

Recognizing the City on Screen: City Mindscape, City Selection, and City Recognition

Jiaze Li*

Columbia University School of the Arts, New York, NY, United States

*jl6920@columbia.edu

Abstract

This article proposes a framework for understanding how spectators recognize cities in films by theorizing the relationship between city mindscape, city selection, and city recognition. Building on Kevin Lynch's notion of the "image of the city" and research on cognitive mapping, the article distinguishes experience-based "city images" from media-based "city illusions," and unifies them as a "city mindscape" – the composite mental landscape through which viewers perceive and remember urban space. On the production side, it introduces "city selection" to describe how filmmakers' sample, weight, and arrange urban elements through seven layers: spatial and temporal sampling, semantic choice, kinetic orchestration, photometric and chromatic design, sound, and editing. City recognition is defined as the moment when these organized selections align with viewers' prior mindscapes, generating a sense of urban identity that depends less on cartographic fidelity than on rhythmic, textural, and narrative plausibility. Methodologically, the article combines theory-driven modeling with close scene analysis. A case study of Hong Kong in Wong Kar-wai's *Fallen Angels* (1995) demonstrates how nocturnal lighting schemes, textures, recurrent spatial anchors, and a characteristic sonic rhythm render the city recognizable even in the absence of explicit landmarks or expository dialogue. The study proposes a portable analytic kit for decomposing cinematic cities into manipulable modules and tracing how films both reflect and reshape public understandings of urban space. It concludes by arguing that city recognition can serve as an interpretive key for analyzing narrative space, while also warning against the stabilization of clichéd urban imaginaries.

Keywords

Cinematic City; City Mindscape; City Selection; City Recognition; Hong Kong; Urban Space on Screen.

1. Introduction

Have you ever had the experience of watching a film and suddenly recognizing a place you have visited, or only previously encountered through other media? At other times, the city on screen may feel strangely at odds with the city you remember from lived experience. This tension between cinematic and lived urban space underlies my interest in how cities are represented on screen: how audiences recognize a city in a film, and why particular city settings carry such weight for a film's broader meaning.

Methodologically, this study adopts a theoretical-hypothetical approach. Chapter 1 formulates the core hypothesis and introduces the key terms. Building on Kevin Lynch's theory of the "image of the city" as the framework's foundation, I examine how mental images of cities-shaped by memory, experience, and media exposure-condition the ways we perceive and interpret cities on screen. I also coin the term "city illusion" to describe media-shaped imaginings of cities that may precede or exceed direct experience.

Chapter 2 turns to the question of city recognition. Rather than assuming that recognition depends on geographic fidelity or cartographic accuracy, I argue that a sense of recognition arises when the way viewers read the city in everyday life (through city images and city illusions) aligns with the way the film organizes urban elements. The chapter examines how recurring visual, sonic, and rhythmic patterns enable viewers to identify a specific city even in the absence of iconic landmarks and develops criteria for articulating this process in terms of observable cinematic choices such as composition, spatial organization, lighting, sound, and movement.

Chapter 3 focuses on Hong Kong as a case study, tracing recurring and recognizable elements across *Fallen Angels* and other Hong Kong films. By analyzing recurrent visual motifs, camera movements, architectural forms, soundscapes, and spatial arrangements, I show how a recognizable “Hong Kong at night” is constructed and reinforced. The chapter demonstrates how such elements both reflect and reshape audiences’ city mindscapes.

Across these chapters, the essay seeks to establish a concise yet adaptable model for understanding how films create recognizable cities—a composite of memory, geography, and cinematic technique. City recognition, I argue, exceeds simple location-spotting and functions as an interpretive key for narrative space and the film’s broader significance.

Analytically, I treat the environment—whether physical or fictional—as the independent variable, and the viewer’s perception as the response. “City” refers here to the perceivable arrangement of built forms, surfaces, and materials; spatial syntax (paths, edges, districts, nodes, landmarks); light and color; movement and sound; and everyday practices that can be organized on screen. I seek those qualities that contribute to mental attributes of identity and structure. The central claim is straightforward: audiences recognize a city when cinematic language—framing, camera movement, lighting and color, sound, and editing—organizes elements of urban form into patterns that align with their prior mindscapes of that city.

A substantial body of scholarship on film and the city underpins this project. While work on “city and cinema” is extensive, it is frequently abstract. For example, the landmark edited volume *The Cinematic City* (Clarke, 1997) collects studies of how cinema and the modern/postmodern city co-produce each other.[2] The essays show that film not only depicts urban spaces but also shapes audiences’ perception, memory, movement, and imagination of cities. Spanning the 1930s to the 1990s and drawing on genres such as documentaries, noir, exposés, art cinema, and science fiction, the volume traces loops between urban form, media technologies, and spectatorship across methods including urban history, cultural studies, and film theory. It offers a powerful conceptual and historical framework, which I draw on throughout this study. Yet, with a few exceptions—such as detailed analyses of the nocturnal noir city—it rarely translates its insights into practical methods for decomposing the big term “city” into manipulable components. My project takes up that task.

The present study offers a detailed, scene-level analysis that deconstructs the cinematic city into analyzable elements and connects them to film language. It breaks urban imagery down into modules such as lighting and exposure, architectural typologies and edges, soundscapes, and material textures, and links them with filmic devices such as shot scale and lensing, camera height and movement, blocking, color timing and grading, and sound design. It asks why certain combinations of cinematic and urban elements trigger city recognition, and others do not. The result is a portable analytic kit rather than a purely descriptive account. This kit provides a disciplined vocabulary for terms such as “city image,” “city illusion,” “city selection,” and “city mindscape,” and argues that films do not simply reproduce cities but assemble clusters of cues that resonate with audience memory. The method specifies these cues and demonstrates how to compare them across films, periods, and cities.

2. City Mindscape and City Selection

2.1. City Mindscape

Humans are small; cities are big. We cannot perceive a city's panorama from a god's-eye view. Instead, we assemble what we see and hear into a composite image. Our sense of the city emerges as a collage of partial views, sounds, and movements rather than as a complete, synoptic map.

In this study, I treat the city not as an administrative boundary or GPS polygon but as a sensory-structural field accessible to perception and cinematography. It includes built forms and surfaces such as skylines, blocks, façades, materials, ground textures, and the spatial syntax of paths, edges, districts, nodes, and landmarks. Light and color matter-glow, reflection, and contrast-as do movement and cadence: crowd tempo, traffic rhythms, escalators, overpasses, and approach-reveal sequences. The sound bed comprises languages, engines, water, reverberation, and environmental noise. Practices and props-silhouettes, postures, umbrellas, phones, kiosks-further characterize the city. I bracket wider social, political, and economic forces except when they become legible in form, such as signage regimes or land-use densities. This formal focus allows me to "uncover the role of form itself" and provide a scene-level account of which cues contribute to identity and how they assemble into cinematic structure.

Once we bracket what we are not analyzing, what remains is an assembled mental picture: a functional map constructed from fragments of walking, riding, reading signs, turning corners, feeling wind, and hearing the city's hum. Kevin Lynch names this composite as the "city image" in his book, *The Image of the City* (Lynch, 1960)[6]. It is not a photographic replica but a practical image-a tool for orientation and action crafted from perception and memory.

Lynch carefully analyzes how such images form. First, image-making is a two-way process: "the environment suggests distinctions and relations, and the observer...selects, organizes, and endows with meaning what he sees" (Lynch, 1960, p. 6).[6] The result is not simply "what is out there," but what is useful to us out there. Urban images become "legible" when three characteristics converge. First, identity: an entity must be distinguishable from others and recognizable as separate (Lynch, 1960, p. 8).[6] Second, spatial relations: the image must include the pattern relation of the object to the observer and to other objects, a fundamentally relational and dual process. Third, meaning: whether practical or emotional, the object must carry significance for the observer (Lynch, 1960, p. 8).[6] Meaning is itself a relation, beyond purely spatial or geometric relations.

For analytic purposes, Lynch organizes the physical supports of the city image into five recurrent elements: paths, edges, districts, nodes, and landmarks (Lynch, 1960, pp. 46–48).[6] First, paths are habitual lines of movement-streets, walkways, rail lines, canals-along which we perceive the city and locate other elements "along the way." Second, edges are boundaries or seams-shores, walls, cuts-that separate or connect areas. Third, districts are medium-to-large areas into which we can mentally "enter," recognizable by characteristic material, use, scale, or texture. Fourth, nodes are strategic points-junctions, squares, lobbies-where decisions concentrate and attention intensifies. Fifth, landmarks are external points of reference, from towers to tiny signs, highlighted by contrast and used for orientation. Crucially, these elements interact: districts are "structured with nodes, defined by edges, penetrated by paths, and sprinkled with landmarks" (Lynch, 1960, p. 48).[6] The image is the organized arrangement of parts, not a mere list of elements. It is a structured experience unfolding over time-precisely the temporal character that cinema shares.

For viewers with lived contact, the "city image" develops through practical engagements with paths, edges, districts, nodes, and landmarks. For viewers without such contact, there is still an image of the city, though Lynch himself says little about it. This image arises primarily from

mediated encounters and can be vivid yet uneven. I term this media-shaped construct “city illusion.”

Yomi Braester’s *Painting the City Red* (2010)[1] extends this point through analyses of films from mainland China and Taiwan. Film, theater, and urban planning operate as intertwined instruments that co-produce images of the city, from socialist display cultures to reform-era spectacles. Such images shape generational stereotypes. The association of red with socialism and “China”—reinforced by propaganda, festivals, and traditional aesthetics—encourages audiences to imagine Chinese cities as saturated with red decorations. For my model, this widens the notion of “city illusion” beyond direct depictions of cityscapes in film, television, radio, newspapers, and advertising, to include interpretations of abstract concepts. Film provides blueprints through which audiences imagine what a city is or could be.

I argue that both “city image” (experience-based) and “city illusion” (media-based) are not “authentic” in the strict cartographic sense. Before anything is “remembered,” it is pre-processed by purposes, affects, and schemata. Cognitive maps may exaggerate or ignore certain parts of the city or reorganize the overall layout. Just as Barbara Tversky writes in “Cognitive Maps, Cognitive Collages, and Spatial Mental Models” (1993), she described that “city illusion” is a collage rather than a coherent map:

“It is unlikely that the pieces of information can or will be organized into a single, Coherent map-like cognitive structure. In these cases, rather than resembling maps, people’s internal representations seem to be more like collages. Collages are thematic overlays of multimedia from different points of view. They lack the coherence of maps, but do contain figures, partial information, and differing perspectives.” (1993, 15)[8]

Lynch’s own metaphors echo this perspective. He likens the cognitive map to a drawing on “an infinitely flexible rubber sheet; directions were twisted, distances stretched or compressed” (Lynch, 1960, p. 47).[6] The internal map cannot be authenticated like a GPS track or a mathematically measured survey. To capture this, I use the umbrella term “city mindscape” to describe these layered, selective, and distorted internal landscapes. The term highlights the limited credibility of cognitive maps while redirecting attention to the “landscape in our minds.”

So far, I have focused on the “hardware” of the city mindscape: built form, texture, light, sound, and spatial syntax. Yet “software” also plays a crucial role. When people describe New York, for example, they often invoke diversity and opportunity. Ethnic heterogeneity becomes part of the city’s image. While some metropolises may look similar in plan or architecture, the degree and visibility of racial and cultural diversity contribute significantly to how they are distinguished. In New York, people from diverse backgrounds live in close proximity, with different histories and habits that continually collide and mix. Spike Lee’s *Do the Right Thing* (1989) provides a vivid example: set in a predominantly Black Brooklyn neighborhood, it stages frictions and solidarities among Black residents, an Italian American pizzeria owner, and Korean shopkeepers. The streets become a melting pot of Black, white, and Asian residents, and the interplay of ethnicities and everyday customs drives the film’s conflict. The film’s celebrated status in American cinema underlines how ethnic “software” can be central to a city’s mindscape.

Costume also contributes to this “software.” People from different regions cultivate particular clothing styles, and visitors to New York quickly encounter a wide array of fashion trends and stylistic hybrids. Individuals from different social backgrounds—corporate, creative, working class, immigrant—mark their identities through dress, producing a dense layering of styles. Compared with Los Angeles, where a relaxed West Coast aesthetic often dominates, East Coast fashion tends to be more structured and tailored. Subways and sidewalks resemble moving runways where subcultures, corporate identities, luxury brands, thrift finds, and diasporic

styles intersect. On screen, wardrobe is not merely decorative; it functions as urban software, selecting and amplifying the city's mindscape.

A third frequently overlooked dimension is texture. By texture I refer to the tactile and visual qualities of materials: glass reflections, the grain of stone, the roughness of concrete, the patina of metal, the slickness of wet asphalt. Different technological conditions and historical periods leave different material legacies. In older European cities such as Rome, much infrastructure retains stone construction; façades often appear rough and weathered. In relatively newer cities such as New York, red brick and steel dominate the urban fabric. From the late 19th century through the post-World War II decades, the United States was the world's largest steel producer, which helps explain why New York feels structurally sharp and materially cold. Even when the same material is used, texture varies: recipes, cooling processes, and manufacturing techniques differ; joints and bends reflect precision and tolerance; weather and use add patina; size, thickness, and maintenance affect roughness or smoothness. On camera, lighting and color grading intensify or transform these textures, making them appear warmer, harsher, softer, or more brittle. Recognizing urban texture thus becomes a key component of recognizing a particular city.

Taken together, these hardware and software dimensions-spatial elements, sensory cues, symbolic associations, and cultural practices-compose the "city mindscape," the internal landscape through which viewers approach and make sense of cities on screen.

2.2. City Selection

So far, I have worked from the side of perception, asking how inhabitants and audiences construct city mindscapes. I now turn to the transmitting side: the role of media, and especially the camera, in deciding what parts of the city are shown and how. Why must city elements be selected at all?

In early cinema, films often presented the cities through the means of view. Audiences were interested in watching 'actualities' to glimpse distant places and modern life. Tom Gunning once commented on the view of the city in this way: "Nearly all early film documents present a *mise en abîme* of audiences filling vaudeville halls from busy city streets in order to see projected on the screen – busy city streets." (as cited in Weihsmann, 1997, p8)[10], which exactly shows the attractions of the view that presents the city. It's fascinating that people on the street flock to the theater to watch films that depict bustling city streets. From my perspective, films offer a more imaginative way to envision a city by flattening its three-dimensional space into a flat image. However, when people become too involved in the city's operations, they may lose its inherent beauty. Therefore, the portrait of the city in the films gives the audience an opportunity to appreciate the city. Although it may or may not be accurate, film shows the aesthetics of the city. Thus, people like this way of watching "their" lives. Just as Zhu Guangqian wrote in the book *Tan Mei (On Beauty)* (1932):

"When we view things with a utilitarian mindset, they are merely tools or obstacles in practical life, capable only of stirring desire or aversion. To perceive the beauty proper to things themselves, we must step beyond utility and, in a spirit of doing something "for its own sake," appreciate their very forms. In short, beauty keeps a distance from practical life; to see a thing's own beauty, we must set it at the proper distance and look from there. (1932, 15) [11]"

Cinematic cityscapes therefore captivate audiences through a productive distance. By selectively stylizing and rhythmically arranging fragments, film invites viewers to project a city mindscape beyond the literal streets. This logic becomes especially clear in the city symphony films of the 1920s. These works assemble urban life into non-narrative yet rhythmically structured films that follow "a day in the city." They do not offer complete maps or single points of view; instead, they organize the city through montage, tempos, and thematic sections-sometimes celebratory, sometimes critical. Some glorify speed and machinery; others

juxtapose boulevards and back alleys to expose class textures. What unites them is not cartographic completeness, but the insistence that the city becomes legible through selective arrangement. Scott McQuire also mentions this strategy in his book, *The Media City: Media, Architecture and Urban Space* (2008), “Vertov (1984: 18–19) outlined the strategy which will be put to work in *Man with the Movie Camera* (1929, Dziga Vertov):

A day of visual impressions has passed. How is one to construct the impressions of the day into an effective whole, a visual study? If one films everything the eye has seen, the result, of course, will be a jumble. If one skillfully edits what has been photographed, the result will be clearer. If one scrapes bothersome waste, it will be better still” (as quoted by McQuire, 2008, p73)[7].

He quoted Vertov’s work to emphasize the importance of the selection. Only the selected material can be organized organically and generate what is to be called a film.

I use the term “city selection” for this systematic sampling, weighting, and sequencing of urban elements into representations. Many media-radio, advertising, television-select city elements in their own ways. In this essay, however, I focus on how films construct city selections; other media appear only insofar as they are incorporated into films as elements (for example, billboards, television screens, or radio voices) rather than as independent systems.

City selection operates across seven interrelated production layers. The first is spatial sampling: decisions about locations, camera positions, camera height, lens choice and focal length, depth of field, and framing. These decisions determine where we look from and what we see, and different spatial samples yield different urban structures and identities. A high-angle telephoto shot can flatten a block grid to emphasize financial density and modern order, suppressing the messiness of street-level alleys. A low camera with a wide-angle lens can immerse us in the crowd, highlighting intimacy or congestion.

The second layer is temporal sampling: time of day, season, weather, and event cadence. Temporal choices are crucial both narratively and affectively. Rainy nights with neon reflections support melancholy or futuristic moods; sunny mornings with high saturation and contrast showcase a garden’s beauty and imply optimism. Temporal sampling shapes how the city feels in time.

The third layer is semantic choice: the foregrounding of paths, edges, districts, nodes, landmarks, bodies, postures, props, vehicles, and signage. Semantic choices communicate what the city is supposed to represent. To depict a city as hyper-modern, filmmakers may emphasize skyscrapers, glass façades, and corporate districts, rather than factories or informal settlements. Levels of illumination, for instance, can signal economic power. Anne Cronin (2010) notes how the illumination of city space at night reorganizes urban experience, lighting up the sky with signs of consumption. In places like Times Square, illuminated signs and billboards become iconic in themselves. Language and text (street names, shop signs, billboards) can signal ethnic and historical layers.[3] Props and postures-umbrellas and masks, mobile phones and selfie sticks-index lifestyle and period. Through such choices, films construct what their cities are taken to stand for.

The fourth layer is kinetic orchestration: the composition of movements of bodies, vehicles, and cameras. Motion patterns provide the city’s heartbeat and influence perceived speed, density, tension, and flow. In *North by Northwest* (Hitchcock, 1959), for example, the opening sequence uses reflections, crowd flows, and elevator and escalator movements to stage New York as dense and hectic before the plot even begins. Intersecting movement lines-pedestrians crossing, commuters entering subways, vehicles threading across the frame-convey bustling urban rhythm. Kinetic orchestration also involves the interplay of linear rhythms (journeys that do not return to their starting point) and cyclical rhythms (day–night shifts, commuter cycles), echoing Lefebvre’s (2004) rhythm analysis.[11]

The fifth layer comprises photometric and chromatic properties: exposure, contrast, fill ratio, white balance, saturation, film grain and glow, halation, and related choices. These parameters determine the perceptual “climate” of a location-cold or warm, soft or hard-and can make the same place feel radically different, or make different places feel surprisingly similar. Her (Jonze, 2013) illustrates this power: although shot in multiple cities, consecutive shots of subway interiors and exteriors read as a continuous urban environment because their exposure, contrast, saturation, and white balance align. Diffusion filters, controlled halation, and carefully tuned warmth and softness integrate Shanghai and Los Angeles into a single, plausible city. Photometric and chromatic decisions can lend a powerful sense of “truth” that exceeds literal geographic accuracy.

What I analyzed above is all in the visual layer, but what cannot be ignored is the sixth layer: sonic bed. By shaping the ambient sound layer, this layer anchors scale, place, and social texture within the city’s auditory environment. While some locations may appear similar, the unique sonic backdrop sets them apart. For instance, despite being a densely populated global city, Shanghai maintains a remarkably quieter atmosphere compared to other metropolitan areas. The absence of the harsh engine roar and motorcycle and car horns on the streets is the primary reason for this. The city has banned gas-powered motorcycles and honking in the core, which has become an integral part of its auditory identity. Adding specific ambient sounds, such as languages, dialects, reverberation, water, wind, podcasts, and radio, will determine the scale of the city and showcase its unique identity.

The seventh layer is film editing. Editing does more than assemble footage into a film; it stitches city fragments into a cinematic mindscape. Shot length, cutting rate, average shot length (ASL), color grading, sound motifs, sound mixing, split screens, and film emulation all influence how fragments are combined. Different editing strategies produce different city illusions: rapid cuts and jump cuts can produce a frenetic, disorienting metropolis; long takes and slow cutting rhythms can produce a contemplative or emptied city.

Because city selection is patterned and repeatable, it carries the risk of cliché. Reliance on a single habitual mode-certain skyline angles, familiar traffic montages, stereotypical night shots-can harden into a narrow city illusion that reinforces stereotypes and circulates misleading or overly selective images. At the same time, city selection interacts reflexively with city mindscapes: films construct exportable city versions that feed back into how audiences remember, navigate, and judge actual places. Repeated viewings crystallize into expectations that subsequent productions either inherit or challenge. Analyzing city selection in relation to city mindscapes therefore allows us to understand cinematic cities as evolving, negotiated representations rather than neutral replicas.

3. City Recognition

When we recognize a city on screen, what exactly is happening? Does recognition depend on a one-to-one replication of geographic reality, or something else? The previous chapter argued that “city mindscapes” do not rely on GPS fidelity but on collages of perception, memory, and media exposure. They are not precise maps of reality but cognitive and affective constructions. The question, then, is whether city recognition in film follows the same standard: does it depend on objective depiction, or on a sense of truth?

I argue that city recognition is anchored less in cartographic coordinates than in this sense of truth. Recognition occurs when the film’s organization of city selections aligns with viewers’ pre-existing city mindscapes, producing a felt sense of identity. This is a dual and historically variable process. As urbanization fragments experience and memory, and as media proliferate, the ways we perceive cities change-and so do the ways films present them. David B. Clarke (1997) notes that a world once perceived “as a living whole” can no longer be experienced as

coherent and complete (p. 4). In such conditions, recognition depends on sequences that accept fragmentation but arrange it logically enough to be intelligible. Viewers use imagination to complete the gaps.

City recognition is not secured by merely piling up “objective” elements-skyscrapers, bridges, or iconic landmarks-but by how those elements are selected, weighted, and composed. What persuades is the orchestration that binds disparate city selections into an intelligible rhythm: camera vantage and angle, combinations of light and weather, cadences of bodies and traffic, textures of sound. These arrangements produce a sense of the city’s “voice,” within which even non-indexical shots can feel appropriate. Conversely, when the rhythm of city selection shifts-when pacing slows, colors desaturate, soundscapes thin out-the city takes on a different character. Recognition is less about discovering fixed boundaries than about accepting a convincing arrangement. A city becomes recognizable not because its coordinates are predetermined but because its elements are woven into a coherent and plausible chain of appearances.

This perspective resonates with Marcel L’Herbier’s description of film as a “dialectical unity of the real and the unreal” (as cited in Virilio, 1991, p. 65) [9]and with Lefebvre’s suggestion that rhythm analysts are more aware of moods and atmospheres than of discrete images or spectacles (Lefebvre, 2004, p. 81).[5] For city recognition, fidelity is less important than a sense of truth-an atmosphere that feels right, even if it is not literally accurate.

In the previous discussion, I emphasized that city recognition is achieved through a structured selection of city elements rather than a precise coordinate match. The most appropriate analytic lens for this is narrative. Filmic sequences naturally align with human perceptual processes. On screen, we assemble scenes by following shot order and causality; in the city, as we walk or drive, we stitch fragments together in a similar manner. As Richard Koeck (2013) observes, “as we walk or drive with open eyes through cities today, we make our way through sequential configurations of space” (p. 40).[4] Buildings and streets are “edited” by movement and looking into continuous units of perception. Explaining city recognition in narrative terms is therefore not an external imposition but a faithful description of how we see and feel in everyday life.

On this basis, I narrow the focus to sequence analysis that can trigger recognition without relying on conspicuous landmarks. I deliberately set aside landmark-based recognition because landmarks are too obvious; their low threshold of identification can obscure the subtler work of narrative sequencing and urban grammar. I analyze sequential spatial narratives as the basic unit and explore four dimensions that generate city recognition: temporality and rhythm, spatial configurations, and semantic shifts.

I conceptualize city recognition as a narrative act. Instead of verifying landmarks one by one, I trace how the camera weaves fragmented city selections into sequences that resemble the pathways we experience in reality. A spatial configuration is the temporary arrangement of space, bodies, light and shadow, traffic, crowds, signage, and sound at a specific moment. A city is not recognized by a single configuration but by a sequence of such configurations unfolding in a particular order and rhythm.

Walking and driving are already stories that the body narrates: a brief halt, a scan, a step off, a crossing, an entrance into an arcade, a tap on a gate, a descent of stairs, a wait on the platform. Cinema does not invent this sequence; it visualizes it. Shot order mirrors action order, and cuts are typically placed at natural event boundaries that we feel in real time. Shot duration tends to correspond to the real-time duration of an act: a quick point-of-view shot for a glance, a longer lateral move for wandering. Rhythm is equally important. Shot length, cutting rate, intervals of footsteps and traffic, and the density of the sound bed together establish a city’s rhythm. Once viewers internalize that rhythm, recognition can be triggered by only a few cues-

a braking hiss, a subway rumble. At the same time, deviations from an established rhythm can produce surprise or alienation. A prolonged shot, a sudden flurry of cuts, or a drop to near-silence can make the city feel hostile or strange. Rhythm thus both normalizes and disrupts urban experience.

This isomorphism extends to spatial logic. We comprehend a city by following coherent routes, relying on stable signage, and anticipating predictable turns. Classical continuity editing—respecting the axis, matching eyelines and actions—is the audiovisual articulation of this cognitive habit. Consider a simple crossing: a wide shot shows a character approaching a corner; a close-up reveals a foot landing on the crosswalk; a reverse wide shot shows the far curb. The sequence parallels how we actually register the crossing: attention moves from the corner, to the step, to the destination. Doors, stairs, escalators, and turnstiles guide us through space; tracking, craning, panning, and lateral moves stitch fragments into visible paths. A push-in invites entry; a tilt down signals descent; a follow shot maintains movement; a side tracking shot aligns with parallel flows. The camera does not invent the route; it visualizes the route the body would already know.

Different cities lend themselves to different narrative logics. In New York, deep focus and dense *mise-en-scène* can encapsulate a complete micro-narrative in a single take. In *Do the Right Thing*, for example, the camera lingers on Sal's pizzeria and the sidewalk: in the foreground, an exchange unfolds; in the midground, an argument intensifies; in the background, neighbors and police circulate. Without cutting, the viewer perceives layered actions and high urban density. Static setups inside diners can maintain a seamless relation between foreground dialogue and background street life. New York's density allows micro-narratives to be "stacked" in one view, and recognition is tied to this simultaneity. In more dispersed Los Angeles spaces, the same micro-narrative often requires segmented coverage: a wide shot of the arterial, a ramp approach, a closer shot of the doorway, then interior medium shots and close-ups, punctuated by dashboard views and signage. A single glance cannot encompass the spatial spread; sequences must be "chained" across time. This formal contrast itself becomes a recognition cue: New York appears compressed, multi-voiced, and dizzying; Los Angeles appears sparse, route-driven, and time-extended.

Semantic elements can rapidly alter a location's meaning without changing its basic spatial frame. A few substitutions can transform a familiar street into a different kind of place. Imagine a church bell replaced by a stock-market bell, bright shop neon replaced by formal notice boards, and a street vendor replaced by a lobby security guard. Anchors such as the doorway, curb, sign position, and camera perspective remain constant, so the street is still recognizable as the same place. Yet shifts in sound and surface details—the bell's tone, the sign's typography, the clothing and posture of figures, the texture of footsteps—alter the location's semantic category. We still recognize the street, but we understand it differently. City recognition operates not only at the level of "what city is this?" but also "what kind of place within this city is this, now?"

4. Hong Kong's City Mindscape

Fallen Angels (1995) intertwines two lonely stories that cross paths in nocturnal Hong Kong. One thread follows a hitman who wishes to retire and his female partner, who secretly loves him and keeps their connection alive by sifting through traces of his daily life. The hitman briefly falls for a "blonde" woman and is eventually killed during a job. The other thread follows mute drifter He Zhiwu, who "reopens" closed shops at night to do business in his own eccentric way, cares for his elderly father, and falls for a young woman named Charlie, obsessed with her ex-boyfriend. The two narratives converge when He Zhiwu gives the female partner a late-night

ride. Together they form a portrait of Hong Kong's night city-neon, rain, speed, desire-in which everyone searches for connection before it slips away.

To examine Hong Kong's cinematic mindscape, I analyze the film from the perspective of city selection, combining the seven layers discussed earlier. The establishment of the city begins from the first frame. The opening monochrome shot uses high-key lighting and a tilted camera to convey tension and uncertainty. In a medium shot, the woman trembles and smokes in the right foreground, partially in shadow, while the man sits behind her, lit more prominently. A wide-angle lens subtly distorts her head, producing a sense of alienation and unease. The high-contrast black-and-white palette makes her oversized diamond ring and his suit pop against the surroundings. Film grain is pronounced, underlining their isolation. This shot sets the tone: solitary, uneasy, stylish. At this point we do not yet know where we are, but the city's atmosphere is already being constructed.

After this black-and-white prelude, a sequence shows the woman returning "home." The first four shots follow her through a subway station: quick handheld cuts, unbalanced compositions, and bold focal lengths disorient the viewer. The architecture's yellow pillars and walls, bright strip lights on the floor, and reflective handrails create a visually striking but unstable space. A medium-back-shot shows her striding through the station; the next long shots add more spatial information, but rapid cuts and blocked sightlines prevent us from stabilizing the layout. She appears on different sides of the frame; pillars interrupt our view. The camera then shifts to a high, wide angle as she rides an escalator, reinforcing instability. On the sonic layer, dialogue gives way to a driving, percussive underscore that echoes a heartbeat, then to the rhythm of high heels and the mechanical clatter of escalators and passing trains. The interplay of quick cuts, canted angles, and percussive sound design creates a kinetic rhythm that we come to recognize as Hong Kong's nocturnal tempo.

Once this atmosphere is established, subsequent sequences reveal the city more explicitly. Following a long back shot of the woman on the escalator, a wide environment shot introduces a leafy tropical plant, peeling paint, Chinese-language billboards under warm light, and a train passing on an elevated track. She retrieves a key through a pitch-black hole in a broken patterned glass window-an iconic texture of Southern Chinese architecture. Two close-ups of the glass emphasize its distinctive pattern. Mirrored shots then show her walking along the corridor: first from behind, then as a reflection in the window. When she enters her room and turns on the light, the clock flickers before settling into a cyan glow produced by a mercury lamp. She kneels in latex and rubber gloves, scrubbing around a red plastic bucket while a train rumbles past outside. Quick cuts turn housekeeping into a tense, almost ritualistic sequence. Mold and condensation stain the walls; orange streetlight seeps through the window and mingles with greenish fluorescent light-a hallmark of Hong Kong nights that distorts space and heightens unease.

A subsequent sequence maintains these visual and rhythmic traits while deepening spatial detail. Wider lenses and steeper angles reveal that a billboard, previously seen in isolation, is fixed to the side of a small building; on the other side, stacks of corrugated material accumulate. In the foreground, the hitman walks across a highly reflective surface that mirrors lights and passing trains, suggesting recent rain. The camera follows him from exterior into corridor, where increased illumination exposes damp floors and mildew-scarred walls. Inside, cramped interiors and low ceilings emphasize the iconic smallness of Hong Kong flats. Half-open patterned windows frame distant trains, which keep slicing through the background. These recurring textures-mildewed walls, corrugated stacks, patterned glass, reflective floors, right-hand-drive buses-accumulate into a collage of Hong Kong.

Interior scenes, such as a laundry room or a small eatery with mahjong tiles, add further cultural specificity. Yet it is the exteriors, especially at night, that ultimately define the city. Wong Kar-wai overwhelmingly favors night; only a single daylight shot punctuates a film otherwise

saturated with darkness, neon, and artificial light. Christopher Doyle's cinematography often uses long exposure and step printing to stretch light trails and blur motion, creating a sense that the city is bustling yet emotionally distant-present but not centered on any one individual. This validates the argument that city recognition does not depend on clear geographic landmarks. What matters is whether the film's audiovisual language resonates with our urban experience.

Different modes of transit anchor the city's identity. When the female partner cannot cope with her ambiguous relationship, she takes the subway home. A melancholic voice-over accompanies the train's arrival, and through her window we see the city reduced to streaks of colored light. Later, after a killing, the hitman rides a bus home. A tilted shot frames him and an old classmate, suggesting an unstable bond, while glimpses through the front windshield reveal right-hand driving and multi-colored neon signs in Chinese. We may not fully read each sign, but the combination of British-style infrastructure and Chinese signage strongly cues Hong Kong's colonial and postcolonial context.

Butcher shops recur throughout the film, marked by hanging slabs of pork and whole pigs, red lighting, and metal shutters. In many Chinese markets, red lighting and red tents are used to make meat look fresher and more appealing. The film's repeated return to these spaces taps into viewers' everyday visual memory of local markets and anchors the city's Chinese character.

Wong Kar-wai also returns intermittently to black-and-white footage. Stripping away neon color emphasizes loneliness and disconnection, while increased contrast makes shutters, staircases, corridors, and reflective metal or glass surfaces stand out graphically. These high-contrast images turn the night city into a stark graphic environment that is easy to remember and re-recognize.

Throughout the film, bold white-balance choices push images toward cyan or amber, creating a distinctive teal-orange split. Neon signage and street advertisements appear hotter or cooler than they might in reality, yet feel "right" for Hong Kong. Production design and costume work in tandem with this grading. The female partner's red leather dress catches sharp highlights under green light; the "blonde" woman's golden hair vibrates against green interiors and red signage. Together, graded white balance, lighting, and costume craft a high-contrast color story that frames Hong Kong as a fashion-forward night city: vivid, graphic, and instantly recognizable.

In *Fallen Angels*, the seven layers of city selection operate in concert. Spatial sampling uses wide lenses, canted frames, dense mise-en-scène, and high or low vantage points to distort space and infuse the city with unease. Temporal sampling confines the narrative mostly to wet nights, rendering mirror-bright streets under sodium orange and green fluorescent light. Semantic choices focus on southern Chinese textures-mildewed walls, peeling paint, patterned windows, corrugated stacks, butcher stalls, and right-hand-drive buses. Kinetic orchestration synchronizes quick walking, escalators, subways, and traffic with step printing and long exposures that smear movement into light streaks. Photometric and chromatic decisions exaggerate saturation and halation so that neon appears to overflow the frame. The sonic bed layers corporate Cantonese voice-over, subway roars, escalator clicks, heel strikes, and bus engines into a fast-living rhythm. Editing uses rapid cuts and jump cuts to mirror the city's pace and reinforce disorientation. From this intricate arrangement, city recognition emerges. Recurring anchors-patterned windows with passing trains, narrow corridors, damp floors, mildewed walls, red butcher lights, Chinese lightboxes, and right-hand traffic conventions-stitch together a cinematic Hong Kong that feels unmistakable, even when the film withholds conventional landmarks.

5. Conclusion

This study set out to answer a deceptively simple question: how do viewers recognize a city on screen? Rather than treating cinematic cities as straightforward reproductions of geographic reality, the article has argued that recognition arises from an encounter between viewers' pre-existing city mindscapes and the film's organized patterns of city selection. City recognition is therefore neither a matter of verifying cartographic accuracy nor of ticking off famous landmarks. It is an interpretive event that occurs when cinematic form, urban form, and individual memory converge into a persuasive sense of urban truth.

On the perceptual side, I proposed the concept of the city mindscape to bring together Lynch's experience-based "city image" and media-generated "city illusions." The mindscape names the layered collage of paths, edges, districts, nodes, landmarks, textures, sounds, practices, and abstract ideas that people accumulate through both lived and mediated contact with urban environments. This concept foregrounds the selective, schematic, and sometimes distorted character of cognitive maps. Rather than lamenting their inaccuracy, the article treats these distortions as central to how cities are actually remembered, imagined, and later reactivated in spectatorship.

On the production side, I introduced the notion of city selection to describe how films sample, weight, and arrange urban elements through seven interrelated layers: spatial sampling, temporal sampling, semantic choice, kinetic orchestration, photometric and chromatic design, sound design, and editing. This layered model shifts attention from questions of "faithful representation" to the compositional labor through which films build exportable versions of cities. By specifying how camera position, lensing, lighting, rhythm, and sonic textures reconfigure paths, edges, districts, and landmarks, city selection connects film form to urban morphology in a more granular way.

Taken together, city mindscape and city selection provide the conditions for city recognition. Recognition was defined not as the identification of a fixed geographic object but as a narrative and rhythmic alignment between viewers' internalized urban grammars and the film's sequential organization of city fragments. The analysis of city recognition therefore focused on spatial configurations, temporal rhythms, and semantic shifts rather than on spectacular monuments. It highlighted how small details of texture, cadence, and sound can anchor a sense of place even when visual information is partial, stylized, or disorienting.

The case study of Hong Kong in Wong Kar-wai's *Fallen Angels* demonstrated how this framework can be operationalized in close reading. By tracking the recurrence of patterned windows, elevated trains, narrow corridors, mildewed walls, butcher stalls, neon halos, and right-hand-drive buses, the analysis showed how the film constructs a nocturnal Hong Kong that is instantly recognizable despite the near absence of overt landmark shots or expository dialogue. The seven layers of city selection converge to produce a humid, high-density urban rhythm, while the narrative stitching of sequences invites viewers to experience Hong Kong as an affective, fragmented, yet coherent night city. The result is a city that feels "true" without ever claiming to be exhaustive, stable, or purely documentary.

Methodologically, the article contributes a portable analytic kit for decomposing cinematic cities into manipulable modules. By specifying how particular constellations of spatial sampling, motion, texture, light, color, and sound cue recognition, the model offers a vocabulary that can travel across different films, periods, and urban contexts. It is intended not as a rigid checklist but as a flexible toolbox that can be adapted, expanded, or contested in dialogue with specific case studies.

Finally, the argument carries a critical implication. Because city recognition depends on repeated patterns of selection, cinema participates in stabilizing powerful clichés of urban life. Familiar clusters of neon, crowding, and rain can make Hong Kong instantly legible, just as

certain views of skyscrapers, bridges, or freeways repeat established imaginaries of New York or Los Angeles. The same mechanisms that enable recognition can therefore also narrow the range of possible urban futures by circulating highly selective and uneven images of cities. By making the compositional labor of city selection visible, the framework developed here encourages both scholars and practitioners to attend to what is excluded as well as what is shown, and to imagine cinematic cities that refresh, complicate, or resist inherited mindscapes rather than simply confirming them.

6. Limitations and Directions for Future Research

Several limitations of this study should be acknowledged. First, the argument is developed primarily through a single extended case study of Hong Kong, with only occasional reference to other cities such as New York or Los Angeles. This focus allows for detailed scene-level analysis but inevitably raises questions of generalizability. Urban histories, planning regimes, and media ecologies differ widely, and the framework has yet to be systematically tested across other cinematic traditions and geographic contexts.

Second, the analysis concentrates on narrative feature films and implicitly assumes an individual, cinema-based spectator. Other media forms-such as television series, streaming platforms, games, virtual tours, or immersive environments-may organize city selection according to different temporal rhythms and address more fragmented, networked viewing practices. The model does not yet account for how binge watching, algorithmic recommendation, or user-controlled navigation reshape the accumulation and revision of city mindscapes.

Third, the study is largely theoretical and interpretive. While it draws on research on cognitive mapping to frame the concept of city mindscape, it does not empirically measure how viewers recognize cities or how their mental images change after exposure to specific films. The claim that recognition arises when cinematic selections resonate with pre-existing mindscapes remains a plausible but untested hypothesis.

These limitations suggest several directions for further research. Comparative projects could apply the analytic kit developed here to multiple cities within a single filmography or to different national cinemas that represent the same city. Longitudinal studies might trace how the recognizable image of a city evolves over decades as urban form and media technologies change. Audience research-using interviews, mapping tasks, or experimental designs-could investigate how prior travel, migration histories, or media exposure shape city mindscapes and recognition thresholds. Finally, extending the concepts of city mindscape, city selection, and city recognition beyond cinema to interactive and platform-based media would help clarify how contemporary screen environments jointly script what cities are imagined to be and for whom they are imagined.

Filmography

- Allen, W. (Director). (1979). *Manhattan* [Film].
- Chan, P. (Director). (1996). *Comrades: Almost a love story* [Film].
- Harron, M. (Director). (2000). *American psycho* [Film].
- Hitchcock, A. (Director). (1959). *North by Northwest* [Film].
- Jonze, S. (Director). (2013). *Her* [Film].
- Kubrick, S. (Director). (1968). *2001: A space odyssey* [Film].
- Lee, S. (Director). (1989). *Do the right thing* [Film].
- Phillips, T. (Director). (2019). *Joker* [Film].

- Scott, R. (Director). (1982). Blade runner [Film].
- Shi, R. (Director). (1998). A beautiful new world [Film].
- Vertov, D. (Director). (1929). Man with the movie camera [Film].
- Wong, K.-w. (Director). (1995). Fallen angels [Film].
- Yuan, M. (Director). (1937). Street angel [Film].

References

- [1] Braester, Y. (2010). Painting the city red: Chinese cinema and the urban contract. Duke University Press.
- [2] Clarke, D. B. (Ed.). (1997). The cinematic city. Routledge.
- [3] Cronin, A. M. (2010). Advertising, commercial spaces and the urban. Palgrave Macmillan.
- [4] Koeck, R. (2013). Cine-scapes: Cinematic spaces in architecture and cities. Routledge.
- [5] Lefebvre, H. (2004). Rhythmanalysis: Space, time and everyday life. Continuum.
- [6] Lynch, K. (1960). The image of the city. MIT Press.
- [7] McQuire, S. (2008). The media city: Media, architecture and urban space. SAGE Publications.
- [8] Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), Spatial information theory: A theoretical basis for GIS (pp. 14–24). Springer.
- [9] Virilio, P. (1991). The aesthetics of disappearance. Verso.
- [10] Weihsmann, H. (1997). The city in twilight: Charting the genre of the 'city film' 1900–1930. In F. Penz & M. Thomas (Eds.), Cinema & architecture: Méliès, Mallet-Stevens, multimedia (pp. 8–27). BFI Publishing.
- [11] Zhu, G. (1932). Tan mei [On beauty]. Commercial Press.