(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2024 년 6 월 27 일 (27.06.2024)

WIPO | PCT

(10) 국제공개번호
WO 2024/136115 A1

(51) 국제특허분류:
G06V 40/20 (2022.01) *G06V 10/774* (2022.01)
G06V 20/40 (2022.01)

(21) 국제출원번호: PCT/KR2023/017770

(22) 국제출원일: 2023 년 11 월 7 일 (07.11.2023)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2022-0182779 2022 년 12 월 23 일 (23.12.2022) KR

(71) 출원인: 한국전자기술연구원 (**KOREA ELECTRONICS TECHNOLOGY INSTITUTE**) [KR/KR]; 13509 경기도 성남시 분당구 새나리로 25, Gyeonggi-do (KR).

(72) 발명자: 김성제 (**KIM, Sung Jei**); 05834 서울특별시 송파구 중대로 24, 234동 1103호, Seoul (KR). 주희영 (**JOO, Hee Young**); 16701 경기도 수원시 영통구 봉영로 1526, 712동 1001호, Gyeonggi-do (KR). 김제우 (**KIM, Je Woo**); 13105 경기도 성남시 수정구 고등로 34, 204동 102호, Gyeonggi-do (KR). 정진우 (**JEONG, Jin Woo**); 05837 서울특별시 송파구 송파대로 111, 105동 1205호, Seoul (KR). 권용훈 (**KWON, Yong Hoon**); 06783 서울특별시 서초구 동산로16길 21, 203호, Seoul (KR).

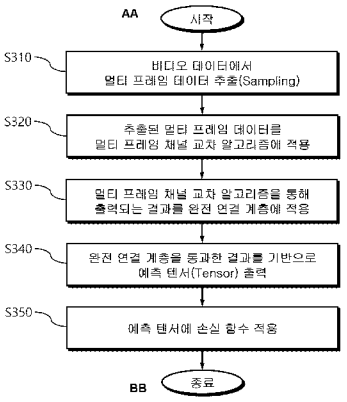
(74) 대리인: 남충우 (**NAM, Choong Woo**); 06226 서울특별시 강남구 언주로 321, 7층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) **Title:** HUMAN MICRO-GESTURE RECOGNITION SYSTEM AND METHOD TO WHICH MULTI-FRAME TIME-AXIS CHANNEL-CROSSING ALGORITHM IS APPLIED

(54) **발명의 명칭:** 멀티프레임 시간 축 채널 교차 알고리즘을 적용한 인간 미세 제스처 인식 시스템 및 방법



S310 ... Sample multi-frame data from video data
S320 ... Apply sampled multi-frame data to multi-frame channel-crossing algorithm
S330 ... Apply, to fully connected layer, result that is output through multi-frame channel-crossing algorithm
S340 ... Output predicted tensors on basis of result that passed through fully connected layer
S350 ... Apply loss function to predicted tensors
AA ... Start
BB ... End

(57) **Abstract:** A human micro-gesture recognition system and method to which a multi-frame time-axis channel-crossing algorithm is applied are provided. The human micro-gesture recognition method according to an embodiment of the present invention comprises: a first step in which a recognition system samples multi-frame data from video data; a second step in which the recognition system applies the sampled multi-frame data to a multi-frame channel-crossing algorithm; a third step in which the recognition system applies, to a fully connected layer, a result that is output through the multi-frame channel-crossing algorithm; a fourth step in which the recognition system outputs predicted tensors corresponding to a preset number of gesture classes on the basis of the result that passed through the fully connected layer; and a fifth step in which the recognition system applies a loss function to the predicted tensors. Therefore, during learning of multi-frame input data, feature channels accumulated and generated in an input order are crossed with two frames in two directions (-1 and +1 directions) with respect to a time axis, and when a zero padding calculation method is replaced with a circular shift method and zero padding calculation is performed, a portion of a channel in a start frame and an end frame is removed to 0 to prevent loss of information and, simultaneously, enable efficient calculation for recognizing a micro-gesture.

WO 2024/136115 A1

[다음 쪽 계속]

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(57) 요약서: 멀티프레임 시간 축 채널 교차 알고리즘을 적용한 인간 미세 제스처 인식 시스템 및 방법이 제공된다. 본 발명의 실시예에 따른 인간 미세 제스처 인식 방법은, 인식 시스템이, 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하는 제1 단계; 인식 시스템이, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용하는 제2 단계; 인식 시스템이, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하는 제3 단계; 인식 시스템이, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서(Tensor)가 출력되도록 하는 제4 단계; 및 인식 시스템이, 예측 텐서에 손실 함수를 적용하는 제5 단계;를 포함한다. 이에 의해, 멀티프레임 입력 데이터가 학습되는 과정에서 입력 순서에 따라 축적되며 생성되는 피처 채널들을 시간 축 기준 양 방향(-1, +1 방향) 두 프레임 교차시키고, Zero padding 연산 방법을 Circular Shift 방법으로 대체하여 Zero padding 연산 수행 시, 시작 프레임과 종료 프레임의 채널 일부를 0으로 제거하여 정보가 손실되는 것을 방지하는 동시에 미세 제스처(Micro Gesture)를 인식하기 위한 효율적 연산을 수행할 수 있다.

명세서

발명의 명칭: 멀티프레임 시간 축 채널 교차 알고리즘을 적용한 인간 미세 제스처 인식 시스템 및 방법

기술분야

[1] 본 발명은 인간 미세 제스처 인식 시스템 및 방법에 관한 것으로, 더욱 상세하게는 멀티프레임 시간 축 채널 교차 알고리즘을 적용한 인간 미세 제스처 인식 시스템 및 방법에 관한 것이다.

배경기술

[2] 인간 행동의 인식하는 인식 기술은 비디오에서 인간의 움직이는 행동을 인식하는 기술로서, 인간 행동을 인식하면 사용자에게 의도에 적합한 서비스를 제공할 수 있으므로 HCI(Human Computer Interaction) 분야에서 행동 인식 기술의 중요성이 언급되어 왔다.

[3] 또한, 행동인식 기술은 컴퓨터나 로봇이 인간과 유사한 방식으로 의사 표현을 하는 데에 필수적인 기술로서, 인간의 행동은 제스처(gesture), 액션(action), 행동(activity) 등으로 세부적으로 분류할 수 있다.

[4] 여기서, 제스처란 인간의 독립된 신체 부위의 움직임을 말하며, 액션은 다양한 신체 부위가 지속적으로 움직이는 것을 의미하고, 행동이란 긴 시간동안 연속적으로 움직이는 것을 말한다.

[5] 행동 인식 세부 분야 중 특히 제스처 인식은 인간의 단일 신체 부위 움직임을 의미하므로 손의 움직임이 주가 되는 수화 인식이 제스처 인식의 대표적 사례라 할 수 있다.

[6] 하지만 움직임의 의도가 명확히 드러나는 수화와 달리 의도가 명백히 드러나지 않는 미세한 제스처 인식의 기술적 필요성 또한 대두되고 있다.

[7] 예를 들어 인사를 하기 위해 손을 양 옆으로 흔드는 동작과 긴장하여 양 손 끝을 비비는 동작을 비교하면 전자의 경우가 후자에 비해 행동 의도가 명확히 드러나며 동작의 폭이 더 크다.

[8] 이와 같이 의도가 명백히 드러나지 않지만 인간 감정을 기저에 두어 세밀하게 드러나는 미묘한 제스처를 인식하는 인공지능 기술은 표현의 변화 폭이 적어 행동 인식 난이도가 높다.

[9] 이러한 인간 미세 제스처(Micro Gesture)인식 기술은 깊은 소통이 필요한 상담이나 교육용 챗봇 등에 무궁무진하게 활용 가능하므로 인간의 지능 뿐 아니라 인간 심리를 파악하고 이를 모사하는 인공지능 기술의 발전을 위해 연구가 필요한 분야이다.

[10] 종래의 인간 행동 인식 기술은 명확히 구분이 가능한 개인의 행동 또는 인간과 다른 물체와의 상호작용이 중심이 된 행동을 분류하는 방향으로 발전하였다.

[11] 인간 행동 오픈 데이터는 적게는 100가지, 많게는 700가지 항목으로 분류된 대용량 비디오 데이터이므로 개인의 미세한 행동이 아닌 다양한 상황 속 연출되는 인간의 두드러진 동작과 관련되어 있다.

[12] 따라서, 종래의 행동 인식 기술은 시간 축에 따른 모션 변화가 큰 행동 특징점(Feature) 추출에 적합하도록 발전하였으나, 인간 깊은 내면의 감정에 기인하여 세밀하게 드러나는 미세 제스처 인식을 위한 데이터 셋은 기존 행동인식 기술 개발에 보편적으로 사용되는 데이터로는 한계가 있다.

[13] 또한, 종래의 모션 특징점 추출 방법은 시간에 따른 인간 행동 변화가 큰 기존 데이터 셋에 적합하도록 개발되어 기존 방법으로는 미세 제스처의 변화가 상대적으로 적은 세밀한 모션을 모델링하는 데에 한계가 있다.

[14] 예를 들면, 시간 축을 포함한 3D 컨볼루션(Convolution) 기반의 모션 피처 추출기는 기본적으로 32 프레임 이상의 입력 비디오 데이터에서 발생하는 모션에 적합하도록 모델링되었기 때문에 25 FPS 영상에서 1초 미만으로 발생하는 마이크로 제스처 피처 추출을 위해서는 이전과는 다른 모델링 기술이 필요하다.

발명의 상세한 설명

기술적 과제

[15] 본 발명은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은, 멀티프레임 입력 데이터가 학습되는 과정에서 입력 순서에 따라 축적되며 생성되는 피처 채널들을 시간 축 기준 양 방향(-1, +1 방향) 두 프레임 교차시키며, Zero padding 연산 수행 시, 시작 프레임과 종료 프레임의 채널 일부를 0으로 제거하여 정보가 손실되는 것을 방지하기 위해, Zero padding 연산 방법을 Circular Shift 방법으로 대체할 수 있는 인간 미세 제스처 인식 시스템 및 방법을 제공함에 있다.

[16] 또한, 본 발명의 다른 목적은, 이웃하는 프레임 뿐 아니라, 현재 프레임 기준 두 프레임 이상 차이가 나는 모든 프레임의 피처 정보를 담은 채널을 포함하는 전략으로 Short term 뿐만 아니라 Long term 정보에 대한 모델링이 가능한 인간 미세 제스처 인식 시스템 및 방법을 제공함에 있다.

과제 해결 수단

[17] 상기 목적을 달성하기 위한 본 발명의 일 실시예에 따른, 인간 미세 제스처 인식 방법은, 인식 시스템이, 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하는 제1 단계; 인식 시스템이, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용하는 제2 단계; 인식 시스템이, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하는 제3 단계; 인식 시스템이, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서(Tensor)가 출력되도록 하는 제4 단계; 및 인식 시스템이, 예측 텐서에 손실 함수를 적용하는 제5 단계;를 포함한다.

[18] 그리고 제1 단계는, 원본 비디오 데이터를 대상으로 세그먼트 기반 추출 기법(Segment Based Sampling Method)을 적용하여 복수의 프레임(멀티 프레임)을 갖는 학습용 입력 데이터 클립(Clip)을 추출할 수 있다.

[19] 또한, 제2 단계는, 제1 단계에서 추출된 클립 단위 데이터를 대상으로 다수 계층(Multi Layer)으로 구성된 컨볼루션(Convolution) 연산을 수행하되, 컨볼루션 계층마다 해당 연산이 수행되기 이전에 해당 계층의 채널 교차 알고리즘을 적용할 수 있다.

[20] 그리고 멀티 프레임 채널 교차 알고리즘은, 이전 컨볼루션 계층을 통과한 후 생성된 출력 피처(feature)를 입력으로 하고, 시간 축을 기준으로 양 방향(+ 방향 및 - 방향)으로, 이웃하는 채널 일부를 각 방향을 기준으로 현재 프레임의 해당 채널과 교차시키는 구조로 형성될 수 있다.

[21] 또한, 멀티 프레임 채널 교차 알고리즘은, 정보가 손실되는 것을 방지하기 위해, 이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동(Circular Shift) 되어, 현재 프레임의 해당 채널과 교차될 수 있다.

[22] 그리고 멀티 프레임 채널 교차 알고리즘은, 이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동되어 현재 프레임의 해당 채널과 교차되는 경우, 현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 모양 크기 조정(Shape Resizing)이 수행될 수 있다.

[23] 또한, 멀티 프레임 채널 교차 알고리즘은, 현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 Shape Resizing이 수행되는 경우, Height 축 및 Width 축 방향을 따라 Shape Resizing의 대상이 되는 채널 피처를 기설정된 값으로 응축시키고, 이때, 기설정된 값은, 1/(프레임 수)가 될 수 있다.

[24] 그리고 제5 단계는, 손실 함수인 Cross-Entropy를 적용하여 예측 텐서와 정답(Ground Truth) 사이의 손실(Loss)을 출력한 후, 손실을 줄이는 방향으로 각 필터 레이어 가중치를 학습할 수 있다.

[25] 한편, 본 발명의 다른 실시예에 따른, 인간 미세 제스처 인식 시스템은, 멀티 프레임 채널 교차 알고리즘이 저장되는 저장부; 및 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하고, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용하며, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하고, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서(Tensor)가 출력되도록 하며, 예측 텐서에 손실 함수를 적용하는 프로세서;를 포함한다.

[26] 그리고 본 발명의 다른 실시예에 따른, 인간 미세 제스처 인식 방법은, 인식 시스템이, 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하는 단계; 및 인식 시스템이, 추출된 멀티 프레임 데이터를 시간 가중치 멀티프레임 피처 추출기에 적용하는 단계;를 포함하고, 시간 가중치 멀티프레임 피처 추출기는, 각 프

레임 피처가 Embedding Layer를 거쳐 LSTM(Long Short Term Memory)에 입력되는 구조로 형성된다.

[27] 그리고 Embedding Layer를 거친 프레임 피처는, 시간 순서에 따라 LSTM에 입력되며, 시간 가중치 멀티프레임 피처 추출기는, 종료 프레임을 기준으로 Short Term뿐 아니라 Long Term으로 입력되는 모든 프레임 정보가 모델링된 최종 피처가 추출될 수 있다.

[28] 또한, 본 발명의 다른 실시예에 따른, 인간 미세 제스처 인식 시스템은, 시간 가중치 멀티프레임 피처 추출기가 저장되는 저장부; 및 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하고, 추출된 멀티 프레임 데이터를 시간 가중치 멀티프레임 피처 추출기에 적용하는 프로세서;를 포함하며, 시간 가중치 멀티프레임 피처 추출기는, 각 프레임 피처가 Embedding Layer를 거쳐 LSTM(Long Short Term Memory)에 입력되는 구조로 형성된다.

발명의 효과

[29] 이상 설명한 바와 같이, 본 발명의 실시예들에 따르면, 멀티프레임 입력 데이터가 학습되는 과정에서 입력 순서에 따라 축적되며 생성되는 피처 채널들을 시간 축 기준 양 방향(-1, +1 방향) 두 프레임 교차시키고, Zero padding 연산 방법을 Circular Shift 방법으로 대체하여 Zero padding 연산 수행 시, 시작 프레임과 종료 프레임의 채널 일부를 0으로 제거하여 정보가 손실되는 것을 방지하는 동시에 미세 제스처(Micro Gesture)를 인식하기 위한 효율적 연산을 수행할 수 있다.

[30] 또한, 본 발명의 실시예들에 따르면, 이웃하는 프레임 뿐 아니라, 현재 프레임 기준 두 프레임 이상 차이가 나는 모든 프레임의 피처 정보를 담은 채널을 포함하는 전략으로 Short term 뿐만 아니라 Long term 정보에 대한 모델링이 가능하도록 할 수 있다.

도면의 간단한 설명

[31] 도 1은, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 시스템의 구성 설명에 제공된 도면,

[32] 도 2는, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 시스템의 동작 설명에 제공된 도면,

[33] 도 3은, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 방법의 설명에 제공된 흐름도,

[34] 도 4는, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘의 설명에 제공된 도면,

[35] 도 5는, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘에 적용되는 멀티 프레임 피처(Multi Frame feature)가 도식화된 모습이 예시된 도면,

[36] 도 6은, 멀티 프레임 피처에 Zero padding 연산 기법이 적용되는 모습이 예시된 도면,

[37] 도 7은, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘에서 멀티 프레임 피처에 순환 이동(Circular Shift) 기법이 적용되는 모습이 예시된 도면, 그리고

[38] 도 8은, 본 발명의 다른 실시예에 따른 시간 가중치 멀티프레임 피처 추출기의 설명에 제공된 도면이다.

발명의 실시를 위한 최선의 형태

[39] 이하에서는 도면을 참조하여 본 발명을 보다 상세하게 설명한다.

[40] 도 1은, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 시스템의 구성 설명에 제공된 도면이고, 도 2는, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 시스템의 동작 설명에 제공된 도면이다.

[41] 본 실시예에 따른 멀티프레임 시간 축 채널 교차 알고리즘을 적용한 인간 미세 제스처 인식 시스템(이하에서는 '인간 미세 제스처 인식 시스템'으로 총칭하기로 함)은, 멀티프레임 입력 데이터가 학습되는 과정에서 입력 순서에 따라 축적되며 생성되는 피처 채널들을 시간 축 기준 양 방향(-1, +1 방향) 두 프레임 교차시키며, Zero padding 연산 수행 시, 시작 프레임과 종료 프레임의 채널 일부를 0으로 제거하여 정보가 손실되는 것을 방지하기 위해, Zero padding 연산 방법을 Circular Shift 방법으로 대체할 수 있다.

[42] 또한, 본 실시예에 따른 인간 미세 제스처 인식 시스템은, 이웃하는 프레임 뿐 아니라, 현재 프레임 기준 두 프레임 이상 차이가 나는 모든 프레임의 피처 정보를 담은 채널을 포함하는 전략으로 Short term 뿐만 아니라 Long term 정보에 대한 모델링을 지원할 수 있다.

[43] 도 1을 참조하면, 본 인간 미세 제스처 인식 시스템은, 통신부(110), 프로세서(120) 및 저장부(130)를 포함할 수 있다.

[44] 통신부(110)는, 비디오 데이터를 획득하기 위해, 외부 시스템과 연결될 수 있다.

[45] 저장부(130)는, 프로세서(120)가 동작함에 있어 필요한 프로그램 및 데이터를 저장하는 저장매체이다.

[46] 예를 들면, 저장부(130)는, 멀티 프레임 채널 교차 알고리즘에 대한 정보가 저장될 수 있다.

[47] 프로세서(120)는, 인간 미세 제스처 인식 시스템의 제반 사항들을 처리할 수 있다.

[48] 예를 들면, 프로세서(120)는 통신부(110)를 통해 획득된 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하며, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용할 수 있다.

[49] 그리고 프로세서(120)는 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하며, 완전 연결 계층을 통과한 결과를 기반으로 예측 텐서(Tensor)가 출력되도록 하고, 예측 텐서에 손실 함

수를 적용하여, 예측 텐서와 정답(Ground Truth) 사이의 손실(Loss)을 출력한 후, 손실을 줄이는 방향으로 각 필터 레이어 가중치를 학습시킬 수 있다.

[50] 도 3은, 본 발명의 일 실시예에 따른 인간 미세 제스처 인식 방법의 설명에 제공된 흐름도이다.

[51] 본 실시예에 따른 인간 미세 제스처 인식 방법은, 도 1 내지 도 2를 참조하여 전술한 인간 미세 제스처 인식 시스템에 의해 실행될 수 있다.

[52] 도 3을 참조하면, 인간 미세 제스처 인식 방법은, 인간 미세 제스처 인식 시스템을 이용하여 비디오 데이터에서 멀티 프레임 데이터를 추출하며(S310), 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용할 수 있다(S320).

[53] 이때, 인간 미세 제스처 인식 시스템은 비디오 데이터에서 멀티 프레임 데이터를 추출하는 경우, 원본 비디오 데이터를 대상으로 세그먼트 기반 추출 기법(Segment Based Sampling Method)을 적용하여 복수의 프레임(멀티 프레임)을 갖는 학습용 입력 데이터 클립(Clip)을 추출할 수 있다.

[54] 그리고 인간 미세 제스처 인식 방법은, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층에 적용하며(S330), 완전 연결 계층을 통과한 결과를 기반으로 예측 텐서가 출력되도록 하고(S340), 예측 텐서에 손실 함수를 적용하여(S350), 예측 텐서와 정답(Ground Truth) 사이의 손실(Loss)을 출력한 후, 손실을 줄이는 방향으로 각 필터 레이어 가중치를 학습시킬 수 있다.

[55] 여기서, 인간 미세 제스처 인식 시스템은 예측 텐서가 출력되도록 하는 경우, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서가 출력되도록 할 수 있다.

[56] 그리고 인간 미세 제스처 인식 시스템은 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스 개수만큼 예측 텐서가 출력되면, 출력된 각 예측 텐서에 Softmax 함수를 적용하여 각 클래스별 확률 값을 출력할 수 있다.

[57] 도 4는, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘의 설명에 제공된 도면이다.

[58] 도 4를 참조하면, 본 실시예에 따른 멀티 프레임 채널 교차 알고리즘은, 다수 계층(Multi Layer)으로 구성될 수 있다.

[59] 이를 통해, 인간 미세 제스처 인식 시스템은 복수의 프레임(멀티 프레임)을 갖는 입력 데이터 클립이 추출되면, 추출된 입력 데이터 클립을 대상으로 다수 계층(L0~LN)으로 구성된 컨볼루션(Convolution) 연산을 수행하되, 컨볼루션 계층마다 해당 연산이 수행되기 이전에 해당 계층의 채널 교차 알고리즘이 적용되도록 할 수 있다.

[60] 그리고 멀티 프레임 채널 교차 알고리즘은, 이전 컨볼루션 계층을 통과한 후 생성된 출력 피처(feature)를 입력으로 하고, 시간 축을 기준으로 양 방향(+ 방향 및 - 방향)으로, 이웃하는 채널 일부를 각 방향을 기준으로 현재 프레임의 해당 채널과 교차시키는 구조로 형성될 수 있다.

[61] 여기서, 도 5는, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘에 적용되는 멀티 프레임 피처(Multi Frame feature)가 도식화된 모습이 예시된 도면이다.

[62] 시간 축은 1차원이므로 교차 방향은 두 개 존재하며, 도 5에서는 이를 - 방향 및 + 방향으로 표기한다.

[63] 도 6은, 멀티 프레임 피처에 Zero padding 연산 기법이 적용되는 모습이 예시된 도면이고, 도 7은, 본 발명의 일 실시예에 따른 멀티 프레임 채널 교차 알고리즘에서 멀티 프레임 피처에 순환 이동(Circular Shift) 기법이 적용되는 모습이 예시된 도면,

[64] 멀티 프레임 채널 교차 알고리즘은, 도 6에 예시된 바와 같이 멀티 프레임 피처에 Naive Shift를 통한 Zero padding 연산 기법이 적용되는 경우, Zero padding 연산으로 채널 일부를 0으로 제거하여 정보를 온전히 활용하지 못하는 문제가 발생할 수 있다.

[65] 이에, 본 멀티 프레임 채널 교차 알고리즘은, Zero padding 연산 수행 시, 시작 프레임과 종료 프레임의 채널 일부를 0으로 제거하여 정보가 손실되는 것을 방지하기 위해, 도 7에 예시된 바와 같이 이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동(Circular Shift) 되어, 현재 프레임의 해당 채널과 교차되도록 할 수 있다.

[66] 그리고 멀티 프레임 채널 교차 알고리즘은, 이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동되어 현재 프레임의 해당 채널과 교차되는 경우, 현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 모양 크기 조정(Shape Resizing)이 수행되도록 할 수 있다.

[67] 구체적으로 예를 들면, 멀티 프레임 채널 교차 알고리즘은, 현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 Shape Resizing이 수행되는 경우, Height 축 및 Width 축 방향을 따라 Shape Resizing의 대상이 되는 채널 피처를 기설정된 값으로 응축시킬 수 있다.

[68] 이때, 기설정된 값은, 1/(프레임 수) 가 될 수 있다.

[69] 이를 통해, 기존 채널 중 일부는, 도 5에 예시된 바와 같이 응축된 멀티프레임 채널 피처로 대체될 수 있다.

[70] 도 8은, 본 발명의 다른 실시예에 따른 시간 가중치 멀티프레임 피처 추출기의 설명에 제공된 도면이다.

[71] 본 실시예에 따른 인간 미세 제스처 인식 시스템은, 도 1을 참조하여 전술한 시스템의 구성은 동일하나, 멀티 프레임 채널 교차 알고리즘을 대체하여 클립 단위 데이터를 시간 가중치 멀티프레임 피처 추출기에 적용할 수 있다.

[72] 시간 가중치 멀티프레임 피처 추출기는, 각 프레임 피처가 Embedding Layer를 거쳐 LSTM(Long Short Term Memory)에 입력되는 구조로 형성될 수 있다.

[73] Embedding Layer를 거친 프레임 피처는, 시간 순서에 따라 LSTM에 입력될 수 있으며, 시간 가중치 멀티프레임 피처 추출기는, 종료 프레임을 기준으로 Short Term뿐 아니라 Long Term으로 입력되는 모든 프레임 정보가 모델링된 최종 피처가 추출되도록 할 수 있다.

[74] 첨언하면, 시간 가중치 멀티프레임 피처 추출기에서 출력된 최종 피처는, 별도로 마련되는 메모리 공간에 저장될 수 있으며, 이후, 멀티 프레임 채널 교차 알고리즘에 적용되는 멀티 프레임 데이터에 Concatenate되는 채널 역할을 수행할 수 있다.

[75] 한편, 본 실시예에 따른 장치와 방법의 기능을 수행하게 하는 컴퓨터 프로그램을 수록한 컴퓨터로 읽을 수 있는 기록매체에도 본 발명의 기술적 사상이 적용될 수 있음은 물론이다. 또한, 본 발명의 다양한 실시예에 따른 기술적 사상은 컴퓨터로 읽을 수 있는 기록매체에 기록된 컴퓨터로 읽을 수 있는 코드 형태로 구현될 수도 있다. 컴퓨터로 읽을 수 있는 기록매체는 컴퓨터에 의해 읽을 수 있고 데이터를 저장할 수 있는 어떤 데이터 저장 장치이더라도 가능하다. 예를 들어, 컴퓨터로 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광디스크, 하드 디스크 드라이브, 등이 될 수 있음은 물론이다. 또한, 컴퓨터로 읽을 수 있는 기록매체에 저장된 컴퓨터로 읽을 수 있는 코드 또는 프로그램은 컴퓨터간에 연결된 네트워크를 통해 전송될 수도 있다.

[76] 또한, 이상에서는 본 발명의 바람직한 실시예에 대하여 도시하고 설명하였지만, 본 발명은 상술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 발명의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안 될 것이다.

청구범위

[청구항 1] 인식 시스템이, 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하는 제1 단계;
인식 시스템이, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용하는 제2 단계;
인식 시스템이, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하는 제3 단계;
인식 시스템이, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서(Tensor)가 출력되도록 하는 제4 단계; 및
인식 시스템이, 예측 텐서에 손실 함수를 적용하는 제5 단계;를 포함하는 인간 미세 제스처 인식 방법.

[청구항 2] 청구항 1에 있어서,
제1 단계는,
원본 비디오 데이터를 대상으로 세그먼트 기반 추출 기법(Segment Based Sampling Method)을 적용하여 복수의 프레임(멀티 프레임)을 갖는 학습용 입력 데이터 클립(Clip)을 추출하는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 3] 청구항 2에 있어서,
제2 단계는,
제1 단계에서 추출된 클립 단위 데이터를 대상으로 다수 계층(Multi Layer)으로 구성된 컨볼루션(Convolution) 연산을 수행하되, 컨볼루션 계층마다 해당 연산이 수행되기 이전에 해당 계층의 채널 교차 알고리즘을 적용하는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 4] 청구항 3에 있어서,
멀티 프레임 채널 교차 알고리즘은,
이전 컨볼루션 계층을 통과한 후 생성된 출력 피처(feature)를 입력으로 하고,
시간 축을 기준으로 양 방향(+ 방향 및 - 방향)으로, 이웃하는 채널 일부를 각 방향을 기준으로 현재 프레임의 해당 채널과 교차시키는 구조로 형성되는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 5] 청구항 4에 있어서,
멀티 프레임 채널 교차 알고리즘은,
정보가 손실되는 것을 방지하기 위해, 이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동(Circular Shift) 되어, 현재 프레임의 해당 채널과 교차되도록 하는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 6] 청구항 5에 있어서,
멀티 프레임 채널 교차 알고리즘은,
이웃하는 채널의 각 프레임 피처가 각 방향(+ 방향 및 - 방향)을 따라 순환 이동되어 현재 프레임의 해당 채널과 교차되는 경우, 현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 모양 크기 조정(Shape Resizing)이 수행되도록 하는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 7] 청구항 6에 있어서,
멀티 프레임 채널 교차 알고리즘은,
현재 프레임에서 두 프레임 이상 간격을 갖는 멀티 프레임의 채널 피처 중 일부 또는 전부를 대상으로 Shape Resizing이 수행되는 경우, Height 축 및 Width 축 방향을 따라 Shape Resizing의 대상이 되는 채널 피처를 기설정된 값으로 응축시키고,
기설정된 값은,
1/(프레임 수)가 되는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 8] 청구항 1에 있어서,
제5 단계는,
손실 함수인 Cross-Entropy를 적용하여 예측 텐서와 정답(Ground Truth) 사이의 손실(Loss)을 출력한 후, 손실을 줄이는 방향으로 각 필터 레이어 가중치를 학습하는 것을 특징으로 하는 것을 인간 미세 제스처 인식 방법.

[청구항 9] 멀티 프레임 채널 교차 알고리즘이 저장되는 저장부; 및
비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하고, 추출된 멀티 프레임 데이터를 멀티 프레임 채널 교차 알고리즘에 적용하며, 멀티 프레임 채널 교차 알고리즘을 통해 출력되는 결과를 완전 연결 계층(Fully Connected Layer)에 적용하고, 완전 연결 계층을 통과한 결과를 기반으로 기설정된 제스처 클래스(Class) 개수만큼 예측 텐서(Tensor)가 출력되도록 하며, 예측 텐서에 손실 함수를 적용하는 프로세서;를 포함하는 인간 미세 제스처 인식 시스템.

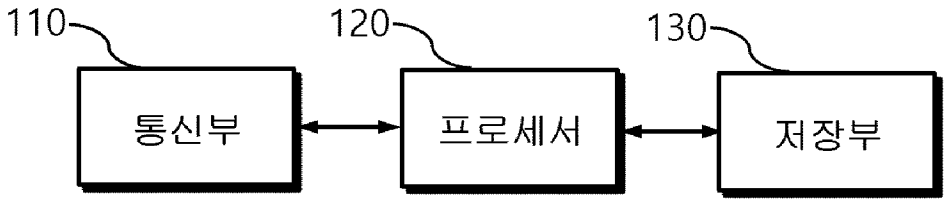
[청구항 10] 인식 시스템이, 비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하는 단계; 및
인식 시스템이, 추출된 멀티 프레임 데이터를 시간 가중치 멀티프레임 피처 추출기에 적용하는 단계;를 포함하고,
시간 가중치 멀티프레임 피처 추출기는,
각 프레임 피처가 Embedding Layer를 거쳐 LSTM(Long Short Term Memory)에 입력되는 구조로 형성되는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 11] 청구항 10에 있어서,
Embedding Layer를 거친 프레임 피처는,

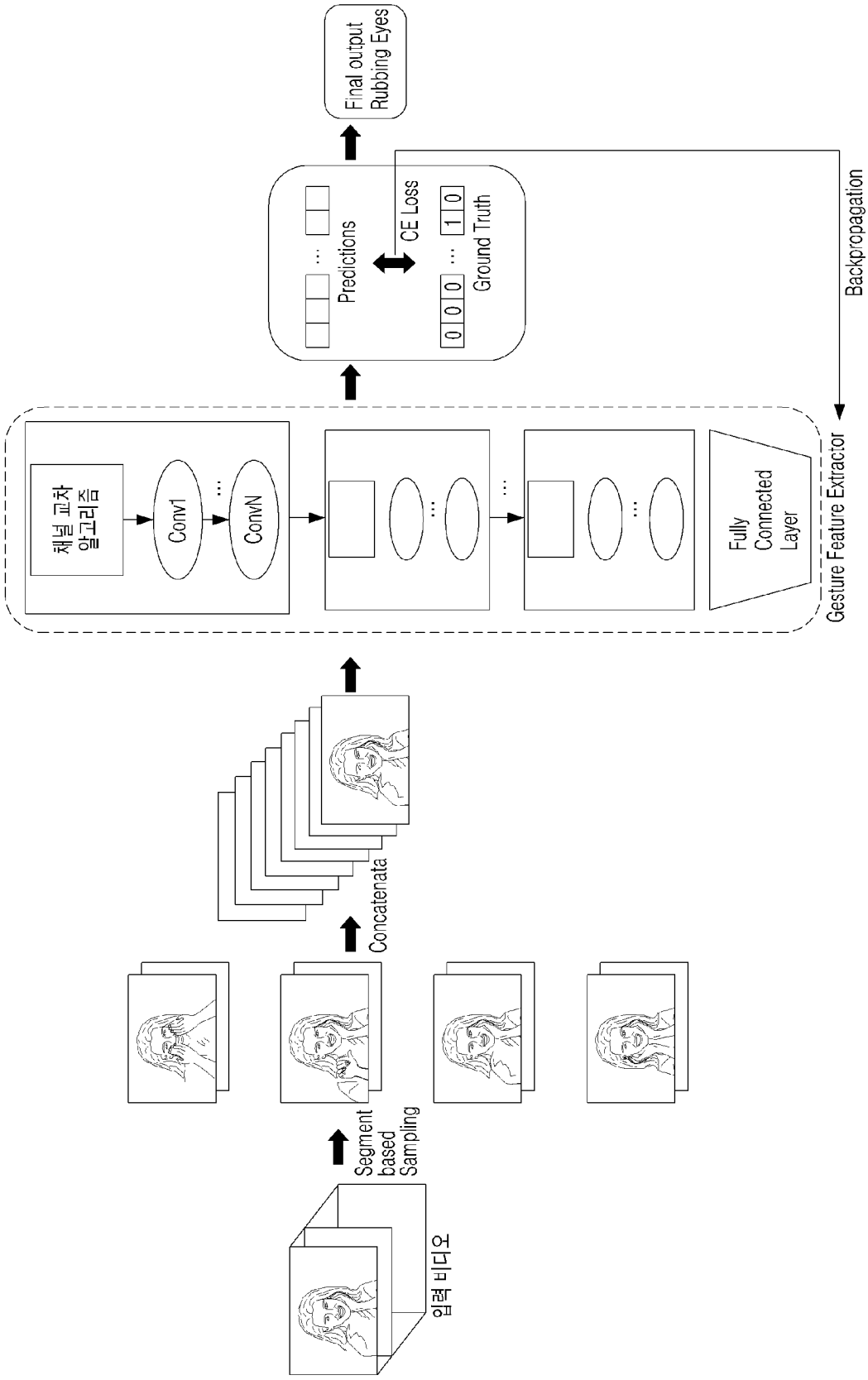
시간 순서에 따라 LSTM에 입력되며,
시간 가중치 멀티프레임 피처 추출기는,
종료 프레임을 기준으로 Short Term뿐 아니라 Long Term으로 입력되는 모든 프레임 정보가 모델링된 최종 피처가 추출되는 것을 특징으로 하는 인간 미세 제스처 인식 방법.

[청구항 12] 시간 가중치 멀티프레임 피처 추출기가 저장되는 저장부; 및
비디오 데이터에서 멀티 프레임 데이터를 추출(Sampling)하고, 추출된 멀티 프레임 데이터를 시간 가중치 멀티프레임 피처 추출기에 적용하는 프로세서;를 포함하며,
시간 가중치 멀티프레임 피처 추출기는,
각 프레임 피처가 Embedding Layer를 거쳐 LSTM(Long Short Term Memory)에 입력되는 구조로 형성되는 것을 특징으로 하는 인간 미세 제스처 인식 시스템.

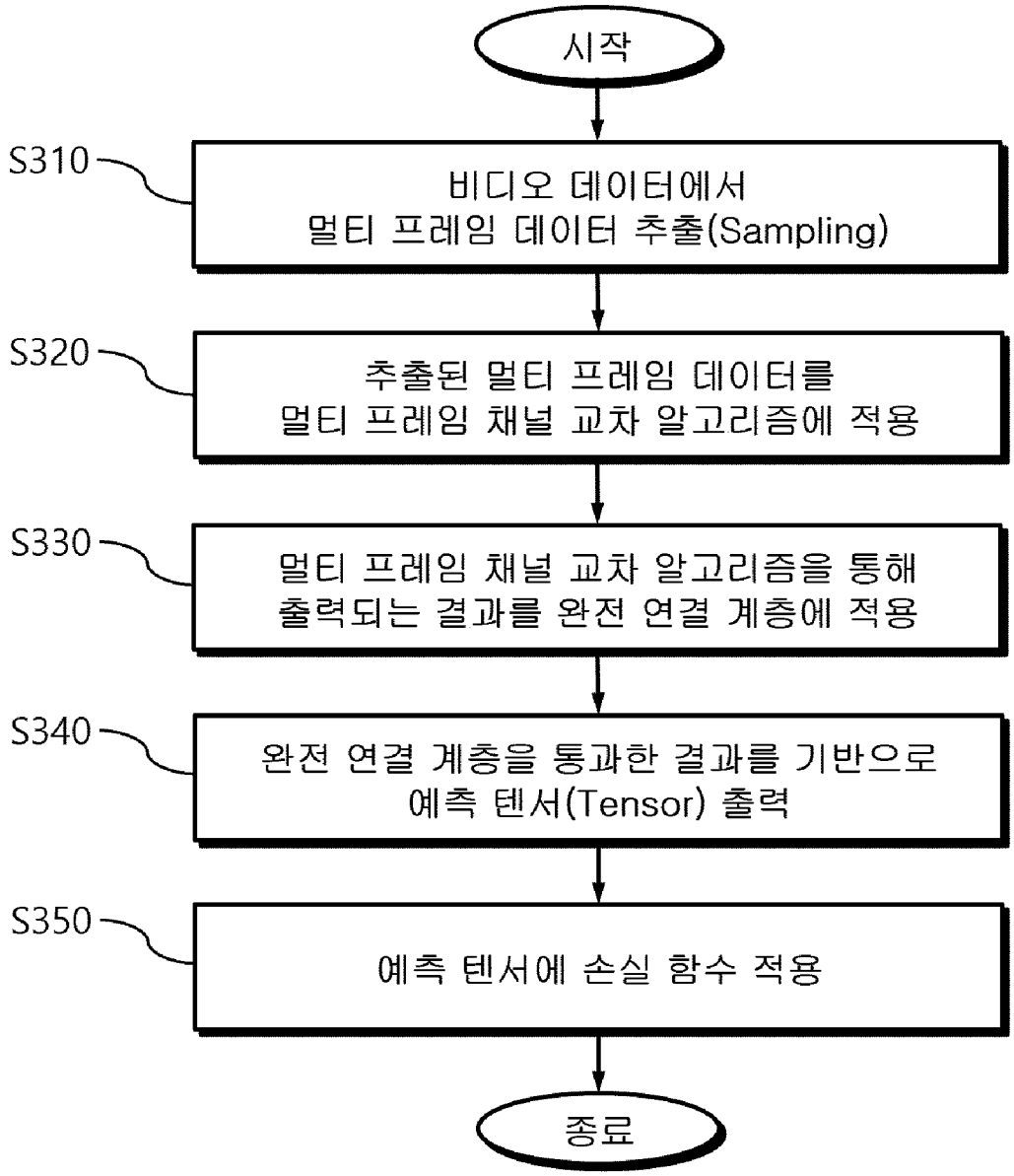
[도1]



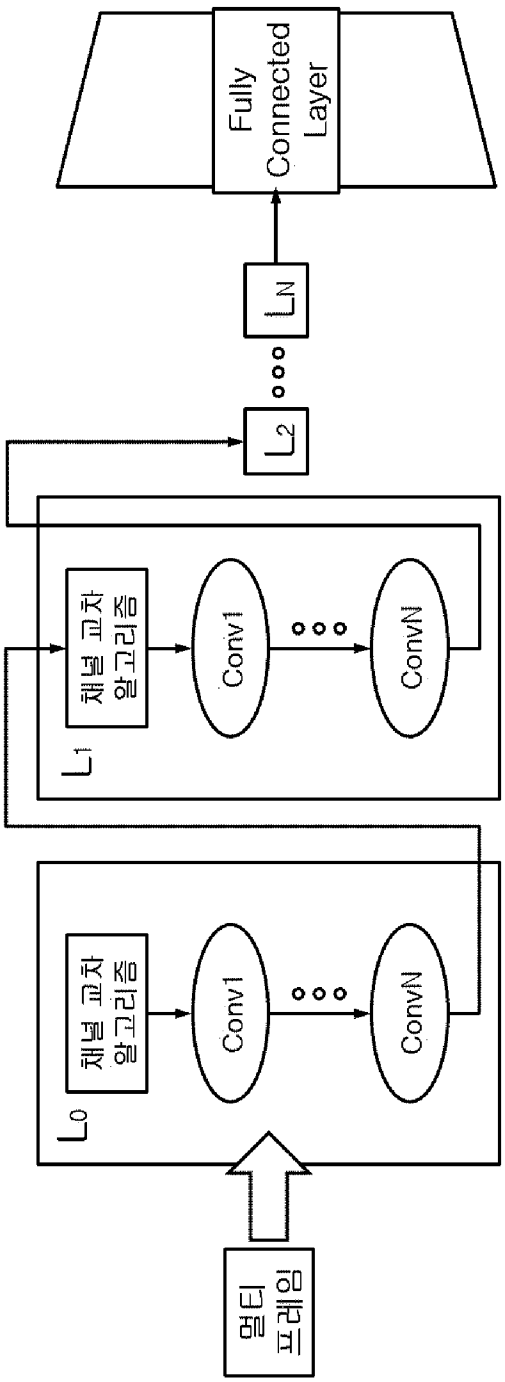
[도2]


Final output
Rubbing Eyes
Predictions
CE Loss
Ground Truth
Backpropagation
채널 교차 알고리즘
Conv1
ConvN
Fully Connected Layer
Gesture Feature Extractor
Concatenata
Segment based Sampling
입력 비디오

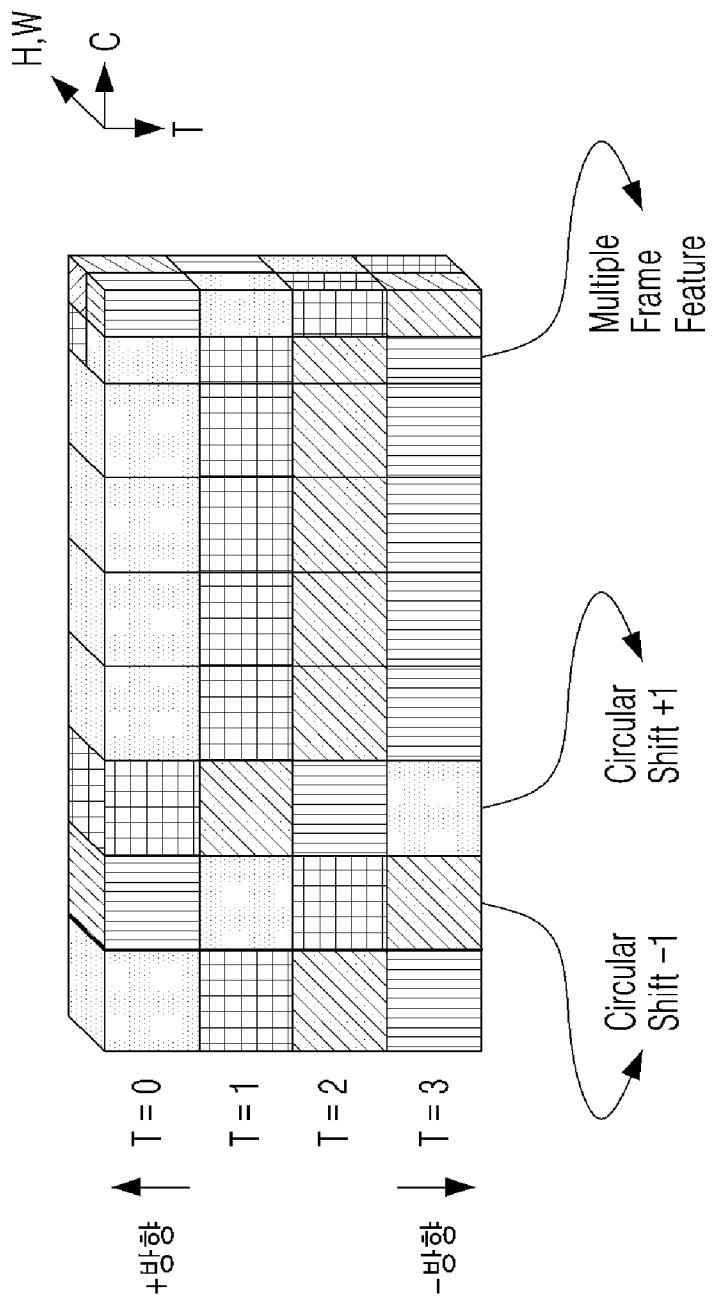
[도3]


시작
S310
비디오 데이터에서
멀티 프레임 데이터 추출(Sampling)
S320
추출된 멀티 프레임 데이터를
멀티 프레임 채널 교차 알고리즘에 적용
S330
멀티 프레임 채널 교차 알고리즘을 통해
출력되는 결과를 완전 연결 계층에 적용
S340
완전 연결 계층을 통과한 결과를 기반으로
예측 텐서(Tensor) 출력
S350
예측 텐서에 손실 함수 적용
종료

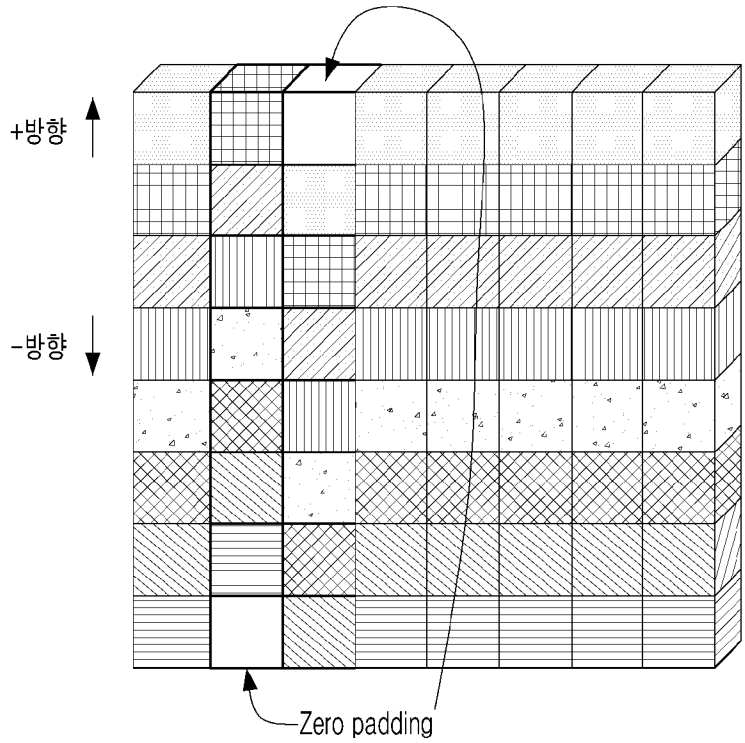
[도4]


Fully Connected Layer
LN
L2
L1
채널 교차 알고리즘
Conv1
ConvN
L0
채널 교차 알고리즘
Conv1
ConvN
멀티 프레임

[도5]

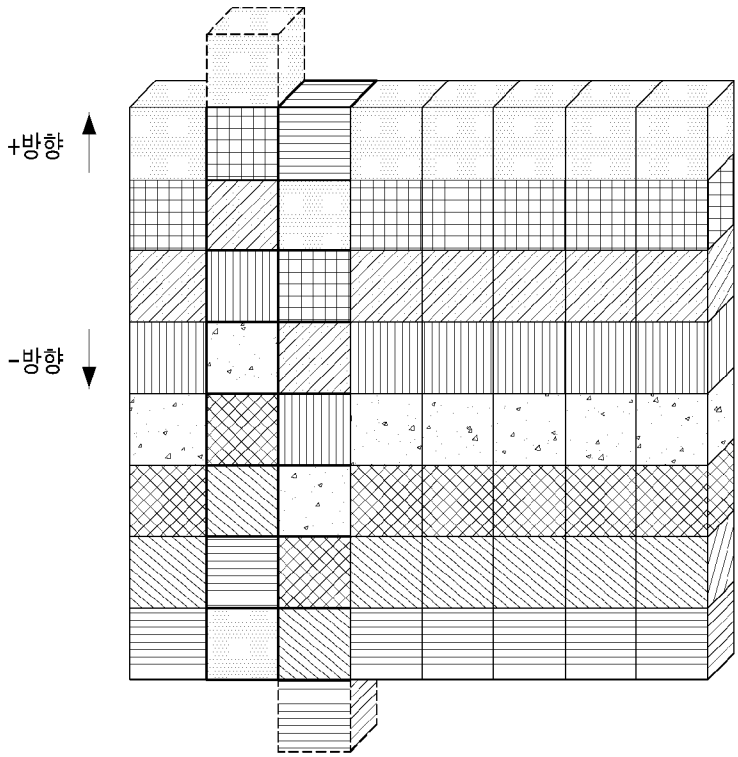

H,W
C
T
Multiple Frame Feature
Circular Shift +1
Circular Shift -1
T = 0
T = 1
T = 2
T = 3
+방향
-방향

[도6]



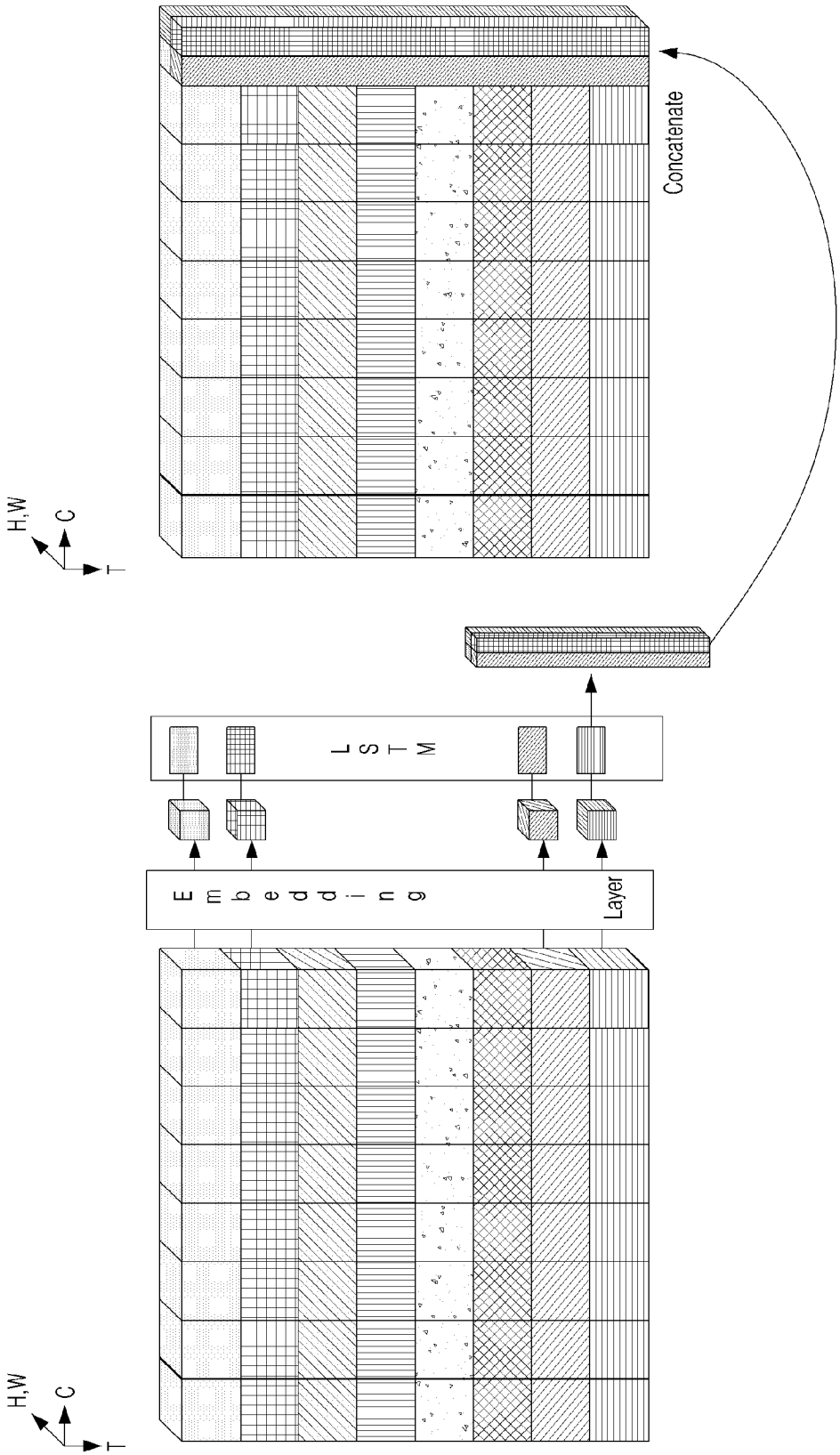
Naive Shift

[도7]



Circular Shift

[도8]



INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2023/017770

A. CLASSIFICATION OF SUBJECT MATTER

G06V 40/20(2022.01)i; **G06V 20/40**(2022.01)i; **G06V 10/774**(2022.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06V 40/20(2022.01); G06K 9/00(2006.01); G06K 9/62(2006.01); G06N 99/00(2010.01); G06V 10/762(2022.01); G06V 20/40(2022.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 제스처 인식(gesture recognition) 프레임(frame), 알고리즘(algorithm), 완전 연결 계층(fully connected layer), 텐서(tensor), LSTM

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 111860117 A (ANHUI BIGENG SOFTWARE CO., LTD.) 30 October 2020 (2020-10-30) See claims 1 and 7-8 and figure 1.	10-12
A		1-9
A	CN 113158909 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 23 July 2021 (2021-07-23) See claims 1 and 7-8 and figures 1-7.	1-12
A	CN 111046798 A (WISDOM VISION (HANGZHOU) TECHNOLOGY DEVELOPMENT CO., LTD.) 21 April 2020 (2020-04-21) See claims 1 and 4 and figures 1-2.	1-12
A	CN 115131710 A (FUZHOU UNIVERSITY) 30 September 2022 (2022-09-30) See claims 1-6 and figure 1.	1-12

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 February 2024	**05 February 2024**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office **Government Complex-Daejeon Building 4, 189 Cheongsa-ro, Seo-gu, Daejeon 35208**	
Facsimile No. **+82-42-481-8578**	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2023/017770

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2019-0080818 A (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 08 July 2019 (2019-07-08) See claims 1-17 and figures 1-8.	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2023/017770

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 111860117 A	30 October 2020	None	
CN 113158909 A	23 July 2021	CN 113158909 B	27 June 2023
CN 111046798 A	21 April 2020	None	
CN 115131710 A	30 September 2022	None	
KR 10-2019-0080818 A	08 July 2019	None	

국 제 조 사 보 고 서

국제출원번호
PCT/KR2023/017770

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06V 40/20(2022.01)i; **G06V 20/40**(2022.01)i; **G06V 10/774**(2022.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06V 40/20(2022.01); G06K 9/00(2006.01); G06K 9/62(2006.01); G06N 99/00(2010.01); G06V 10/762(2022.01); G06V 20/40(2022.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 제스처 인식(gesture recognition) 프레임(frame), 알고리즘(algorithm), 완전 연결 계층(fully connected layer), 텐서(tensor), LSTM

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	CN 111860117 A (ANHUI BIGENG SOFTWARE CO., LTD.) 2020.10.30 청구항 1, 7-8 및 도면 1	10-12
A		1-9
A	CN 113158909 A (INSTITUTE OF AUTOMATION, CHINESE ACADEMY OF SCIENCES) 2021.07.23 청구항 1, 7-8 및 도면 1-7	1-12
A	CN 111046798 A (WISDOM VISION (HANGZHOU) TECHNOLOGY DEVELOPMENT CO., LTD.) 2020.04.21 청구항 1, 4 및 도면 1-2	1-12
A	CN 115131710 A (FUZHOU UNIVERSITY) 2022.09.30 청구항 1-6 및 도면 1	1-12
A	KR 10-2019-0080818 A (한국과학기술원) 2019.07.08 청구항 1-17 및 도면 1-8	1-12

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☑ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌
"D" 본 국제출원에서 출원인이 인용한 문헌
"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌
"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌
"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌
"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌
"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌
"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.
"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.
"&" 동일한 대응특허문헌에 속하는 문헌

국제조사의 실제 완료일	국제조사보고서 발송일
2024년02월05일(05.02.2024)	**2024년02월05일(05.02.2024)**

ISA/KR의 명칭 및 우편주소	심사관
대한민국 특허청 **(35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사)**	이강하
팩스 번호 **+82-42-481-8578**	전화번호 **+82-42-481-5003**

국 제 조 사 보 고 서
대응특허에 관한 정보

국제출원번호
PCT/KR2023/017770

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
CN 111860117 A	2020/10/30	없음	
CN 113158909 A	2021/07/23	CN 113158909 B	2023/06/27
CN 111046798 A	2020/04/21	없음	
CN 115131710 A	2022/09/30	없음	
KR 10-2019-0080818 A	2019/07/08	없음	