

Breakfast in Paris. Lunch in Tokyo. Dinner in Times Square.

One continuous shot. Three continents. **Zero flights.**

Here is the entire prompt, the reference images, and the two rules that make it work.

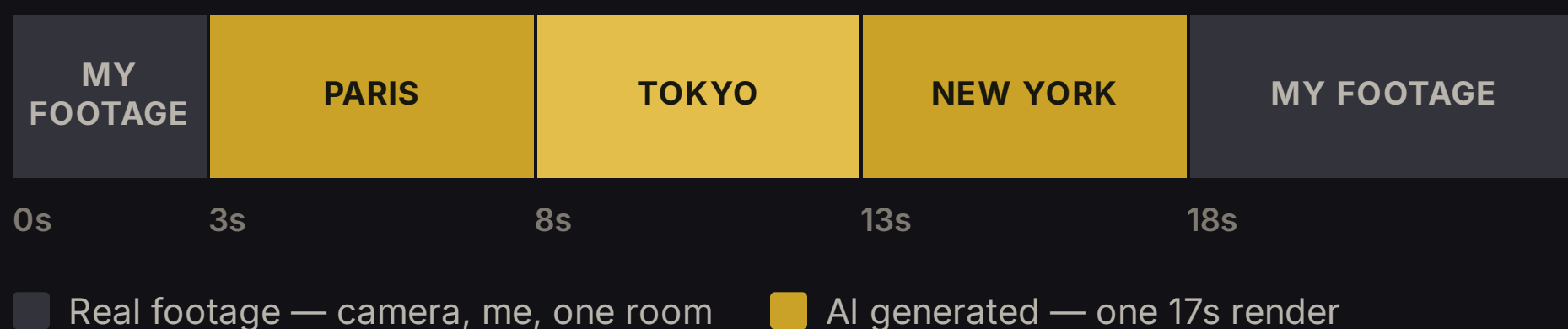


YASER GHULOOM
@YASERGHM

SWIPE →

Only the **middle 15 seconds** are AI.

I shot myself talking in one plain room. The AI never touches the start or the end — it rebuilds the world in between, and my voice runs unbroken across all of it.

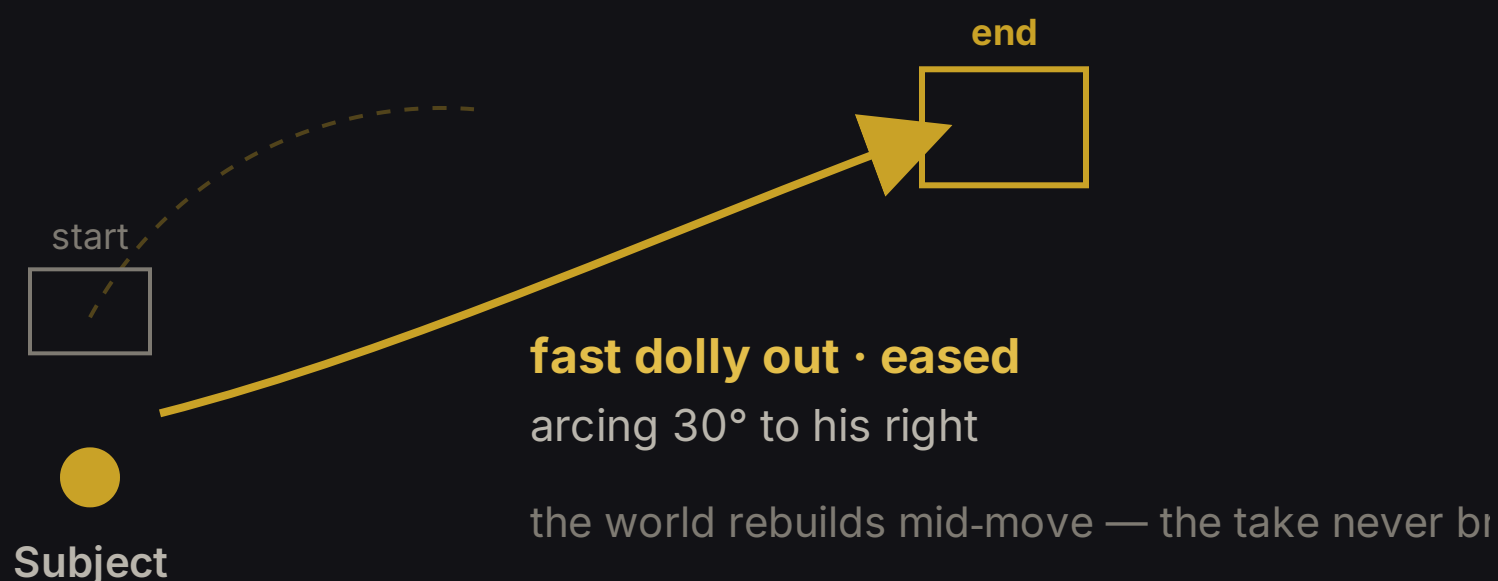


Every city change is a **magical environment swap** inside a single unbroken camera move. No cuts, no dissolves, no transitions bought from a template pack. The room itself rebuilds around me while the camera keeps travelling.

The whole piece is 24 seconds. One AI render covers 15 of them.

The trick is not the locations. It is the **refusal to cut**.

Anyone can generate a clip of Paris. What makes this read as one journey instead of six stock clips is a single repeating camera move that every environment change hides inside.



Write that move once, then fire it at every scene change. Consistency of motion is what buys you the illusion of one take.

Ten images. Every one has a job.

Six are locations, three are outfits, one is the food. The prompt calls each of them by number, so the model knows exactly which picture belongs to which second of the video.



@Image1
Paris café interior



@Image2
Paris street



@Image3
Japanese restaurant



@Image4
Japan street



@Image5
Times Square street



@Image6
Times Square tables



@Image7
Paris wardrobe



@Image8
Japan wardrobe



@Image9
Noodle soup



@Image10
New York wardrobe

Order matters. The images map to **@Image1 ... @Image10** by the order you upload them — not by filename. Get the order wrong and your kimono ends up in Times Square.

Use clean, high-resolution references. Screenshots with watermarks or search-engine icons baked in will show up in your footage.

Upload **yourself**, twice.

Images give you places and clothes. Only video holds your face steady across six environments — and tells the model exactly what frame to leave from and land on.



@VIDEO01 · 8 SECONDS

The opening clip

Cut from **00:00–00:08** of your original. Carries your face, build, hair and clothing, plus the room framing the AI has to grow out of.



@VIDEO02 · 4 SECONDS

The closing clip

Cut from **00:18–00:22**. This is the exact frame the journey must collapse back into, so the return to your real footage is invisible.

Export them, do not stream-copy them. A fast trim that only copies the stream keeps the original timestamps, so a 3-second file still reports itself as 24 seconds and gets rejected. Re-encode and the duration is honest.

Strip the audio out of both clips while you are at it. The model is reading your face, not your voice.

Two limits. Both bite at once.

This is where most people give up. The error message only tells you about the first one.

What people try first

23s — rejected

10s opening clip + 13s closing clip



The white line is the 15-second ceiling for all reference video *combined*.

What actually runs

12s — accepted

8s + 4s



Leaves room for a 17-second output.

Limit 1 All your reference clips **added together** must be 15 seconds or less. Not 15 seconds each.

Limit 2 Reference length **plus** output length must stay under 30 seconds. Every extra second of reference costs you a second of video.

12 seconds of reference + 17 seconds of output = 29. That is why the clips are 8 and 4.

Six steps, **start to finish.**

1

Shoot one plain take

Stand in a neutral room and talk to camera for the full length. Keep the framing locked. This is your anchor at both ends.

2

Collect your references

Six locations, three outfits, one prop. Clean images, no watermarks. Note the order you will upload them in.

3

Cut two clips

8 seconds from the head, 4 seconds from the tail. Re-encode rather than stream-copy so the durations report honestly.

4

Paste the prompt

Swipe on for the full text. Change the cities and clothes to yours — keep the structure exactly as it is.

5

Test at the cheapest setting

480p, 9:16, 17 seconds. Look at identity, transitions and timing before you spend on a final render.

6

Finish in your editor

Drop the render between your real footage. Your original voiceover stays on its own track, untouched, end to end.

Copy this **exactly**.

Part 1 of 5 — who and what the model is looking at.

REFERENCES

@Video1 = opening reference AND primary identity reference. 00:00-00:08 of the source. The man in a plain ordinary interior room, everyday home clothing, talking to camera. His face, build, hair, skin tone and clothing must match this in every shot. Also the source of truth for the room's framing, lens height and light at the start.

@Video2 = closing reference. 00:18-00:22 of the same continuous take. Same man, same plain room, same framing. The video must land back on this exactly.

@Image7 = wardrobe for the Paris scenes.

@Image8 = wardrobe for the Japan scenes.

@Image10 = wardrobe for the New York scenes.

@Image1 = Paris restaurant interior.

@Image2 = Paris street.

@Image3 = Japanese restaurant interior.

@Image9 = bowl of noodle soup with chopsticks.

@Image4 = Japan street.

@Image6 = New York Times Square outdoor restaurant table.

@Image5 = New York Times Square street.

ONE CONTINUOUS TAKE. No cuts anywhere in the video. Vertical 9:16. Cinematic 35mm look, natural skin tones, shallow depth of field throughout. The same man from @Video1 - identical face, build and hair - in every single shot.

Copy this **exactly**.

Part 2 of 5 — the camera rule, then Paris.

THE SIGNATURE MOVE, repeated at every environment change: a very fast dolly-out with eased acceleration and eased settle, arcing 30 degrees to the man's right, never stopping. The world dissolves and rebuilds around him mid-move - walls, light, people, furniture and his clothing reassembling in real time - while the camera keeps travelling. The environment changes, the take never breaks. No wipes, no dip to black, no cross-dissolve between separate shots: the space itself transforms.

Shot 1 [0-1s]: The man stands in a plain ordinary interior room in simple everyday home clothing, facing camera. Medium-close framing, flat neutral indoor light, static camera. Framing, wardrobe and light match @Video1.

Shot 2 [1-2s]: Signature move fires. The plain room peels away and rebuilds into the Parisian cafe in @Image1 - table, coffee cups, wicker chairs, awning, street beyond - and his clothing changes in the same sweep to @Image7. He settles seated at the cafe table with coffee and breakfast in front of him, relaxed, glancing slightly off-camera. Medium shot. Warm Paris morning light, soft haze.

Shot 3 [2-3s]: Without cutting, the camera pushes smoothly in to a close-up as he lifts the coffee cup, sips, and sets it down with a small satisfied breath. Steam off the coffee, creamy background bokeh.

Shot 4 [3-4.5s]: Signature move fires. The cafe dissolves and rebuilds into the Paris street in @Image2 - buildings, pavement, shopfronts, passers-by - while he rises straight out of the chair into a walking stride, outfit @Image7 unchanged. The camera settles into a smooth backward track leading him as he walks toward it at an easy confident pace.

Copy this **exactly.**

Part 3 of 5 — Tokyo, and into New York.

Shot 5 [4.5-6s]: Still tracking backward, the camera eases in to a close-up on his face. He glances up at the buildings and smiles faintly, light flickering across him as he passes under awnings.

Shot 6 [6-7.5s]: Signature move fires. The Paris street folds away and rebuilds into the Japanese restaurant in @Image3 - low wooden table, paper screens, warm lanterns - and his clothing changes in the same sweep to @Image8. The camera drops to near floor level and he comes to rest seated on the floor on the white cushion on the right side of the low table, legs folded, the steaming bowl of noodle soup from @Image9 in front of him, chopsticks in hand. Medium shot, warm low-key lantern light.

Shot 7 [7.5-9s]: Without cutting, the camera pushes in close as the chopsticks lift a tangle of noodles from the bowl, steam curling upward, and he eats and closes his eyes for a beat, enjoying it.

Shot 8 [9-10s]: Signature move fires. The restaurant dissolves and rebuilds into the street in @Image4 - signage, storefronts, pedestrians, evening light - while he rises from the floor straight into a walking stride, outfit @Image8 unchanged. The camera rises from floor level to standing eye level and settles into a backward track leading him.

Shot 9 [10-11s]: Still tracking backward, the camera eases in close on his face. Neon and signage streak across him in soft coloured bokeh as he looks around, relaxed.

Shot 10 [11-12s]: Signature move fires. The Japanese street folds away and rebuilds into the outdoor restaurant table at New York Times Square in @Image6 - cafe table, barriers, towering illuminated billboards, crowds - and his clothing changes in the same sweep to @Image10. He comes to rest seated at the table, lifting a large burger with both hands. Medium shot across the table, shifting billboard light.

Copy this **exactly**.

Part 4 of 5 — Times Square, and the landing.

Shot 11 [12-13s]: Without cutting, the camera pushes in close as he bites into the burger with his eyes shut, savouring it, then gives a slow satisfied nod.

Shot 12 [13-15s]: Signature move fires. The table dissolves and rebuilds into the heart of Times Square in @Image5 - billboards, crowds, yellow cabs, wet reflective pavement at night - while he rises from the chair straight into a walking stride, outfit @Image10 unchanged. The camera tracks backward leading him through the crowd, then eases in close on his face as he looks up at the billboards and smiles toward camera.

Shot 13 [15-16s]: One last fast eased push in toward his face as the entire city peels away and collapses inward - billboards, crowd and noise swallowed - rebuilding into the same plain ordinary interior room from Shot 1, his clothing returning to the simple everyday home clothing.

Shot 14 [16-17s]: He stands in the plain room facing camera, calm, flat neutral indoor light, static camera. Framing, light and wardrobe match @Video2 exactly.

Copy this **exactly**.

Part 5 of 5 — the sound design.

SOUND

This video will be laid under an existing continuous voiceover in the edit, so the generated audio must be **ambience and music ONLY**, mixed low, with clear headroom left across the whole track for a voice to sit on top.

Environmental ambience changes with each location and crossfades on every transition: quiet plain room tone at the start; Parisian street ambience with distant scooters, cutlery on porcelain, muted cafe chatter and light breeze; Paris street footsteps on stone, traffic hum, a bicycle bell; quiet Japanese restaurant with soft broth simmer, ceramic clinks and a distant sliding door; Japanese street footsteps, crowd murmur, shop jingles; dense Times Square crowd with taxi horns, distant sirens and street performer drums; then everything collapsing back to quiet room tone.

A soft airy whoosh on every environment change.

Music: a continuous low instrumental bed that shifts with each city - accordion in Paris, koto and shakuhachi in Japan, brass and drums in New York - building to its fullest in Times Square, then cutting away on the final transition. Keep the music well under the ambience.

No dialogue throughout the entire video. No background narration. No voices speaking. No singing. No lyrics. The man's mouth does not move as if talking in any shot after Shot 1.

Three countries, **one score.**

Generated instrumental, 24 seconds, starting the moment Paris appears.

STYLE

Cinematic travel-montage instrumental. Three national themes in rapid succession with hard musical cuts and no fades between them: French musette accordion cafe waltz, then Japanese koto, shakuhachi and soft taiko, then New York big-band brass with street drumline. Each theme lasts roughly five seconds and switches instantly. One steady tempo and a light continuous percussion pulse carries straight through all three changes so the piece reads as a single journey rather than three clips. Bright, playful, confident, film-trailer energy. Arranged with open midrange and space for a voiceover sitting on top. Fully instrumental.

STRUCTURE

[0:00 - FRANCE - solo accordion musette waltz, upright bass, brushed snare, warm Parisian cafe swing]
[0:05 - HARD SWITCH - JAPAN - koto and shakuhachi over a soft taiko heartbeat, sparse and elegant, same tempo]
[0:10 - HARD SWITCH - AMERICA - big band brass stabs, walking bass, snare drumline, bold New York swagger]
[0:15 - OUTRO - brass thins to a single sustained chord, warm resolve, settle to near silence by 0:21]

EXCLUDE

vocals, singing, choir, lyrics, spoken word, rap, ambient drone, slow ballad, lo-fi, sad, minor key dirge

It will not land the switches on the exact second. Generate three versions, pick the closest, and slide it in your editor.

Never pay full price for a **first attempt**.

The expensive mistake is rendering at maximum quality to find out your face drifts in the third scene. Climb the ladder instead — every rung answers a different question.

480p

Does the idea hold?

Is it one face the whole way through? Do the transitions land where you wrote them? Is anyone talking who should not be? You can read all of that at the lowest setting.

720p

Does the detail hold?

Check wardrobe, hands, food, signage and the seams where AI meets your real footage. Only climb here once the 480p pass is right.

1080p

Render once.

By now you are paying to finish something you already know works, not to find out whether it does.

Change one thing per attempt. If you rewrite the prompt, swap two images and change the duration all at once, you learn nothing from the result — you just paid for confusion.

AI Video Generation: Better Results, Lower Cost

A hands-on course on directing AI video like a filmmaker — and paying a fraction of what most people burn getting there.

I

Hold one face across a dozen worlds

Build reference sets that lock identity instead of hoping the model remembers what you look like.

II

Write camera movement in words

The specific vocabulary models actually respond to — and the phrasing that quietly gets ignored.

III

Cut your render bill

A testing ladder that finds your mistakes at the cheapest setting, before they cost you anything.

IV

Make the seam disappear

Match AI footage to real footage on framing, light and grade so nobody can find the join.

V

Direct sound, not just picture

Generate ambience and score that sit under your voice instead of fighting it.

VI

Ship a finished piece

Leave with one complete vertical video, built end to end, ready to post.

Taught the way I work: real briefs, real budgets, real client deadlines.

EARLY REGISTRATION

Register early, study for less.

Dates and full curriculum are announced shortly. Everyone on the early list gets the discounted seat — and gets it before registration opens publicly.

- Discounted rate for early registered students
- First access before seats open to everyone
- The full prompt library from every breakdown I post

Six outcomes. One finished video.

Join the early list

Tap the **link in my bio** to register your interest.

Save this post so you have the prompt when you need it.



YASER GHULOOM · YG MEDIA
@YASERGHM

Photography & media
production training