특집논문 (Special Paper)
방송공학회논문지 제31권 제4호, 2026년 7월 (JBE Vol.31, No.4, July 2026)
https://doi.org/10.5909/JBE.2026.31.4.700
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

생성형 AI 영상의 시간 구성 방식에 관한 사례 연구 - Kling 기반 바운싱 볼 시퀀스 분석을 중심으로 -

김 소 영[a)‡]

An Exploratory Case Study on Temporal Construction in Generative AI Video: Focusing on Kling-Based Bouncing Ball Sequences

Soyoung Kim[a)‡]

요 약

본 연구는 전통적 애니메이션의 프레임 기반 시간 구성 원리를 이론적 비교 기준으로 삼고, 생성형 AI 영상에서 시간적 사건이 어떻게 형성되고 변동되는지를 Kling 기반 바운싱 볼 시퀀스 사례를 통해 탐색적으로 분석하였다. 전통적 애니메이션은 프레임 단위를 통해 시간을 외부에서 설계하는 반면, 생성형 AI 영상은 시공간 데이터를 학습한 모델의 확률적 연산 과정 속에서 시간적 연속성이 형성된다. 동일한 프롬프트로 생성한 AI 영상 사례 분석 결과, 바닥 접촉 시점과 반동 양상이 샘플별로 다르게 나타나 시간 구조의 변동성이 확인되었다. 이를 통해 본 연구는 생성형 AI 영상의 시간성을 '생성적 시간성'으로 개념화하였다. 여기서 생성적 시간성은 단순한 결과물의 변동성이 아니라, 시간적 사건의 발생 시점, 간격, 지속, 운동 궤적이 창작자의 프레임 단위 설계가 아닌 모델의 확률적 생성 과정에서 형성되는 성질을 의미한다.

Abstract

This study examines, through an exploratory case analysis of Kling-based bouncing ball sequences, how temporal events are formed and vary in generative AI video, using the frame-based temporal construction of traditional animation as a theoretical reference point. Traditional animation externally designs time through the segmentation and arrangement of frames, whereas generative AI video forms temporal continuity through probabilistic computations based on spatiotemporal data. In particular, an analysis of AI-generated video cases created using identical prompts revealed variations in ground-contact timing and rebound patterns across outputs, confirming the variability of temporal structure. Based on these findings, this study defines generative temporality as an analytical concept referring to the way in which the timing, interval, duration, and trajectory of temporal events are formed through the probabilistic generation process of generative AI video, rather than being directly fixed through frame-by-frame design.

Keyword : Generative AI Video, Generative Temporality, Temporal Construction, Frame-based Animation, Temporal Event

Ⅰ. 서 론

최근 생성형 AI의 발전은 디지털 영상 제작 방식에 구조적인 변화를 야기하고 있다. 텍스트 프롬프트를 기반으로 시간 축을 포함한 영상 시퀀스를 생성하는 기술은 기존의 촬영 기반 영상 제작 방식뿐만 아니라, 프레임(frame)을 기본 단위로 설정하고 그 배열을 통해 움직임을 형성하는 애니메이션 제작 방식과도 구별되는 생성 원리를 갖는다. 특히 이러한 방식은 개별 프레임을 직접 설계하거나 배열하는 과정 없이 시각적 연속성을 형성한다는 점에서 기존 영상 제작 방식과는 다른 생성 원리를 보인다.

영상 매체의 역사에서 이러한 구조적 변화는 필름에서 디지털로의 전환 과정에서도 확인된 바 있다. 필름 기반 영상이 현실을 프레임 단위로 기록하는 방식이었다면, 디지털 환경에서는 이미지가 데이터로 변환되어 저장, 가공, 변형이 가능한 구조로 변화하였다. 이에 따라 프레임은 물리적 흔적이 아닌 정보 단위로 전환되었으며, 영상 제작 방식 역시 기록에서 처리로 그 중심이 이동하였다. 그러나 이러한 변화에도 불구하고 프레임은 여전히 영상 구성의 기본 단위로 기능해 왔다. 특히 애니메이션은 프레임 단위의 구성 원리를 가장 명확하게 드러내는 매체로 이해되어 왔다. 애니메이션은 정지된 이미지들을 시간의 흐름에 따라 배열함으로써 움직임을 구성하는 방식에 기반하며, 이는 곧 움직임이 시간적으로 어떻게 조직되는가의 문제와 직결된다. 다시 말해, 애니메이션은 단순히 움직임을 재현하는 매체가 아니라, 분절된 상태들을 배열함으로써 시간과 운동을 구성하는 매체로 이해될 수 있다.

그러나 생성형 AI 기반 영상은 이러한 프레임 중심의 구성 원리와는 다른 방식으로 작동한다. 생성형 AI는 개별 프레임을 설계하거나 배열하는 과정을 거치지 않고, 시공간적 데이터를 학습한 모델이 확률적 연산 과정을 통해 영상 시퀀스를 생성한다. 이 과정에서 시간적 연속성은 사전에 설계되는 것이 아니라 생성 과정 속에서 형성되는 결과로 나타난다. 즉, 프레임은 창작의 기본 단위가 아니라, 생성된 결과를 기술적으로 분절하는 출력 형식으로 기능하게 된다.

이러한 변화는 영상에서 시간성이 어떻게 구성되는가에 대한 문제를 새롭게 제기한다. 전통적 애니메이션에서 시간은 프레임의 분절과 배열을 통해 외부에서 조직되는 구조를 가지는 반면, 생성형 AI 영상에서는 시간적 연속성이 모델 내부의 연산 과정 속에서 형성된다. 이는 두 매체가 모두 움직임을 제시하지만, 그 시간 구성 원리는 서로 다를 수 있음을 보여준다.

따라서 본 연구는 전통적 애니메이션의 프레임 기반 시간 구성 원리를 이론적 비교 기준으로 삼고, 생성형 AI 영상에서 시간적 사건이 어떠한 방식으로 형성되고 변동되는지를 탐색적 사례 분석을 통해 검토하고자 한다. 본 연구에서는 동일한 텍스트 프롬프트와 기본 생성 조건에서 짧은 영상을 반복 생성할 수 있는 상용 생성형 AI 영상 모델인 Kling을 사례 모델로 선정하였다. 이후 Kling으로 생성한 바운싱 볼 시퀀스를 분석하여 동일한 운동 과제에서도 바닥 접촉 시점과 반동 양상이 어떻게 달라지는지를 검토하고, 이를 생성형 AI 영상의 ‘생성적 시간성’ 개념을 통해 설명하고자 한다.

Ⅱ. 이론적 배경

1. 애니메이션의 본질: 프레임 기반 시간 구성

애니메이션은 전통적으로 정지된 이미지를 시간적으로 배열함으로써 움직임의 환영을 생성하는 시각 매체로 정의되어 왔다. 이러한 정의는 단순한 시각적 착시에 기반한 것이 아니라, 이산적인 이미지들 간의 관계를 통해 연속적인 운동을 구성하는 지각적, 시간적 구조에 기반한다. 즉, 애니메이션은 연속적인 움직임을 기록하는 매체가 아니라, 분절된 상태들을 배열함으로써 움직임을 구성하는 매체이다.

Wells는 애니메이션을 직접 촬영된 움직임이 아닌, 프레임 단위의 조작을 통해 구성된 인공적 운동으로 정의하며, 애니메이션의 본질을 ‘움직임의 생성 과정’에 두고 있다[1].

a) 경일대학교 만화애니메이션학부(School of Manhwa&Animation, Kyungil University)
‡ Corresponding Author : 김소영(Soyoung Kim)
E-mail: yollha@naver.com
Tel: +82-53-600-5315
ORCID: https://orcid.org/0009-0006-2181-8639
· Manuscript received July 6, 2026; Revised July 9, 2026; Accepted July 10, 2026.

또한 Wells에 따르면 McLaren은 애니메이션을 “움직이는 그림이 아니라, 그려진 움직임의 예술”로 규정하며, 프레임 간 변화 자체가 애니메이션의 본질임을 강조한다[1]. 이는 애니메이션이 개별 이미지의 집합이 아니라, 이미지 간 변화와 관계를 설계하는 시간적 매체임을 시사한다. 이는 프레임(frame)을 기본 단위로 하는 창작 방식에서 가장 명확하게 드러난다. 하나의 프레임은 특정 시점의 정지 이미지를 의미하며, 이들이 시간적으로 배열될 때 움직임이 형성된다. 특히 전통적인 애니메이션 제작에서는 주요 상태를 정의하는 키프레임과 그 사이를 연결하는 중간 프레임을 통해 움직임이 구성되며, 이는 애니메이션이 개별 프레임의 설계와 그 사이의 관계 설정을 통해 형성된다는 점을 보여준다. 이 과정에서 중요한 것은 개별 프레임의 내용이 아니라 프레임 간 간격과 변화 방식이다. 동일한 시작과 끝 상태를 가지더라도 프레임 간 간격에 따라 전혀 다른 운동이 형성되며, 이는 애니메이션이 시간의 재현이 아니라 시간의 설계에 기반한 매체임을 의미한다. Lasseter는 이러한 원리가 3D 컴퓨터 애니메이션에서도 동일하게 적용됨을 강조하며, 타이밍과 스페이싱이 움직임 형성의 핵심 요소임을 지적한다[2]. 디지털 환경에서도 이러한 프레임 기반 구조는 유지되었는데, Manovich는 디지털 미디어가 수치적 데이터와 모듈성에 기반하여 구성되며, 영상 또한 분절된 단위들의 집합으로 존재한다는 점을 강조하였다[3]. 이는 디지털 영상 역시 근본적으로 이산적 구조를 유지하고 있음을 의미하며, 애니메이션의 프레임 기반 시간 구성 원리가 기술 변화에도 불구하고 지속되고 있음을 보여준다. 이러한 논의들은 애니메이션이 이산적 프레임의 배열과 프레임 간 관계 설정을 통해 시간과 움직임을 구성하는 매체임을 보여준다. 따라서 애니메이션은 창작자가 프레임 단위의 관계를 의도적으로 설계함으로써 작동하는 ‘프레임 단위 시간 체계’로 이해될 수 있다.

2. 생성형 AI 영상의 시간 생성 구조

생성형 AI 기반 영상은 전통적인 애니메이션 제작 방식과 달리, 개별 프레임을 직접 설계하고 배열하는 방식이 아니라, 데이터로부터 학습된 확률적 분포를 기반으로 시각적 결과를 생성한다. 특히 최근의 영상 생성 모델은 시공간적 정보를 통합적으로 학습함으로써, 시간 축을 포함한 연속적인 영상 시퀀스를 생성하는 특징을 가진다.

확산 모델(diffusion model)의 경우, 생성 과정은 노이즈 상태에서 시작하여 점진적으로 데이터를 복원하는 방식으로 이루어진다. Ho, Jain, and Abbeel은 이미지 확산 모델의 기본 원리를 제시하며, 이러한 생성 과정이 확률적 복원 절차와 반복적인 노이즈 제거 과정에 기반함을 설명하였다[4]. 이후 Ho et al.은 이를 영상 생성으로 확장하여 고정된 수의 비디오 프레임을 3D U-Net 기반 확산 모델로 생성하는 Video Diffusion Models를 제안하였다[5]. 이 연구는 비디오 데이터를 프레임의 단순 나열이 아니라, 프레임 축을 포함한 시공간적 구조로 모델링할 수 있음을 보여준다. 따라서 생성형 AI 영상은 최종적으로 프레임 시퀀스의 형태로 출력되지만, 생성 과정에서는 프레임 간 관계와 시간 축을 포함한 모델링을 통해 시간적 연속성이 형성된다고 이해할 수 있다.

한편, 다른 텍스트-비디오 생성 모델들 역시 시간 구조를 처리하는 방식에서 차이를 보인다. Meta에서 발표한 텍스트-비디오 생성 모델인 ‘Make-A-Video’는 기존 이미지 생성 모델을 시간 차원으로 확장하여 시공간적 구조를 통합적으로 모델링하고, 시간적 어텐션(temporal attention)과 프레임 보간(frame interpolation)을 활용하여 다수의 프레임을 생성하고 시간적 연속성을 형성한다[6]. Google에서 발표한 텍스트-비디오 생성 모델인 Phenaki는 텍스트 입력을 조건으로 비디오를 이산 토큰 시퀀스로 표현하고, 시간 방향의 인과적 구조를 유지하면서 다수의 비디오 토큰을 병렬적으로 예측하는 방식을 사용한다[7]. 이러한 모델들은 모두 시간 정보를 포함한 프레임 간 관계를 모델링하며, 생성 과정에서 시간적 연속성과 시퀀스의 일관성을 산출한다. 생성형 AI 영상에서 시간은 사전에 완결된 형태로 주어지는 것이 아니라, 반복적이고 단계적인 연산 과정 속에서 점진적으로 구성되는 특성을 보인다.

결과적으로 생성형 AI 영상은 프레임을 독립적인 설계 단위로 다루는 전통적 애니메이션과 달리, 시공간적 상관관계를 학습한 모델이 확률적으로 영상을 구성한다는 점에서 차이를 가진다. 이때 시간은 외부에서 직접 설계되는 대상이라기보다, 데이터 분포와 생성 알고리즘의 내부 연산 과정 속에서 재구성되는 요소로 이해될 수 있다. 이는 두

매체가 시간과 움직임을 구성하는 방식에서 서로 다른 구조적 원리를 전제하고 있음을 시사한다.

3. 영상 매체에서의 시간성

영상 매체에서 시간은 단순히 흐르는 물리적 차원이 아니라, 매체의 구성 방식과 생성 원리에 따라 서로 다른 형태로 조직되고 경험되는 개념이다. 이러한 시간성은 고정된 실체라기보다, 이미지가 생성되고 배열되는 방식에 따라 달라지는 구조적 속성으로 이해될 필요가 있다.

먼저, Bergson은 시간을 균질적으로 분할 가능한 물리적 단위가 아니라, 질적으로 지속되는 흐름으로 보았다[8]. 그는 시간을 공간처럼 동일한 단위로 나눌 수 있는 것이 아니라, 의식 속에서 연속적으로 변화하는 경험으로 이해하였으며, 이러한 관점은 시간의 본질이 연속성에 있음을 강조한다. 이와 같은 논의에 따르면, 전통적 애니메이션은 본래 연속적인 시간의 흐름을 프레임이라는 이산적 단위로 분절하고 이를 재배열함으로써 움직임을 구성하는 매체로 해석될 수 있다. 즉, 애니메이션은 지속으로서의 시간을 해체하고, 이를 다시 설계된 구조로 조직하는 방식의 시간 구성 체계를 가진다. 한편, Deleuze는 영화 이미지를 '운동-이미지(movement-image)'와 '시간-이미지(time-image)'로 구분하며, 영상 매체에서 시간이 드러나는 방식을 설명하였다[9][10]. 운동-이미지는 인과적 연결과 연속적인 움직임을 통해 시간을 간접적으로 구성하는 방식으로, 고전적 영화 및 애니메이션의 구조와 밀접하게 관련된다. 반면 시간-이미지는 움직임의 연쇄에 의존하지 않고 시간 자체가 직접적으로 드러나는 형태를 의미한다. 이러한 구분은 영상 매체에서 시간이 단순히 재현되는 것이 아니라, 특정한 구성 원리에 따라 다르게 나타날 수 있음을 보여준다. 이러한 시간 개념은 디지털 미디어 환경에서 또 다른 방식으로 확장된다. Manovich는 디지털 미디어가 수치적 데이터와 모듈성에 기반하여 구성된다고 보았으며, 영상 또한 분절 가능한 단위들의 집합으로 존재한다고 설명하였다[3]. 이는 디지털 환경에서도 프레임 기반 구조가 유지되며, 시간이 데이터 단위로 분할되고 조작될 수 있음을 시사한다. 즉, 디지털 영상 역시 기본적으로는 분절된 요소들의 배열을 통해 시간적 흐름을 구성하는 체계를 유지하고 있다. 생성형 AI 기반 영상에서는 이러한 시간 구성 방식이 다른 방식으로 나타난다. 전통적 애니메이션이 프레임의 배열을 통해 외부에서 시간을 조직하는 '구성된 시간(constructed temporality)'에 기반한다면, 생성형 AI 영상에서는 시간적 연속성이 모델 내부의 확률적 연산 과정에서 형성된다. 이는 시간의 흐름이 창작자의 설계에 의해 단계적으로 구성되는 것이 아니라, 모델의 확률적 생성 과정 속에서 형성되는 시간 구성 방식으로 이해될 수 있음을 의미한다.

결과적으로 본 연구는 전통적 애니메이션의 시간을 프레임의 분절과 배열을 통해 외부에서 조직되는 구성적 시간성으로, 생성형 AI 영상의 시간을 모델의 확률적 생성 과정에서 형성되는 생성적 시간성으로 구분한다. 여기서 '생성적 시간성(generative temporality)'은 시간적 사건의 발생 시점, 간격, 지속, 운동 궤적이 창작자의 프레임 단위 설계가 아니라 모델의 생성 과정 속에서 형성되는 성질을 의미한다. 따라서 결과물의 변동성은 생성적 시간성 자체가 아니라, 시간적 사건이 생성 과정 안에서 산출되고 있음을 보여주는 관찰 지표로 이해할 수 있다.

Ⅲ. 비교 분석 및 사례 분석

1. 시간 구성 방식의 비교 기준

전통적 애니메이션과 생성형 AI 영상은 모두 최종적으로 프레임의 연속을 통해 움직임을 제시하지만, 시간 구조가 형성되는 방식에는 차이가 있다. 전통적 애니메이션에서 시간은 창작자가 프레임 단위로 직접 설계하고 배열하는 방식으로 구성된다. 애니메이터는 키프레임과 중간 프레임을 설정하고, 타이밍(timing)과 스페이싱(spacing)을 조정함으로써 움직임의 속도, 무게감, 접촉, 반동, 리듬을 의도적으로 통제한다. 따라서 바운싱 볼과 같은 기본 운동 사례에서 바닥 접촉 시점, 압축 변형, 반동 타이밍은 창작자가 프레임 단위로 조절할 수 있는 시간적 사건에 해당한다.

반면 생성형 AI 영상에서 창작자는 개별 프레임의 위치나 프레임 간 운동 간격을 직접 지정하기보다, 텍스트 프롬프트, 영상 길이, 화면비, 해상도와 같은 생성 조건을 설정한다. 최종 결과물은 프레임 시퀀스의 형태로 출력되지만,

각 프레임 사이의 운동 관계와 시간적 사건의 발생 시점은 모델의 시공간적 생성 과정 속에서 형성된다. 따라서 생성형 AI 영상에서 바닥 접촉, 반동, 상승과 같은 사건은 창작자의 직접적인 프레임 설계 결과라기보다, 프롬프트와 생성 조건을 바탕으로 모델이 산출한 결과로 이해할 수 있다.

본 연구에서는 이러한 차이를 바탕으로 전통적 애니메이션의 프레임 단위 시간 구성을 이론적 비교 기준으로 설정하고, 생성형 AI 영상에서 시간적 사건이 어떻게 형성되고 변동되는지를 사례 분석을 통해 검토한다. 특히 바운싱 볼 시퀀스는 애니메이션 교육과 제작에서 타이밍, 스페이싱, 접촉, 반동을 설명하는 기본 운동 사례이므로, 생성형 AI 영상의 시간 구성 방식을 관찰하기 위한 분석 대상으로 적합하다. 이에 본 연구는 동일한 바운싱 볼 운동 과제를 생성형 AI 모델에 반복적으로 제시하고, 각 결과물에서 최초 바닥 접촉 시점과 반동 양상이 어떻게 달라지는지를 분석함으로써 생성형 AI 영상의 시간 구성 특성을 살펴본다.

2. 전통적 애니메이션과 생성형 AI 영상의 시간 구성 방식의 차이

앞서 정리한 비교 기준에 따라 전통적 애니메이션과 생성형 AI 영상의 시간 구성 방식은 시간의 단위, 시간 통제 방식, 창작 개입 수준에서 구분될 수 있다. 이러한 차이는 [표 1]과 같이 정리할 수 있다.

[표 1]에서 보듯이 전통적 애니메이션은 창작자가 프레임 간 간격과 변화를 직접 설계함으로써 시간 구조를 구성하는 반면, 생성형 AI 영상은 프롬프트와 생성 조건을 바탕으로 모델이 시간적 사건을 산출한다. 따라서 두 매체의 차이는 단순히 제작 기술의 차이에 그치지 않고, 시간과 움직임이 형성되는 방식의 차이로 확장된다. 본 연구는 이러한 차이를 바탕으로, 생성형 AI 영상에서 바닥 접촉과 반동 같은 시간적 사건이 동일한 조건에서도 어떻게 달라지는지를 사례 분석을 통해 검토한다.

3. 사례 분석: 공 운동 시퀀스를 통한 시간 구성 방식 비교

3.1 실험 조건 및 분석 기준

전통적 애니메이션과 생성형 AI 영상의 시간 구성 방식 차이를 구체적으로 검토하기 위해, 본 연구에서는 생성형 AI 영상에 대한 사례 분석을 중심으로 논의를 진행하였다. 전통적 애니메이션의 시간 구성 방식은 기존 이론과 제작 원리를 통해 비교적 명확하게 설명될 수 있는 반면, 생성형 AI 영상의 경우 동일한 조건에서도 결과가 매번 다르게 생성되는 특성을 보인다. 따라서 생성형 AI 영상의 시간 구조가 어떠한 방식으로 형성되는지를 사례적으로 확인하는 것이 필요하다고 판단하였다. 이에 동일한 공 튀기기 장면을 대상으로 반복 생성 분석을 수행하고, 그 결과를 바탕으로 시간 구성 방식의 변동성을 분석하였다. 본 연구는 생성형 AI 영상에서 시간적 사건이 어떠한 방식으로 형성되고 변동되는지를 확인하기 위해 Kling Video Model 3.0을 사용하여 바운싱 볼 시퀀스를 반복 생성하였다. Kling은 동일한 텍스트 프롬프트와 기본 생성 조건을 유지한 상태에서 짧은 영상을 반복 생성할 수 있다는 점에서, 동일한 운동 과제에 대한 샘플 간 시간 구조의 차이를 탐색적으로 관찰하는 데 적합하다. 다만 본 연구는 Kling의 내부 구조나 생성 원

표 1. 프레임 단위 애니메이션과 생성형 AI 영상 간 차이
Table 1. Differences Between Frame-based Animation and Generative AI Video

Category	Frame-based Animation	Generative AI Video
Nature of Time	Time externally designed and constructed	Time formed through probabilistic generation
Structural Unit	Frame sequence directly arranged by the animator	Frame sequence generated through learned spatiotemporal modeling
Time Control	Direct control through timing and spacing	Indirect control through prompts and generation settings
Creative Intervention	Frame-level design and adjustment	Condition-setting and result selection
Characteristics	Reproducible motion based on intentional frame design	Variability and unpredictability across generated outputs

리를 일반화하려는 것이 아니라, 특정 모델에서 관찰되는 시간 구성 양상을 사례적으로 분석하는 데 목적이 있다.

생성 방식은 텍스트-투-비디오(text-to-video) 방식이며, 동일한 프롬프트를 반복 입력하여 총 10개의 영상을 생성하였다. 생성된 10개의 영상은 모두 분석 대상으로 포함하였다. 영상 길이는 3초로 설정하였으며, 분석 대상 영상의 프레임 속도는 24 fps, 해상도는 720p, 화면비는 16:9이다. 별도의 seed 고정, 스타일 프리셋, 추가 설정값은 적용하지 않았으며, 다운로드 이후 영상 편집, 속도 조절, 프레임 보간 등의 후처리는 수행하지 않았다. 이러한 생성 조건은 [표 2]와 같다.

분석 기준은 '공이 최초로 바닥에 닿는 프레임'으로 설정하였다. 본 연구에서 바닥 접촉 프레임은 공의 최하단 윤곽이 화면 내 바닥 선과 처음으로 맞닿거나, 공의 압축 변형이 시작되는 최초 프레임으로 정의하였다. 생성된 영상은 프레임 단위로 확인하였으며, 연구자가 각 시퀀스의 전후 프레임을 비교하여 접촉 여부를 수동 판정하였다. 바닥 접촉 여부가 모호한 경우에는 공의 최하단 위치, 압축 변형의 시작 여부, 이후 반동 움직임의 발생 여부를 함께 검토하여 최초 접촉 프레임을 결정하였다. 해당 시점은 공의 운동이 명확하게 드러나는 최초의 시점으로, 이후 반동과 궤적 형성에도 직접적인 영향을 미친다. 따라서 동일한 프롬프트 조건에서 생성된 영상 간 시간 구조의 일관성과 변동성을 정량적으로 비교할 수 있는 유효한 지표로 판단하였다. 또한 판정의 일관성을 확보하기 위해 동일 영상을 일정 시간 간격을 두고 2회 반복 확인하였으며, 1차와 2차 판정 결과

표 2. 생성형 AI 영상 생성 조건
Table 2. Generation Conditions for Generative AI Video Samples

Category	Condition
Model	Kling Video Model 3.0
Generation Type	Text-to-video
Number of Generations	10
Analyzed Samples	All 10 generated videos
Video Length	3 seconds
Frame Rate	24 fps
Resolution	720p
Aspect Ratio	16:9
Seed	Not fixed
Style Preset	Not applied
Additional Settings	Default settings
Post-processing	None
Prompt	A simple red ball bouncing once on a plain white background. Fixed camera. No motion blur. The ball falls vertically, hits the ground, squashes slightly, then bounces upward. Minimal style, clean 2D animation look.

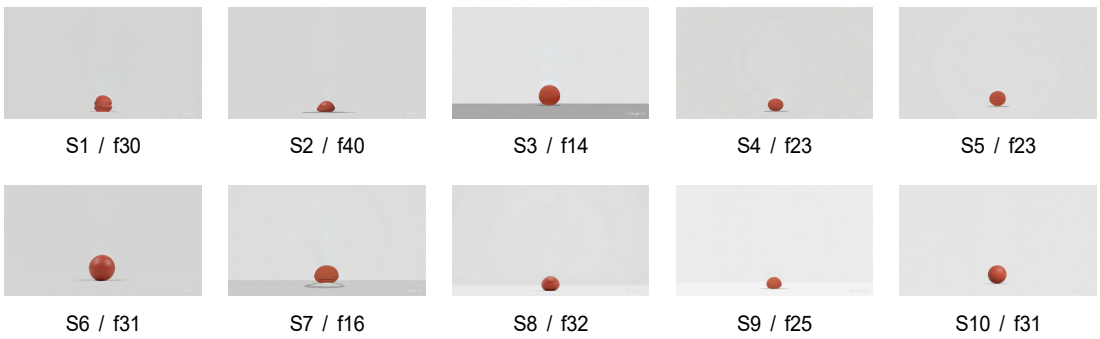

그림 1. 생성형 AI 영상 샘플별 바닥 접촉 프레임
Fig. 1. Ground Contact Frames of Generated Video Samples

가 다른 경우에는 해당 구간의 전후 프레임을 재검토하여 최종 프레임을 확정하였다.

3.2 분석 결과

3.1에서 제시한 프롬프트를 활용하여 총 10개의 생성형 AI 영상을 Kling 모델을 통해 생성하였으며, 그 결과는 [그림 1] 및 [표 3]과 같다.

표 3. 생성형 AI 영상 샘플별 바닥 접촉 시점 분석 결과
Table 3. Analysis Results of Ground Contact Timing in Generated Video Samples

Sample	Contact Frame	Contact Time (s, 24 fps 기준)[1]	Rebound Pattern
S1	30	1.25	Clear rebound
S2	40	1.67	Clear rebound
S3	14	0.58	Weak rebound
S4	23	0.96	Clear rebound
S5	23	0.96	Multiple rebounds
S6	31	1.29	Clear rebound
S7	16	0.67	Clear rebound
S8	32	1.33	Weak rebound
S9	25	1.04	Clear rebound
S10	31	1.29	Clear rebound

1) Contact time was calculated by dividing the contact frame by 24 fps.

[표 3]에 제시한 바와 같이, 동일한 프롬프트와 동일한 생성 조건을 적용하였음에도 각 샘플의 최초 바닥 접촉 시점은 일정하게 나타나지 않았다. 10개 샘플의 접촉 시점은 최소 14프레임에서 최대 40프레임까지 분포하였으며, 이는 24 fps 기준 약 0.58초에서 1.67초에 해당한다. 또한 접촉 이후의 반동 양상에서도 차이가 확인되었다. 대부분의 샘플에서는 최초 접촉 이후 명확한 반동이 나타났으나, S3과 S8에서는 반동이 상대적으로 약하게 나타났으며, S5의 경우에는 프롬프트에서 제시한 'once' 조건과 달리 접촉과 반동이 여러 차례 반복되는 양상을 보였다.

형태 변형의 경우 일부 샘플에서 접촉 시점 전후로 공의 압축 또는 일시적인 형태 변화가 관찰되었다. 다만 본 연구의 핵심 분석 대상은 공의 형태 변화 정도가 아니라, 바닥 접촉과 반동이라는 시간적 사건이 어느 시점에서 발생하는가에 있다. 또한 생성 영상마다 공의 크기, 위치, 압축 정도가 달라 형태 변형을 정량적 기준으로 비교하기에는 한계가 있으므로, 본 연구에서는 형태 변형을 독립적인 분석 항목으로 설정하지 않고 접촉 및 반동 판정을 보조하는 관찰 요소로만 활용하였다.

[그림 2]는 각 샘플의 최초 바닥 접촉 프레임과 이를 24 fps 기준으로 환산한 시간을 시각화한 것이다. 이를 통해 동일한 생성 조건에서도 바닥 접촉이라는 시간적 사건의 발생 시점이 샘플별로 다르게 나타남을 확인할 수 있다.

바닥 접촉 시점의 변동성을 보다 명확히 제시하기 위해, [표 3]의 접촉 프레임 값을 바탕으로 기술통계를 산출하였다. 접촉 시간은 24 fps 기준으로 접촉 프레임을 초 단위로 환산한 값이며, 기술통계에는 평균, 표준편차, 최소값, 최대

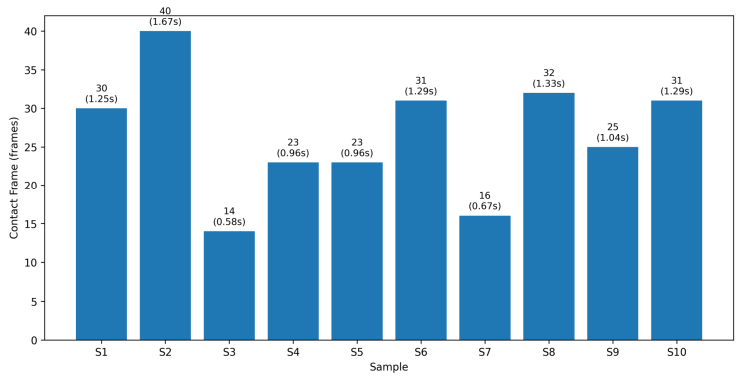

그림 2. 샘플별 최초 바닥 접촉 프레임 및 시간 변동성
Fig. 2. Variability in First Ground Contact Frame and Time Across Samples
Note: Contact time was calculated by converting the contact frame based on 24 fps

표 4. 바닥 접촉 시점의 프레임 및 시간 기술통계
Table 4. Descriptive Statistics of Ground Contact Frame and Time

Measure	Contact Frame	Contact Time (s)
N	10	10
Mean	26.50	1.10
Standard Deviation	7.88	0.33
Minimum	14	0.58
Maximum	40	1.67
Median	27.50	1.15

값, 중앙값을 포함하였다. 그 결과는 [표 4]와 같다.

분석 결과, 최초 바닥 접촉 시점의 평균은 26.50프레임으로 나타났으며, 이는 24 fps 기준 약 1.10초에 해당한다. 표준편차는 7.88프레임, 약 0.33초로 나타났고, 최소값은 14프레임, 최대값은 40프레임으로 확인되었다. 이러한 결과는 동일한 프롬프트와 동일한 생성 조건을 적용하였음에도 불구하고, 바닥 접촉이라는 시간적 사건의 발생 시점이 샘플별로 일정하게 고정되지 않았음을 보여준다.

이러한 변동성은 생성적 시간성 그 자체라기보다, 생성적 시간성이 사례 분석에서 관찰되는 방식으로 이해할 수 있다. 즉, 동일한 운동 과제가 제시되었음에도 바닥 접촉 시점과 반동 양상이 다르게 나타났다는 점은 시간적 사건의 발생 시점과 운동 간격이 프레임 단위로 사전에 고정된 것이 아니라, 모델의 생성 과정 속에서 산출되었음을 보여주는 사례적 근거이다. 전통적 애니메이션에서는 바닥 접촉 시점과 반동 타이밍이 프레임 단위로 명확하게 결정되며, 동일한 시간 구조를 반복적으로 재현할 수 있다. 반면 생성형 AI 영상에서는 동일한 프롬프트를 사용하더라도 시간적 사건의 발생 시점이 영상마다 달라지며, 이는 시간 구조가 창작자의 직접적 설계에 의해 고정되지 않고 모델의 확률적 생성 과정 속에서 가변적으로 형성될 수 있음을 시사한다.

Ⅳ. 결 론

본 연구는 전통적 애니메이션의 프레임 기반 시간 구성 원리를 이론적 비교 기준으로 삼고, Kling 기반 바운싱 볼 시퀀스 사례를 통해 생성형 AI 영상에서 시간적 사건이 어떻게 형성되고 변동되는지를 검토하였다. 전통적 애니메이션은 프레임 단위를 기반으로 시간의 흐름을 외부에서 분절하고 배열함으로써 움직임을 구성하는 반면, 생성형 AI 영상은 시공간적 데이터를 학습한 모델의 연산 과정 속에서 시간적 연속성이 형성되는 특성을 보인다.

사례 분석 결과, 생성형 AI 영상은 동일한 입력 조건에서도 바닥 접촉과 반동 같은 시간적 사건의 발생 시점이 일정하게 유지되지 않고 샘플별로 다르게 형성되는 경향을 보였다. 이는 생성형 AI 영상의 시간이 사전에 고정된 구조로 설계되기보다, 모델의 생성 과정 속에서 가변적으로 구성됨을 보여준다. 따라서 두 매체의 차이는 단순한 표현 방식의 차이를 넘어, 시간과 움직임이 형성되는 구조적 원리의 차이로 이해될 수 있다.

본 연구는 생성형 AI 영상의 시간 구조가 전통적 애니메이션과는 구별되는 방식으로 형성된다는 점을 확인하고, 이를 통해 영상 매체에서 시간이 구성되는 방식에 대한 이해를 확장할 필요가 있음을 시사한다. 다만 본 연구는 특정 모델(Kling)과 단일 프롬프트 조건에 기반한 탐색적 사례 분석에 한정되어 있으므로, 그 결과를 생성형 AI 영상 일반의 보편적 특성으로 확대하기에는 한계가 있다. 또한 본 연구에서 제시한 '생성적 시간성'은 확정된 이론 개념이라기보다, Kling 기반 바운싱 볼 시퀀스에서 관찰된 시간적 사건의 변동성을 설명하기 위한 분석적 관점이다. 따라서 향후 연구에서는 다양한 생성 모델, 프롬프트 조건, 운동 유형을 대상으로 생성형 AI 영상의 시간 구조를 다각적으로 검토하고, 생성적 시간성 개념의 타당성과 적용 가능성을 보다 구체화할 필요가 있다.

참 고 문 헌 (References)

[1] P. Wells, *Understanding Animation*, Routledge, London, 1998.
[2] J. Lasseter, "Principles of Traditional Animation Applied to 3D Computer Animation," ACM SIGGRAPH Computer Graphics, Vol.21, No.4, pp.35-44, Jul. 1987.
doi: https://doi.org/10.1145/37402.37407
[3] L. Manovich, *The Language of New Media*, MIT Press, Cambridge, MA, pp.1-354, 2001.
[4] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," Advances in Neural Information Processing Systems, Vol.33, pp.6840-6851, 2020.

[5] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, “Video Diffusion Models,” arXiv preprint arXiv:2204.03458, 2022. doi: https://doi.org/10.48550/arXiv.2204.03458
[6] U. Singer, A. Polyak, T. Hayes, X. Yin, J. An, S. Zhang, Q. Hu, H. Yang, O. Ashual, O. Gafni, D. Parikh, S. Gupta, and Y. Taigman, “Make-A-Video: Text-to-Video Generation without Text-Video Data,” arXiv preprint arXiv:2209.14792, 2022. doi: https://doi.org/10.48550/arXiv.2209.14792
[7] R. Villegas, M. Babaeizadeh, P.-J. Kindermans, H. Moraldo, H. Zhang, M. T. Saffar, S. Castro, J. Kunze, and D. Erhan, “Phenaki: Variable Length Video Generation from Open Domain Textual Descriptions,” arXiv preprint arXiv:2210.02399, 2022. doi: https://doi.org/10.48550/arXiv.2210.02399
[8] H. Bergson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, Dover Publications, Mineola, NY, pp.1-255, 2001.
[9] G. Deleuze, *Cinema 1: The Movement-Image*, University of Minnesota Press, Minneapolis, pp.1-244, 1986.
[10] G. Deleuze, *Cinema 2: The Time-Image*, University of Minnesota Press, Minneapolis, pp.1-344, 1989.

저 자 소 개

김 소 영

- 중앙대학교 첨단영상대학원 영상학과, 영상학 박사
- 현재 : 경일대학교 만화애니메이션학부 조교수
- ORCID : https://orcid.org/0009-0006-2181-8639
- 주관심분야 : 애니메이션, 생성형 AI, 산업, 정책