

비지도 학습을 이용한 생체 모방 동작 인지 기반의 동작 순서 인식

Bio-mimetic Recognition of Action Sequence using Unsupervised Learning

김진옥^{1*}

Kim Jin Ok

요약

대상의 동작을 잘 예측하는 것은 사회적 상호작용과 의사결정 컨텍스트 이해의 핵심이다. 본 연구는 동작 인식 과정에서 인간 뇌 시각인지 과정을 모방한 방법으로 관절 동작의 동작 순서 패턴을 학습하는 컴퓨팅 모델을 제안하였다. 제안 방법의 핵심은 뇌에서 동작 인지 자극을 처리하는 신경생리학적 IT, MT, STS의 피질 기능과 특정 시각 신경 회로 네트워크 기능을 모방하여 비지도 방법으로 동작 순서를 학습한 후 동작을 예측, 인식하는 것이다. 실험을 통해 제안 모델이 어떻게 연속적으로 입력되는 비디오에서 의미있는 동작 스냅샷 뿐 아니라 중요한 동작 패턴을 자동으로 선택하는 지를 제시하였다. 이 핵심 움직임은 학습 네트워크가 정적 시점에서 정지 영상에 합쳐진 동작을 인식하는지를 증명하는데 이용하는 관절 자세이다. 또한 STS 피질 영역에서 어떻게 정지와 움직임 입력을 통합하는지를 모방하여 학습하고, 학습한 피드백 연결이 차후 동작의 입력 순서를 어떻게 예측하는지를 제시하였다. 네트워크 시뮬레이션을 통해 동작 인식에 대한 제안 모델의 우수성을 입증하였다.

☞ 주제어 : 동작 순서 인식, 동작 분류, 생체모방 동작 인지, 비지도 학습

ABSTRACT

Making good predictions about the outcome of one's actions would seem to be essential in the context of social interaction and decision-making. This paper proposes a computational model for learning articulated motion patterns for action recognition, which mimics biological-inspired visual perception processing of human brain. Developed model of cortical architecture for the unsupervised learning of motion sequence, builds upon neurophysiological knowledge about the cortical sites such as IT, MT, STS and specific neuronal representation which contribute to articulated motion perception. Experiments show how the model automatically selects significant motion patterns as well as meaningful static snapshot categories from continuous video input. Such key poses correspond to articulated postures which are utilized in probing the trained network to impose implied motion perception from static views. We also present how sequence selective representations are learned in STS by fusing snapshot and motion input and how learned feedback connections enable making predictions about future input sequence. Network simulations demonstrate the computational capacity of the proposed model for motion recognition.

☞ keyword : Action sequence detection, Categorical action classification, Bio-inspired Action perception, Unsupervised learning

1. 서론

인간의 뇌에서 일어나는 시각적 자극 대비 반응 과정에 대해 실험심리학에서는 자극이 주어졌을 때 인간이 이를 어떻게 인지하느냐로 설명하고 신경생리학에서는 주어진 자극을 어떻게 뇌에서 처리하느냐를 다뤄 있다. 점차 이를 세분화하여 자극요인 처리 분야는 신경생물학이, 자극요인을 인지하는 데는 정신물리학 분야가 담당하면서 뇌에서 동작을 인지하고 해석하는 과정은 신경물

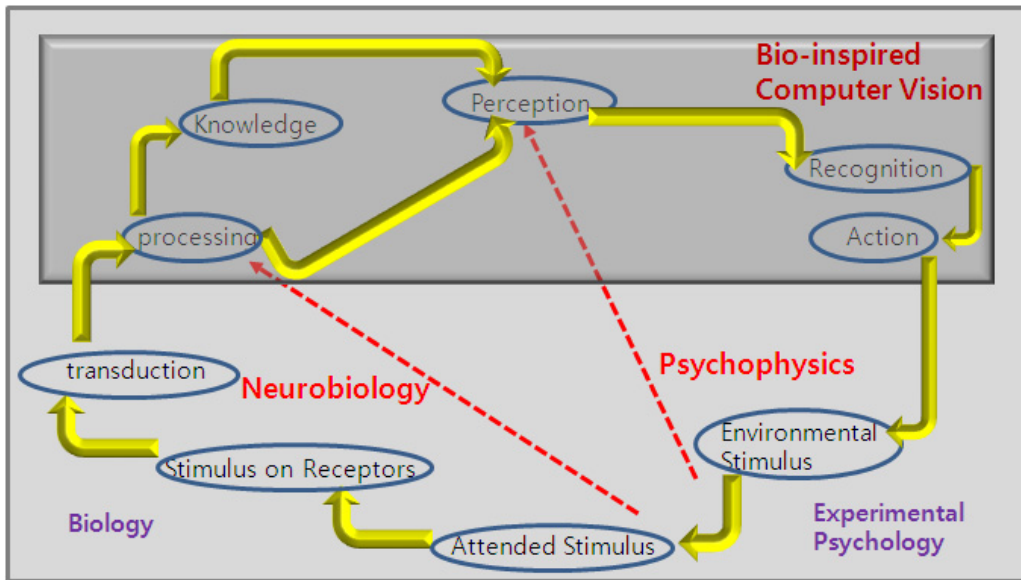
리학적으로 모델로 구체화되고 있다[1].

최근 컴퓨터비전 영역에서는 그림 1과 같이, 정신물리학 및 신경생리학 연구에서 도출된, 인간의 뇌가 자극을 입력받아 반응을 출력하기까지의 과정을 컴퓨터가 데이터를 수용하여 정보로 처리하는 과정에 모방하여 컴퓨터 계산 모델로 설명하는 생체 모방 인지 연구가 새롭게 부상하고 있다[1]. 그림 1에서 Transduction은 자극 에너지를 다른 형태로 변환하는 과정, Perception은 의식적인 센싱 경험, Recognition은 대상을 카테고리에 위치시키는 식별능력, Action은 Perception 과정에 대응해 신체의 움직임과 같은 동작을 행동으로 나타내는 것이며 Knowledge는 다른 대상을 식별하는 개인 능력, Stimulus on the receptors는 눈 망막의 감수영역에 대상의 이미지가 감지되는 것을 의미한다. Environmental Stimulus는 주

¹ Faculty of Mobile Contents, Daegu Haany University, Geongbuk, 712-715, Korea

* Corresponding author (bit@dhua.ac.kr)

[Received 04 March 2014, Reviewed 19 March 2014, Accepted 05 June 2014]



(그림 1) 시각인식 관련 학문과 컴퓨터비전과의 관계

(Figure 1) Relation between visual Information Processing related studies and computer vision

위환경에서 전해오는 자극이고 그중 주의를 기울이는 자극이 **Attended Stimulus**가 된다. 이는 인간의 뇌를 대상으로 하는 생물학과 생리학, 심리학 연구 결과를 컴퓨터 인공지능 분야에서 응용, 신경생물학적 모델을 계산 가능한 상태로 만들어 컴퓨터 비전 응용 연구 및 학제간 융합 연구로 확장한 과정을 설명한 것이다.

인간의 뇌 시각정보 처리는 배측 동작 에너지 경로(dorsal pathway)와 복측 형태/모양 경로(ventral pathway) 두 가지 스트림으로 나뉘어 이루어지고 있다[2]. 신경생물학적 시각 자극 인지과정 중 Figure 1에 제시된 것처럼 감수영역 자극 입력에서 처리가 시작되어 인지 단계로 가는 것을 상향식 처리, 주변에서 가져온 지식을 이용하여 인지를 시작하는 처리는 하향식 처리 방향이라 한다.

컴퓨터 비전 연구자들은 이와 같은 신경생물학의 뇌 정보처리 과정이 의견이 컴퓨팅 모델의 객체 인식 과정과 유사하다고 이해하여 생체 시각정보처리를 모방한 방식으로 동작 인식을 수행하는 연구를 지속하고 있다 [3][4][5]. 이 중, Gieso와 Poggio[6]는 생물학 기반의 계산 매커니즘 기반의 점광 보행자 방식을 이용하여 다단계 순방향 네트워크 구조를 제안했다. 그들이 제안한 계산 프레임워크는 배측 경로와 복측 경로로 구분된 생체 모방 시각경로를 이용하여 대상의 형태와 동작을 처리하는 과정을 설명하였다. 계산 프레임워크 다단계 구조의 최

상위 계층에서는 순서를 결정하는 동작 프로토타입을 구축하여 스냅샷 순서에서 나타나는 광류 패턴 특징에 대한 프로토타입 형 동작 패턴을 학습하였으며 동작 패턴의 평균 값을 계산하여 출력을 결정하여 제시함으로써 인간 신경계와 유사한 시각 정보처리과정을 컴퓨터 모델로 보여주었다[7].

또한 생물학에서 시도한 방식과 유사한 처리과정을 적용하여 대상의 형태와 동작 처리과정을 조합한 방법 [8][9]과 생물학의 동작 인식 과정을 반영한 동작 인식 방법도 제안되었다[10].

하지만 기존 생물학의 인식처리 방식을 모방한 컴퓨터 비전 연구에서는 뇌의 시각 담당 피질 중 MST(Medial superior temporal area)에서 동작을 어떻게 표현하는지, IT(Inferior temporal cortex)에서 어떻게 표현 방식 요소를 형성하는지, STS(Superior temporal sulcus)에서 어떻게 순서 선택적 패턴들이 상호작용하는지, 어떤 특징들이 학습에 사용되는지를 명확하게 모델링한 내용은 제시하지 못하고 있다. 또한 복측과 배측 경로간의 영역 간 정보의 전달과 주변에서 가져온 지식으로 인식을 시작하는 정보 전달의 하향식 방식이 끼치는 영향에 대한 구체적인 연구 내용은 아직 없는 상태이다.

한편, 신경생물학적 동작 인식은 제시된 자극의 시간 순서에 종속되어 있다. 시간진행 순서가 뒤섞인 영화를

보면 내용 이해도가 떨어지는 것도 그 때문이다. 그렇기 때문에 동작 인식 모델 역시 동작의 순서를 선택할 수 있는 신경망 매커니즘을 필수적으로 반영해야 한다. 하지만 동작 순서를 이해하는 생체 모방 기반의 동작 인식 연구는 많지 않은 상태이다.

본 연구에서는 컴퓨터비전 측면에서 신경생물학적 정보처리 방식을 모방한 비지도 학습 기반의 동작 인식 모델을 제안하고 자세 스냅샷의 동작 순서를 분류하여 동작을 인식하는데 적용하고자 한다. 이를 위해 대상의 형태와 동작 에너지 경로에 대한 프로토타입을 설정하여 변형된 헤비안(Hebbian) 학습법[11]으로 비지도 학습을 한 다음 입력되는 비디오 스트림에서 스냅샷 자세 즉 정지자세 프로토타입을 자동으로 선택한다. 선택한 스냅샷은 학습 네트워크가 정적 시점에서 정지 영상에 함축된 동작을 인식하는지를 증명하는데 이용하는 관절 자세 템플릿으로 이용한다.

이를 위해 대뇌에서 시각 인지과정을 담당하는 V1, IT, MT, MST, STS 피질 영역의 역할과 처리과정을 제안 모델의 처리 과정에 반영하였으며 특히 다음과 같은 생물학적 시각 처리 특징을 고려하여 동작의 순서를 인식하도록 한다.

- 뇌의 STS 피질영역에서 관절체 동작을 순서대로 설명하는 것은 동작과 형태 프로토타입을 같이 입력하고 활성화시켜야 가능하다.
- 역방향 연결을 학습시켜 STS 뉴런이 동작의 형태를 선택하는 IT 영역과 동작에 민감한 MST 영역에서 기대한 입력을 예측하도록 한다.
- 연속 동작이 없는 정적 관절 자세를 처리할 때 STS에서의 동작 표현은 해당 스냅샷 프로토타입 활성화를 통해 가능하다[12].
- STS는 함축된 동작 자극에 대해 배측 경로에서 차례차례 신경 반응을 만들어내면서 복측 경로에 피드백을 보낸다[13].

2. 관련 연구

생물학적 시각에서 동작 검출과 분석은 신경계 작용의 연속으로 시각장의 한정된 영역 내에서 부분 동작 신호를 검출하여 전체적인 대상 동작의 방향과 순서를 추론하는 과정으로 이루어져 있다.

동작 인식에 생물학 기능을 모방한 계산모델을 적용한 것은 Van Santen[14]과 Adelson[15]이 여러 공간과 크

기로 다양한 방향의 부분 동작을 통합하여 전역적 동작으로 해석해 내는 연구에서 시작되었다. 80년대 Brook[16]가 생물학에서 영향을 받아 상황에 맞춰 센서모터 경험을 능동적으로 구축하는 로봇 에이전트 연구를 발표한 이래로 Sekuler[17]는 동작 검출과 분석을 생물학적 시각에서 신경작용의 연관관계로 보고 시각장의 한정된 영역 내에서 부분 동작 신호를 검출한 다음 대상 동작의 방향과 속도를 전역적으로 통합하여 대상의 동작을 판별하였다. 이외, 부분적으로는 분산되어 있지만 자동조직화를 통해 수렴하면서 신경협력 및 경쟁적 상호작용을 하는 V1뉴런의 기능을 이용한 연구[18], V1과 MT간의 협력/경쟁형태로 상호작용하는 MT 뉴런 기능에 대한 연구[19] 및 일관적 동작과 자체운동 동작 검출에 MST의 기능을 이용한 연구도 제안되었다[20].

Hubel[21]에 의하면 시각 인지를 위한 첫 번째 피질 분석은 V1 뉴런 영역에서 이루어지며 확장된 감수영역으로 대조 감도를 확인한다. V1 뉴런은 주로 배측 경로를 통해 확대되어 선호하는 동작 방향으로 바뀌고 필터링으로 불리는 동작 에너지의 부분 분석을 수행한다. 그의 연구에서 방향 선택 셀은 시공간적 필터로 모델링하며 시각 감수영역은 시공간상으로 활성화된 상호작용으로 모델링하였다. Derrington[22]은 협력/경쟁의 대조적인 뉴런 검출로 동작 방향을 확인하여 방향 선택적 프리미티브를 통해 동작을 추출하는 시각적 매커니즘을 구축하였다.

기존 연구 내용을 반영하여 동작 그 자체만으로 동작 자세의 순서를 분류하는 것이 가능하다는 가정을 바탕으로 두 가지 매커니즘을 설정할 수 있다. 첫 번째 방법은 완벽하게 시각장(명시적 스냅샷)을 파악하여 2D 스냅샷 감수영역을 활성화시키는 것이고 두 번째 방법은 순서(암묵적 스냅샷)를 완벽하게 코딩하도록 뉴런의 분포도를 넓게 하는 것이다[14].

본 연구에서는 스냅샷 감수 영역을 활성화시킨 명시적 스냅샷을 통해 순서를 분류하는 모델을 설정하여 새로운 시각적 동작 순서 인식 방법을 제안한다. 특히 배측과 복측경로를 통해 시각 인지가 이루어지는 과정에 두 경로에서 담당하는 일을 조합하여 비디오의 순서를 효과적으로 결정하는 방법을 설명한다. 형태를 인식하는데 이용하는 정적 스냅샷을 학습과정에 이용하여 동작의 시간적 순서를 학습하고 순방향, 역방향으로 정보를 전달하여 시간 순서를 정확히 이해하도록 함으로써 단순 스냅샷 상태만을 대상으로 동작을 인식하는 방법에서 탈피하여 동작 순서를 이해하는 새로운 동작 인식 방식을 제시한다.

3. 모델 구조

3.1 동작인식을 위한 시각인지 영역과 처리 구조

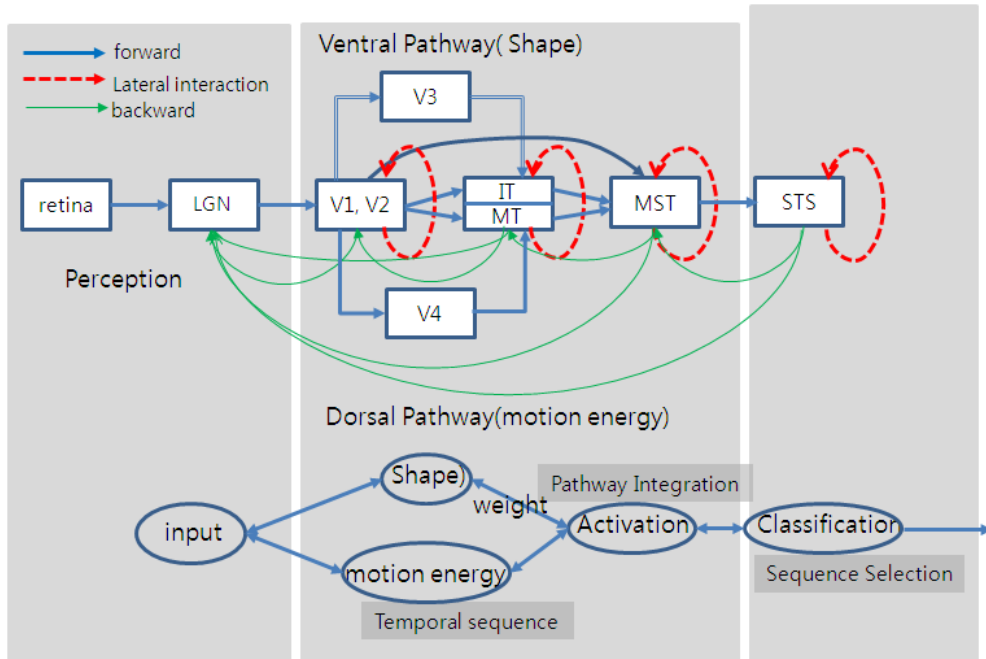
신경생물학적으로는 시각 자극을 처리하기 위해 먼저 공간적으로 세분화된 윈도우형태의 감수영역이 있는 V1에서 망막 이미지의 부분 동작을 뉴런으로 추출한다. 그 다음 뉴런이 특정 방향으로 움직이는지 검출한다. 측면 상호작용 형태의 부분적 처리가 동작 검출과 예측의 첫 번째 단계가 된다. 하지만 동작의 시각적 인지가 온전히 신경계 감수영역의 부분 반응으로만 결정되는 것은 아니며 V1에서 검출정보를 수집하여 조합한 다음 MT에서 한꺼번에 모은 후에 속도 정보를 확인하는데 감수영역의 반응을 이용한다.

이는 조합한 신호를 이용하여 개별 뉴런 반응의 불확실성을 해결하는 방법으로 시각 정보는 시각처리를 담당하는 각 영역에 순방향(feedforward)으로 정보를 전달하고 해당 영역에서 측면 상호작용을 통해 정보를 모은 다음 역방향(feedback)을 통해 시각 인지 처리 결과를 확인하는 과정으로 이루어져 있다.

V1(primary visual area) 피질영역에서 이루어지는 원

데이터 입력은 적응과 방향 선택 필터링 초기 단계이다. V1의 반응 결과는 정적 형태 표상을 담당하는 피질들(V2, IT 영역)의 복측 경로와 특정 시간 흐름 패턴을 담당하는 피질들(MT-middle temporal area, MST)의 배측 경로에 제공되어 두 경로 중 하나를 선택하게 된다. 시각 인지를 담당하는 복측과 배측 두 경로는 세 단계의 뉴런 모델로 구성되며 해당 모델 구조는 그림 2와 같다.

시각은 시간적 패턴에 의존한다. 눈으로 들어오는 패턴들이 시간이 흐르면서 계속 변하고 순서가 결정되며 그 순서대로 기억되기 때문에 최종 피질 단계에서 동작을 판단하기 전에 감각 영역에서 운동영역으로 이어지는 순방향 연결에서 입력을 받은 영역으로 역방향 정보를 보내야 한다. 그래서 실제로 일어나는 것은 위로 흐르고 일어날 것을 기대하는 것은 아래로 흐른다. 그래서 순방향 뿐 아니라 역방향 및 자체 피질 내 피드백도 고려해야 한다. 배측 경로를 통해 MT영역에 전달된 뉴런은 동작 방향을 확인하여 동작 에너지를 부분 분석한다. MT영역의 방향 선택적 감수영역은 시공간상 흥분과 억제 상태로 뉴런을 모델링하고 두 상태를 대조하여 검출한 결과로 동작 방향을 확인한다. STS 영역에서는 광류 패턴을 검출하여 동작을 분류한다. 복측 경로를 통해 V2 영역에



(그림 2) 시각 인지 피질 영역의 정보처리 구조

(Figure 2) Information Processing Structure for Cortex Area of Visual Perception

전달된 뉴런은 색상을, IT 영역에서는 형태를 검출한다. STS에서 최종 패턴을 분류하여 동작을 판별함으로써 인간의 시각 매커니즘은 방향 선택적 뉴런을 중심으로 시각 자극을 처리함을 알 수 있다[23]. 이러한 생체 시각 인지 구조를 컴퓨터 비전의 주요 단계와 비교하면 인지과정은 비디오시퀀스의 입력, 그리고 각 시각 피질에서 감수영역을 통해 포착하는 형태와 동작방향 특징의 전달과정은 특징 검출과 추출단계 그리고 시각 피질 마지막 단계에서 전달과정을 통합해 동작을 식별하는 것은 컴퓨터 비전의 패턴 분류과정과 매핑할 수 있다. 생체 시각인지 과정에서 자연스럽게 이루어지는 시간적 순서 구분과 식별을 컴퓨터비전의 인식 과정에 반영한 것이 본 연구의 기여점이다.

3.2 제안 모델 구조

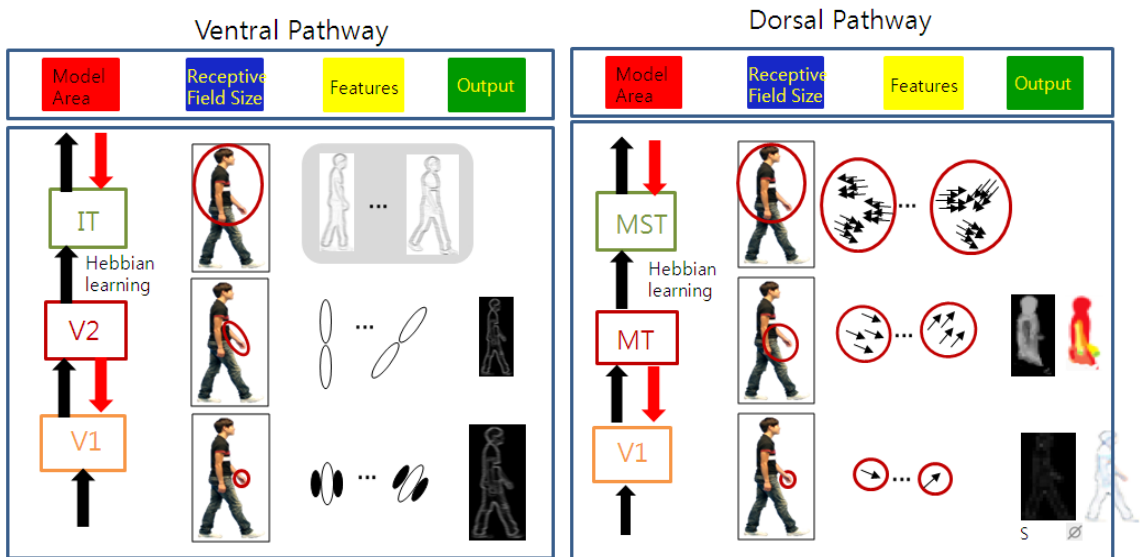
동작 순서 식별을 하기 위한 제안 모델 구조는 그림 3과 같이 [14]의 연구를 바탕으로 비디오 시퀀스 내 대상의 형태와 동작에 대한 처리를 분리하여 배측과 복측 경로, 두 개의 개별 시각경로에서 정보를 처리하도록 구성하며 피질 기능을 모방한 다단계 순방향과 역방향 정보 전달 과정으로 이루어진다.

제안 모델은, 복측 경로의 경우 움직임 이미지가 V1 피질에 인지되면 감수 영역은 부분적인 특징 부위에 반

응한다. 단순한 선분과 기초적인 특징으로 개략적 윤곽을 만들고 이 윤곽은 V2로 넘겨진다. V2는 그 이미지를 더 정교하게 분석한 다음 V4에 넘긴다. 대상의 인식은 입력이 최상위에 놓인 IT에 도달해야 이루어진다. V1의 감수영역은 특징 부위에 반응하고 IT 영역의 감수영역은 움직임 전체에 걸쳐 반응 부위를 넓힌다. 배측 경로도 동작자세의 방향에 복측경로와 유사하다. 동작 자세에서 V1에서는 세부 특징 부위의 방향을 감수 영역에서 반응하고 이 방향을 MT, MST 피질로 넘기면서 점차 전체 움직임 영역의 느린 변화를 보이는 부위로 확대해 간다.

전체 모델의 각 피질 계층에서 정보를 받아 들이는 감수 영역의 크기는 단계별로 다르다. 모델 영역간의 역방향 처리는 여러 단계를 거치는 원래 입력 값이 특징 처리를 통해 변형되지 않고 안정화되는 역할을 한다. 비지도 헤비안 학습으로 형태와 동작 프로토타입 반응을 구축하고 구축 결과는 출력으로 나타낸다.

제안 모델에서 망막을 통해 시각 자극을 인지하는 중간 단계인 IT(inferotemporal area)영역에서는 시각 정보에 대한 표현 내용을 형성하고 MST(middle superior area) 영역에서는 수렴하는 연결강도 특성이 있는 변형된 헤비안(Hebbian) 학습을 통해 프로토타입형 광류를 구축한다. 헤비안 학습은 시냅스 양쪽의 뉴런이 동시에 또 반복적으로 활성화되면 그 두 뉴런 사이의 연결강도는 강화된다는 관찰에 근거한 것으로 시냅스의 변화가 인접한 두



(그림 3) 복측(좌)과 배측(우) 경로에서 다단계 순방향과 역방향 정보처리 구조
(Figure 3) Feed Forward and Feed Backward Structure in ventral/dorsal pathway

뉴런의 상태에 의해서만 결정되기 때문에 학습이 부분적으로 이루어진다. 초기 연결강도는 0에 가까운 작은 값으로 하고 출력 값을 학습신호로 사용한다. 헤비안 학습법은 순방향 신경망에만 적용할 수 있기 때문에 본 연구에서는 이를 변형하여 순방향과 역방향 모두에 적용하도록 한다.

동작 인식 매커니즘은 형태 경로 즉 복측 경로를 통해 비디오 입력 스트림에서 자동으로 인식 대상의 스냅샷을 선택한다. 프로토타입의 형태와 동작 셀 활성화는 복잡한 STS 모델에서 수렴하여 특정 순서와 상관관계가 있는 동작 자세를 학습한다. 그리고 V1 피질에서 IT, MST 피질과 향하는 상향과 IT, MST 피질에서 V1 피질로 향하는 하향 정보처리과정을 조합하여 순서 선택적 표현을 구축하는데 상향, 하향 두 과정은 모두 헤비안 방식으로 학습한다.

3.3 형태와 동작 프로토타입의 비지도 학습

3.3.1 형태와 동작 경로에서의 헤비안 학습

프로토타입 표현 학습에 필요한 이미지 영역을 선택하기 위해 비디오 시퀀스 내 인식 대상을 바운딩 박스를 표시하여 관심영역(ROI)을 나타낸다. 대상의 동작 특징은 형태와 동작 경로에서 순방향 연결강도를 학습하여 선택하며 이 과정은 변형된 헤비안 학습 방법을 적용한다. 원래 헤비안 학습은 식 1과 같다. 두 뉴런에 거의 동시에 신경 흥분이 일어나면서 두 뉴런 간의 연결강도가 세지는 과정이다.

$$w_{ij}(t+1) = w_{ij}(t) + \eta y_j(t) x_i(t) \quad (1)$$

뉴런 i, j 의 출력이 x_i, y_j 이며 이들은 시냅스 w_{ij} 로 연결되며 η 는 학습율이다. 헤비안 학습은 시냅스의 변화가 인접한 두 뉴런의 상태에 의해서만 결정되며 순방향에만 적용하기 때문에 본 연구에서는 순방향과 역방향 모두를 고려하도록 식 2와 같이 변형한다.

$$\Delta w_{ij}^{FF,s} = \eta_s \cdot \bar{u}_i^{post} \cdot (w_j^{pre} - \bar{v}_i^{post} \cdot w_{ij}^{FF,s}) \quad (2)$$

$\Delta w_{ji}^{FF,s}$ 는 학습율 η_s 과 강화 연결의 효과를 이산화된 변화율로 나타내며 $s \in \{\text{형태}, \text{동작}\}$ 은 형태와 동작 경로에서 학습할 때 같은 학습 매커니즘을 적용함을 의미한다. 변수 $w_j^{pre} = f(x_j)$, $v_i^{post} = f(y_i)$ 는 전후 시냅스

셀의 전위차(membrane potential)로 구동되는 점화율로 여기서는 셀의 활성화를 의미한다. 사후 시냅스 셀 활성화 \bar{v}_i 은 $\bar{v}_i^t = (1 - \lambda) \bar{v}_i^{t-1} + \lambda v_i^t, 0 \leq \lambda \leq 1$ [24]의 시간 추적 규칙으로 계산한다.

이와 같은 시간 추적과 특정 뉴런으로 들어가는 연결 강도만 변경하는 인스타(Instar) 학습[25]조합에서 적응 조건 $1 - \lambda$ 의 가중치 커널은 사후 시냅스 활성화에 따라 조절되고 자동 가중치 정규화 과정을 통해 가파른 하강 학습을 구현한다. 개별 입력 강화 학습을 개시하는 사후 시냅스 셀은 연속력과 최적 정합 반응으로 입력 가중치에 적응하기 위해 뉴런 경쟁 계층에 정렬된다.

3.3.2 강화된 스냅샷 학습

Giese-Poggio 모델[6]은 생물학적 동작 인식에서 순서 선택은 정적 스냅샷 순서에 따라 이루어진다고 제안했다. 이 모델은 시간상 규칙적으로 샘플링한 스냅샷에 의존하지만 본 연구에서는 관절을 강하게 사용하는 자세의 스냅샷을 자동 선택하도록 하여 배쪽 경로인 MT/MST에서 동작에 부가적으로 가한 가중치 신호를 이용하여 형태 채널에서 스냅샷 표현을 학습하도록 한다. 주어진 이웃 셀 전체에 대한 동작 에너지의 가중된 통합은 식 3과 같다.

$$m_e = \int_{\Omega} \underline{u}_{\varphi}(x) \cdot A(x) dx d\varphi \quad (3)$$

$A(\cdot)$ 는 방향선택성 φ 가 있고 공간 위치 x 에서 동작 반응 $\underline{u}_{\varphi}(\cdot)$ 의 상대적 기여도에 대한 가중치를 표시하는 공간커널을 의미한다. 동작 에너지 신호 그 자체는 형태 경로에서 학습을 조절하는데 사용하는 시간합수이다.

본 연구에서는 정적 형태나 스냅샷을 다른 하위분포로 표현하는 방법을 제안하여 시각 영역 모델들이 강하고 약한 관절 자세를 취하는 자세 모두를 포함할 수 있도록 학습한다. 이를 위해 팔을 최대한 뻗은 신호로 강하게 관절을 이용하는 자세 스냅샷을 이용한다. 팔 움직임 에너지는 관절을 강하게 사용하는 단계에서 동작의 방향이 바뀌면 급격히 저하된다. 그러므로 $\partial_t m_e = 0, \partial_i m_e > 0$ 와 같이 부분 동작의 에너지 최저치에서 동시에 나타나는 형태 입력을 선호하는 스냅샷 학습에 주의하기 위해 함수 $g(\cdot)$ 를 포함시킨다. 이에 따라 식 2의 $\Delta w_{ji}^{FF,s}$ 강화 적용 단계에서 학습율은 동작 중속적으로 강화되는 값 $\eta_s \cdot g(m_s)$ 에 따라 달라진다.

3.4 순서 선택적 표현의 비지도 학습

이전 단계에서 학습을 한 IT와 MST의 형태와 동작 경로에서 카테고리 표현은 자신들의 활성화 상태를 STS에 순방향으로 전달한다. 그리고 입력신호가 부분적으로 손실되지만 표현과 활성화 분포를 안정시키기 위해 STS의 순서 선택 표현은 해당 입력 단계에 하향 신호를 보낸다.

3.4.1 순서 선택 동작 표현의 순방향 학습

시공간 순서 선택성이 있는 프로토타입 표현은 형태와 동작 프로토타입 학습과 유사하게 변형된 헤비안 학습 매커니즘을 이용하여 학습한다.

$$\Delta w_{ji}^{in,FF} = \eta_{seq} \cdot \bar{u}_i^{post} \cdot (u_j^{pre} - \bar{v}_i^{post} \cdot w_{ji}^{in,FF}) \quad (4)$$

연결강도 커널 $\Delta w_{ji}^{in,FF}$ 은 $IT \rightarrow STS$ 수렴을 나타내고 $MST \rightarrow STS$ 상향식 방향으로

STS 사후 시냅스 셀에 입력된다. η_{seqFF} 는 학습율, u_j, v_i 는 전후 시냅스 뉴런의 점화율이다. 사후 시냅스 활성화는 시간추적 매커니즘으로 다시 계산한다. 사전 시냅스 활성화는 $u = u^{IT} \cup u^{MST}$ 로 형태와 동작 출력 활동을 연결하여 생성한다.

3.4.2 학습 역방향 연결

STS에서 순서 선택적 프로토타입은 형태와 동작 프로토타입 표현을 분리하기 위해 출력 연결강도를 $STS \rightarrow IT + MST$ 과 같이 차례차례 거꾸로 되돌려 학습한다. 순방향 학습 매커니즘과는 달리 역방향 학습은 사전 시냅스 셀에서 처리하여 하향식 연결강도로, 특정 뉴런에서 나가는 연결강도만을 변경하는 아웃스타(outstar) 매커니즘으로 처리한다[25]. 학습은 식 5에 따라 이루어지며 식 4의 상향식 학습과 같은 성분을 이용한다.

$$\Delta w_{ji}^{in,FB} = \eta_{seq} \cdot \bar{u}_i^{pre} \cdot (u_j^{post} - w_{ji}^{out,FB}) \quad (5)$$

상향식과 하향식 학습구조는 경쟁적 개념에서 다른 체계로 이루어진다. 결과적으로 연결강도의 합 $\sum_i w_{ji}^{out,FB}$ 은 평균 활성화 값 u_j^{post} 에 가까워지면서 순서 선택적 단위 v_i^{pre} 는 입력을 구성하는 기대치 패턴을 기억하게 된다.

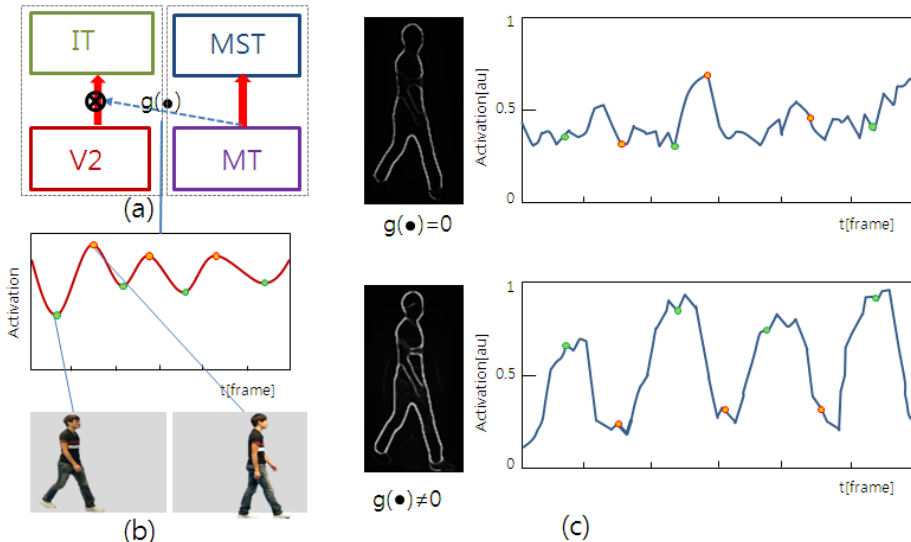
그림 4는 가중치 신호를 이용하여 IT 프로토타입을 학습시키는 과정을 설명한 것이다. 그림 4 (a)의 두가지 다른 특징 처리 경로에서 배측 경로의 동작 방향 특징 값을 복측 경로의 형태 특징에 피드백하여 추가하면 동작 신호가 강화, 활성화되어 인식율을 높인다. 그림 4 (b)의 최소, 최대치 동작 에너지가 각각 비관절과 관절체 상태에 해당한다. IT 프로토타입을 연속적으로 학습시키면 그림 4 (c)와 같이 낮은 관절상태를 활성화하여 끌어 올리게 되고 이를 통해 동작을 강화시키면 걸음걸이와 동기를 맞춘 자세의 신호 스냅샷이 잘 활성화된 IT 프로토타입을 확인할 수 있으므로 동작 인식이 쉬워진다.

4. 실험 및 결과

본 연구의 첫 번째 실험을 통해 입력 비디오 스트림의 정적 관절체 동작에서 어떤 속성으로 스냅샷을 선택하는지를 설명한다. 활발한 관절 동작 자세는 자세에 어떤 동작을 함축하고 있는지를 나타내고 그 스냅샷 자세를 통해 다음 동작을 예측한다는 점에서 스냅샷 속성의 아이디어를 얻었다. 실험 대상 입력 비디오는 오른쪽에서 왼쪽으로 걷는 3명의 보행자 동작을 이용하였다.

그림 4는 강화 신호를 사용하거나 사용하지 않음으로써 IT 프로토타입을 훈련시키는 과정이다. 그림 4 (b)의 동작 에너지의 최소, 최대 값은 관절체와 비관절체 자세에 해당한다. 최소값이 나타나면 그림 4 (a)와 같이 배측 경로인 MI/MST에서 동작에 부가적으로 가한 강화 신호를 형태 채널에 이용하여 강화된 스냅샷 표현을 학습하도록 하였다. 헤비안 학습으로 시냅스 양쪽 뉴런이 동시에 반복적으로 활성화되면 두 뉴런간의 연결강도는 강화되므로 동작 에너지의 최소 값이 전체 신체 모양 프로토타입을 학습하기 위한 강화 신호가 된다. IT 프로토타입에 대한 시간차에 따른 반응신호를 통해 가중치 $g(m_e) = 0$ 일 때는 강화가 이루어지지 않으며, $g(m_e) \neq 0$ 일때 강화가 이루어진다. 그림 4 (c) 위쪽과 같이 IT 프로토타입에 강화가 이루어지지 않으면 스냅샷이 활성화가 명확하지 않은 상태가 되어 동작 선택이 어렵지만 그림 4 (c) 아래쪽과 같이 동작을 구동하는 강화 과정을 통하면 신호 스냅샷이 걸음걸이와 동시성을 이루면서 동작이 명확해진다.

다음으로는 STS 표상의 반응 속성과 동작 순서 선택을 설명하기 위해 먼저 순서 선택 표현은 왼쪽에서 오른쪽을 향해 걷는 보행자에 대한 STS 단계를 학습한다. 그



(그림 4) 동작 강화 신호를 학습한 IT 프로토타입
(Figure 4) IT prototype learnt by Activation Enforcement Signal

리고 형태와 동작 경로에서 입력표현은 IT와 MST 모델에서 나타나도록 설정하고 학습 단계와 같은 방향으로 걷는 보행자, 반대 방향으로 걷는 보행자, 역방향으로 거꾸로 걷는 보행자 등 순차적으로 세 개의 다른 동작 시나리오를 제시하였다.

그 결과, 형태와 동작 프로토타입과 순서 표현은 학습 과정과 같은 테스트 비디오에서 가장 잘 나타났고 반대 방향으로 걷는 경우에는 형태와 동작 프로토타입은 최저 값을 보였으며 순서 선택 셀에서도 같은 결과를 보였다. 역방향으로 거꾸로 걷는 경우 형태 프로토타입은 관절 사용 자세 입력에 대해 높은 값을 보였으나 동작에 대해서는 최저 값을 나타냈다.

결과적으로, 그림 5와 같이 순서 선택 표현은 중간단계인 IT에서 반응을 했다. 이 결과는 [26]와 [27]의 실험 결과와 일맥상통한다. 1m/s 속도로 오른쪽 방향을 향해 걷는 보행자를 대상으로 학습한 내용을 참조하여 움직임 방향과 속도를 달리하는 여러 보행자를 설정하고 테스트 단계에서는 보행 방향을 $\pm \{5^\circ, 10^\circ, 20^\circ, 40^\circ, 80^\circ\}$ 로 회전시켰다. 모델 시뮬레이션 결과 약 $\pm 45^\circ$ 씩 진폭 방향이 반으로 회전하였다.

속도를 달린 보행자의 가다가 반대방향으로 걷는 동작 역시 관찰하였다. 학습한 IT/MST와 STS 순방향/역방향 연결이 전체적으로 잘 이루어진 모델을 참고로 사용하였다. 동작 입력(MST) 또는 스냅샷 입력(IT)에서 상

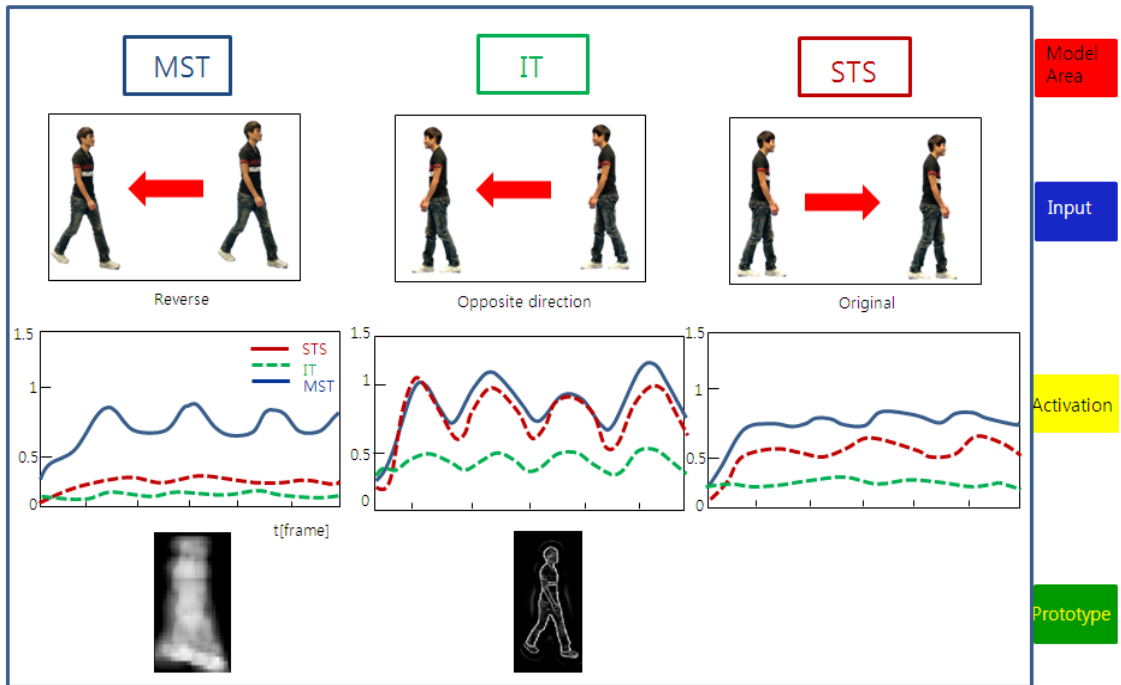
향 연결하여 STS에서 순서 선택 뉴런 반응을 차단했을 때 반응 진폭은 반으로 낮아졌다. STS에서 역방향으로 전달된 정보는 IT와 MST 표상에서 진폭 활성화를 저하시켰다.

IT 단독으로 한 순방향 활성화는 STS 순서 뉴런을 구동시켜 처리하였다. IT의 스냅샷 표현은 STS 순서 뉴런을 구동시키고 차례로 IT, MST 프로토타입 표상에 역방향 신호를 보냈다. 동작 경로에서 이런 역방향 신호전달을 하면 전체 시냅스 활성화를 증가시킨다.

그림 5는 왼쪽에서 오른쪽으로 걸어가는 보행자가 있는 비디오 입력으로 학습한 결과로 IT 스냅샷 뉴런, MST 동작 패턴 뉴런, 순서 선택적인 STS 셀의 반응을 나타냈다. 학습 순서와 같은 방향으로 걷는 보행자, 반대 방향으로 걷는 보행자, 거꾸로 걷는 보행자 동작 등 다른 입력조건에 따라 모델 영역의 활성화가 다양하게 나타나는 결과를 확인할 수 있다.

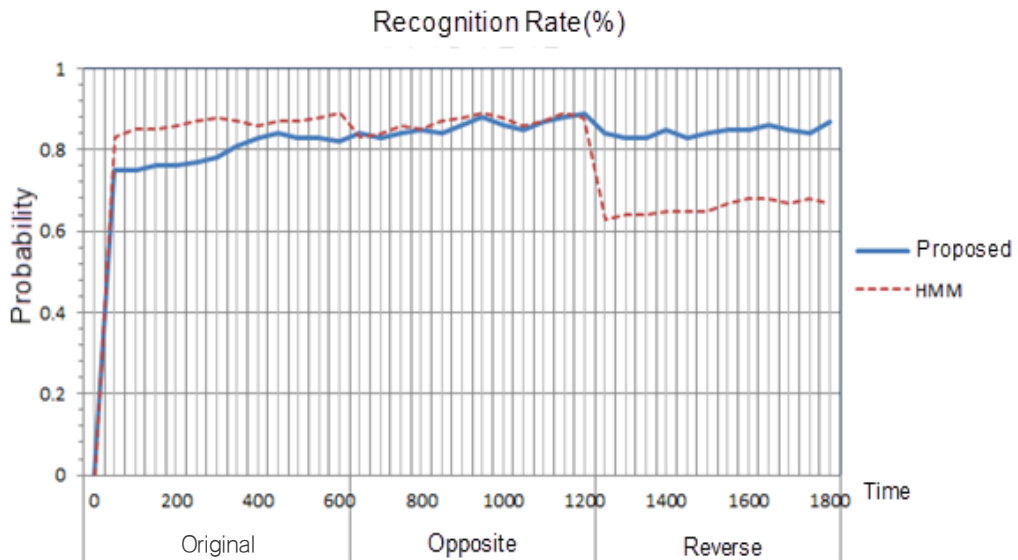
왼쪽에서 오른쪽을 향해 걷는 보행자, 반대 방향으로 걷는 보행자, 거꾸로 걷는 보행자 동작 시나리오를 설정하고 제안 방법을 동작 인식에 가장 많이 적용되는 HMM 패턴 분류 방식에 비교한 결과는 그림 6과 같다.

각각의 동작 시퀀스를 일정 시간동안 선행적 입력 값으로 제시하여 제안 방법과 HMM으로 동작 인식을 수행한 결과 왼쪽에서 오른쪽으로 걸어가는 체원 동작에 대해서는 HMM이 더 우수한 분류 결과를 보였으나 거꾸로 걷는 동작에 대해서는 제안 방법이 훨씬 높은 인식률을



(그림 5) IT 스냅샷 뉴런, MST 동작 패턴 뉴런, 순서 선택적 STS 셀의 반응

(Figure 5) Responses of IT snapshot neurons, MST action pattern neurons and Sequence-selective STS cell



(그림 6) 제안 방법과 HMM 동작 인식을 비교

(Figure 6) Comparing of Motion Recognition between Proposed and HMM

보였다. HMM을 통해 학습한 동작 인식은 학습한 방향과 같은 방향으로 걷는 재현동작에 대해서는 제안 방법보다 높은 인식률을 보였고 반대방향을 향해 순차적으로 걷는 동작은 제안 방법과 비슷한 인식률을 보였다. 그러나 거꾸로 걷는 동작에 대해서는 HMM이 낮은 인식률을 보여 입력데이터의 시퀀스를 구분하는데 제안 방법이 더 우수함을 알 수 있다.

표 1은 제안 방법과 HMM이 주어진 시간동안 입력 값에 대해 지속적으로 정확하게

인식하고 있는지를 비교한 것이다. 왼쪽에서 오른쪽으로 가는 원래 동작은 HMM이 더 우수한 지속율을 보였으나 반대 방향으로 가는 경우와 동작을 거꾸로 했을 때의 정확한 동작 인식 지속율은 제안 방법이 더 나은 결과를 보였다.

(표 1) 정확 지속율 결과 비교
(Table 1) Comparing of correct duration rate

Scenario	Correct Duration Rate(%)	
	Proposed	HMM
Original	78	87
Opposite direction	88	87
Reverse	87	66

5. 결 론

본 연구에서는 동작을 인식하는 방법으로 비지도 학습을 이용한 생체 모방 동작 인지 방식을 제안하였다. 제안 모델은 대뇌 피질 구조와 특정 뉴런 표현 기능으로 구성된 신경생리학의 시각 자극인지 과정을 기반으로 스냅샷 자세에 내포된 동작을 인식하기 위해, 형태와 동작을 각각 분리하여 처리하는 뇌의 생물학적 시각처리 경로에서 순방향과 역방향 처리과정을 통해 형태와 동작 프로토타입을 학습하였다. 그리고 형태와 동작 처리 경로를 상호 교차 작용시켜 동작 시퀀스에서 핵심 자세를 선택하도록 하였다. 이러한 동작 인식 매커니즘은 많은 관절을 사용하는 동작의 스냅샷 자세 표현을 더 정확히 인식하는 데 도움이 된다.

[5][6]과 같은 이전 방식과는 달리 본 연구에서는 형태와 동작 반응에 대해 STS 단계에서 순서 선택 표현을 학습하도록 하여 동작 인식에서 중요한 순서를 찾도록 했다. 헤비안 학습을 통해 형태와 동작 모두를 조합하여

STS에서 순서 선택 표현을 학습함으로써 스냅샷과 동작 패턴 프로토타입이 개별적으로 순서를 선택하지 않아도 되며 전체 비디오 시퀀스에서 특정 움직임의 순서를 이해하도록 했다. 또한 입력 값을 더 정확히 표현하기 위해 모델 영역간의 정보 전달을 역방향으로도 연결하여 비지도 학습하도록 하고 $IT/MTS \rightarrow STS$ 와 $STS \rightarrow IT/MST$ 네트워크를 이용하여 광류 패턴과 관련 스냅샷 순서를 예측하도록 했다. 이는 출력을 되돌려 입력으로 삼아 모든 뉴런들에 보내고, 피드백을 수행하면 패턴들은 서로의 순서를 학습하는 신경생물학적 방법을 컴퓨터비전에 적용한 것으로 제안 내용이 컴퓨터 비전 연구에 기여하는 바이다.

결과적으로 시간적 동작 순서를, 주기적으로 연결되는 IT, MST, STS가 구성하는 분산네트워크로 표현하여 관절 동작의 특징과 규칙성을 강건하게 재현(recall)할 수 있도록 했다. 또한 1장에서 제시한 바와 같이 생물학적 시각정보 인식 특징을 고려하였으며 [12]에서 관측된 것과 같이 제안 모델은 이전에 학습한 순서만으로 정적 중심 자세를 표현하여 STS 순서 선택 뉴런의 활성화를 도모하였다. 그리고 정적 관절 자세 표현으로 동작 예측이 가능하며 동작 경로에서 동작 자세 특징별로 인식 영역을 활성화할 수도 있다[13]. 동작 인식을 위한 자세 입력이 형태 경로에서는 초기에 미미한 수준으로만 반응하지만 STS 움직임 선택 뉴런을 활성화시킨다는 것을 예측할 수 있다. 그리고 STS 셀에서 역방향으로 전달되는 정보를 통해 인식 성능 활성화가 이루어지며 IT 스냅샷 반응을 개선함을 확인하였다.

동작 인식을 위한 대표적 모델 구조인 HMM과 제안 방법을 비교한 결과 HMM은 입력비디오에서 순서 인식에는 아주 낮은 인식률을 보였지만 제안 방법은 거꾸로 걷는 형태의 동작비디오도 잘 인식함에 따라 스냅샷 순서 인식에 강건함을 알 수 있다.

동작을 통해 사용자의 의도와 감정을 인식하는 사전 연구[28][29][30]을 바탕으로 한 본 연구를 발전시켜 향후에는 순서 선택 프로토타입의 동작 방향 및 속도를 조절하는 내용과 인식 대상의 성별, 연령별로 제안 모델이 기질적 변화를 선택할 수 있는 방법을 지속적으로 연구할 계획이다. 이외, 신경생물학적인, 학습한 표현간의 활성화 흐름을 컴퓨터 비전 시각으로 구현하는 과정도 다룰 예정이다. 또한 다양한 동작 비디오를 대상으로 제안 방법을 테스트하여 연구 신뢰도를 높이도록 할 것이다.

참 고 문 헌(Reference)

- [1] B. McCane, K. Nobings, D. Grannitch, B. Galvin. "On benchmarking optical flow." *Computer Vision and Image Understanding*, 2003, pp. 126-143,
- [2] Jeff Hawkins, Sandra Baksless, "On Intelligence-How a new understanding of the brain will lead to the creation of truly intelligent machines", Owl Books, NY, 2005.
- [3] H. Jhuang, T. Serre, L. Wof, T. Poggio, "A Biologically inspired system for Action Recognition," *ICCV(Inter. Conf. Com. Vision)*, 2007, pp. 1-8.
- [4] R. J. Peters, L. Itti, "Beyond bottom-up : Incorporating task-dependent influences into a computational model of spatial attention," *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
- [5] V. Navalpakkam, L. Itti, "An Integrated Model of Top-down and Bottom-up Attention for Optimal Object Detection, *CVPR(Computer Vision and Pattern Recognition)*, 2006, pp. 2049-2056.
- [6] M. A. Giese, T. Poggio. "Neural mechanisms for the recognition of biological movements." *Nat. Rev. Neurosci*, vol. 4, pp. 179 - 192, 2003.
- [7] A. Casile, M. A. Giese. "Roles of motion and form in biological motion recognition." *Lecture Notes in Computer Science*, vol. 2714, pp. 854 - 862, 2003.
- [8] H. Jhuang, T. Serre, L. Wolf, T. Poggio. "A biologically inspired system for action recognition." *IEEE Intl. Conf. on Computer Vision*, 2007, pp. 14-20,
- [9] K. Schindler, L. Van Gool. "Action snippets: How many frames does human action recognition require?" *IEEE Conf. on Computer Vision and Pattern Recognition*, 2008, pp. 22-24.
- [10] M. J. Escobar, G. S. Masson, T. Vieville, P. Kornprobst. "Action recognition using a bio-inspired feedforward spiking network." *Intl. Journal of Computer Vision*, vol. 83, 2009, pp. 284 - 301.
- [11] C. M. Bishop. "Neural Networks for Pattern Recognition." Oxford University Press, 1995.
- [12] T. Jellema, D.I. Perrett. "Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion?" *Neuropsychologia*, Vol. 41, pp.1728 - 1737, 2003.
- [13] Z. Kourtzi, N. Kanwisher. "Activation of human MT/MST by static images with implied motion." *Journal of Cognitive Neuroscience*, vol. 12, pp.48 - 53, 2000.
- [14] J. P. Van Santen, G. Sperling. "Temporal covariance model of human motion perception," *Journal of the Optical Society of America*, vol. 1, pp. 451, 1984.
- [15] E. H. Adelson, J. Bergen. "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 284-299, 1985.
- [16] R. A. Brooks. "A robot that walks; emergent behaviors from a carefully evolved network," *Neural Computation*, vol. 1, no. 2, pp. 253-262, 1989.
- [17] R. Sekuler, S. N. J. Watamaniuk, R. Blake. "Motion Perception," *Steven's Handbook of Experimental Psychology*, vol. 1, pp. 121-176, 1998.
- [18] W. A. Fellez, J. G. Talyor. "Establishing retinotopy by lateral-inhibition type homogeneous neural fields," *Neurocomputing*, vol. 48, pp. 313-322, 2002.
- [19] E. Mingolla. "Neural models of motion integration and segmentation," *Neural Networks*, vol. 16, pp. 939-945, 2003.
- [20] C. Pack, S. Grossberg, E. Mingolla, "A neural model of smooth pursuit control and motion perception by cortical area MST," *Technical Report CAS/CNR-TR 99-023*, Department of Cognitive and neural systems and Center for Adaptive Systems, MIT, 2000.
- [21] M. J. Escobar, G. S. Masson, T. Vieville, P. Kornprobst, "Action recognition using a bio-inspired feedforward spiking network," *International Journal of Computer vision*, vol. 82. no. 3, pp. 284-297, 2009.
- [22] D. H. Hubel, T. N. Weisel. "Receptive Fields, binocular iteration and functional architecture in the cats visual cortex," *Journal of Physiology*, vol. 160, pp. 106-154, 1962.
- [23] A. M. Derrington, G. B. Henning. "Detecting and discriminating the direction of motion of luminance and colour gratings," *Visual Research*, vol. 33, pp. 799-811, 1993.
- [24] P. Foldiak. "Learning invariances from transformation sequences." *Neural Computation*, vol. 3, pp. 194 - 200, 1991.

- [25] G. A. Carpenter, S. Grossberg. "Pattern recognition by self-organizing," Neural networks. MIT Press, 1991.
- [26] M. W. Oram, D. I. Perrett. "Integration of form and motion in the anterior superior temporal polysensory area (STPA) of the macaque monkey," Journal of Neurophysiology, vol. 76, no. 1, pp. 109 - 129, 1996.
- [27] J. M. Singer, D. L. Sheinberg, "Temporal cortex neurons encode articulated actions as slow sequences of integrated poses," Journal of Neuroscience, vol. 30, no. 8, pp. 3133 - 3145, 2008.
- [28] Jinok Kim, "A Study on Visual Perception based Emotion Recognition using Body-Activity Posture," The KIPS Transactions, Part B, vol. 18, no. 5, pp. 305-314, 2011.
- [29] Jinok Kim, "Agent's Activities based Intention Recognition Computing", Journal of Korean Internet Society, vol. 13, no. 2, pp. 87-98, 2012.
- [30] Jinok Kim, "Effective Pose-based Approach with Pose Estimation for Emotional Action Recognition", The KIPS Transactions: Part B, vol. 2, no. 3, pp. 1-10, 2013.

● 저 자 소 개 ●



김 진 옥 (Kim Jin Ok)

1989년 성균관대학교 졸업(학사)

1998년 성균관대학교 대학원 정보통신공학과 졸업(석사)

2002년 성균관대학교 대학원 전기전자 및 컴퓨터공학과 졸업(박사)

2004년~현재 대구한의대학교 디자인공학대학 모바일콘텐츠학부 교수

관심분야 : 멀티미디어공학, 패턴인식, 영상처리, HCI

E-mail : bit@dhu.ac.kr