

딥러닝 기반 옵티컬 플로우 추정을 사용한 시각 정보의 촉각 대체 기술

신유정^{1,2}, 김무섭², 정치윤²

¹한밭대학교 컴퓨터공학과

²한국전자통신연구원 인공지능연구소

tlsdbwjd0911@naver.com, gomskim@etri.re.kr, iamready@etri.re.kr

Tactile Vision Substitution Method using Deep Learning-based Optical Flow Estimation

Yujeong Shin^{1,2}, Mooseop Kim², Chi Yoon Jeong²

¹Dept. of Computer Engineering, Hanbat National University

²Artificial Intelligence Research Lab., ETRI

요 약

감각대체기술은 손상된 감각 정보를 다른 감각으로 전환하여 전달하는 기술로써 기존의 시각장애인을 위한 시각 정보의 촉각 대체 기술은 주로 거리 정보나 물체의 윤곽선 정보를 전달하여 사용자가 주변 환경을 이해하는 데 어려움이 있었다. 이를 해결하기 위해 본 논문에서는 딥러닝을 사용하여 사용자 주변의 모션 정보를 분석하고, 이를 촉각 정보로 전달함으로써 사용자가 주변 상황 정보를 인지할 수 있는 방법을 제안하였다. 제안 방법과 기존의 윤곽선 정보 전달 방법을 사용자 실험을 통하여 비교하였을 때, 제안 방법이 영상 속 물체의 움직임 정보를 이해하는 데에 더욱 효과적임을 확인하였다.

1. 서론

감각대체기술은 손상되거나 기능이 저하된 감각 정보를 다른 형태의 감각으로 전환하여 사용자에게 전달하는 기술이다[1]. 이러한 기술은 손상된 감각 능력을 향상시켜 지속적인 경제 활동과 삶의 질을 향상시킬 수 있기 때문에 많은 연구가 진행되고 있다. 그 중, 시각장애인을 위해 시각 정보를 청각으로 대체하는 기술들[2-4]이 가장 활발히 연구되고 있지만, 시각 정보의 청각 대체 기술은 일상생활에서 소음의 영향을 많이 받기 때문에 최근에는 시각 정보를 촉각으로 대체하는 전달하는 기술에 관한 연구들이 진행되고 있다[5-7].

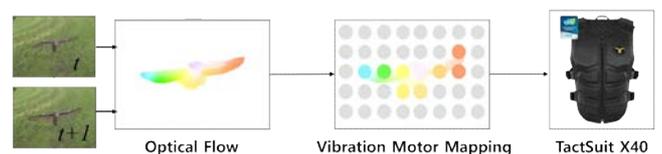
시각 정보의 촉각 대체 기술은 물체의 윤곽선 정보를 추출하여 햅틱 슈트(Haptic suit)로 전달하는 연구[5]와 깊이 카메라를 사용하여 사용자와 물체 간의 거리 정보를 추출한 후, 햅틱 슈트로 전달하는 연구[6]가 진행되었다. 또한, 사용자의 움직임을 분석하여 물체와의 거리 정보를 추정 후, 햅틱 헬멧을 통해 전달하는 연구[7]가 진행되었다. 하지만 거리 정보는 단순히 앞에 존재하는 물체와의 거리만을 알 수 있고, 물체의 움직임 정보를 지각할 수 없다.

또한, 기존의 윤곽선 정보 전달 방법은 모든 물체의 윤곽선을 검출하여 전달하기 때문에 주변 환경을 이해하는 데에 방해가 된다. 이러한 문제점들을 해결하기 위해 본 논문에서는 전체 영상에서 움직이는 물체들의 정보를 전달하는 방법을 제안한다.

본 논문에서 제안하는 모션 정보를 활용한 시각 정보의 촉각 대체 방법은 모션 정보를 통하여 사용자가 물체의 움직임과 거리를 지각할 수 있게 하여 주변 상황을 이해하는 데 도움을 줄 수 있다.

2. 모션 정보 기반 시각 정보의 촉각 대체 기술

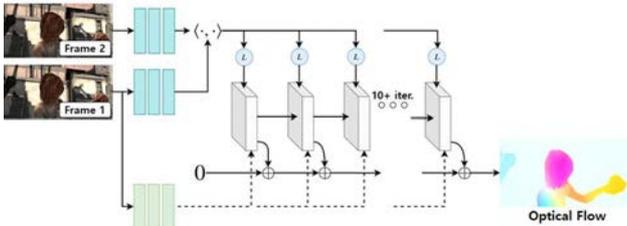
본 논문에서 제안하는 방법의 전체적인 흐름은 그림 1과 같다. 먼저, 이전 영상과 현재 영상을 사용해 옵티컬 플로우(Optical flow)를 추정하고 모션 정보를 생성한다. 생성된 모션 정보는 제안된 매핑 방법을 통해 햅틱 슈트의 진동으로 전달하여 사용자가 촉각을 통해 모션 정보를 지각할 수 있게 하였다.



(그림 1) 제안 방법의 전체적인 흐름도.

모션 정보를 분석하기 위하여 연속된 두 영상을 비교하여 각 픽셀의 모션 벡터를 계산하는 옵티컬 플로우 추정 방법을 사용하였다. 전통적인 옵티컬 플로우 추정 방법은 작고 빠르게 움직이는 객체를 인지하지 못하고 광원의 움직임 변화를 정확히 구별하지 못하는 단점이 있다[8]. 이를 해결하기 위하여 최근에는 딥러닝 기반의 옵티컬 플로우 방법들이 연구되고 있다.

제안 방법은 옵티컬 플로우 추정을 위하여 종단간 기계학습(End-to-end machine learning)을 사용한 RAFT(Recurrent All-Pairs Field Transforms) 모델[8]을 사용하여 모션 정보를 분석하였다. RAFT 모델은 연속된 두 영상을 통해 장면 간 유사도를 계산하고, 반복적인 연산 학습 구조를 통해 플로우 필드(Flow field)의 정확도를 향상시켰다. RAFT 모델의 네트워크 구조는 그림 2와 같으며, 입력으로 받은 연속된 두 영상으로부터 특징을 추출하는 과정, 연속된 두 영상의 유사도를 계산하는 과정, 회귀적 연산을 통해 매 학습단계마다 정확도를 높이는 과정으로 구성된다.



(그림 2) RAFT 네트워크 구조도[8].

RAFT 모델을 통하여 모션 정보 분석의 정확도를 향상시키면 물체가 움직이는 영역의 정보만을 검출할 수 있으며, 이를 햅틱 슈트로 전달하면 움직임이 없는 배경은 진동을 느끼지 못하고 움직임이 있는 물체는 진동으로 느낄 수 있다.

영상의 모션 정보는 영상의 각 픽셀에 대해서 생성되는 반면, 햅틱 슈트의 모터의 개수는 제한적이다. 본 논문에서 사용된 햅틱 슈트의 경우 bHaptics사의 TactSuit X40 모델로써, 앞면과 뒷면 각각 20개씩 총 40개의 진동모터가 내장되어 있다. 햅틱 슈트의 40개 진동모터는 5x8의 매트릭스 형태로 배치되어 있기 때문에 높은 해상도의 모션 정보를 5x8의 매트릭스로 매핑하는 방법이 필요하다.

이를 위하여 RAFT 모델에서 생성한 영상 내 각 픽셀의 모션 벡터 크기(Magnitude)를 계산한다. 그리고 입력 영상을 총 40개의 영역으로 분할 한 후, 식 1과 같이 각 영역에 존재하는 모션 벡터들의 크

기 합을 계산하여 정규화한다.

$$T_{fg} = \frac{1}{M} \times \left(\sum_{j=m.f}^{m(f+1)-1} \sum_{i=n.g}^{n(g+1)-1} F_{ij} \right) \quad (1)$$

식 1에서 m 과 n 은 영상을 40개 영역으로 분할하기 위해 영상 크기(w, h)를 각각 8과 5로 나눈 값이다. f 와 g 는 각각 $0 \leq f \leq 4$, $0 \leq g \leq 7$ 의 범위를 갖으며, F_{ij} 는 각 픽셀에서 모션 벡터의 크기를 나타낸다. M 은 모션 벡터 크기 중 가장 큰 값으로, 정규화를 위해서 사용된다. T_{fg} 는 정규화된 40개의 값들을 원소로 가지는 5x8 크기의 행렬이며, 햅틱 슈트가 제공하는 진동모터의 최대 진동세기(K)에 매핑하기 위해서 K 를 곱해준다. 본 논문에서 K 값으로 100을 사용하였다.

사용자가 영상 영역에 대응하는 햅틱 슈트 영역을 직관적으로 인지할 수 있도록 그림 3과 같이 영상의 중앙 부분은 햅틱 슈트 앞면에, 영상의 양 끝 부분은 뒷면에 매핑하였다. 따라서 영상 중앙(그림 3의 $F1F4$ 영역)에서의 물체 움직임은 사용자의 가슴과 배에서, 영상 양 끝에 있는 물체의 움직임은 사용자의 등에서 느낄 수 있도록 설계하였다.



(그림 3) 영상 영역과 햅틱슈트의 진동모터 매핑도.

제안된 방법을 사용하면 물체의 움직임뿐만 아니라, 물체가 가까워질수록 한 영역에서 차지하는 움직이는 픽셀의 양이 많아져 더욱 센 진동 신호로 전달할 수 있으므로 사용자가 물체의 움직임과 거리를 직관적으로 지각할 수 있다.

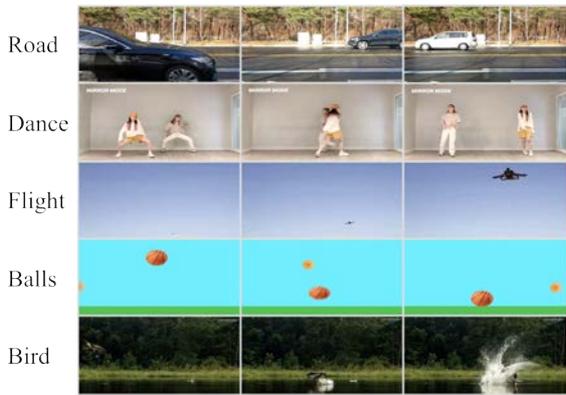
3. 실험 결과

본 논문에서 제안한 방법의 성능을 알아보기 위하여 기존의 윤곽선 정보를 전달하는 방법[5]과 비교하는 사용자 실험을 진행하였다.

사용자 실험은 시각 정보의 촉각 대체 기술에 대한 이해를 돕기 위한 과정과 두 가지 방법으로 시각 정보를 촉각으로 전달할 때 시각 정보와 일치하는 방법을 선택하는 과정으로 구성된다. 첫 번째 과정에서는 햅틱 슈트의 진동 자극을 가로, 세로로 전달하면서 진동모터와 영상 위치를 매핑하여 표시하고 한 개의 데모 영상을 통해 움직이는 물체에 대한 진

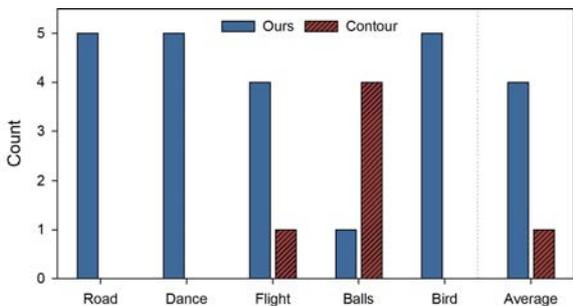
동 자극을 체험할 수 있게 하였다. 두 번째 과정에서는 사용자가 동영상을 시청하면서 랜덤 순서로 전달되는 두 가지 방법의 촉각 자극 중 영상 정보와 더 일치하는 방법을 선택하도록 하였다.

실험 참가자 수는 남성 2명, 여성 3명으로 평균 연령은 24.8세이다. 실험에 사용된 동영상은 그림 4와 같이 도로, 춤, 비행기, 공, 새 등 총 5개이다.



(그림 4) 실험에 사용된 5개의 동영상의 샘플 영상.

실험 결과는 그림 5와 같으며, 대다수가 제안 방법이 전달하는 촉각 정보가 기존 방법보다 영상 정보와 더 일치한다고 선택하였다. 실험에 사용된 동영상 중 컴퓨터 그래픽으로 생성된 공 영상의 경우, 배경이 심플하고 물체의 윤곽선 추출이 용이하여 다수 사용자들이 윤곽선 기반의 촉각 정보가 시각 정보와 더 일치한다고 선택하였다. 하지만 실제 환경에서 취득되는 영상은 배경이 복잡하고 물체의 윤곽선을 정확히 추출하기 어렵다. 따라서 실제 환경에서는 제안 방법이 더 효과적일 것으로 판단된다.



(그림 5) 사용자 실험 결과.

4. 결론

본 논문에서는 기존의 시각 정보의 촉각 대체 기술의 문제점을 해결하기 위하여 모션 정보를 활용한 시각 정보의 촉각 대체 방법을 제안하였다. 제안 방법은 딥러닝 기반의 RAFT 모델을 통하여 옵티컬 플로우를 추정함으로써 정확한 모션 정보를 생성할

수 있다. 또한 40개의 진동모터를 갖는 햅틱 슈트로 모션 정보를 전달하는 방법을 제안하였으며, 사용자 실험 결과 제안 방법이 기존 윤곽선 기반의 방법보다 영상의 시각 정보와 일치하는 촉각 정보를 전달함을 확인하였다. 제안 방법을 사용하면 사용자가 물체의 움직임과 거리를 지각할 수 있게 하여 주변 상황을 이해하는 데 도움을 줄 수 있다.

사사문구

이 논문은 한국전자통신연구원 연구운영비지원사업(기본사업)으로 수행한 인간중심의 자율지능시스템 원천기술연구 과제(22ZS1200)의 논문입니다.

참고문헌

[1] K.D. Moon et al., “Technological Trends in Sensory Substitution,” *Electronics and Telecommunications Trends*, vol. 32, no. 4, 2019.

[2] S. Abboud et al., “EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution,” *Restor. Neurol. Neurosci.*, vol. 32, no. 2, 2014.

[3] S. Caraiman et al., “Computer Vision for the Visually Impaired: the Sound of Vision System,” *Proc. ICCVW*, 2017.

[4] M. Kim et al., “Analysis and Validation of Cross-Modal Generative Adversarial Network for Sensory Substitution,” *Int. J. Environ. Health Res.*, vol. 18, no. 12, 2022

[5] J. Wu, J. Zhang, J. Yan, W. Liu and G. Song, “Design of a Vibrotactile Vest for Contour Perception”, *International Journal of Advanced Robotic System*, vol. 9, no. 5, 2012.

[6] P. Wacker et al., “VibroVision: An On-Body Tactile Image Guide for the Blind”, *Proc. CHI EA*, 2016.

[7] S. Morita et al., “Environment understanding during walking via modality conversion from visual to haptic information: localization experiment with vibro-stimuli simulated by optical flow”, *Proc. IWAIT*, 2020.

[8] Z. Teed and J. Deng, “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow,” *Proc. ECCV*, 2020.