



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0103918
(43) 공개일자 2024년07월04일

- (51) 국제특허분류(Int. Cl.)
G06V 40/20 (2022.01) G06F 3/01 (2006.01)
G06V 10/44 (2022.01) G06V 10/75 (2022.01)
G06V 10/764 (2022.01) G06V 10/778 (2022.01)
G06V 10/94 (2022.01) G06V 20/40 (2022.01)
- (52) CPC특허분류
G06V 40/28 (2022.01)
G06F 3/011 (2022.02)
- (21) 출원번호 10-2023-0033723
(22) 출원일자 2023년03월15일
심사청구일자 2023년03월15일
- (30) 우선권주장
1020220185388 2022년12월27일 대한민국(KR)
- (71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
- (72) 발명자
이규호
울산광역시 울주군 언양읍 유니스트길 50
- 김승빈
울산광역시 울주군 언양읍 유니스트길 50
- (74) 대리인
특허법인태백

전체 청구항 수 : 총 16 항

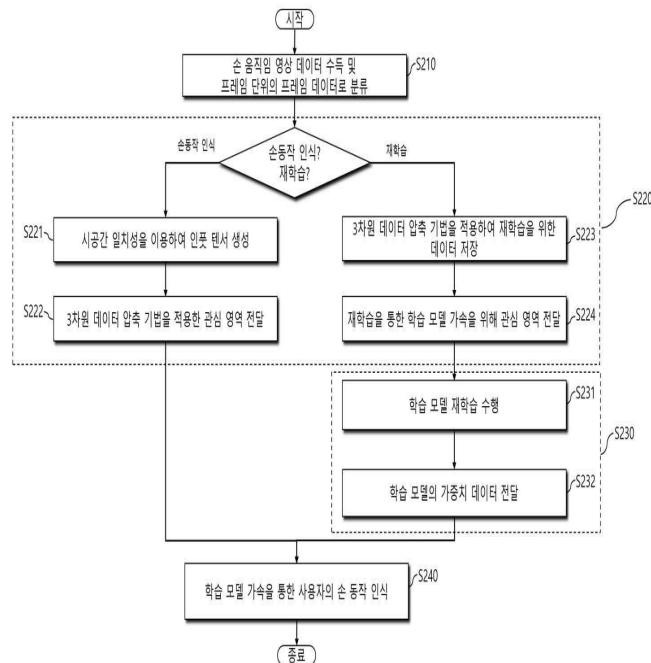
(54) 발명의 명칭 시공간 일치성을 이용한 손동작 인식 방법 및 이를 이용한 최소성 기반 3차원 합성곱 신경망 가속 장치

(57) 요약

본 발명은 시공간 일치성을 이용한 손동작 인식 방법 및 이를 이용한 최소성 기반 3차원 합성곱 신경망 가속 장치에 관한 것이다. 본 발명에 따르면, 시공간 일치성을 이용한 손동작 인식 방법에 있어서, 인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하는

(뒷면에 계속)

대표도 - 도2



단계; 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 프레임 데이터의 희소성을 증가시키는 단계; 복수의 영상 데이터를 이용하여 학습 모델 학습 및 재학습을 수행하는 단계; 및 획득된 인식 대상자의 손 움직임 영상 데이터를 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식하는 단계를 포함한다.

이와 같이 본 발명에 따르면, 입력된 영상 데이터의 시공간 일치성을 이용하여 3차원 합성곱 신경망의 희소성을 증가시키고, 희소성을 기반으로 연산을 수행함으로써 종래의 3차원 합성곱 신경망 모델이 요구하는 연산 환경, 막대한 연산 및 메모리 전송 요구를 줄일 수 있다.

(52) CPC특허분류

G06F 3/017 (2013.01)

G06V 10/454 (2023.08)

G06V 10/751 (2023.08)

G06V 10/764 (2023.08)

G06V 10/778 (2023.08)

G06V 10/955 (2022.01)

G06V 20/49 (2022.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711194743
과제번호	00212222
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	소형 인공지능을 위한 복합 다차로직 기반 초저전력, 초고효율 프로세서 구조 개발
기 여 율	80/100
과제수행기관명	울산과학기술원
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193983
과제번호	2020-0-01336-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(울산과학기술원)
기 여 율	20/100
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

시공간 일치성을 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치에 의해 수행되는 손동작 인식 방법에 있어서,

인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 상기 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하는 단계;

상기 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 상기 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시키는 단계;

복수의 영상 데이터를 이용하여 학습 모델 학습 및 재학습을 수행하는 단계; 및

상기 취득된 인식 대상자의 손 움직임 영상 데이터를 상기 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식하는 단계를 포함하는 손동작 인식 방법.

청구항 2

제1항에 있어서,

상기 프레임 데이터의 희소성을 증가시키는 단계는,

프레임 데이터 사이의 시공간 일치성을 이용하여 높은 희소성 비율을 가지도록 인풋 텐서를 구성하거나,

희소성을 이용한 관심 영역 압축기법을 적용하여 상기 학습 모델의 재학습을 위한 데이터를 저장하는 손동작 인식 방법.

청구항 3

제2항에 있어서,

상기 프레임 데이터의 희소성을 증가시키는 단계는,

상기 생성된 인풋 텐서의 첫번째 프레임 데이터에서 손의 초기 위치 정보를 획득하고,

두 번째 프레임 데이터부터 직전 프레임 데이터와 현재 프레임 데이터 사이의 차이로부터 상기 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시킴으로써 높은 희소성 비율을 가지도록 인풋 텐서를 구성하는 손동작 인식 방법.

청구항 4

제1항에 있어서,

상기 학습 모델 학습 및 재학습을 수행하는 단계는,

상기 입력된 영상 데이터의 인풋 텐서의 희소성을 이용하여 학습 모델을 경량화하는 손동작 인식 방법.

청구항 5

제1항에 있어서,

상기 학습 모델은,

하나의 3D 합성곱 레이어(Convolution layer), 복수의 3D 잔차 레이어(Residual layer), 복수의 3D 맥스-풀 레이어(Max-pool layer) 및 전연결 레이어(Fully-connected layer)로 구성되고,

상기 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시키는 손동작 인식 방법.

청구항 6

제5항에 있어서,

상기 학습 모델은,

글로벌 프루닝(Global pruning) 기법을 적용하여 상기 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어 및 전연결 레이어 전반에 걸쳐 가중치 희소성 비율을 증가시키는 손동작 인식 방법.

청구항 7

시공간 일치성을 기반으로 하는 손동작 인식 알고리즘을 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치에 있어서,

인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 상기 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하는 데이터 입력부;

상기 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 상기 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시키는 데이터 전처리부;

복수의 영상 데이터를 이용하여 학습 모델 학습 및 재학습을 수행하는 데이터 학습부; 및

수득된 인식 대상자의 손 움직임 영상 데이터를 상기 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식하는 손동작 인식부를 포함하는 3차원 합성곱 신경망 가속 장치.

청구항 8

제7항에 있어서,

상기 데이터 전처리부는,

프레임 데이터 사이의 시공간 일치성을 이용하여 높은 희소성 비율을 가지도록 인풋 텐서를 구성하거나,

희소성을 이용한 관심 영역 압축기법을 적용하여 상기 학습 모델의 재학습을 위한 데이터를 저장하는 3차원 합성곱 신경망 가속 장치.

청구항 9

제8항에 있어서,

상기 데이터 전처리부는,

상기 생성된 인풋 텐서의 첫번째 프레임 데이터에서 손의 초기 위치 정보를 획득하고,

두 번째 프레임 데이터부터 직전 프레임 데이터와 현재 프레임 데이터 사이의 차이로부터 상기 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시킴으로써 높은 희소성 비율을 가지도록 인풋 텐서를 구성하는 3차원 합성곱 신경망 가속 장치.

청구항 10

제7항에 있어서,

상기 데이터 학습부는,

상기 입력된 영상 데이터의 인풋 텐서의 희소성을 이용하여 학습 모델을 경량화하는 3차원 합성곱 신경망 가속 장치.

청구항 11

제7항에 있어서,

상기 학습 모델은,

하나의 3D 합성곱 레이어(Convolution layer), 복수의 3D 잔차 레이어(Residual layer), 복수의 3D 맥스-풀 레

이어(Max-pool layer) 및 전연결 레이어(Fully-connected layer)로 구성되고,
상기 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시키는 3차원 합성곱 신경망 가속 장치.

청구항 12

제11항에 있어서,

상기 학습 모델은,

글로벌 프루닝(Global pruning) 기법을 적용하여 상기 3D 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어 및 전연결 레이어 전반에 걸쳐 가중치에 희소성 비율을 증가시키는 3차원 합성곱 신경망 가속 장치.

청구항 13

제7항에 있어서,

상기 손동작 인식부는,

코너 탐색기(Corner detector);

글로벌 스케줄러(Global scheduler); 및

복수의 희소성 인식 연산기, 활성화 레지스터(Activation register) 및 가중치 메모리로 구성되는 희소성 인식 연산기 모듈을 포함하는 3차원 합성곱 신경망 가속 장치.

청구항 14

제13항에 있어서,

상기 코너 탐색기는,

상기 희소성 인식 연산기 모듈에서 연산한 출력 특징 맵(output feature map)의 관심 영역을 탐색하는 3차원 합성곱 신경망 가속 장치.

청구항 15

제13항에 있어서,

상기 글로벌 스케줄러는,

상기 관심 영역을 포함한 기 설정된 크기의 관심 영역 텐서를 결정하고,

상기 관심 영역 텐서에서 관심 영역을 제외한 구역에 있는 희소성과 관심 영역 내부의 희소성으로만 구성된 제로 타일에 대한 연산을 제외시키는 3차원 합성곱 신경망 가속 장치.

청구항 16

제13항에 있어서,

상기 가중치 메모리는,

상기 학습 모델의 가중치 데이터에서 비희소성 요소가 희소성을 이용한 3차원 데이터 압축기법을 적용하여 W-값, W-마스크 및 W-인덱스로 저장된 3차원 합성곱 신경망 가속 장치.

발명의 설명

기술 분야

[0001]

본 발명은 희소성 기반 3차원 시공간 일치성을 이용한 손동작 인식 방법 및 이를 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치에 관한 것으로서, 더욱 상세하게 설명하면 연산 자원이 제한되는 환경인 스마트 디바이스(예를 들어, 스마트폰, 태블릿 PC 등)를 제어하는 시공간 일치성을 이용한 손동작 인식 방법 및 이를 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치에 관한 것이다.

배경 기술

- [0002] 스마트 디바이스들을 제어하기 위한 인터페이스 방식(Human-computer interface)는 PC의 보급에 따른 키보드, 마우스에서부터 스마트폰 보급에 따른 터치스크린 기반의 인터페이스로 발전해왔다. 최근, 코로나-19(Covid-19)로 인한 팬데믹(Pandemic) 현상은 접촉 인터페이스 방식에 대한 위생 문제를 대두시켰고, 이에 따라 비접촉 인터페이스 방식이 요구되고 있다.
- [0003] 이와 관련하여, 손동작 인식 기술(Hand gesture recognition, HGR)의 경우 주변 소음의 영향을 받지 않으며, 물리적으로 공간을 적게 차지한다는 점에서 다수의 연구가 진행되고 있다.
- [0004] 종래에는 손동작 인식을 위해 2D 합성곱 레이어(2D Convolution layer)를 이용하여 공간적 정보를 추출하고, 장단기 메모리(Long short term memory, LSTM)를 이용하여 시간 정보를 추출하는 특징을 가진 알고리즘을 사용했다. 하지만, 종래의 알고리즘은 시공간의 정보를 개별적으로 추출한다는 특징 때문에 높은 정확도를 달성하지 못한다는 한계가 있다.
- [0005] 최근 손동작 인식 분야에서 주목을 받고 있는 3차원 합성곱 신경망 (3D Convolution neuron network) 알고리즘은 기존의 2D 합성곱 신경망과는 다르게 3D 합성곱 레이어(3D Convolution layer)로 구성되었다. 3D 합성곱 레이어는 연산 과정에서 커널이 공간상의 2D 방향뿐만 아니라, 시공간상의 3D 방향으로 이동하며 합성곱 연산을 진행한다. 2D 합성곱 레이어와는 다르게 시공간 정보를 동시에 추출하고, 이를 통해 3D 합성곱 신경망은 종래의 손동작 알고리즘과 대비하여 약 10%의 높은 정확도를 달성할 수 있다.
- [0006] 하지만, 3차원 합성곱 신경망 알고리즘은 막대한 연산과 메모리 전송을 요구함으로써 고성능 컴퓨터 혹은 서버와 같은 환경에서만 제한적으로 실시간 사용이 가능하고, 연산 자원이 제한되는 모바일 환경에서는 실시간 가속이 불가능하다는 단점이 있어 이를 극복하는 기술이 부족한 실정이다.
- [0007] 본 발명의 배경이 되는 기술은 대한민국 등록특허공보 제10-2179999호(2020.11.17 등록)에 개시되어 있다.

발명의 내용

해결하려는 과제

- [0008] 이와 같이 본 발명에 따르면, 연산 자원이 제한되는 환경인 스마트 디바이스(예를 들어, 스마트폰, 태블릿 PC 등)를 제어하는 시공간 일치성을 이용한 손동작 인식 방법 및 이를 이용한 회소성 기반 3차원 합성곱 신경망 가속 장치를 제공하기 위한 것이다.

과제의 해결 수단

- [0009] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따르면 시공간 일치성을 이용한 회소성 기반 3차원 합성곱 신경망 가속 장치에 의해 수행되는 손동작 인식 방법에 있어서, 인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 상기 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하는 단계; 상기 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 상기 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 상기 프레임 데이터의 회소성을 증가시키는 단계; 복수의 영상 데이터를 이용하여 학습 모델 학습 및 재학습을 수행하는 단계; 및 상기 취득된 인식 대상자의 손 움직임 영상 데이터를 상기 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식하는 단계를 포함한다.
- [0010] 상기 프레임 데이터의 회소성을 증가시키는 단계는, 프레임 데이터 사이의 시공간 일치성을 이용하여 높은 회소성 비율을 가지도록 인풋 텐서를 구성하거나, 회소성을 이용한 관심 영역 압축기법을 적용하여 상기 학습 모델의 재학습을 위한 데이터를 저장할 수 있다.
- [0011] 상기 프레임 데이터의 회소성을 증가시키는 단계는, 상기 생성된 인풋 텐서의 첫번째 프레임 데이터에서 손의 초기 위치 정보를 획득하고, 두 번째 프레임 데이터부터 직전 프레임 데이터와 현재 프레임 데이터 사이의 차이로부터 상기 시공간 일치성을 이용하여 상기 프레임 데이터의 회소성을 증가시킴으로써 높은 회소성 비율을 가지도록 인풋 텐서를 구성할 수 있다.
- [0012] 상기 학습 모델 학습 및 재학습을 수행하는 단계는, 상기 입력된 영상 데이터의 인풋 텐서의 회소성을 이용하여 학습 모델을 경량화할 수 있다.

[0013] 상기 학습 모델은, 3차원 신경망의 일종으로, 하나의 3D 합성곱 레이어(Convolution layer), 복수의 3D 잔차 레이어(Residual layer), 복수의 3D 맥스-풀 레이어(Max-pool layer) 및 전연결 레이어(Fully-connected layer)로 구성되고, 상기 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시킬 수 있다.

[0014] 상기 학습 모델은, 글로벌 프루닝(Global pruning) 기법을 적용하여 상기 3D 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어 및 전연결 레이어 전반에 걸쳐 가중치 희소성 비율을 증가시킬 수 있다.

[0015] 본 발명의 다른 실시 예에 따르면, 시공간 일치성을 기반으로 하는 손동작 인식 알고리즘을 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치에 있어서, 인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 상기 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하는 데이터 입력부; 상기 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 상기 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 상기 프레임 데이터의 희소성을 증가시키는 데이터 전처리부; 복수의 영상 데이터와 영상 데이터의 인풋 텐서를 각각 이용하여 학습 모델 학습 및 재학습을 수행하는 데이터 학습부; 및 상기 취득된 인식 대상자의 손 움직임 영상 데이터를 상기 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식하는 손동작 인식부를 포함한다.

발명의 효과

[0016] 이와 같이 본 발명에 따르면, 입력된 영상 데이터의 시공간 일치성을 이용하여 3차원 합성곱 신경망의 희소성을 증가시키고, 희소성을 기반으로 연산을 수행함으로써 종래의 3차원 합성곱 신경망 모델이 요구하는 연산 환경, 막대한 연산 및 메모리 전송 요구를 줄일 수 있다.

[0017] 또한, 연산 자원이 제한되는 환경(예를 들어, 모바일 또는 IoT 등)에서 3차원 합성곱 신경망 가속을 통해 실시간 손동작 인식을 수행할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 실시예에 따른 3차원 합성곱 신경망 가속 장치의 구성도이다.

도 2는 본 발명의 실시예에 따른 3차원 합성곱 신경망 가속 장치에 의해 수행되는 손동작 인식 방법의 순서도이다.

도 3은 본 발명의 실시예에 따른 입력된 영상 데이터의 희소성을 증가시키는 방법을 구체적으로 설명하는 예시도이다.

도 4는 본 발명의 실시예에 따른 희소성을 이용한 3차원 데이터 압축기법을 도시하는 도면이다.

도 5는 본 발명의 실시예에 따른 학습 모델을 도시하는 도면이다.

도 6은 본 발명의 실시예에 따른 손동작 인식부를 구체적으로 설명하는 예시도이다.

도 7은 본 발명의 실시예에 따른 손동작 인식부의 코너 탐색기를 도시하는 도면이다.

도 8은 본 발명의 실시예에 따른 손동작 인식부의 글로벌 스케줄러를 도시하는 도면이다.

도 9는 본 발명의 실시예에 따른 손동작 인식부의 희소성 인식 연산기 모듈을 도시하는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0019] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시 예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0020] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0021] 첨부한 도면을 참고로 하여 본 발명의 실시 예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다.

[0022] 후술하는 실시예에서는 3차원 합성곱 신경망 가속 장치(100)는 ToF(Time of flight) 카메라를 통해 촬영된 사용

자의 손 동작을 포함한 영상 즉, 연속적인 뎀스 이미지(depth image)를 이용하여 손 동작 인식을 수행하는 것으로 구체적인 예를 들어 설명한다. 그러나, 본 발명이 반드시 이로 제한되는 것은 아니며, 다른 종류의 카메라를 통해 촬영된 영상을 이용할 수 있다.

- [0023] 도 1은 본 발명의 실시예에 따른 3차원 합성곱 신경망 가속 장치의 구성도이다.
- [0024] 도 1을 참조하면, 본 발명의 실시예에 따른 3차원 합성곱 신경망 가속 장치(100)는 데이터 입력부(110), 데이터 전처리부(120), 데이터 학습부(130) 및 손동작 인식부(140)를 포함한다.
- [0025] 먼저, 데이터 입력부(110)는 인식 대상자로부터 손의 움직임 영상 데이터를 취득하고, 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류한다.
- [0026] 다음으로, 데이터 전처리부(120)는 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 연속되는 프레임 데이터간 시공간 일치성을 이용하여 프레임 데이터의 희소성(Sparsity)을 증가시킨다.
- [0027] 구체적으로, 데이터 전처리부(120)는 데이터 입력부(110)에서 전달된 연속적인 프레임 데이터를 3차원 합성곱 신경망의 입력(Input)으로 활용하기 위해 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성할 수 있다. 여기서, 인풋 텐서의 첫 번째 프레임 데이터는 원본 이미지를 그대로 활용함으로써 손의 초기 위치 정보를 제공한다. 또한, 두 번째 프레임 데이터부터 직전 프레임 데이터와 현재 프레임 데이터 사이의 각 픽셀 값 차이를 활용한다. 이때, 희소성은 픽셀 값이 0인 데이터이다.
- [0028] 다음으로, 데이터 학습부(130)는 복수의 영상 데이터를 이용하여 학습 모델 학습 및 재학습을 수행한다.
- [0029] 구체적으로, 데이터 학습부(120)는 입력된 영상 데이터를 활용하여 학습 모델을 학습하고, 대용량 손동작 데이터세트(Dataset)을 기반으로 학습된 학습 모델에 대해 사용자의 데이터에 대한 재학습을 통해 높은 손동작 인식률을 달성할 수 있다. 이때, 데이터 학습부(120)는 학습 모델을 학습하거나 재학습을 수행하기 위해 수많은 복잡한 연산 및 메모리 전송이 필요하므로 많은 연산 자원이 제공 가능한 고성능 컴퓨터 혹은 클라우드 서버에서 수행될 수 있다.
- [0030] 여기서, 학습 모델은 3차원 합성곱 신경망의 일종으로, 하나의 3D 합성곱 레이어(Convolution layer)와 복수의 3D 잔차 레이어(Residual layer), 복수의 3D 맥스-풀 레이어(Max-pool layer) 및 하나의 전연결 레이어(Fully-connected layer)로 구성되어 있고, 전연결 레이어에서 최종적으로 손동작 분류할 수 있다.
- [0031] 또한, 학습 모델은 글로벌 프루닝(Global pruning) 기법을 적용하여 3D 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어 및 전연결 레이어 전반에 걸쳐 가중치 희소성 비율을 증가시킬 수 있다.
- [0032] 다음으로, 손동작 인식부(140)는 취득된 인식 대상자의 손 움직임 영상 데이터를 이용하여 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식한다.
- [0033] 구체적으로, 손동작 인식부(140)는 취득된 인식 대상자의 손 움직임 영상 데이터를 이용하여 학습 및 재학습이 완료된 학습 모델에 적용하여 인식 대상자의 손동작을 인식할 수 있다.
- [0034] 또한, 손동작 인식부(140)는 코너 탐색기(Corner detector)(141), 글로벌 스케줄러(Global scheduler)(142) 및 희소성 인식 연산기 모듈(143)을 포함할 수 있다. 이때, 희소성 인식 연산기 모듈(143)은 복수의 희소성 인식 연산기, 활성화 레지스터(Activation register) 및 가중치 메모리로 구성될 수 있다.
- [0035] 이하에서는 도 2 내지 도 9를 통하여 본 발명의 실시예에 따른 시공간 일치성을 이용한 희소성 기반 3차원 합성곱 신경망 가속 장치(100)에 의해 수행되는 손동작 인식 방법에 대하여 상세하게 설명한다.
- [0036] 도 2는 본 발명의 실시예에 따른 3차원 합성곱 신경망 가속 장치에 의해 수행되는 손동작 인식 방법의 순서도이다.
- [0037] 도 2에 도시된 바와 같이, 데이터 입력부(110)는 인식 대상자로부터 손 움직임 영상 데이터를 취득하고, 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류한다(S210).
- [0038] 구체적으로, 데이터 입력부(110)는 손의 움직임 영상 데이터를 취득하고, 취득된 손 움직임 영상 데이터를 프레임 단위의 프레임 데이터로 분류하여 데이터 전처리부(120)으로 전달할 수 있다.
- [0039] 다음으로, 데이터 전처리부(120)는 분류된 연속적인 프레임 데이터를 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 생성 및 저장하고, 연속되는 프레임 데이터 사이의 시공간 일치성을 이용하여 프레

임 데이터의 희소성을 증가시킨다(S220).

- [0040] 구체적으로, 데이터 전처리부(120)는 프레임 데이터 사이의 시공간 일치성을 이용하여 높은 희소성 비율을 가지도록 적어도 하나의 인풋 텐서를 구성하거나(S221), 희소성을 이용한 3차원 데이터 압축기법을 적용하여 학습 모델의 재학습을 위한 데이터를 저장한다(S223). 여기서, 데이터 전처리부(120)는 학습 모델의 재학습을 위한 데이터를 서버(미도시)와 같은 별도의 메모리에 저장하고, 학습 모델의 재학습을 수행할 시 저장된 데이터를 불러와 재학습을 수행할 수 있다.
- [0041] 또한, 데이터 전처리부(120)는 데이터 입력부(110)에서 전달받은 연속적인 프레임 데이터를 3차원 합성곱 신경망의 인풋으로 활용하기 위해 도 3과 같이 기 설정된 개수의 프레임마다 축적함으로써 하나 이상의 인풋 텐서를 구성할 수 있다.
- [0042] 도 3은 본 발명의 실시예에 따른 입력된 영상 데이터의 희소성을 증가시키는 방법을 구체적으로 설명하는 예시 도이다.
- [0043] 도 3을 참조하면, 데이터 전처리부(120)는 생성된 인풋 텐서의 파란색으로 표시된 첫 번째 프레임 데이터(참조 프레임)에서 손의 초기 위치 정보를 획득하고, 붉은 색으로 표시된 두 번째 프레임 데이터부터 직전 프레임 데이터와 현재 프레임 데이터 사이의 차이(차 이미지 프레임)로부터 시공간 일치성을 이용하여 프레임 데이터의 희소성을 증가시킴으로써 높은 희소성 비율을 가지도록 인풋 텐서를 구성할 수 있다.
- [0044] 즉, 입력된 영상 데이터의 시공간 일치성으로 인하여 연속적인 프레임 간의 변화하는 복수의 픽셀 값만이 손동작 정보를 가지고 있고, 변화하지 않는 복수의 픽셀 값은 손동작 인식 방법에서 불필요한 데이터이다.
- [0045] 또한, 데이터 전처리부(120)는 시공간 일치성을 이용하여 프레임 데이터의 희소성을 증가시키기 위해 비희소성 픽셀의 패턴을 관심 영역(Region-of-interest, ROI) 내부로 제한하여 생성시킬 수 있다. 이때, 관심 영역은 희소성이 존재하지 않는 최소 면적의 사각형 형태의 영역이며, 희소성은 픽셀 값이 0인 데이터이고, 입력된 영상 데이터의 관심 영역 패턴은 손 움직임이 존재하는 구역에서만 발생한다.
- [0046] 즉, 희소성에 대한 합성곱 연산은 0으로 정해지기 때문에 종래의 손동작 인식 과정에서 희소성에 대한 데이터를 저장하거나 연산하는 과정은 불필요한 소비전력, 면적 및 메모리 점유를 발생시킨다. 따라서, 데이터 전처리부(120)는 관심 영역 내부의 데이터만을 저장하는 관심 영역 압축기법을 통해 방대한 양의 비디오 데이터를 적은 메모리를 사용하여 저장할 수 있다.
- [0047] 또한, 데이터 전처리부(120)는 관심 영역 압축기법을 적용하여 저장된 관심 영역을 학습 모델의 인식 대상자의 손동작 인식을 위해 손동작 인식부(140)로 전달하거나(S222), 재학습을 통한 학습 모델의 가속을 위해 데이터 학습부(130)로 전달할 수 있다(S224). 이를 통해 3차원 합성곱 신경망의 높은 메모리 전송량을 감소시킬 수 있다.
- [0048] 도 4는 본 발명의 실시예에 따른 희소성을 이용한 3차원 데이터 압축기법을 도시하는 도면이다.
- [0049] 도 4를 참조하면, 희소성을 이용한 3차원 데이터 압축기법은 가중치의 비희소성 요소만을 W-값(W-value), W-인덱스(W-index) 및 W-마스크(W-mask)를 활용하여 메모리에 저장하는 방법이다.
- [0050] 종래의 희소성 압축기법(예를 들어, Zero run length, Coordinate list 및 Compressed sparse row)은 1D 벡터(Vector) 또는 2D 매트릭스(Matrix) 상의 데이터만을 압축할 수 있어, 합성곱 연산을 위해 별도의 데이터 복원 과정을 요구하였다. 그러나 본 발명의 실시예에 따른, 희소성을 이용한 3차원 데이터 압축기법은 3차원 텐서 형태의 가중치를 압축할 수 있어, 압축된 데이터의 복원 없이 3차원 합성곱 연산이 가능하다.
- [0051] 부연하자면, 커널의 사이즈가 $3 \times 3 \times 3$ 인 경우, W-값은 픽셀 값을 저장되고, W-인덱스는 기 설정된 범위(0 내지 26)에서 하나의 값을 가지고, 너비(Width)-높이(Hight)-깊이(Depth)에 따라 순차적으로 저장될 수 있다.
- [0052] 또한, W-마스크는 각 커널을 구분하는 값으로 커널의 첫 번째 비희소성 요소를 의미하는 '1'이나 첫 번째 비희소성 요소를 제외한 다른 요소를 의미하는 '0'이 저장될 수 있다. 이로 인해, 별도의 데이터를 저장하거나 할당된 주소를 확인하는 동작없이 데이터를 읽어 알맞은 activation과 합성곱 연산을 수행할 수 있다.
- [0053] 예를 들어, 학습 모델의 가중치 데이터에 3차원 데이터 압축기법을 적용하여 $3 \times 3 \times 3$ 커널에 저장하는 경우, W-값에는 가중치 데이터의 값 자체가 저장되며, W-인덱스는 가중치 데이터의 커널 내부에서 순서가 기 설정된 범위(0 내지 26)의 값으로 저장될 수 있다.

- [0054] 다음으로, 데이터 학습부(130)는 복수의 영상 데이터를 이용하여 학습 모델에 대한 학습 및 재학습을 수행한다(S230).
- [0055] 도 5는 본 발명의 실시예에 따른 학습 모델을 도시하는 도면이다.
- [0056] 도 5에 도시된 바와 같이, 데이터 학습부(130)에서 학습된 학습 모델은 하나의 3D 합성곱 레이어(Convolution layer), N 개의 3D 잔차 레이어(Residual layer), M 개의 3D 맥스-풀 레이어(Max-pool layer) 및 전연결 레이어(Fully-connected layer)로 구성되고, 영상 데이터의 복수의 시공간 특징점을 추출할 수 있다. 여기서, N 및 M은 0보다 큰 정수이고, 전연결 레이어는 복수의 영상 데이터의 복수의 시공간 특징점을 활용하여 최종적으로 손동작을 분류한다.
- [0057] 또한, 데이터 학습부(130)에서 학습된 학습 모델은 시공간 일치성을 이용하여 프레임 데이터의 희소성을 증가시킬 수 있다. 이때, 학습 모델은 글로벌 프루닝(Global pruning) 기법을 적용하여 3D 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어 및 전연결 레이어 전반에 걸쳐 레이어의 가중치 희소성 비율을 증가시킬 수 있다.
- [0058] 또한, 데이터 학습부(130)는 많은 연산 자원 제공이 가능한 고성능 컴퓨터 또는 클라우드 서버를 통해 학습 모델에 대한 학습 및 재학습을 수행하며(S231), 학습이 완료된 학습 모델을 가속하기 위해 학습 모델의 가중치 데이터를 손동작 인식부(140)로 전달한다(S232). 그러나, 본 발명이 이에 제한되는 것은 아니다. 여기서, 가중치 데이터는 3차원 데이터 압축기법을 통해 압축될 수 있고, 희소성 인식 연산기의 3차원 합성곱 연산 수행을 위해 희소성 인식 연산기 모듈(143)에 저장될 수 있다.
- [0059] 이와 같이, 데이터 학습부(130)의 학습 모델 재학습 과정을 통해 인식 대상자에 따른 손동작 인식을 감소를 최소화할 수 있다.
- [0060] 다음으로, 손동작 인식부(140)는 인식 대상자의 손 움직임 영상 데이터를 구축된 학습 모델에 적용하여 인식 대상자의 손동작을 인식한다(S240).
- [0061] 구체적으로, 손동작 인식부(140)는 취득된 인식 대상자의 손 움직임 영상 데이터를 이용하여 학습 및 재학습이 완료된 학습 모델에 적용하여 인식 대상자의 손동작을 인식할 수 있다.
- [0062] 또한, 손동작 인식부(140)는 취득된 인식 대상자의 손 움직임 영상 데이터의 프레임 데이터, activation 및 학습 모델의 가중치 데이터의 희소성을 이용하여 소비전력, 면적, 메모리를 적게 함으로써 학습 모델을 실시간으로 가속할 수 있다.
- [0063] 도 6은 본 발명의 실시예에 따른 손동작 인식부를 구체적으로 설명하는 예시도이다.
- [0064] 도 6을 참조하면, 손동작 인식부(140)는 코너 탐색기(Corner detector)(141), 글로벌 스케줄러(Global scheduler)(142) 및 희소성 인식 연산기 모듈(143)을 포함하고, 코너 탐색기(141), 글로벌 스케줄러(142) 및 희소성 인식 연산기 모듈(143)은 내부 데이터 네트워크를 통해 연결되며, 내부 데이터 네트워크를 컨트롤러를 통해 제어될 수 있다. 이때, 희소성 인식 연산기 모듈(143)은 복수의 희소성 인식 연산기, 활성화 레지스터(Activation register) 및 가중치 메모리로 구성될 수 있다.
- [0065] 또한, 코너 탐색기(141), 글로벌 스케줄러(142) 및 희소성 인식 연산기 모듈(143)은 각각 연속적인 프레임 데이터 사이의 시공간 일치성을 통해 증가된 손 움직임 영상 데이터의 프레임 데이터 각각의 희소성과 학습 모델을 프루닝 기법을 적용하여 발생한 가중치 데이터의 희소성을 이용하여 연산 지연속도를 감소시킬 수 있다.
- [0066] 기 구축된 학습 모델을 이용하여 합성곱 연산을 수행하기 위해, activation에 가중치 커널을 적용하여 출력 특성 맵(Output feature map)을 연산하는 과정에서 activation과 가중치는 원소 간(Elementary-wise) 곱셈연산 및 그 결과들의 덧셈연산으로 구성된다. 여기서, activation과 가중치가 0인 경우, 곱셈연산 및 덧셈연산은 수행되지 않고, 출력 특성 맵을 연산할 수 있으므로 손동작 인식부(140)가 손동작 인식을 위해 수행하는 연산의 수, 연산 지연속도 및 소비 전력이 감소한다.
- [0067] 또한, 코너 탐색기(141)는 희소성 인식 연산기 모듈(143)에서 연산한 출력 특성 맵(Output feature map)의 관심 영역을 탐색한다.
- [0068] 도 7은 본 발명의 실시예에 따른 손동작 인식부의 코너 탐색기를 도시하는 도면이다.
- [0069] 도 7을 참조하면, 코너 탐색기(141)는 3D 합성곱 레이어, 복수의 3D 잔차 레이어, 복수의 3D 맥스-풀 레이어에

서 연산된 출력 특성 맵(Output feature map)을 다음 레이어의 activation으로 활용할 때 변화하는 activation의 관심 영역 코너의 좌표값으로 합성곱 연산 이후 인식할 수 있다.

- [0070] 또한, 코너 탐색기(141)는 회소성 인식 연산기 모듈(143)이 연산한 출력 특성 맵의 관심 영역을 탐색하고, 다음 레이어의 activation의 관심 영역만을 전달할 수 있다. 이때, 관심 영역은 2차원 행렬 형태로, 회소성이 존재하지 않는 최소 면적의 사각 형태의 영역이다.
- [0071] 다음으로, 글로벌 스케줄러(142)는 관심 영역 텐서를 결정하고, 관심 영역 텐서 내부의 회소성으로만 구성된 제로 타일(Zero tile)에 대한 연산을 제외시킬 수 있다.
- [0072] 도 8은 본 발명의 실시예에 따른 손동작 인식부의 글로벌 스케줄러를 도시하는 도면이다.
- [0073] 도 8을 참조하면, 글로벌 스케줄러(142)는 프레임 데이터 또는 채널별로 크기 및 코너 좌표가 상이한 관심 영역으로 구성된 activation의 관심 영역 텐서를 결정한다. 여기서, 관심 영역 텐서는 관심 영역을 포함하는 기 설정된 크기의 최소한의 연산이 수행 가능한 연산 범위이다.
- [0074] 또한, 글로벌 스케줄러(142)는 관심 영역 텐서를 기 지정된 크기의 타일로 분할하여 3차원 합성곱 연산을 수행할 수 있다. 이때, 타일의 모든 데이터가 '0'으로 설정된 제로 타일(Zero tile)인 경우, 글로벌 스케줄러(142)는 제로 타일에 대한 연산을 생략할 수 있다.
- [0075] 더 나아가, 글로벌 스케줄러(142)는 코너 디텍터로부터 전달된 관심 영역의 코너 정보를 이용하여 제로 타일을 선제적으로 선별함으로써 계층적으로 회소성 인식할 수 있다.
- [0076] 다음으로, 회소성 인식 연산기 모듈(143)은 회소성 인식 연산기 모듈(143)은 복수의 회소성 인식 연산기, 활성화 레지스터(Activation register) 및 가중치 메모리를 포함할 수 있다. 여기서, 가중치 메모리는 학습 모델의 가중치 데이터에서 비회소성 요소가 회소성을 이용한 3차원 데이터 압축기법을 적용하여 W-값(W-value), W-마스크(W-mask) 및 W-인덱스(W-index)로 저장하고, 회소성 인식 연산기는 복수의 W-값, W-마스크 및 W-인덱스에 저장된 각각의 값을 기초로 연산한다.
- [0077] 또한, 회소성 인식 연산기는 가중치 메모리에 저장된 가중치 데이터를 복원하지 않고 합성곱에 사용함으로써 가중치 데이터의 회소성을 인식하는 연산을 수행할 수 있다.
- [0078] 도 9는 본 발명의 실시예에 따른 손동작 인식부의 회소성 인식 연산기 모듈을 도시하는 도면이다.
- [0079] 도 9를 참조하면, 회소성 인식 연산기 모듈(143)은 W-인덱스의 값을 기준으로 활성화 레지스터(Activation register)로부터 대응하는 데이터 값을 선택하고, W-값과 대응하는 데이터 값을 이용하여 곱셈연산을 할 수 있다.
- [0080] 회소성 인식 연산기 모듈(143)은 W-인덱스 값과 대응하는 데이터 값의 곱셈연산 값은 이에 대응하는 결과 레지스터의 값과 덧셈연산을 수행하고, 덧셈연산의 값은 결과 레지스트에 재저장할 수 있다.
- [0081] 더 나아가, 회소성 인식 연산기 모듈(143)은 W-마스크의 값을 기준으로 결과 레지스터에서 불러오는 위치를 바꿈으로써 서로 대응하는 요소 간의 덧셈 연산을 수행한다.
- [0082] 이와 같이 본 발명의 실시예에 따르면, 입력된 영상 데이터의 시공간 일치성을 이용하여 3차원 합성곱 신경망의 회소성을 증가시키고, 회소성을 기반으로 연산을 수행함으로써 종래의 3차원 합성곱 신경망 모델이 요구하는 연산 환경, 막대한 연산 및 메모리 전송 요구를 줄일 수 있다.
- [0083] 또한, 연산 자원이 제한되는 환경(예를 들어, 모바일 또는 IoT 등)에서 3차원 합성곱 신경망 가속을 통해 실시간 손동작 인식을 수행할 수 있다.
- [0084] 또한, 본 발명의 실시예에 따르면, 학습 모델은 종래의 신경망(C3D, R3D)와 비교하여 약 94% 적은 가중치 데이터를 사용함으로써 경량화할 수 있다.
- [0085] 또한, 본 발명의 실시예에 따르면, 학습 모델은 손동작 인식을 위한 종래의 3D 합성곱 신경망 대비 적은 3D 합성곱 레이어로 구성되어 높은 정확도를 달성하였다.
- [0086] 또한, 본 발명의 실시예에 따르면, 학습 모델에 프루닝 기법을 적용함으로써 정확도의 감소 없이 높은 가중치 데이터의 회소성 비율을 달성하였다.
- [0087] 또한, 본 발명의 실시예에 따르면, 데이터 학습부(130)는 학습 모델의 학습 및 가속을 수행하고, 손동작 인식부

(140)는 실시간으로 손동작 인식을 수행함으로써, 연산 자원이 제한되는 환경에서도 실시간 손동작 인식 기능을 효율적으로 제공할 수 있다.

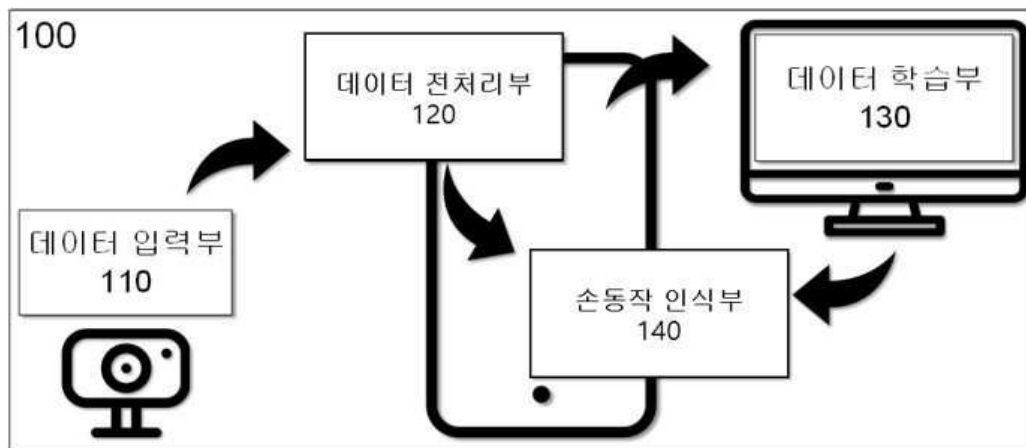
[0088] 본 발명은 도면에 도시된 실시예를 참고로 하여 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술이 속하는 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 아래의 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

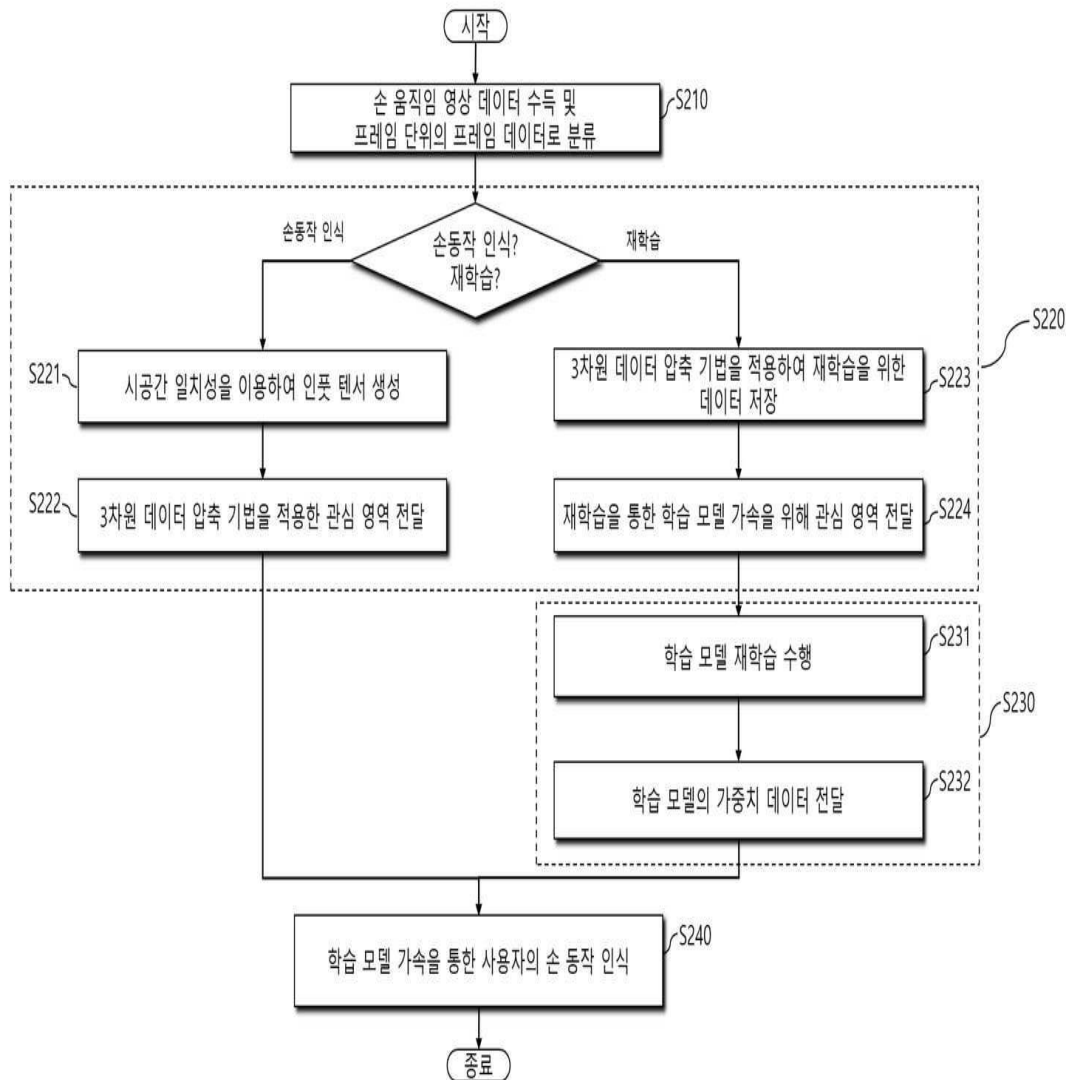
[0089] 100: 3차원 합성곱 신경망 가속 장치
110: 데이터 입력부
120: 데이터 전처리부
130: 데이터 학습부
140: 손동작 인식부

도면

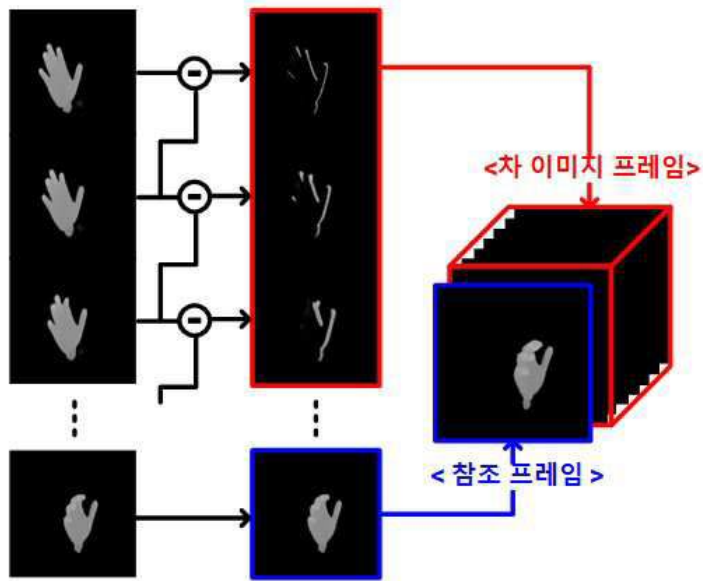
도면1



도면2



도면3



도면4

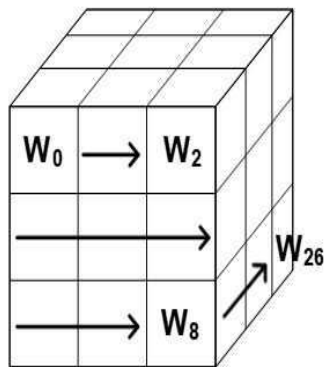
원본 가중치 데이터

W_0	0	W_2	0	0	0	0	0	W_8
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	W_{26}

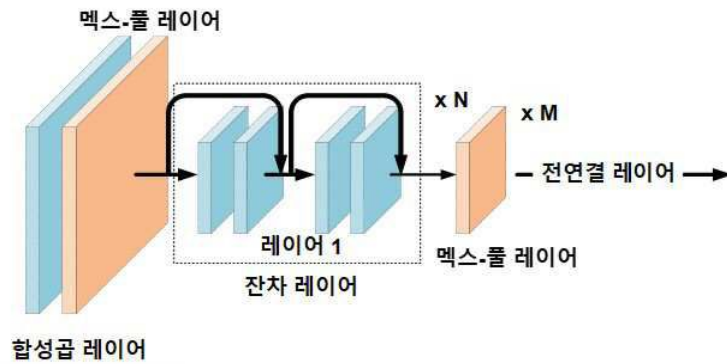
희소성을 이용한
3차원 데이터압축 기법

W_0	W_2	W_8	W_{26}	W-Value
1	0	0	0	W-Mask 2비트
0	2	8	26	W-Index 5비트

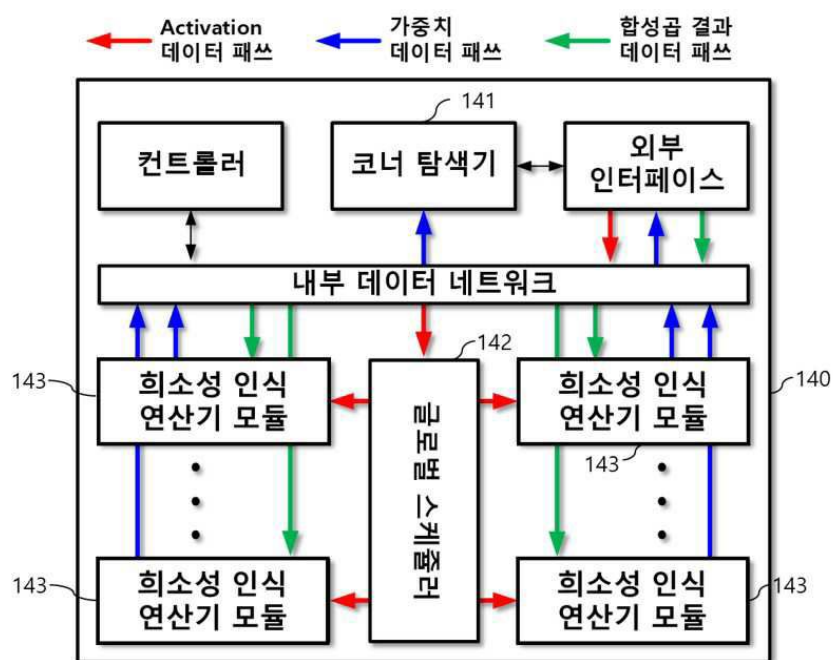
3 x 3 x 3
가중치 커널



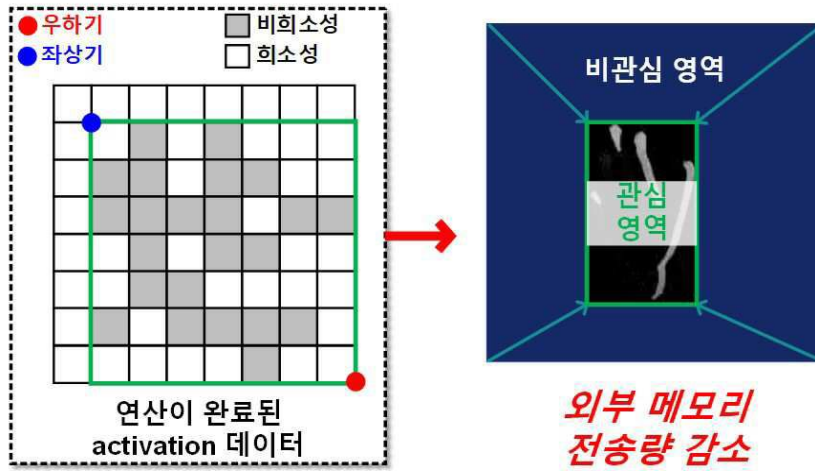
도면5



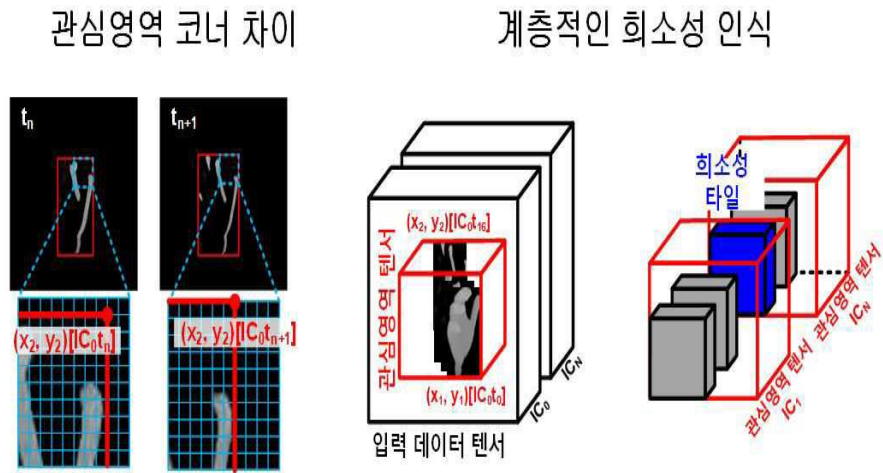
도면6



도면7



도면8



도면9

