

적응형 신호 처리 및 경량화 알고리즘 기반의 고정밀 제어 시스템

서우진, 박시환, 심지훈, 김용강*

국립 공주 대학교

ca1101760@gmail.com, pswbest2371@naver.com, jihun4510@naver.com, ygkim@kongju.ac.kr*

High-Precision Gesture Control System

via Adaptive Signal Processing and Lightweight Algorithms

Woojin Seo, Sihwan Park, Jihun Shim and Yonggang Kim*

Kongju National University

요약

본 논문은 연산 비효율성과 입력 신호 불안정성 문제를 해결하기 위해 고비용의 딥러닝 모델에 의존하지 않고, 이원화 아키텍처와 1 Euro Filter 및 FSM 기반의 경량 신호 처리·제어 기법을 활용한 입력 신호 안정화 및 제스처 해석 접근을 제안한다. 이를 통해 시스템 부하를 최소화하면서도 저사양 하드웨어 환경에서 적용 가능한 실시간 비접촉 제어 프레임워크를 설계하고, 연속 제어와 이벤트 기반 제어를 모두 지원하는 범용적인 손 제스처 인터페이스 구조를 제시한다.

I. 서론

물리적 접촉이 제한되는 의료 및 산업 현장에서 비접촉 인터페이스는 필수적인 대안으로 주목받고 있다. 특히 팬데믹 이후 비접촉 기술에 대한 수요가 급증함에 따라, 사용자의 손짓을 실시간 명령으로 변환하여 공간을 디지털 작업 영역으로 확장하는 기술의 실용적 가치가 더욱 부각되고 있다. 기존 가상 마우스 연구들은 단순 스무딩 기법에 의존하여 환경 요인에 따른 신호 불안정성을 해결하지 못했거나, 딥러닝 모델의 과도한 연산량으로 인한 반응성 저하 문제를 노출하였다[1]. 실제 사용자 환경에서는 고도화된 학습 모델보다 시스템의 경량화와 실시간 최적화가 자연 없는 즉각적인 반응성을 위해 우선되어야 한다. 따라서 본 연구는 센싱과 제어 로직을 분리하는 이원화 아키텍처와 기하학적 연산 기반의 경량 알고리즘을 도입하여, 고사양 장비 없이도 마우스 수준의 정밀 제어가 가능한 비접촉 인터페이스 시스템을 제안한다. 이를 통해 저사양 환경에서도 운용 가능한 고성능 시스템 구현이 가능함을 입증하고자 한다.

II. 본론

1. 시스템 구성

본 논문은 실시간성을 보장하기 위해 그림 1과 같이 4단계 데이터 흐름 프라인으로 구성된 이원화 아키텍처를 채택한다. 비전 인식과 제어 로직을 분리함으로써 각 처리 단계의 특성에 최적화된 실행 환경을 적용하며, 비전 처리에는 Google MediaPipe와의 높은 호환성을 갖는 Python 환경을 활용하여 CPU 환경에서도 21개의 3D 손 랜드마크를 안정적으로 추출한다[2]. 추출된 손 랜드마크 데이터는 WebSocket을 통해 Java 백엔드로 실시간 전송되며, 저지연 양방향 통신을 통해 60FPS 이상의 데이터 전달을 보장한다. 수신된 데이터는 입력 제스처 유형에 따라 O(1)의 시간 복잡도로 처리되어 연산 지연을 최소화하며, 멀티스레드 안정성과 구조적 확장성이 강점을 지닌 Java 환경을 통해 제어 로직의 신뢰성을 확보한다. 또한 제어 로직과 UI 렌더링을 분리하여 시스템 전반의 안정성을 강화하고, 3D 지도 기반 테스트 베드 상에서 제안 시스템의 제어 성능을 시각적으로 검증한다.

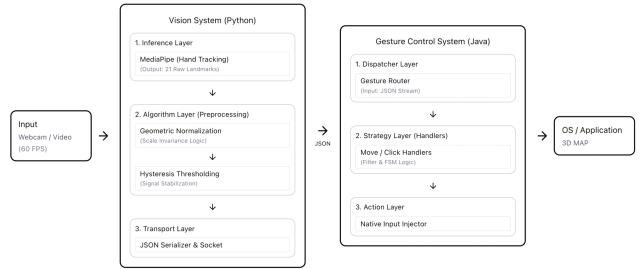


그림 1. 제안하는 비접촉 제스처 인터페이스의 전체 시스템 아키텍처

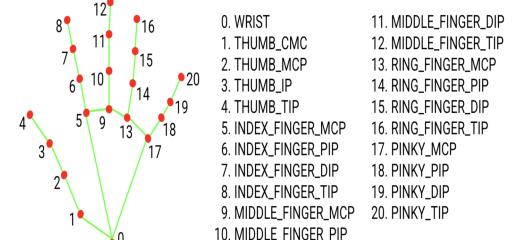


그림 2. MediaPipe 기반 손 랜드마크 정의 및 인덱싱 구조

2. 기하학적 정규화 및 히스테리시스

입력으로 사용되는 손 랜드마크는 카메라와 사용자 거리 변화에 따라 스케일 의존성을 갖는다. 이를 해결하기 위해, 절대 좌표가 아닌 비율 기반의 정규화를 적용하였다. 기준이 되는 손의 크기 L은 손목(0)에서 네 손가락 끝점(8,12,16,20)까지의 유클리드 거리 평균으로 정의한다.

$$L_{\text{handsize}} = \frac{1}{4} \sum_{i \in \{8, 12, 16, 20\}} \text{Distance}(P_0, P_i)$$

수식 1. 유클리드 거리 기반 손 스케일 정교화

제스처 판별은 각 동작에 해당하는 손가락 끝점 사이의 거리를 수식 1에서 정의한 L_{handsize} 로 나눈 비율을 기반으로 수행된다.

3. 적응형 신호 안정화

웹캠 기반 인터페이스의 고질적인 문제인 지터와 지연 간의 상충 관계를 해결하기 위해, 본 연구는 속도 기반의 1 Euro Filter를 적용하였다[3]. 이 필터는 물체의 속도 $\dot{|X|}$ 에 따라 차단 주파수 f_c 를 동적으로 조절한다. 즉,

정지 상태의 미세한 손 떨림은 강하게 필터링하여 억제하고, 빠른 움직임에 대해서는 필터링을 최소화하여 지연 없는 반응성을 보장한다.

$$f_c = f_{c_{\min}} + \beta |\dot{X}|$$

수식 2. 1 Euro Filter 기반 차단 주파수 조절

$$\hat{X}_i = \alpha X_i + (1 - \alpha) \hat{X}_{i-1}$$

수식 3. 지수 이동 평균(Exponential Moving Average)

수식 2와 같이 손의 이동 속도가 빠를수록 f_c 를 증가시켜 지연을 최소화하고, 정지 상태에서는 f_c 를 낮춰 고주파 노이즈를 효과적으로 제거한다. 이때 계산된 f_c 는 다시 평활화 계수 α 로 변환되어 수식 3의 지수 이동 평균 필터에 실시간으로 반영된다. 이를 통해 고사양 하드웨어 없이도 매우 수준의 부드러운 트레킹을 구현하였다.

4. 상태 관리 및 백엔드 최적화

단일 펀치 동작만으로 클릭과 드래그를 구분하기 위해 유한 상태 머신을 도입하였다. 시스템은 IDLE, PRESSED, DRAGGING 상태를 관리하며, 200ms의 임계 시간 내 펀치 해제 여부에 따라 클릭과 드래그 이벤트를 분기한다. 또한, 다수의 제스처 핸들러를 효율적으로 관리하기 위해 전략 패턴과 해시맵 기반의 라우팅 아키텍처를 구축하였다. 이를 통해 조건문 분기 없이 O(1)의 시간 복잡도로 핸들러를 호출함으로써 백엔드 연산 부하를 최소화하였다.

III. 실험 및 결과

본 논문의 핵심 기술인 입력 신호 안정화 성능을 검증하기 위해, 웹캠 센서로부터 수신된 원시 데이터와 적응형 1 Euro Filter를 거친 보정 데이터의 채적을 비교 분석하였다. 실험결과, 그림 2와 같이 필터 적용 전 관측되던 고주파 노이즈가 효과적으로 제거되어 매끄러운 곡선 채적을 그리는 것을 확인하였으며, 정지 상태에서의 떨림 억제와 빠른 움직임에서의 지연 최소화 사이의 상충 관계를 성공적으로 완화되었음을 입증하였다.

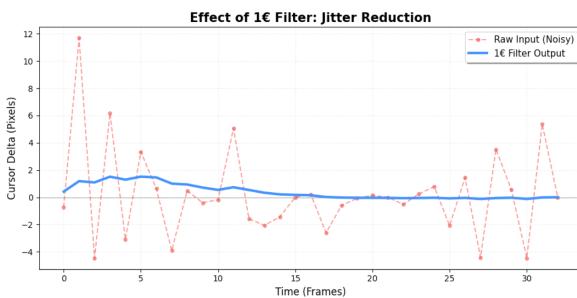


그림 3. 1 Euro Filter 적용 전후의 좌표 채적 비교

입력 신호 안정화 기법과 제스처 해석 로직이 결합된 제안 시스템의 실제 동작 특성을 확인하기 위해, 3D 지도 애플리케이션을 테스트 배드로 구축하여 연속 제어 동작을 실험하였다. 그림 4는 손가락 간 거리 변화에 기반한 확대·축소 제어 동작을 나타내며, 안정화된 입력 신호가 화면 스케일 변화로 자연스럽게 반영되는 과정을 보여준다. 또한 그림 5는 손의 상대적 위치 이동을 이용한 시점 이동 및 카메라 각도 조절 동작의 예로, 연속적인 손 움직임이 시점 변화로 매끄럽게 매핑됨을 확인할 수 있다. 아울러 드래그 및 포인터 제어와 같은 이벤트 기반 입력 동작 또한 안정적으로 구현되어, 제안하는 시스템이 다양한 연속 제어 및 이벤트 기반 제어를 안정적으로 처리할 수 있음을 확인하였다.



그림 4. 오른손 랜드마크 4, 8, 12번(P4, P8, P12)을 이용한 확대·축소

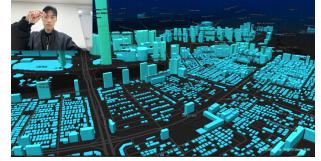


그림 5. 왼손 랜드마크 4, 8번(P4, P8)을 이용한 카메라 시점 변경

Resource / Mode	Standard Mouse Input	Gesture Interface (Our System)	Note
Frontend (Chrome)	110% (High)	110% (High)	<i>3D Rendering burden(Equal)</i>
Backend (Java)	N/A	< 1.2% (Very Low)	<i>Negligible logic processing overhead</i>
Sensor (Python)	N/A	About 41%	<i>Computer Vision(Essential Resources)</i>
Visual FPS	120FPS	120FPS	<i>Equal/Interpolated</i>
Error Rate	~0% (Reference)	< 2.3% (977/1000 Success)	<i>Measured over 1000 click/drag trials</i>

표 1. 표준 마우스 입력과 제안하는 제스처 인터페이스의 성능 비교

총 1,000회의 반복 제스처 테스트를 수행한 결과, 표 1과 같이 기하학적 정규화 및 히스테리시스 알고리즘을 통해 97.7%의 인식 성공률과 2.3% 미만의 오차율을 달성하였다. 아울러 데이터 센싱과 UI 랜더링을 분리하는 디커플링 아키텍처와 해시맵 기반의 백엔드 최적화를 통해 리소스 오버헤드를 1.2% 미만으로 최소화하였으며, 이를 통해 시각적 프레임을 120Hz로 유지하며 입력 지연 없는 실시간 제어 성능을 확보하였다.

IV. 결론

본 논문은 고비용 연산이나 특수 장비 없이, 웹캠과 경량 알고리즘만으로 마우스 수준의 정밀 제어가 가능한 이원화 아키텍처 기반의 인터페이스를 제안하였다. 이를 통해 기존 비접촉 시스템의 한계인 신호 불안정성과 연산 비효율성을 개선하고, 다양한 환경에 적용 가능한 범용성을 확보하였다. 향후 연구에서는 단일 RGB 센서의 구조적 한계인 폐색 현상을 보정하기 위해 기하학적 필터링을 고도화하고, 양손 제스처 도입을 통해 상호 작용의 범위를 확장할 계획이다.

ACKNOWLEDGMENT

본 결과물은 2025년도 교육부 및 충청남도의 지원으로 충남RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과임 (2025-RISE-12-003).

참고 문헌

- [1] S. Zhang, Y. Li, and H. Wang, "Deep vision based real time hand gesture recognition: a review," PeerJ Computer Science, vol. 11, Art. no. e1245, pp. 1–25, Jan. 2025.
- [2] C. Lugaressi et al., "MediaPipe: A Framework for Building Perception Pipelines," in Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR), LongBeach, CA, USA, 2019.
- [3] G. Casiez, N. Roussel, and D. Vogel, "1€ Filter: A Simple Speed-based Low-pass Filter for Noisy Input in Interactive Systems," in Proc.ACM CHI Conf. Hum. Factors Comput. Syst., Austin, TX, USA, May2012, pp. 2527 - 2530