장치의 일 예를 나타낸다.

[0065] 도 4를 참조하면, 학습 장치(330)는 제1 신경망(410) 및 제2 신경망(430)을 포함할 수 있다.

[0066] 제1 신경망(410)은 인코더(encoder)(425-1, 및 425-3), 자기 어텐션 모듈(self attention module)(450-1), 미

래 인식 어텐션 모듈(future aware attention module)(450-3), 및 디코더(deccoder)(490-1)를 포함할 수 있다.

[0067] 학습 장치(330)는 제1 정보(405-1), 및 제2 정보(405-3)에 기초하여 제1 신경망(410)을 통해 제1 어텐션 계수(455)를 생성할 수 있다. 학습 장치(330)는 제1 정보(405-1)를 인코딩(encoding) 하여 제1 정보(405-1)에 대응하는 제1 잠재 벡터(latent vector)(435-1)를 생성할 수 있다. 예를 들어, 학습 장치(330)는 인코더(encoder)(425-1)를 통해 제1 정보(405-1)를 인코딩 하여, 제1 잠재 벡터(435-1)를 생성할 수 있다. 제1 잠재 벡터(435-1)는 제1 정보(405-1)가 저차원의 벡터로 표현된 것일 수 있다. 제1 잠재 벡터(435-1)는 과거 시점의 복수의 차량들(주변 차량들(815) 및 목표 차량(805)를 포함)간의 상호 작용과 관련된 것일 수 있다.

[0068] 학습 장치(330)는 제2 정보(405-3)를 인코딩하여, 제2 정보(405-3)에 대응하는 제2 잠재 벡터(435-3)를 생성할 수 있다. 예를 들어, 학습 장치(330)는 인코더(425-3)를 통해 제2 정보(405-3)를 인코딩 하여, 제2 잠재 벡터(435-3)를 생성할 수 있다. 제2 잠재 벡터(435-3)는 제2 정보(405-3)가 저차원의 벡터로 표현된 것일 수 있다. 제2 잠재 벡터(435-3)는 미래 시점의 복수의 차량들(주변 차량들(815) 및 목표 차량(805)를 포함)간의 상호 작용과 관련된 것일 수 있다.

[0069] 학습 장치(330)는 제1 잠재 벡터(435-1), 및 제2 잠재 벡터(435-3)을 멀티 임베딩(multi-embedding)하여 임베딩 벡터(embedding vector)를 생성할 수 있다. 임베딩 벡터는 제1 잠재 벡터(435-1), 및 제2 잠재 벡터(435-3) 간의 유사성을 측정(또는 계산)하기 위한 것일 수 있다. 임베딩 벡터는 과거 시점, 및 미래 시점에서의 복수의 차량들(주변 차량들(815) 및 목표 차량(805)를 포함) 간의 상호 작용의 유사성을 측정(또는 계산)하기 위한 것일 수 있다.

[0070] 학습 장치(330)는 임베딩 벡터를 키(key)(예: 도 5의 530) 및 밸류(value)(예: 도 5의 550)로 사용하고, 제1 잠재 벡터(435-1)를 쿼리(query)(예: 도 5의 510)로 사용하는 크로스 어텐션 레이어(cross attention layer)(예: 도 5의 570)를 이용하여 상기 제1 어텐션 계수(455)를 생성할 수 있다. 이에 대해서는, 도 5를 참조하여 상세히 설명하도록 한다.

[0071] 제2 신경망(430)은 인코더(425-1), LAM(lane-guided attention module, LAM) 기반의 자기 어텐션 모듈(470-1), 자기 어텐션 모듈(470-3), 및 디코더(490-3)를 포함할 수 있다.

[0072] 학습 장치(330)는 제1 정보(405-1), 및 제1 어텐션 계수(455)에 기초하여, 제2 신경망(430)을 통해 제2 어텐션 계수(미도시)를 생성할 수 있다. 학습 장치(330)는 제1 정보를 인코딩하여 제1 잠재 벡터(435-1)를 생성할 수 있다. 이에 대해서는 위에서 상세히 설명한바, 생략하도록 한다.

[0073] 학습 장치(330)는 제1 잠재 벡터(435-1)에 기초하여 목표 차량(805 또는 도 7의 705)이 주행 가능한 경로(예: 도 7의 750-1과 750-5 사이의 경로, 및 750-1과 750-3 사이의 경로)의 차선(예: 도 7의 750-1, 750-3, 및 750-5)에 관한 정보를 추출할 수 있다. 학습 장치(330)는 차선(750-1, 750-3, 및 750-5)에 관한 정보에 신뢰도 값(confidence value)(예: 도 7의 715)을 할당하여 제1 잠재 벡터(435-1)를 업데이트할 수 있다. 학습 장치(330)는 제1 잠재 벡터(435-1)를 쿼리(예: 도 6의 610) 및 밸류(예: 도 6의 650)로 사용하고, 업데이트된 제1 잠재 벡터를 키(예: 도 6의 630)로 사용하는 크로스 어텐션 레이어(예: 도 6의 670)를 이용하여 제2 어텐션 계수(미도시)를 생성할 수 있다. 이에 대해서는, 도 6 및 도 7을 참조하여 상세히 설명하도록 한다.

[0074] 학습 장치(330)는 제1 어텐션 계수(455) 및 제2 어텐션 계수(미도시)에 기초하여 제2 신경망(430)을 학습시킬 수 있다.

[0075] 학습 장치(330)는 제1 어텐션 계수(455) 및 제2 어텐션 계수(미도시)에 기초하여 증류 손실(distillation loss)을 계산할 수 있다. 증류 손실은 제2 신경망(430)의 학습을 위한 손실함수를 의미할 수 있다. 예를 들어, 증류 손실은 하기의 수학식 1과 같이 정의될 수 있다.

[0077] [수학식 1]

$$\mathcal{L}_{\text{distill}} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}_{KL}\left(\mathcal{A}_i \,||\, \mathbb{A}_i\right)$$

[0078]

[0080] 수학식 1에서, $L_{distill}$은 증류 손실을 나타내고, N은 차량들(목표 차량(805), 및 주변 차량(815)을 포함)의 개수를 나타내고, $\mathbb{A}_i$는 제1 어텐션 계수(455)를 나타내고, $\mathcal{A}_i$는 제2 어텐션 계수(미도시)를 나타낸다.

[0082] 다른 예를 들어, 손실 함수는 하기의 수학식 2 내지 4를 통해 정의될 수 있다.

[0084] [수학식 2]

[0085]
$$\mathcal{L} = \mathcal{L}_{\text{traj}} + \mathcal{L}_{\text{conf}} + \mathcal{L}_{\text{distill}}$$

[0087] [수학식 3]

[0088]
$$\mathcal{L}_{\text{traj}} = \mathcal{L}_{\text{LAM}} + \mathcal{L}_{\text{final}}$$

[0090] [수학식 4]

[0091]
$$\mathcal{L}_{LAM}(p_i, \hat{p}_i) = \mathcal{L}_{final}(p_i, \hat{p}_i) = \frac{1}{\sqrt{2b^2}} \sum_{i=1}^{n} |p_i - \hat{p}_i| \, e^{-\frac{|p_i - \hat{p}_i|}{b}}$$

[0093] 수학식 2 내지 수학식 4에서, $\mathcal{L}$은 손실 함수를 나타내고, L_{traj}은 회귀 손실(regression loss)을 나타내고, L_{conf}은 각각의 주행 경로(예: 도 7의 790-1, 및 790-3)에 대한 신뢰 손실(confidence loss)을 나타내고, L_{LAM}은 LAM 기반의 자기 어텐션 모듈(470-1)에서 사용된 각 경로에 대한 회귀 손실을 나타내고, L_{final}은 라플라스 분포(laplace distribution)에 대한 음의 로그 확률 함수(negative log likelihood function)를 사용한 최종 출력에서 회귀손실을 나타내고, p_i는 i번째 주변 차량(815)의 실제 미래 위치를 나타내고, $\hat{p}_i$는 i번째 주변 차량(815)의 예측된 위치를 나타내고, b는 스케일 파라미터(scale parameter)를 나타낸다.

[0095] 학습 장치(330)는 제1 신경망(410)으로부터 지식 증류 방식으로(knowledge distillation) 제2 신경망(430)을 학습시킬 수 있다. 학습 장치(330)는 제1 신경망(410)을 티처 네트워크(teacher network)로 하고, 제2 신경망(430)을 학생 네트워크(student network)로 할 수 있다. 학습 장치(330)는 제1 신경망(430)을 통해 생성한 지식(예: 제1 어텐션 계수(455))을 제2 신경망(430)에 전달할 수 있다. 예를 들어, 학습 장치(330)는 제1 신경망(410)을 통해 생성한 제1 어텐션 계수(455)와 제2 신경망(430)을 통해 생성한 제2 어텐션 계수(미도시)간의 차이가 최소가 되도록 제2 신경망(430)을 학습시킬 수 있다. 이때, 학습 장치(330)는 증류 손실(또는 수학식 2의 손실 함수)에 기초하여, 종단간 방식으로(end-to-end) 제2 신경망(430)을 학습시킬 수 있다.

[0097] 도 5는 도 4에 도시된 미래 인식 어텐션 모듈의 일 예를 나타낸다.

[0098] 도 5를 참조하면, 미래 인식 어텐션 모듈(450-3)은 멀티 임베딩 레이어(multi embedding layer)(590), 및 크로스 어텐션 레이어(cross attention layer)(570)을 포함할 수 있다.

[0099] 멀티 임베딩 레이어(590)는 제1 잠재 벡터(435-1)(예: s_i') 및 제2 잠재 벡터(435-3)(예: o_i)을 멀티 임베딩하여 임베딩 벡터(예: $s_j' + o_j$)를 생성할 수 있다. 멀티 임베딩 레이어(590)는 임베딩 벡터($s_j' + o_j$)를 크로스 어텐션 레이어(570)로 출력할 수 있다. 예를 들어, 학습 장치(330)는 멀티 임베딩 레이어(590)를 통해 제1 잠재 벡터(435-1), 및 제2 잠재 벡터(435-3)로부터 임베딩 벡터($s_j' + o_j$)를 생성할 수 있다.

[0100] 크로스 어텐션 레이어(570)는 멀티 임베딩 레이어(590)로부터 임베딩 벡터($s_j' + o_j$)를 수신할 수 있다. 크로스 어텐션 레이어(570)는 임베딩 벡터($s_j' + o_j$) 및 제1 잠재 벡터(435-1)를 이용하여 어텐션 연산(attention operation)을 수행할 수 있다. 크로스 어텐션 레이어(570)는 임베딩 벡터($s_j' + o_j$)를 키(530) 및 밸류(550)로 사용하고, 제1 잠재 벡터(435-1)를 쿼리(510)로 하여 제1 어텐션 계수(455)를 생성할 수 있다.

[0101] 예를 들어, 크로스 어텐션 레이어(570)는 하기의 수학식 5와 같이 쿼리(510), 키(530), 및 밸류(550)를 정의하고, 어텐션 연산을 수행하여 제1 어텐션 계수(455)를 생성할 수 있다.

[0103] [수학식 5]

[0104] $$Q_i' = W_Q' \mathbf{s}_i', \; K_{ij}' = W_K' \phi(\mathbf{s}_j' + \mathbf{o}_j), \; V_{ij}' = W_V' \phi(\mathbf{s}_j' + \mathbf{o}_j)$$

[0106] 수학식 5에서, Q_i', K_{ij}', 및 V_{ij}'는 각각 쿼리(510), 키(530), 및 밸류(550)를 나타내고, W_Q', W_K', W_V'는 각각 쿼리(510), 키(530), 및 밸류(550)의 학습 가능한 파라미터를 나타내고, ϕ는 MLP 레이어를 나타내고, s_i'는 제1 잠재 벡터(435-1)를 나타내고, $s_j' + o_j$는 임베딩 벡터를 나타낸다.

[0108] 도 6은 도 4에 도시된 LAM 기반의 자기 어텐션 모듈의 일 예를 나타내고, 도 7은 도 6에 도시된 LAM 레이어의 동작을 설명하기 위한 도면이다.

[0109] 도 6을 참조하면, LAM 기반의 자기 어텐션 모듈(470-1)은 LAM 레이어(lane-guided attention module layer)(690), 및 크로스 어텐션 레이어(670)를 포함할 수 있다.

[0110] 도 7을 참조하면, 그림(710)은 LAM 레이어(690)가 목표 차량(710)의 위치를 기준으로 차선(750-1, 750-3, 및 750-5)에 관한 정보를 추출하는 동작을 나타낸다. 그림(730)은 LAM 레이어(690)가 차선(750-1, 750-3, 및 750-5)에 관한 정보에 신뢰도 값(715)을 할당하는 동작을 나타낸다.

[0111] LAM 레이어(690)는 제1 잠재 벡터(435-1)에 기초하여 목표 차량(705)이 주행 가능한 경로(750-1과 750-3 사이의 경로, 750-3과 750-5 사이의 경로)의 차선(750-1, 750-3, 및 750-5)에 관한 정보를 추출할 수 있다.

[0112] LAM 레이어(690)는 목표 차량(705)의 위치를 기준으로 특정 범위(750-7)에 있는 차선(750-1, 및 750-3)에 관한 정보를 추출할 수 있다. 예를 들어, 특정 범위(750-7)는 목표 차량(705)의 위치를 중심으로 반지름이 D 미터인 원일 수 있다. 특정 범위(750-7)에 차선(750-1, 및 750-3)이 포함될 수 있다. 학습 장치(330)는 특정 범위에 포함된 차선(750-1, 및 750-3)에 대해서만 정보를 추출할 수 있다.

[0113] 제1 잠재 벡터(435-1)로부터 추출된 차선(750-1, 및 750-3)에 관한 정보는 목표 차량(705)이 주행 가능한 경로(750-1과 750-3 사이의 경로, 및 750-3과 750-5 사이의 경로)의 차선(750-1, 750-3, 및 750-5)에 관한 정보만을 포함할 뿐, 목표 차량(705)이 주행할 가능성에 따라 식별(또는 구분, 분류)되지 않은 정보일 수 있다.

[0114] LAM 레이어(690)는 목표 차량(705)이 주행 가능한 경로(750-1과 750-3 사이의 경로, 및 750-3과 750-5 사이의 경로)의 차선(750-1, 750-3, 및 750-5)에 대해 신뢰도 값(715)을 할당할 수 있다. 신뢰도 값(715)은 목표 차량(705)이 주행할 가능성에 관한 것일 수 있다. 신뢰도 값(715)은 각 차선(750-1, 750-3, 및 750-5)에 대해 할당되는 것일 수 있다. LAM 레이어(690)는 주행 가능성이 높은 경로의 차선에 높은 신뢰도 값(715)을 할당할

수 있다.

[0115] 예를 들어, LAM 레이어(690)는 목표 차량(705)의 위치에 기초하여 목표 차량(705)이 주행 가능한 경로(750-1과 750-3 사이의 경로 및 750-3과 750-5 사이의 경로) 별로 목표 차량(705)이 주행할 가능성을 계산할 수 있다. LAM 레이어(690)는 경로(790-3)보다 경로(790-1)가 목표 차량(705)의 주행 가능성이 높은 것으로 계산(또는 결정)할 수 있다. LAM 레이어(690)는 경로(750-1과 750-3 사이의 경로)의 차선(750-1, 및 750-3)에 경로(750-3과 750-5 사이의 경로)의 차선(750-3 및 750-5)보다 높은 신뢰도 값(715)를 할당할 수 있다.

[0116] 동일 차선(750-1)의 경우에도, 목표 차량(705)의 주행 가능성에 기초하여 서로 다른 신뢰도 값(770-1, 및 770-5)이 할당될 수 있다. 예를 들어, 목표 차량(705)이 경로(790-1) 또는 경로(790-3) 중 어느 경로를 주행하는 경우에도, 목표 차량(705)은 차선(750-1) 중 신뢰도 값(770-1)이 부여된 차선까지는 동일하게 주행할 수 있다. 경로(차선(750-1) 중 신뢰도 값(770-1)이 부여된 차선이 포함된 경로)가 경로(차선(750-1) 중 신뢰도 값(770-5)이 부여된 차선이 포함된 경로)보다 목표 차량(705)이 주행할 가능성이 높을 수 있다. LAM 레이어(690)는 동일 차선(750-1)에 대해 서로 다른 신뢰도 값(770-1, 및 770-5)을 할당할 수 있다.

[0117] LAM 레이어(690)는 신뢰도 값이 할당된 정보에 기초하여, 제1 잠재 벡터(435-1)를 업데이트 할 수 있다. 업데이트된 제1 잠재 벡터는 목표 차량(705)이 주행 가능한 경로 및 목표 차량(705)이 주행 가능한 경로 별로 목표 차량(705)의 위치에 따른 주행할 가능성의 수준(또는 레벨)에 관한 정보를 포함할 수 있다.

[0118] 구체적으로, LAM 레이어(690)는 하기의 수학식 6 및 수학식 7을 통해 제1 잠재 벡터(435-1)를 업데이트할 수 있다.

[0120] [수학식 6]

[0121]
$$C_{i\xi} = \sum_{k=1}^{6} \begin{cases} \Pi^k, & \text{if} \quad \exists \hat{p}_t^k \,|\, |\hat{p}_t^k - p_{\xi,0}| < \mathcal{D} \\ 0, & \text{otherwise} \end{cases}$$

[0123] [수학식 7]

[0124]
$$F_{i\xi} = \phi_{lane}\left(\left[R_i^T\left(p_{\xi,1} - p_{\xi,0}\right), R_i^T\left(p_{\xi,0} - p_i^{T_c}\right), C_{i\xi}, a_\xi\right]\right)$$

[0126] 수학식 6 및 수학식 7에서, $\hat{p}_t^k$는 예측된 주행 경로에서 t번째 단계(step)의 i번째 차량(목표 차량(705))의 k번째 모드(mode)의 위치를 나타내고, $p_{\xi,0}$는 ξ번째 차선의 시작 위치를 나타내고, Π^k는 i번째 차량(목표 차량(705))의 예측된 경로에서 k번째 차선의 신뢰도 값을 나타내고, $C_{i\xi}$는 i번째 차량(목표 차량(705))의 ξ번째 차선의 신뢰도 값을 나타낸다. 또한, ϕ_{lane}은 MLP 레이어를 나타내고, a_ξ는 ξ번째 차선의 시맨틱 속성(semantic attribute)을 나타내고, $p_{\xi,1}$는 ξ번째 차선의 종료 위치를 나타내고, R_i는 i번째 차량(목표 차량(705))의 중심 좌표로 변환을 위한 2 x 2 회전 행렬(rotation matrix)을 나타내고, $F_{i\xi}$는 업데이트된 제1 잠재 벡터를 나타낸다.

[0127] 크로스 어텐션 레이어(670)는 LAM 레이어(690)로부터 업데이트된 제1 잠재 벡터를 수신할 수 있다. 크로스 어텐션 레이어(670)는 업데이트된 제1 잠재 벡터 및 제1 잠재 벡터(435-1)를 이용하여 어텐션 연산을 수행할 수 있다. 크로스 어텐션 레이어(670)는 제1 잠재 벡터(435-1)를 쿼리(610) 및 밸류(650)로 사용하고, 업데이트된 제1 잠재 벡터를 키(630)로 사용하여, 제2 어텐션 계수(미도시)를 생성할 수 있다.

[0128] 예를 들어, 크로스 어텐션 레이어(670)는 하기의 수학식 8과 같이 쿼리(610), 키(630), 및 밸류(650)를 정의하고, 어텐션 연산을 수행하여 제2 어텐션 계수(미도시)를 생성할 수 있다.

[0130] [수학식 8]

[0132] $$Q_i = W_Q F_i,\ K_{i\xi} = W_K F_{i\xi},\ V_i = W_V F_i$$

[0134] 수학식 8에서, Q_i, $K_{i\xi}$ 및 V_i는 각각 쿼리(610), 키(630), 및 밸류(650)를 나타내고, W_Q, W_K, W_V는 각각 쿼리(610), 키(630), 및 밸류(650)의 학습 가능한 파라미터를 나타내고, F_i는 제1 잠재 벡터를 나타내고, $F_{i\xi}$는 업데이트된 제1 잠재 벡터를 나타낸다.

[0136] 도 8은 일 실시예에 따른 차량의 주행 경로 예측에 있어 종래 기술과의 비교를 설명하기 위한 일 도면이다.

[0137] 도 8을 참조하면, 차량의 주행 경로를 예측하는 방법들(810, 830, 및 850)이 시나리오(870-1, 870-3, 및 870-5) 별로 목표 차량(805)의 주행 경로를 얼마나 잘 예측할 수 있는지 비교할 수 있다.

[0138] 시나리오(870-1)는 교차로에서 주변 차량(815)의 위치가 변경된 경우를 나타낸다. 시나리오(870-3) 및 시나리오(870-5)는 주변 차량(815)이 제거 또는 추가된 경우를 나타낸다.

[0139] 방법(810)은 종래 기술에 의한 차량의 주행 경로를 예측하는 방법이고, 방법(830)은 학습 장치(330)에 의해 학습된 제2 신경망(430)을 통해 차량의 주행 경로를 예측하는 방법이고, 방법(850)은 제1 신경망(410)을 통해 차량의 주행 경로를 예측하는 방법이다. 주행 경로의 예측의 정확도는 신뢰도로 표현될 수 있다. 신뢰가 1.0에 가까울수록(예측된 주행 경로의 색상이 진할수록) 정확도가 더 높을 수 있다.

[0140] 예를 들어, 시나리오(870-1)에서, 방법(810)에 의해 예측된 목표 차량(805)의 주행 경로 보다 방법(830)에 의해 예측된 목표 차량(805)의 주행 경로의 신뢰도 값이 더 큰 것을 알 수 있다.

[0142] 도 9는 일 실시예에 따른 차량의 주행 경로 예측에 있어 종래 기술과의 비교를 설명하기 위한 다른 도면이다.

[0143] 도 9를 참조하면, 차량의 주행 경로를 예측하는 방법들(810, 및 830)이 시나리오(910, 930, 950, 970, 및 990) 별로 목표 차량(805)의 주행 경로를 얼마나 잘 예측할 수 있는지 비교할 수 있다.

[0144] 시나리오(910) 및 시나리오(930)은 목표 차량(805)이 좌회전 시, 주변 차량(815)이 적은 경우와 주변 차량(815)이 많은 경우를 나타낸다. 시나리오(950) 및 시나리오(970)은 목표 차량(805)이 우회전 시, 주변 차량(815)이 적은 경우와 주변 차량(815)이 많은 경우를 나타낸다. 시나리오(990)은 주변 차량(815)이 많은 상황에서, 목표 차량(990)이 직진하는 경우를 나타낸다.

[0145] 방법(810)은 종래 기술에 의한 차량의 주행 경로를 예측하는 방법이고, 방법(830)은 학습 장치(330)에 의해 학습된 제2 신경망(430)을 통해 차량의 주행 경로를 예측하는 방법이고, 방법(850)은 제1 신경망(410)을 통해 차량의 주행 경로를 예측하는 방법이다. 주행 경로의 예측의 정확도는 신뢰도로 표현될 수 있다. 신뢰가 1.0에 가까울수록(예측된 주행 경로의 색상이 진할수록) 정확도가 더 높을 수 있다.

[0146] 예를 들어, 시나리오(930)에서, 방법(810)에 의해 예측된 목표 차량(805)의 주행 경로 보다 방법(830)에 의해 예측된 목표 차량(805)의 주행 경로의 신뢰도 값이 더 큰 것을 알 수 있다.

[0148] 도 10은 일 실시예에 따른 차량의 주행 경로를 결정하기 위한 학습 방법의 흐름도의 일 예를 나타낸다.

[0149] 도 10을 참조하면, 동작 1010 내지 동작 1050은 순차적으로 수행될 수 있지만, 이에 한정되는 것은 아니다. 예를 들어, 둘 이상의 동작들이 병렬적으로 수행될 수 있다. 동작 1010 내지 동작 1050은 도 1 내지 도 9를 참조하여 설명한 학습 장치(예: 도 3의 학습 장치(330))의 동작과 실질적으로 동일할 수 있다. 이에, 상세한 설명은 생략하도록 한다.

[0150] 동작 1010에서, 학습 장치(330)는 복수의 차량들(310)의 과거 주행 경로에 관한 제1 정보(405-1), 및 복수의 차

량들(310) 중 주변 차량들(815)에 관한 제2 정보(405-3)에 기초하여, 제1 신경망(410)을 통해 제1 어텐션 계수(455)를 생성할 수 있다.

[0151] 동작 1030에서, 학습 장치(330)는 제1 정보(405-1), 및 제1 어텐션 계수(455)에 기초하여, 제2 신경망(430)을 통해 제2 어텐션 계수(미도시)를 생성할 수 있다.

[0152] 동작 1050에서, 학습 장치(330)는 제1 어텐션 계수(455) 및 제2 어텐션 계수(미도시)에 기초하여, 제2 신경망(430)을 학습시킬 수 있다.

[0154] 도 11은 일 실시예에 따른 장치의 일 예를 나타낸다.

[0155] 도 11을 참조하면, 장치(1100)는 메모리(1110) 및 프로세서(1130)를 포함할 수 있다. 장치(1100)는 도 3의 학습 장치(330)를 포함할 수 있다.

[0156] 메모리(1110)는 프로세서(1130)에 의해 실행가능한 인스트럭션들(예: 프로그램)을 저장할 수 있다. 예를 들어, 인스트럭션들은 프로세서(1130)의 동작 및/또는 프로세서(1130)의 각 구성의 동작을 실행하기 위한 인스트럭션들을 포함할 수 있다.

[0157] 프로세서(1130)는 메모리(1110)에 저장된 데이터를 처리할 수 있다. 프로세서(1130)는 메모리(1110)에 저장된 컴퓨터로 읽을 수 있는 코드(예를 들어, 소프트웨어) 및 프로세서(1130)에 의해 유발된 인스트럭션(instruction)들을 실행할 수 있다.

[0158] 프로세서(1130)는 목적하는 동작들(desired operations)을 실행시키기 위한 물리적인 구조를 갖는 회로를 가지는 하드웨어로 구현된 데이터 처리 장치일 수 있다. 예를 들어, 목적하는 동작들은 프로그램에 포함된 코드(code) 또는 인스트럭션들(instructions)을 포함할 수 있다.

[0159] 예를 들어, 하드웨어로 구현된 데이터 처리 장치는 마이크로프로세서(microprocessor), 중앙 처리 장치(central processing unit), 프로세서 코어(processor core), 멀티-코어 프로세서(multi-core processor), 멀티프로세서(multiprocessor), ASIC(Application-Specific Integrated Circuit), FPGA(Field Programmable Gate Array)를 포함할 수 있다.

[0160] 메모리(1110)는 휘발성 메모리 장치 또는 비휘발성 메모리 장치로 구현될 수 있다.

[0161] 휘발성 메모리 장치는 DRAM(dynamic random access memory), SRAM(static random access memory), T-RAM(thyristor RAM), Z-RAM(zero capacitor RAM), 또는 TTRAM(Twin Transistor RAM)으로 구현될 수 있다.

[0162] 비휘발성 메모리 장치는 EEPROM(Electrically Erasable Programmable Read-Only Memory), 플래시(flash) 메모리, MRAM(Magnetic RAM), 스핀전달토크 MRAM(Spin-Transfer Torque(STT)-MRAM), Conductive Bridging RAM(CBRAM), FeRAM(Ferroelectric RAM), PRAM(Phase change RAM), 저항 메모리(Resistive RAM(RRAM)), 나노 튜브 RRAM(Nanotube RRAM), 폴리머 RAM(Polymer RAM(PoRAM)), 나노 부유 게이트 메모리(Nano Floating Gate Memory(NFGM)), 홀로그래픽 메모리(holographic memory), 분자 전자 메모리 소자(Molecular Electronic Memory Device), 또는 절연 저항 변화 메모리(Insulator Resistance Change Memory)로 구현될 수 있다.

[0163] 도 3의 학습 장치(330)는 메모리(1110)에 저장되어 프로세서(1130)에 의해서 실행되거나 프로세서(1130)에 임베디드될 수 있다. 프로세서(1130)는 도 1 내지 도 10을 참조하여 학습 장치(330)의 동작을 실질적으로 동일하게 수행할 수 있다. 이에 상세한 설명은 생략하도록 한다.

[0165] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만,

해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0166] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0167] 실시예에 따른 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 저장할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(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.

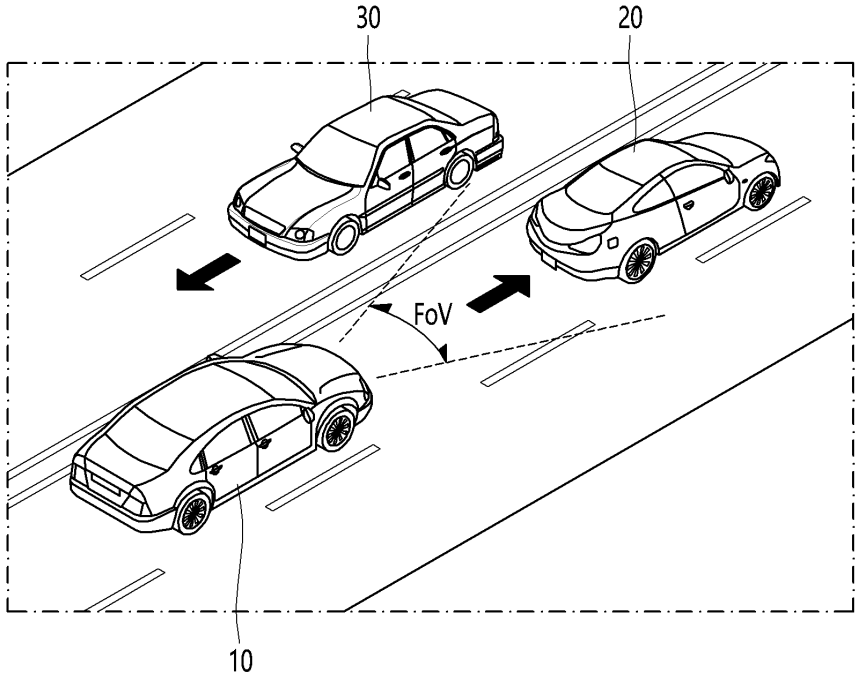
[0168] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0169] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

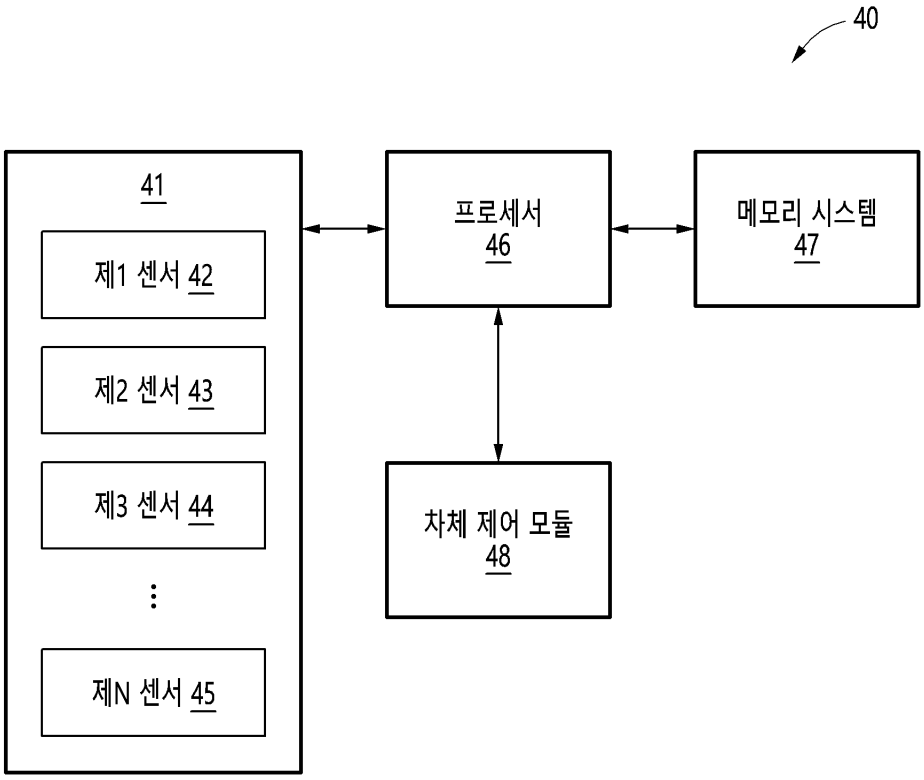
[0170] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

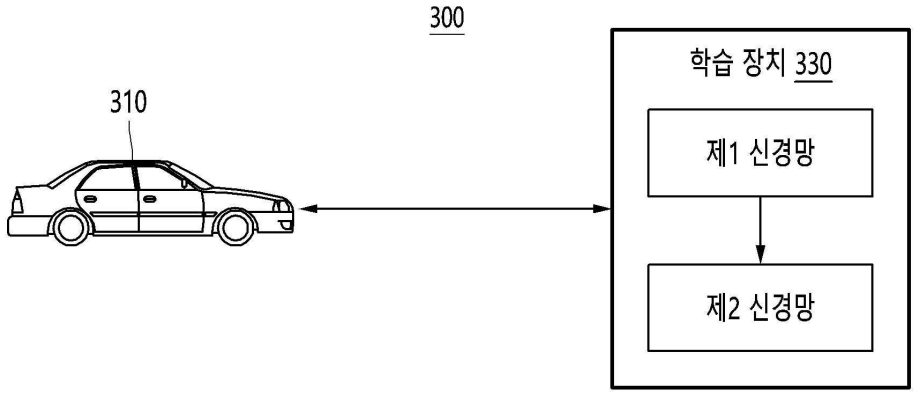
도면1


30
20
FoV
10

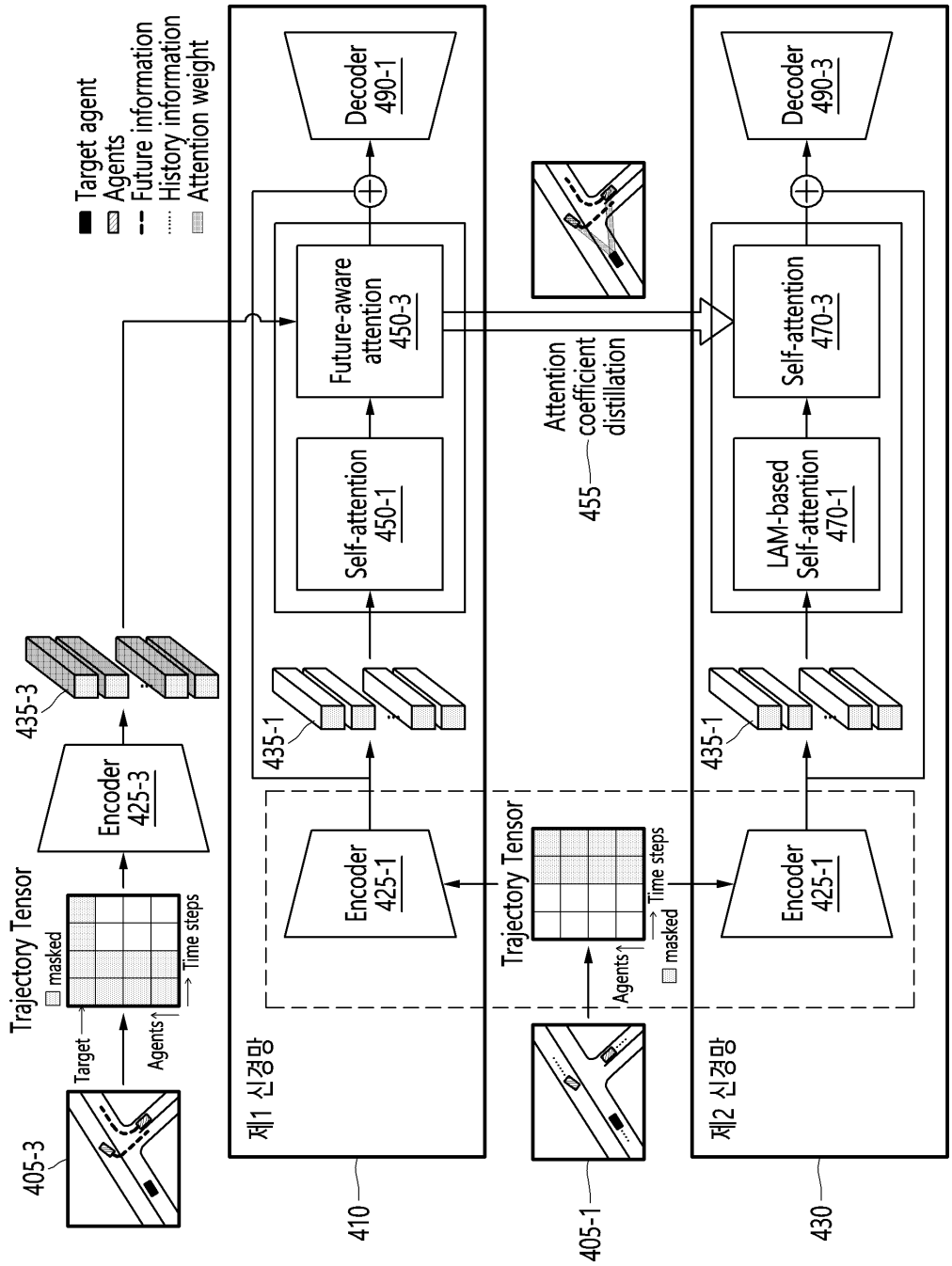
도면2


40
41
제1 센서 42
제2 센서 43
제3 센서 44
제N 센서 45
프로세서
46
메모리 시스템
47
차체 제어 모듈
48

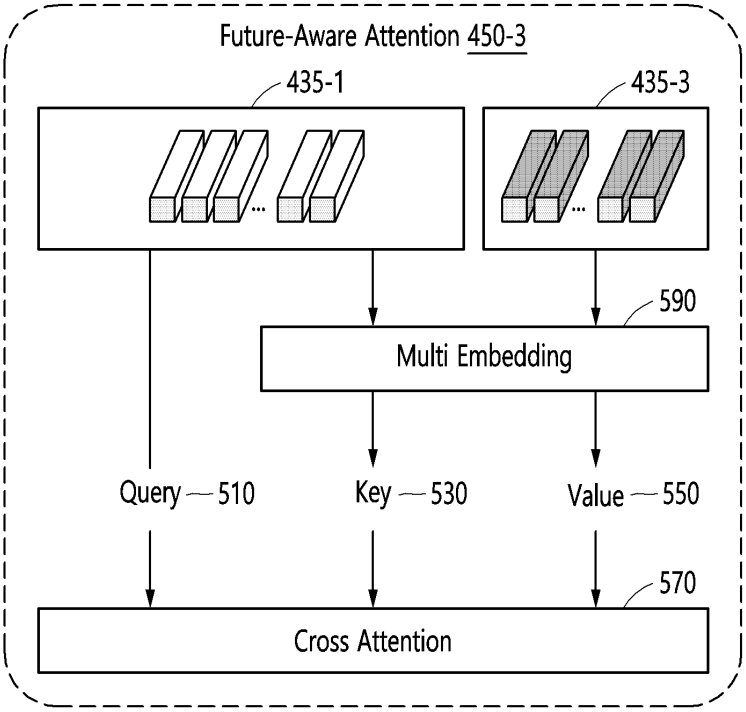
도면3



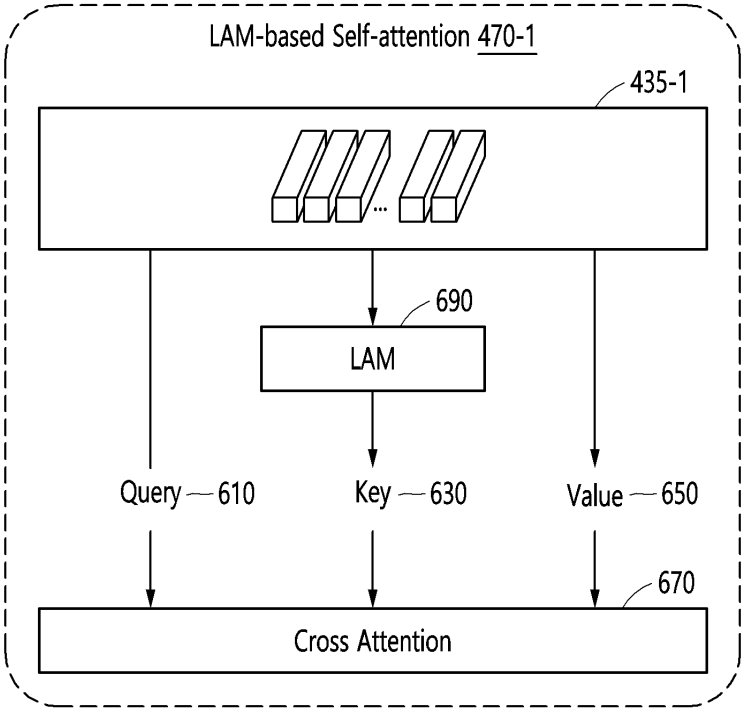
도면4


Target agent
Agents
Future information
History information
Attention weight
Decoder 490-1
Future-aware attention 450-3
Self-attention 450-1
Attention coefficient distillation
455
Self-attention 470-3
LAM-based Self-attention 470-1
Decoder 490-3
435-3
435-1
435-1
Encoder 425-3
Encoder 425-1
Trajectory Tensor
Trajectory Tensor
Encoder 425-1
masked
Time steps
Target
Agents
Agents
Time steps
masked
405-3
405-1
제1 신경망
제2 신경망
410
430

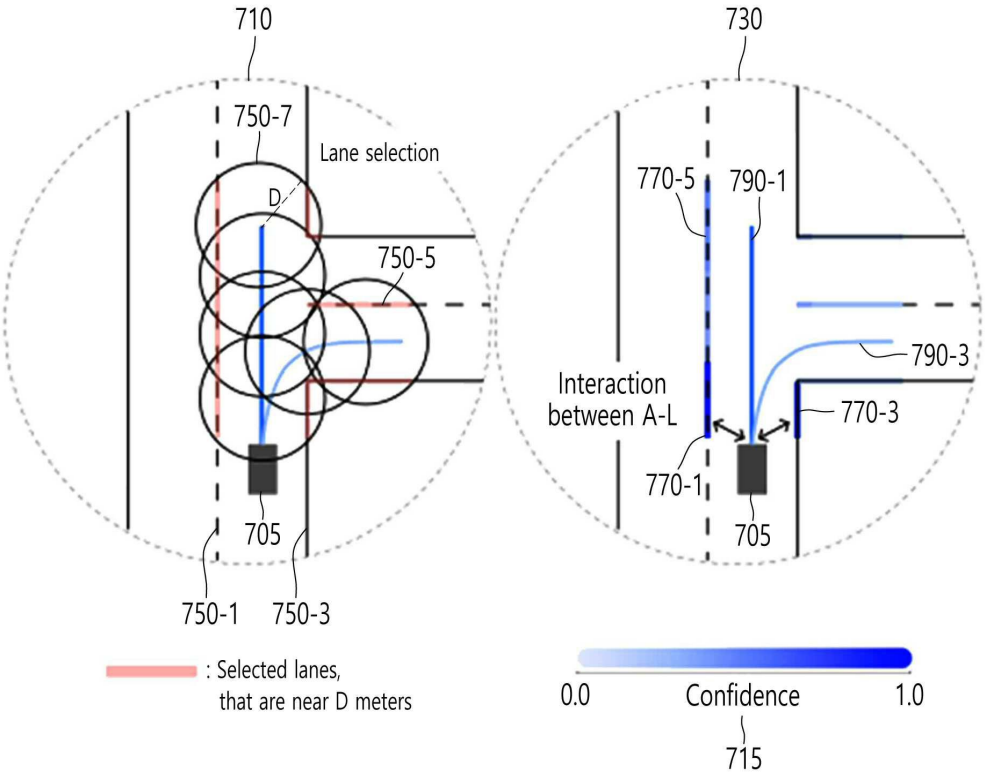
도면5


Future-Aware Attention 450-3
435-1
435-3
590
Multi Embedding
Query 510
Key 530
Value 550
570
Cross Attention

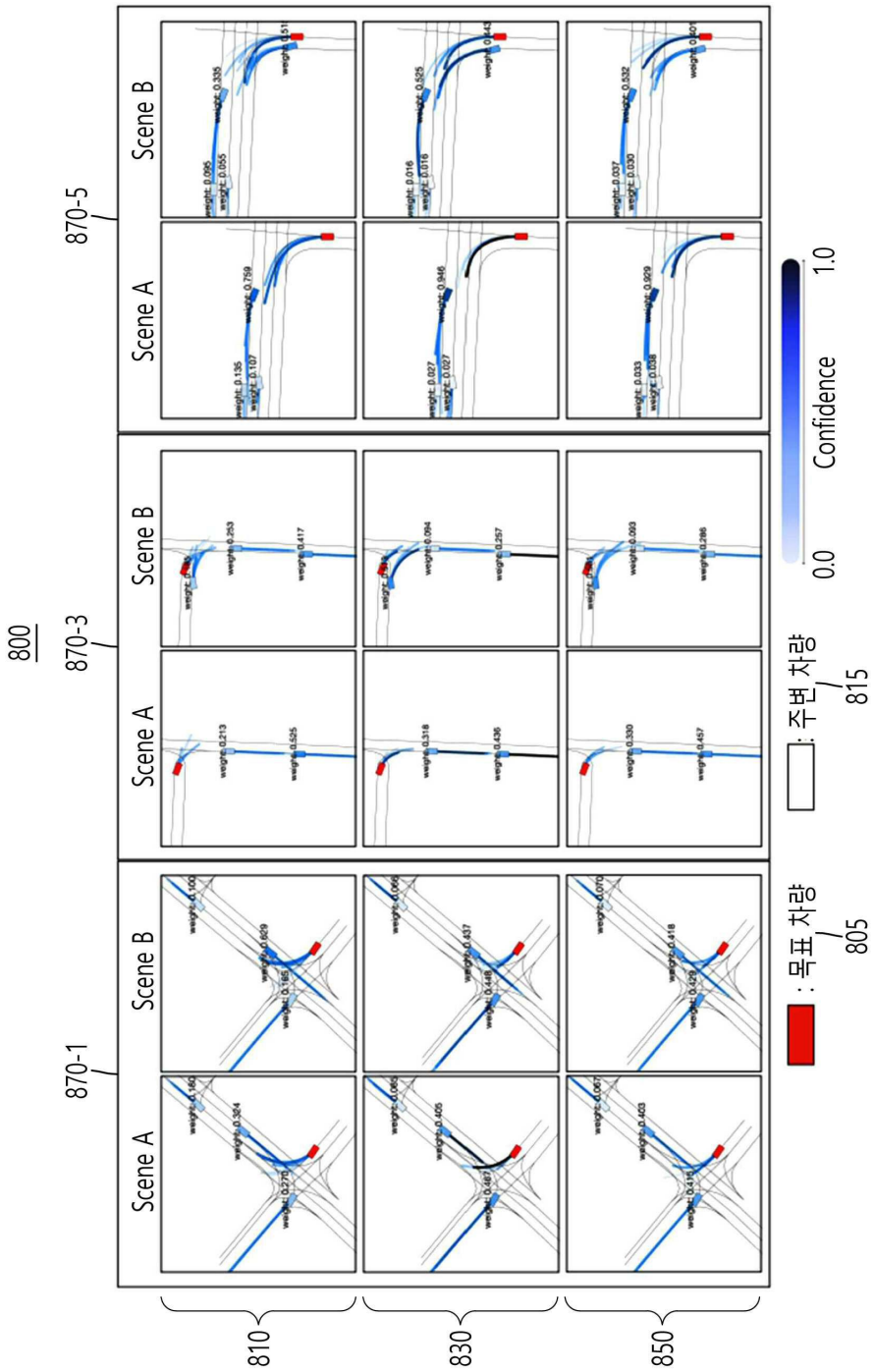
도면6


LAM-based Self-attention 470-1
435-1
690
LAM
Query 610
Key 630
Value 650
670
Cross Attention

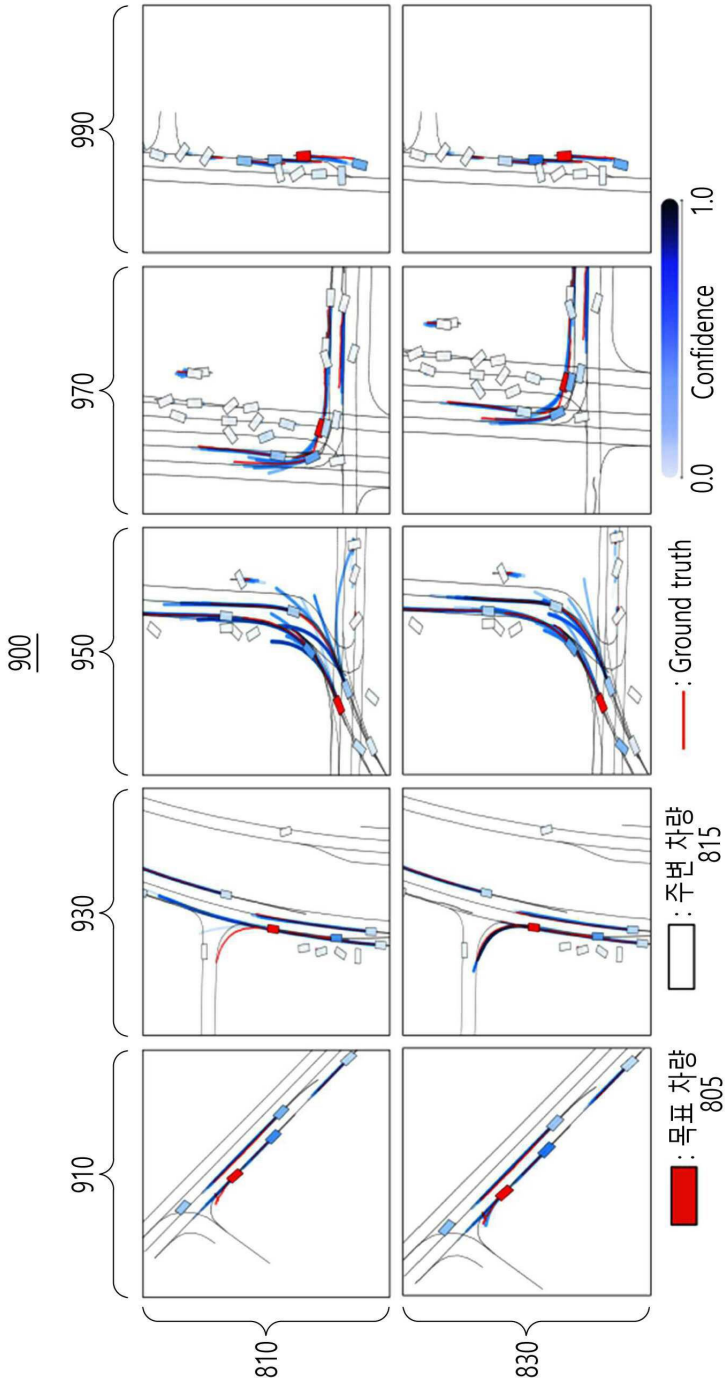
도면7


710
750-7
Lane selection
D
750-5
705
750-1
750-3
: Selected lanes,
that are near D meters
730
770-5
790-1
790-3
Interaction
between A-L
770-3
770-1
705
0.0
Confidence
1.0
715

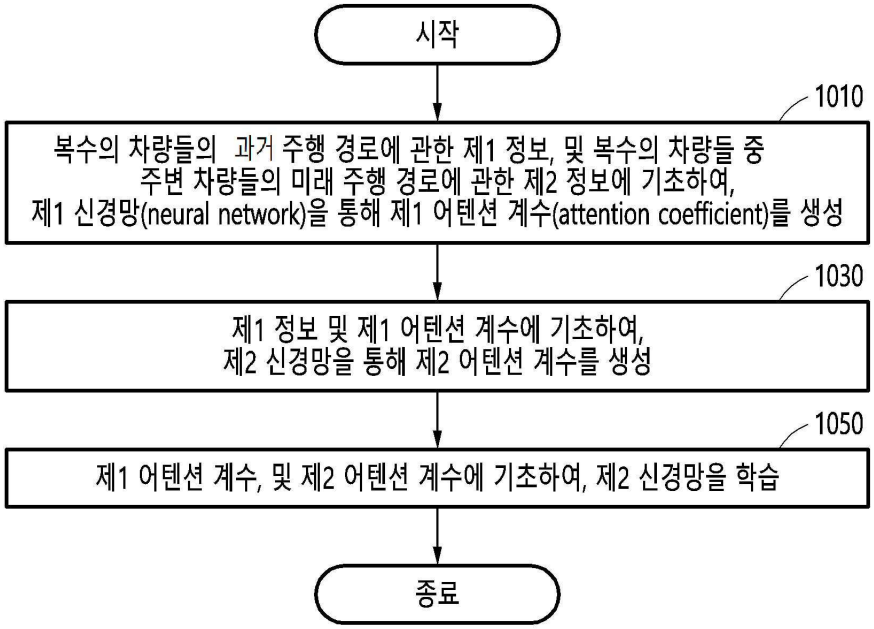
도면8


800
870-1
870-3
870-5
Scene A
Scene B
810
830
850
: 목표 차량
805
: 주변 차량
815
Confidence
0.0
1.0

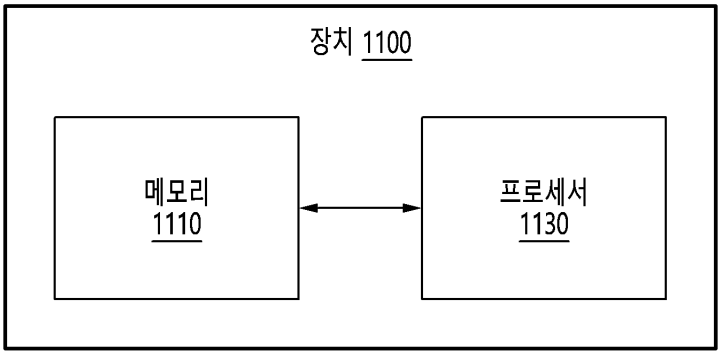
도면9


900
990
970
950
930
910
810
830
1.0
Confidence
0.0
: Ground truth
: 주변 차량
815
: 목표 차량
805

도면10


시작
1010
복수의 차량들의 과거 주행 경로에 관한 제1 정보, 및 복수의 차량들 중
주변 차량들의 미래 주행 경로에 관한 제2 정보에 기초하여,
제1 신경망(neural network)을 통해 제1 어텐션 계수(attention coefficient)를 생성
1030
제1 정보 및 제1 어텐션 계수에 기초하여,
제2 신경망을 통해 제2 어텐션 계수를 생성
1050
제1 어텐션 계수, 및 제2 어텐션 계수에 기초하여, 제2 신경망을 학습
종료

도면11


장치 1100
메모리
1110
프로세서
1130