

논문 2025-4-13 <http://dx.doi.org/10.29056/jsav.2025.12.13>

생성형 AI와 키프레임 방식의 캐릭터 애니메이션 연기 표현 비교 연구 - 비언어적 연기 요소를 중심으로

박헌진*†

A Comparative Study on Character Animation Acting Expression between Generative AI and Keyframe Methods - Focusing on Non-verbal Acting Elements

Hunjin Park*†

요 약

본 연구는 생성형 AI 기술이 영상 제작의 효율성을 향상시켰음에도 불구하고, 캐릭터 애니메이션의 핵심인 '연기' 영역에서는 인간 애니메이터를 대체하기 어려운 한계를 지니고 있음을 고찰한다. 3D 키프레임 애니메이션은 레퍼런스를 분석하고 정제하여 불필요한 요소를 제거하고 의미를 명확히 전달하는 단순함과 명료성의 예술이다. 반면 생성형 AI는 확률적 데이터 조합에 의존하므로 서사적 맥락과 캐릭터의 의도를 이해하지 못하며, 제스처와 서브텍스트 표현에서 패턴화된 결과만을 생성한다. 본 논문은 비언어적 연기 요소를 체계적으로 분류하고, 키프레임 애니메이션과 AI 생성 영상을 비교 분석하여 AI 영상이 시각적 사실성은 확보했으나 연기적 호소력을 구현하지 못하는 원인을 규명한다. 향후 AI는 제작 보조 도구로 활용되며, 최종적인 연기 디렉팅은 인간 애니메이터의 고유 영역으로 남을 것을 제안한다.

Abstract

This study examines the limitations of generative AI in character animation's core aspect of 'acting', despite its efficiency improvements. 3D keyframe animation is an art of simplicity and clarity that refines references to convey meaning precisely. In contrast, generative AI relies on probabilistic data combinations, failing to understand narrative context and character intent, producing only patterned gestures and subtext. This paper systematically classifies non-verbal acting elements and compares keyframe animation with AI-generated videos to identify why AI achieves visual realism but lacks acting appeal. It proposes that AI will serve as a production support tool, while final acting direction remains the exclusive domain of human animators.

한글키워드 : 시각콘텐츠, 생성형 인공지능, 캐릭터 애니메이션, 제스처, 컴퓨터그래픽

keywords : Visual contents, Generative AI, Animation, Gesture, Computer Graphic

* 중부대학교 문화콘텐츠학부

† 교신저자: 박헌진(email: hunjinn@gmail.com)

접수일자: 2025.12.01. 심사완료: 2025.12.11.

게재확정: 2025.12.20.

1. 서론

인공지능(Artificial Intelligence)은 지난 수년간 텍스트와 정지 이미지의 영역에서 인간의 고유

영역이라 믿어졌던 창의성의 벽을 허물었다. 대규모 언어 모델(LLM)은 코딩을 작성하고 철학적인 대화를 나누며, 미드저니(Midjourney)와 같은 이미지 생성 AI는 붓 터치 하나 없이도 정교한 화풍을 모방[1]해낸다. 이러한 흐름 속에서 등장한 소라(Sora), 런웨이(Runway) 등의 영상 생성 AI(Video Generative AI)는 텍스트 프롬프트 한 줄로 고품질의 영상을 만들어 내며, 애니메이션 및 영화 산업 전반에 제작 공정의 혁신과 인간 창작자의 위기라는 양가적인 충격을 던지고 있다.

그러나 움직이는 이미지(Motion Picture)를 생성하는 기술적 성취와, 관객의 공감을 이끌어내는 살아있는 캐릭터를 창조하는 예술적 성취는 별개의 문제이다. 현재의 생성형 AI는 언어적 구조와 시각적 묘사력에서는 비약적인 발전을 이루었으나, 시간의 흐름 속에서 캐릭터의 내면을 드러내는 연기(Acting)의 영역에서는 여전히 답보 상태에 머물러 있다. 특히 애니메이션은 현실을 그대로 베끼는 것이 아니라, 애니메이터의 해석을 통해 동작을 디자인하고 생명력을 불어넣는 예술이다. AI가 만들어 낸 영상 속 캐릭터는 얼핏 보기에 유려하게 움직이는 듯하지만, 그 움직임에는 서사적 맥락(Context)과 캐릭터의 의도(Intent)가 결여되어 있다.

본 연구는 생성형 AI의 알고리즘적 원리를 공학적으로 분석하는 대신, 영화와 애니메이션 연출의 관점에서 AI 영상의 미학적 한계를 고찰하고자 한다. 연구의 초점은 캐릭터 연기의 핵심인 비언어적 커뮤니케이션에 맞춘다.

현시점에서 Google의 Veo 3.1, Open AI의 Sora2, Kling 2.1등 수많은 생성형 인공지능 모델들이 경쟁적으로 서비스 하고 있다. 각 모델들은 조금씩 차별점을 가지고 있으나 캐릭터 애니메이션의 미학적 관점에서 공통적인 문제점을 보이고 있다. 이 연구에서는 각 모델별 차별점에 따른 결과는 다루지 않는다. 이 연구에서는 고품질의

이미지 제작에 특화된 미드저니(Midjourney)와 Google Gemini (Image Generation Model) 그리고 Veo 3.1을 사용해 생성형 인공지능이 만들어 내는 캐릭터 애니메이션의 공통적인 결과를 고찰하였다.

알리바바의 Wan 2.2와 Wonder Dynamics의 VFX 솔루션 같이 원본 소스(live-action footage)를 입력하여 조금 더 정확한 동작을 제어하는 방법도 존재하지만 그런 방식은 정제된 디자인이라기보다 모션캡처나 로토스코핑에 가까운 제작 방식이므로 연구 대상에서 제외했다.

애니메이션과 영화 같은 시각 예술 분야에서는 언어적 정보보다 화면의 구도, 미장센, 그리고 캐릭터의 표정, 제스처, 호흡, 시선 처리와 같은 비언어적[2] 신호가 서사 전달에 있어 더욱 결정적인 역할을 수행한다. 본 논문은 숙련된 애니메이터가 키프레임(Keyframe) 방식으로 캐릭터를 연기시킬 때 사용하는 비언어적 요소들을 체계화하여 표로 정리하고, 이를 기준으로 생성형 AI가 만들어 낸 결과물을 비교 분석한다. 이를 통해 AI 영상이 시각적 리얼리즘(Realism)을 확보했음에도 불구하고 왜 연기적 불쾌한 골짜기(Performance Uncanny Valley)를 넘지 못하는지를 규명하고, 향후 애니메이션 제작 환경에서 인간과 AI의 바람직한 공존 방식을 제안하는 데 목적이 있다.

2. 3D 캐릭터 애니메이션의 연기 미학: 단순함과 명료성

2.1. 레퍼런스 기반의 연기 검증과 그래픽적 정제 (Refining)

3D 컴퓨터 애니메이션[2]이 2D 셀 애니메이션을 대체하며 주류로 자리 잡은 이후, 애니메이터의 역할은 그림을 그리는 화가에서 캐릭터를 조

중하는 디지털 배우로 확장되었다. 그러나 이것이 애니메이터가 실제 배우의 메소드 연기(Method Acting)를 그대로 답습한다는 의미는 아니다. 오히려 애니메이션 연기의 핵심은 실제 연기를 수행해 본 뒤, 그것을 철저히 분석하고 정제하는 과정에 있다.

애니메이터는 작업을 시작하기 전, 직접 거울을 보거나 영상을 촬영하며 레퍼런스(Reference)를 만든다. 이 과정에서 애니메이터는 신체 역학(Body Mechanics)과 감정의 흐름을 직접 체험하며 동작의 타당성을 검증한다. 하지만 촬영된 레퍼런스 영상은 그 자체로 애니메이션이 될 수 없다. 실제 사람의 움직임에는 불필요한 무게 중심의 이동, 무의미한 손가락의 떨림, 옷의 지저분한 구겨짐 등 시각적 노이즈가 포함되어 있기 때문이다.

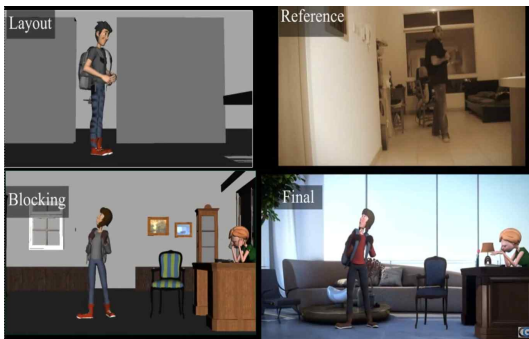


그림 1. 키프레임 애니메이션 제작과정
Fig. 1. Keyframe Animation Production Process

그림 1의 이미지는 애니메이터 스스로가 직접 연기한 레퍼런스 영상을 분석하고 정제해 키포즈와 타이밍을 중심으로 Blocking 과정을 수행한 후, 마지막 Polishing 단계에서 미묘한 손동작이나 표정 같은 디테일한 제스처를 다듬는 과정을 보여주고 있다. 애니메이션은 이러한 정제되지 않은 데이터에서 핵심이 되는 키 포즈(Key Pose)만을 추출하고, 나머지는 과감하게 생략하

거나 과장(Exaggeration)하여 그래픽적인 명료함을 부여하는 예술이다. 관객의 입장에서 볼 때, 캐릭터의 감정과 의도는 0.1초의 짧은 순간에도 명확하게 읽혀야 한다(Readability)[6]. 따라서 애니메이터는 그래프 에디터(Graph Editor)를 통해 움직임의 곡선을 매끄럽게 다듬고, 캐릭터의 실루엣을 단순화하여 시각적 정보를 디자인한다. 즉, 3D 애니메이션의 연기는 사실적 재현이 아닌, 철저히 계산되고 디자인된 이상적 움직임을 추구한다.

2.2. 생성형 AI 영상의 생성 원리와 맥락 부재

반면, 현재 주류를 이루는 영상 생성 AI의 핵심 기술인 확산 모델(Diffusion Model)은 본질적으로 다른 메커니즘[7]을 갖는다. 이 모델은 무작위적인 노이즈(Noise) 상태에서 학습된 데이터의 패턴을 찾아 점진적으로 이미지를 복원(Denoising)하는 방식으로 작동한다. 이 과정은 철저히 확률적 통계에 기반한다. AI는 수억 장의 이미지와 영상 데이터에서 픽셀 간의 연관성을 학습했을 뿐, 그 영상 속에 담긴 인물의 인과관계나 심리 상태를 이해하는 것은 아니다.

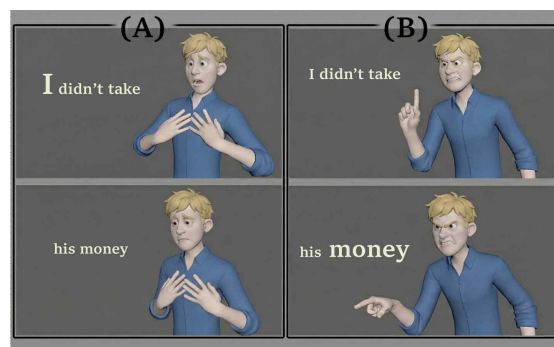


그림 2. 상황별 제스처와 키포즈 디자인
Fig. 2. Context-Specific Gesture and Key Pose Design

이러한 생성 방식은 맥락(Context)의 부재라는

결정적인 한계를 낳는다. 그림 2와 같이 애니메이터는 “I didn’t take his money”라는 대사를 두고 고도 (A)와 (B)의 분명한 제스처와 다른 서브텍스트(Subtext)를 디자인할 수 있다. 애니메이터는 이야기와 상황의 맥락을 고려해 다양한 제스처를 적절히 필요한 장소에 정확히 만들지만, AI는 단순히 무의미한 손동작이나 표정으로 제스처를 확률적으로 만들어 낸다.

그림 3은 프롬프트(Medium shot. A young woman sitting on a cozy sofa, holding an open book. She looks extremely bored and sleepy, her eyelids are heavy and drooping, and she is slouching. The lighting is warm and soft. Despite her sleepy expression, she speaks clearly. Dialogue: ‘interesting’. Realistic, cinematic style.)와 여성 이미지를 이용해 인공지능(Veo 3.1)으로 생성한 결과물이다. 영상을 살펴보면, 확률적으로 비슷한 무드는 만들어 내지만 의미 없는 대사와 눈동자의 움직임이 발견된다. 이는 한편으로 단순한 Text to Video나 Image to Video 방식으로는 정확한 제스처를 컨트롤하는 것이 효과적이지 않다는 것을 말해준다.

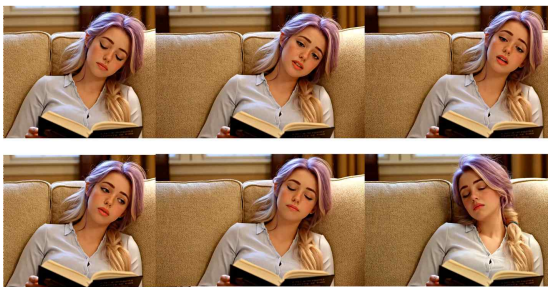


그림 3. 인공지능으로 생성된 영상 (Vevo 3.1)
Fig. 3. AI-generated Video (Vevo 3.1)

2.3. 비언어적(Non-verbal) 연기 요소의 분류
애니메이터가 캐릭터의 생명력을 구현하기 위해 통제하는 비언어적 요소는 매우 광범위하다.

본 연구는 제스처를 분류[8]해 분석의 틀을 마련했다. 이는 후술할 AI 영상의 문제점을 분석할 적절한 기준점을 제공한다.

표 1. 제스처 분류

Table 1. Classification of Non-verbal Gestures

| 분류 | 설명 |
|---------------------|------------------------------------|
| Incidental gestures | 본래 목적은 다른데, 그 동작이 부수적으로 신호가 되는 제스처 |
| Expressive gestures | 감정·정서 상태를 직접적으로 드러내는 생물학적 제스처 |
| Mimic gestures | 무엇인가를 그대로 흉내 내어 의미를 전달하는 모방 제스처 |
| Schematic gestures | 모방 제스처를 단순화·축약한 도식적 제스처 |
| Symbolic gestures | 사회적으로 합의된 상징을 통해 아이디어·기분을 나타내는 제스처 |
| Technical gestures | 특정 직업·기술·전문 분야에서 사용하는 제스처 |

표 1에서 제시된 요소들은 모두 ‘맥락’에 종속된다. 애니메이터는 “지금 캐릭터가 화가 났는가?”를 넘어 “왜 화가 났으며, 그것을 숨기려 하는가, 드러내려 하는가?”라는 질문을 던지며 이 요소들을 조합한다.

3. 생성형 AI와 키프레임 애니메이션의 연기 구현 비교 분석

3.1. 연출된 명료함(Staged Clarity) vs 확률적 사실성(Probabilistic Realism)



그림 4. 실루엣 디자인
Fig. 4. Silhouette design

애니메이션 12원칙 중 스테이징(Staging)과 호소력(Appeal)[9]은 복잡한 상황을 단순 명료하게 전달하여 관객을 몰입시키는 핵심 기법이다. 훌륭한 3D 애니메이션은 캐릭터의 실루엣(Silhouette)만 보더라도 그가 현재 어떤 감정 상태인지, 다음 동작이 무엇일지 예측 가능해야 한다. 그림 4와 같이 이를 위해 애니메이터는 캐릭터의 포즈를 디자인할 때 네거티브 스페이스(Negative Space)를 확보하고, 시선의 흐름을 방해하는 요소를 제거한다. 이는 그래픽 디자인의 조형미를 캐릭터 연기에 적용한 것이다.

그러나 그림 5와 같이 생성형 AI가 만든 영상은 높은 렌더링 퀄리티와 사실적인 질감을 자랑함에도 불구하고, 이러한 가독성 측면에서 실패하는 경우가 빈번하다. AI는 화면의 주연과 조연, 배경을 위계적으로 구분하지 않고 동등한 수준의 디테일로 묘사하려 든다. 이로 인해 관객의 시선은 캐릭터의 눈이나 손 같은 중요 연기 포인트에 집중되지 못하고 분산된다.



그림 5. 생성형 AI로 생성한 캐릭터 키포즈와 실루엣 예시

Fig. 5. Character Key Poses and Silhouette Examples Generated by AI

또한, 시간적 일관성(Temporal Consistency)의 부족은 연기의 명료함을 해치는 주범이다. 키프레임 애니메이션은 리깅(Rigging)된 고정된 뼈대를 기반으로 움직이기 때문에 캐릭터의 부피감과

형태가 완벽하게 유지된다. 반면 AI 영상은 프레임이 진행될수록 캐릭터의 이목구비 위치가 미세하게 미끄러지거나(Sliding), 표정 근육이 액체처럼 흐르는 현상이 발생한다. 이러한 미세한 불쾌감은 관객이 캐릭터를 살아있는 존재가 아닌 그래픽 덩어리로 인식하게 만든다. 이는 애니메이션이 가장 경계하는 군더더기이자, 몰입을 깨트리는 요소이다.

3.2. 맥락 기반의 서브텍스트(Subtext) 구현 vs 패턴화된 감정 표현

캐릭터 액팅의 정점은 겉으로 드러난 대사나 행동(Text) 이면에 숨겨진 속마음, 즉 서브텍스트(Subtext)[10]를 표현하는 데 있다. 인간 애니메이터는 캐릭터가 거짓말을 할 때 입으로는 웃고 있지만 눈썹은 미세하게 찌푸리게 하거나, 대답하기 직전 찰나의 침묵(Pause)을 주어 갈등하는 내면을 묘사한다. 또한, 행동하기 전에 생각하는 과정(Thought Process)을 보여주기 위해 눈동자를 굴리는(Eye Darting) 타이밍을 프레임 단위로 조절한다. 이러한 디테일은 철저히 시나리오의 맥락과 캐릭터의 성격을 분석한 결과이다.



그림 6. 프롬프트: 웃고 있는 소녀(미드저니)

Fig. 6. Prompt: A girl with a smile(Midjourney)

이에 반해 생성형 AI는 텍스트 프롬프트에 의존하여 1차원적인 감정만을 출력한다. 예를 들어 그림 6과 같이 생성형 AI 미드저니에서 "행복하

게 웃고 있는 소녀"라는 프롬프트를 입력하면, AI는 학습 데이터 중 가장 일반적인 Expressive gesture인 '활짝 웃는 표정'을 생성한다. 하지만 해당 장면이 '이별을 앞두고 애써 웃음 짓는' 복합적인 상황이라면, 그림 7과 같이 AI의 '패턴화된 미소'는 상황에 맞지 않는 기이한 연기가 된다.

AI는 감정의 빌드업(Build-up) 과정을 이해하지 못한다. 애니메이션은 감정이 서서히 고조되어 클라이맥스에 다다르는 리듬(Timing & Spacing)이 중요하다. 그러나 AI 영상은 처음부터 끝까지 일정한 텐션으로 움직이거나, 뜬금없는 타이밍에 급격한 표정 변화를 일으킨다. 이는 AI가 현재 프레임과 이전 프레임의 연관성은 계산하지만, 영상 전체를 관통하는 '서사의 흐름'을 인지할 능력이 없음을 시사한다. 결국 서브텍스트가 없는 연기는 영혼 없는 인형의 움직임과 다를 바 없다.

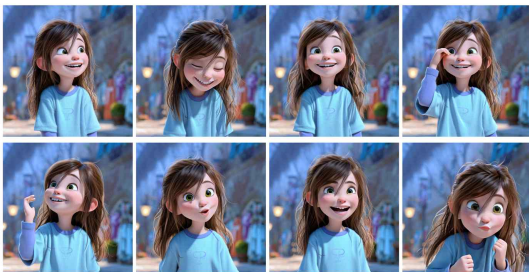


그림 7. 프롬프트: 이별을 앞두고 애써 웃음 짓는 표정과 제스처(미드저니)

Fig. 7. Prompt: A girl forcing a smile and posture of feigned cheerfulness while facing an imminent farewell, subtle sadness in the eyes(Midjourney)

4. 결론

본 연구는 생성형 AI 기술이 영상 제작의 효

율을 비약적으로 높였음에도 불구하고, 애니메이션의 본질인 '캐릭터 연기' 영역에서는 여전히 인간 애니메이터를 대체하기 어려운 한계를 지니고 있음을 고찰하였다. 생성형 인공지능을 이용해 원하는 움직임을 어느 정도 목표하는 근사치에 가깝게 만들 수는 있겠지만 만족스러운 결과물이 나올 때까지 수없이 반복하는 과정은 분명히 비효율적이고 실제 제작에 쓰이기에는 문제가 있다. 3D 키프레임 애니메이션은 실제의 모방이 아닌, 애니메이터의 심미적 판단에 의해 불필요한 요소를 제거하고 의미를 명확히 전달하는 '단순함과 명료성(Clarify)'의 예술이다. 애니메이터는 레퍼런스를 분석하고 정제하여 관객이 캐릭터의 의도를 직관적으로 이해할 수 있도록 그래픽을 디자인한다. 반면, 생성형 AI는 데이터의 확률적 조합에 의존하므로 서사적 맥락을 이해하지 못한다. 이로 인해 맥락에 맞는 적절한 제스처를 표현하지 못하고, 시각적 노이즈와 무의미한 디테일을 양산하여 연기의 가독성을 떨어뜨린다. 결론적으로 AI가 만들어 내는 영상은 사실적(Realistic)일 수는 있으나, 애니메이션이 추구하는 호소력(Appealing) 있는 연기와는 거리가 멀다.

본 연구의 비교 분석을 통해 확인한 바와 같이, 생성형 AI는 물리적인 움직임은 모방할 수 있으나 그 움직임을 발생시키는 동인(Motivation), 즉 '심리'와 '맥락'은 모방하지 못하고 있다. 이는 현재의 AI 학습 방식이 겉으로 드러난 픽셀 데이터의 패턴 인식에 머물러 있기 때문이다. 사람의 미묘한 연기를 AI에게 가르치려면, 단순히 표정과 포즈 데이터의 양을 늘리는 것만으로는 불가능하다. 맥락(Context) + 비언어 신호(Non-verbal Cues) + 의도/심리 상태(Intent)가 모두 상세히 주석(Annotation) 처리된 고품질의 멀티모달(Multi-modal) 데이터셋이 필요하다. 예를 들어, 특정 표정이 어떤 사건 이후

에 나왔는지, 숨겨진 의도가 무엇인지를 AI가 인지할 수 있도록 학습 목표(Objective Function)를 재설계해야 한다. 이는 공학자뿐만 아니라 연기 이론가, 심리학자, 예술가들의 협업이 필요한 영역이다.

생성형 인공지능의 향후 기술의 발전 속도를 고려할 때, AI가 애니메이션 제작 파이프라인의 상당 부분을 변화시킬 것은 자명하다. 배경 텍스처 생성, 군중 시뮬레이션, 혹은 초기 아이디어를 시각화하는 프리비즈(Pre-visualization) 단계에서 AI는 강력한 도구가 될 것이다.

캐릭터의 생명력을 불어넣는 ‘주연 배우’로서의 역할은 여전히 인간의 영역으로 남을 것이다. 기술이 고도화될수록, 역설적으로 쏟아지는 무작위 데이터 속에서 가장 적합한 표정과 타이밍을 골라내는 인간 애니메이터의 ‘선택하는 능력(Selective Eye)’과 ‘덜어내는 미학(Less is More)’은 더욱 중요한 가치를 지니게 될 것이다. 미래의 애니메이터는 단순한 동작 생성 기술자를 넘어, AI를 도구로 활용하되 최종적인 연기의 뉘앙스와 그래픽적 명료함을 결정하고 통제하는 ‘총괄 연기 디렉터’로서 그 역할과 역량을 확장해야 한다. 인공지능은 흉내 낼 수 없는 ‘의도’와 ‘영혼’을 불어넣는 작업이야말로 애니메이션이 기술을 넘어 예술로 남을 수 있는 유일한 길이기 때문이다.

감사의 글

이 논문은 2025년도 중부대학교 학술연구비 지원에 의하여 이루어진 것임

참고 문헌

- [1] S. Vinothkumar, S. Varadhaganapathy, R. Shanthakumari, S. Dhanushya, S. Guhan and P. Krisvanth, “Utilizing Generative AI for Text-to-Image Generation”, 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-6, 2024. DOI: 10.1109/ICCCNT61001.2024.10725454
- [2] R. L. Birdwhistell, “Kinesics and Context: Essays on Body Motion Communication”, University of Pennsylvania Press, 1970, ISBN 978-0-8122-1012-5
- [3] George Maestri, “Digital Character Animation 3”, New Riders, 2006, ISBN 978-0-321-37600-6
- [4] Andy Beane, “3D Animation Essentials”, John Wiley & Sons, 2012, ISBN 978-1-118-14748-1
- [5] Frank Thomas and Ollie Johnston, “The Illusion of Life: Disney Animation”, Disney Editions, 1995, ISBN 978-0-7868-6070-8
- [6] Harold Whitaker, John Halas, and Tom Sito, “Timing for Animation”, CRC Press, 2022, ISBN 978-0-367-52778-3
- [7] Jonathan Ho, Ajay Jain, and Pieter Abbeel, “Denoising Diffusion Probabilistic Models”, Advances in Neural Information Processing Systems (NeurIPS), Vol. 33, pp. 6840-6851, 2020, URL: <https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>
- [8] Desmond Morris, “Manwatching: A Field Guide to Human Behavior”, Harry N. Abrams, 1977, ISBN 978-0-8109-1310-3
- [9] Tosiya L. Kunii and Takao Maeda, “On the silhouette cartoon animation”, Proceedings Computer Animation '96, pp. 110-117, June 1996, DOI: 10.1109/CA.1996.540493
- [10] Eric Goldberg, “Character Animation Crash Course!”, Silman-James Press, 2008, ISBN 978-1-879505-97-1

————— 저 자 소 개 —————



박헌진(Hunjin Park)

2002.2 홍익대학교 미술대학 졸업

2007.2 Academy of Art University 석사

2016.6-현재 : 중부대학교 만화애니메이션
전공 교수

<주관심분야> 3D 애니메이션, 애니메이션
제작, 문화콘텐츠