

바이브 메이커 · 무료 자료

클로드 영상 자동화 스킬 팩

클로드에게 영상을 보여주는 법과,
거기서 토큰이 새는 자리.

설치 /watch --detail 실측 프레임 예산 하이퍼프레임

목차

01 · 토큰은 길이가 아니라 프...	프레임 한 장이 이미지 한 장입니다.	03
02 · 설치	쓰는 도구에 따라 넣는 자리가 다릅니다.	04
03 · 부르는 법과 하는 일	주소를 던지고 궁금한 걸 물으면 됩니다.	05
04 · --detail 네 단계 (실측)	같은 영상, 설정만 바꿔 직접 잰 숫자입니다.	06
05 · 더 줄이는 손잡이	기본 예산은 길이가 정합니다.	07
06 · 만드는 쪽 — 하이퍼프레임	이쪽은 방향이 반대입니다.	08
07 · 언제 무엇을 쓰나	네 단계를 이렇게 고릅니다.	09

프레임 한 장이 이미지 한 장입니다.

클로드에게 영상을 보여준다는 건 **프레임을 이미지로 넣는다**는 뜻입니다. 그래서 비용은 영상 길이가 아니라 **몇 장을 넣었느냐**로 정해집니다. 문서 원문도 같은 말을 합니다 — **"Token cost is dominated by frames."**

≈197

프레임 한 장의 이미지 토큰. 앤트로픽 계산식 **(가로 × 세로) ÷ 750** 에 기본 폭 512px 을 넣으면 720p 프레임이 512×288 이 됩니다. **--resolution 1024** 로 올리면 대략 네 배가 됩니다.

그래서 이 문서가 다루는 것

1. **bradautomates/claude-video**
   클로드가 **영상을 보게** 해주는 스킬. 오픈소스(MIT). 자막을 먼저 찾고, 필요한 만큼만 받고, 프레임을 뽑아 클로드에게 넘깁니다.

2. **heygen-com/hyperframes**
   반대 방향. **웹 문서를 영상으로** 렌더합니다. 오픈소스(Apache-2.0).

3. **--detail 다이얼**
   토큰을 정하는 손잡이. **같은 영상으로 잰 실측표**가 5쪽에 있습니다.

💡 이 문서의 수치는 전부 **각 저장소 README 원문**에서 옮겼습니다. 기준은 **2026-09-16** 이고, 도구가 업데이트되면 달라질 수 있습니다.

쓰는 도구에 따라 넣는 자리가 다릅니다.

어디에	어떻게
클로드코드	`/plugin marketplace add bradautomates/claude-video` 뒤에 `/plugin install watch@claude-video`
커서·코덱스 등	`npx skills add bradautomates/claude-video -g` — `-g` 는 사용자 전체, 빼면 이 프로젝트만
claude.ai 웹	릴리스에서 `watch.skill` 을 받아 설정 → 기능 → 스킬 → **+**

클로드코드에 넣기

두 줄이면 끝납니다

```
/plugin marketplace add bradautomates/claude-video
```

```
/plugin install watch@claude-video
```

1. 나중에 갱신은 `/plugin update watch@claude-video`

✅ **준비물은 신경 안 써도 됩니다** — `yt-dlp` 와 `ffmpeg` 는 **첫 실행에 알아서 설치**됩니다(맥은 brew, 리눅스·윈도우는 명령을 그대로 찍어줍니다). 자막이 있는 영상은 **무료**로 처리되고, **Whisper 키는 자막이 없는 영상에만** 필요합니다.

주소를 던지고 궁금한 걸 물으면 됩니다.

이렇게 부릅니다 주소 또는 로컬 파일 경로

```
/watch https://youtu.be/… 30초 지점에서 무슨 일이 일어나나요?
```

```
/watch https://youtu.be/… 이 사람 훅을 어떻게 열었나요?
```

```
/watch bug-repro.mov 뭐가 잘못되고 있나요?
```

"설명을 보고"가 아니라 **화면을 보고** 답합니다.

부르면 안에서 벌어지는 일

- **1 자막부터** — 자막이 있으면 영상을 안 받고, 없으면 필요한 만큼만 받습니다
- **2 프레임 추출** — 키프레임 또는 장면 전환 프레임, 기본 폭 512px
- **3 전사** — 원본 자막 우선, 없으면 Whisper 로 보충
- **4 클로드가 읽는다** — 프레임마다 **t=MM:SS** 가 붙어 시점을 짚어 답합니다

같은 영상, 설정만 바꿔 **직접 잰** 숫자입니다.

아래는 **49분 08초** 유튜브 영상(1280×720, 영어 자동자막) 한 편을 네 설정으로 각각 돌린 결과입니다. 거의 안 움직이는 화면 녹화라 **상한을 가장 심하게 때리는 경우**예요.

설정	프레임	추출 시간	이미지 토큰
transcript	0장 (자막만)	약 4.5초	**0**
efficient	50장	약 0.5초	**약 9,800**
balanced	100장	약 20.9초	약 19,700
token-burner	116장 (상한 없음)	약 21.0초	약 22,800

⚡ **efficient 는 "적게"가 아니라 "빠르게"입니다**

키프레임만 복원하므로 장면 검출보다 **약 40배 빠릅니다**. 움직임이 적은 영상에서는 키프레임이 장면 전환보다 많아서 **balanced 보다 프레임이 더 나올 수도** 있어요. **token-burner 는 상한을 넘을 때만** balanced 와 달라집니다.

💡 **자막만 쓰는 transcript 는 이미지 토큰이 0** 입니다. 다만 전사 자체가 텍스트 토큰으로 들어가고(이 영상은 약 26,600), 긴 영상에서는 **그쪽이 더 큰 비용**이 되기도 합니다.

기본 예산은 길이가 정합니다.

아무 것도 안 건드리면 영상 길이에 따라 프레임 예산이 자동으로 잡힙니다. **10분이 넘어가면 「듬성듬성 훑었다」는 경고**가 붙어요.

영상 길이	기본 프레임 예산	밀도
30초 이하	약 30장	거의 모든 순간
30초 ~ 1분	약 40장	촘촘
1 ~ 3분	약 60장	넉넉
3 ~ 10분	약 80장	듬성하지만 쓸 만
10분 초과	100장 (상한)	**경고 — 구간을 좁혀라**

토큰을 더 줄이는 네 가지

구간 지정 — `--start` · `--end` — "2분 30초쯤"처럼 지점을 알면 그 구간만. **구간 안에서는 더 촘촘하게** 잡아줍니다(최대 초당 2장).

상한 낮추기 — `--max-frames N` — 예산 자체를 더 조입니다.

중복 제거 — **기본으로 켜져 있습니다.** 한 슬라이드를 90초 잡고 있는 화면 녹화는 거의 같은 프레임을 열두 장 만드는데, 그걸 미리 버립니다. **예산은 중복을 버린 뒤에 적용**되니 남은 예산이 서로 다른 화면에 쓰여요. 끄려면 `--no-dedup`.

해상도 — `--resolution 1024` 는 화면의 **글자를 읽어야 할 때만**. 슬라이드·터미널·코드가 그 경우이고, 대신 토큰이 대략 네 배가 됩니다.

이쪽은 방향이 반대입니다.

클로드 비디오가 **영상을 읽는** 쪽이라면, 하이퍼프레임은 **영상을 만드는** 쪽입니다. HTML·CSS·미디어·애니메이션을 **결정론적인 MP4** 로 렌더합니다. 같은 입력이면 같은 결과가 나와요.

설치 에이전트에 스킬로 붙입니다

```
npx skills add heygen-com/hyperframes
```

1. 고르는 창이 뜨면 **Core Skills** 만으로 충분합니다. 나머지는 필요할 때 알아서 받아옵니다.
2. 에이전트가 자동으로 돌릴 때는 `npx hyperframes skills update` 를 쓰세요 (창 없이 핵심 세트만 설치).

부르는 법 `/hyperframes` 가 길잡이입니다

/hyperframes 로 10초짜리 제품 소개를 만들어 주세요. 제목은 페이드인, 배경 영상, 잔잔한 배경음악.

✅ **/hyperframes 를 먼저 부르는 이유** — 그게 길잡이라 요청을 보고 알맞은 작업 흐름을 골라줍니다. 제품 소개·개념 설명·깃허브 PR 요약·자막 얹기·모션그래픽·덱까지 갈래가 나뉘어 있어요.

네 단계를 이렇게 고릅니다.

이럴 때	이 설정
말한 내용만 필요할 때 (요약·검색·인용)	**transcript** — 이미지 토큰 0
화면을 봐야 하지만 빨리 훑고 싶을 때	**efficient** — 기본값으로 쓰기 좋음
장면이 자주 바뀌는 편집물·광고	**balanced** — 장면 단위로 잡는다
한 프레임도 놓치면 안 될 때	**token-burner** — 상한 없음, 비쌈
"2분 30초쯤" 처럼 지점을 알 때	설정보다 **--start / --end** 가 먼저

알고 쓰기

1 긴 영상은 설정보다 구간이 먼저입니다
30분짜리를 듬성듬성 훑느니 **궁금한 30초를 촘촘히** 보는 편이 답도 좋고 쌉니다.

2 자막이 없으면 Whisper 키가 필요합니다
자막이 있는 공개 영상은 그냥 됩니다. 없을 때만 음성을 떼어 보내므로 **그때만** 키를 준비하면 돼요.

3 둘 다 커뮤니티·오픈소스입니다
claude-video 는 MIT, hyperframes 는 Apache-2.0. **업데이트 뒤에는 무엇이 바뀌었는지 한 번 보고** 쓰는 습관이 안전합니다.