

다양한 고수준 사용자 입력을 위한 잠재변수 매칭 기반 물리 캐릭터 제어 프레임워크*

이창호⁰, 이윤상
한양대학교 컴퓨터 소프트웨어학과
changholee@hanyang.ac.kr, yoonsanglee@hanyang.ac.kr

A Latent Matching Framework for Physics-Based Character Control from Diverse High-Level Inputs

Changho Lee⁰, Yoonsang Lee
Dept. of Computer Science, Hanyang University

요약

본 연구는 자연어, 목표 위치, 실시간 인터랙션 등 다양한 고수준 사용자 입력을 하나의 일관된 메커니즘으로 처리하여, 추가적인 저수준 모델 재학습 없이 물리 기반 캐릭터 제어에 연결할 수 있는 통합 제어 프레임워크를 제안한다. 서로 다른 형태의 고수준 입력을 공통된 의미적 중간 표현(semantic IR)으로 변환하고, 이를 Latent Matching [1] 기반 물리 캐릭터 제어와 결합하여 다양한 입력과 태스크에 대해 일반화 가능한 제어 방식을 제공한다. 또한 EMM [2]의 타원 기반 장애물 회피 기법을 GPU 배치 연산으로 구현하여 통합함으로써 장애물 환경을 고려한 안전한 컨트롤이 가능해지는 캐릭터 제어를 위한 실용적인 시스템을 구축한다.

1. 서론

물리 기반 환경에서 시뮬레이션 되는 휴머노이드 캐릭터를 제어하는 최근의 연구는 크게 두 흐름으로 나뉜다. 강화학습 기반 방법 [3, 4]은 정밀한 물리 제어를 가능하게 하지만, 특정 태스크 또는 입력 형태에 특화된 정책을 학습하기 때문에 새로운 태스크에 대응하려면 재학습이 필요하다는 한계가 있다. 반면 diffusion 기반 방법 [5, 6]은 cost gradient guidance를 통해 런타임에 새로운 태스크를 수행할 수 있는 유연성을 가지지만, 다양한 고수준 사용자 입력을 통합적으로 수용하는 구조를 갖추고 있지 않다.

한편, Latent Matching [1]은 CALM의 사전학습 모델을 활용하여 추가적인 고수준 정책 학습 없이 캐릭터를 제어할 수 있음을 보였다. 그러나 현재는 조이스틱과 같은 제한된 실시간 입력만 지원하며, 환경의 장애물을 고려하지 않는다는 한계가 있다.

본 연구는 다양한 고수준 입력을 통합적으로 처리할 수 있는 범용 제어 프레임워크를 구축함으로써, 기존 방법들이 해결하지 못한 입력 다양성과 시스템 확장성 문제를 해결하고자 한다. 서로 다른 형태의 고수준 입력(자연어, 목표 위치, 실시간 컨트롤러)을 공통된 의미적 중간 표현(Semantic Intermediate Representation)으로 변환하는 통합 입력 처리 파이프라인을 설계하고, EMM [2]의 타원 기반 장애물 회피 기법을 Latent Matching과 통합하여 환경을 고려한 안전한 경로 생성이 가능한 프레임워크를 제안한다.

2. 다양한 입력의 중간표현 변환

기존 잠재변수 매칭은 실시간 조이스틱 입력만을 지원하여, "테이블까지 걸어가" 같은 자연어 명령이나 "앞으로 10m 걸어가" 같은 거리 기반 명령을 처리할 수 없었다. 본 연구는 서로 다른 형태의 입력을 semantic IR로 변환하여 이 문제를 해결한다.

semantic IR은 복수 스테이지 구조로 설계된다. 각 스테이지는 동작 명령 (이동 방향, 바라볼 방향, 특정 위치로 이동), 종료 조건 (시간, 이동거리, 물체에 도달), 다음 스테이지로의 전환 규칙으로 구성된다. 예를 들어 "앞으로 10m 걸어갔다가 오른쪽으로 5초 동안 뛰어가"라는 명령은 스테이지 1: {이동 방향: {1, 0}, 속력: 0.3, 종료: 거리 10m}, 스테이지 2: {이동 방향: {0, -1}, 속력: 1.0, 종료: 시간 5초} 이러한 구조는 거리/시간 기반의 정량적 조건과 도착/근접 기반의 상태 조건을 유연하게 표현할 수 있다.

자연어 명령의 경우, LLM을 통해 자연어를 semantic IR로 변환하며 목표 위치 입력은 현재 위치와의 기하학적 관계로부터 필요한 바라보는 방향과 거리를 계산하여 semantic IR로 변환한다. 조이스틱이나 키보드 입력

* 구두 발표논문, 요약논문 (Extended Abstract)

* 본 논문은 요약논문 (Extended Abstract) 으로서, 본 논문의 원본 논문은 현재 타 학술대회(논문지)에 제출 준비중임.

* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업 (RS-2024-00399136) 및 과학기술정보통신부 및 정보통신기획평가원의 가상융합지능화핵심기술개발사업 (RS-2026-25508404)의 지원을 받아 수행되었음.

은 입력 축 값을 이동 방향과 바라볼 방향으로 직접 매핑한다. 이렇게 생성된 IR로부터 잠재변수 매칭에 사용될 쿼리 피처를 만들고, 잠재변수 매칭을 이용해 저수준 모델에 입력으로 주어질 잠재변수 z 를 얻는다. 이러한 설계를 통해 새로운 입력 형태가 추가되더라도 해당 입력을 IR로 변환하는 함수만 추가하면 되므로, 시스템의 나머지 부분은 변경 없이 높은 확장성을 유지할 수 있다.

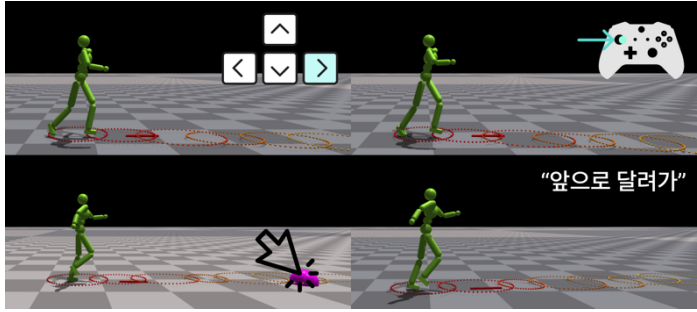


그림 1: 다양한 고수준 입력을 처리할 수 있는 컨트롤러.

3. 환경 인식을 통한 장애물 회피

본 연구는 Environment-aware Motion Matching [2]의 타원 기반 장애물 회피 기법을 우리의 물리 기반 시뮬레이션 환경으로 확장한다. 해당 연구에서 제안한 바와 같이, 일반적인 모션 매칭 피처에 미래 3시점 각각의 캐릭터 점유 타원 9차원을 추가해 장애물과의 미래 충돌 패널티를 계산한다. 각 타원의 장축은 이동 방향, 단축은 직교 방향으로 설정되며, 각 축의 반지름은 해당 시점에서 캐릭터의 관절을 투영한 최대 거리로 계산된다.

이렇게 확장된 피처를 활용하여 장애물을 회피하는 경로를 선택한다. 먼저 정적 피처 거리를 계산하고 각 후보에 대해 장애물 패널티를 적용한다. 타원과 장애물 간

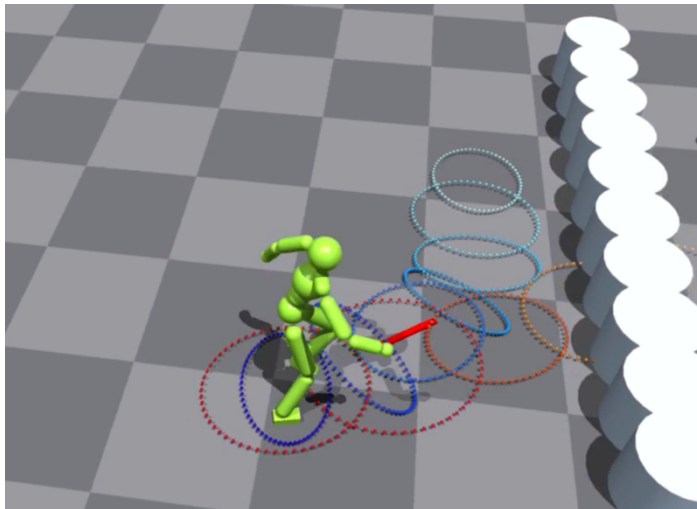


그림 2: 장애물 회피 적용. 충돌 패널티가 없을 때의 미래 궤적(빨간색)과 비교해 충돌 패널티를 적용시켰을 때 미래 궤적(파란색)이 장애물과의 충돌을 최소화함을 보인다.

거리가 임계값보다 가까우면 큰 패널티를 부여하여 해당 경로를 회피하도록 한다. 이후 상위 후보들에 대해 전체 피처 거리를 계산하여 최종 잠재변수를 선택한다. 또한 장애물 근접도와 목표 속도에 따라 미래 궤적의 가중치를 동적으로 조정하여 상황에 적합한 회피 동작을 생성한다.

4. 결론

본 연구는 다양한 고수준 사용자 입력(자연어, 목표 위치, 실시간 컨트롤러)을 공통된 의미적 중간 표현 (semantic IR)으로 통합하는 프레임워크를 제안했다. 기존 강화학습 기반 방법이 특정 태스크에 특화되어 재학습이 필요한 반면, 본 프레임워크는 추가적인 저수준 모델 재학습 없이 사전학습된 Latent Matching만으로 다양한 입력과 태스크를 처리한다. [2]의 장애물 회피 기법을 통합함으로써 실제 환경에서도 안전한 경로 생성이 가능하며, 입력 형태가 변화하더라도 변환 함수만 추가하면 되므로 높은 확장성을 유지한다.

실험 결과, 본 방법은 다양한 입력 형태를 semantic IR로 변환하여 일관되게 처리할 수 있었으며, 장애물을 자연스럽게 우회하며 목표에 도달했다.

향후 연구로는 복수 원 표현을 통한 정밀한 장애물 모델링, 정량적 성능 평가(경로 효율, 충돌률), LLM 기반 IR 생성의 신뢰성 향상, 그리고 더 복잡한 조건부 동작과 의도 표현을 위한 semantic IR 확장을 계획하고 있다. 또한 본 프레임워크는 게임, 메타버스, 로봇틱스 등 다양한 분야의 인터랙티브 캐릭터 제어 시스템에 적용될 수 있을 것으로 기대된다.

참고문헌

- [1] Minsu Kim, Yoonsang Lee, Latent Matching for Simulated Humanoid Control, KCGS 2024.
- [2] Jose Luis Ponton, Sheldon Andrews, Carlos Andujar, Nuria Pelechano, Environment-aware Motion Matching, SIGGRAPH Asia 2025
- [3] Tessler, C., Kasten, Y., Guo, Y., Mannor, S., Chechik, G., & Peng, X. B, CALM: Conditional Adversarial Latent Models for Directable Virtual Characters, SIGGRAPH 2023
- [4] Tessler, C., Guo, Y., Nabati, O., Chechik, G., & Peng, X. B, MaskedMimic: Unified Physics-Based Character Control Through Masked Motion Inpainting, SIGGRAPH Asia 2024
- [5] Huang, X., Truong, T., Zhang, Y., Yu, F., Sleiman, J. P., Hodgins, J., Sreenath, K., & Farshidian, F, Diffuse-CLoC: Guided Diffusion for Physics-based Character Look-ahead Control, ACM Transactions on Graphics (TOG), 2025
- [6] Yuan, Y., Song, J., Iqbal, U., Vahdat, A., & Kautz, J, PhysDiff: Physics-Guided Human Motion Diffusion Model, ICCV 2023