Improved Cinematic-Guided Camera Language Transfer in 3D Scene

Supplementary Material

8. More Details about Cinematic Feature and Camera Language Losses Definition

In this work, we systematically define the objectives of camera language transfer, grounding them in professional cinematography literature. we conduct an in-depth review of cinematography literature to identify eight key cinematic features and encode them into five novel camera language losses. In film theory, camera language is expressed through camerawork, which includes shot size, framing, camera angle, camera movement, lens choice, camera position, focus, and zooming[44][9].

- **Shot Size** Shot size describes how much of the subject(typically a person) appears within the frame. Directors use shot size to manipulate visual composition and guide audience perception, thereby enhancing narrative expression.
- **Framing** Framing refers to the subject's position and spatial layout within the frame. It is used to guide narrative emphasis, evoke emotional responses, and maintain visual coherence.
- **Camera Angle** Camera angle means the specific location and orientation of the camera, which directly affects the viewer's perceived relationship with on-screen characters or space.
- **Camera Movement** Camera movement introduces temporal variation across frames by shifting the camera's position during a shot, thereby influencing temporal movement.
- **Lens Choice** Lens choice allows filmmakers to manipulate camera perspective[28](see Section 11.9) to enhance their visual storytelling, thereby influencing the audience's perception of the scene's depth cues and, ultimately, the construction of filmic space[7].
- **Camera Position** While camera position contributes to determining shot size, framing, camera angle and camera movement, its effects are already implicitly captured in these components and is therefore not modeled separately in our formulation.
- **Zooming** Zooming used for temporal changes in shot size, is also functionally subsumed under our shot size modeling and thus not treated as an independent factor.
- **Focus** Focus, which is closely tied to depth of field, determines which parts of the scene appear sharp or blurred, guiding attention and suggesting emotional or narrative focus.

8.1. Shot Size loss

As defined in [28] (see Section 11.2), shot size refers to the degree to which a subject (typically a character) dominates the frame.

According to cinematic conventions[28] (see Section 11.2 for details), shot sizes can be categorized as follows:

- **Extreme Long Shot (ELS)**: Often an establishing shot, dominated by the setting or landscape.
- **Long Shot (LS)**: Shows the full subject within the context of the background.
- **Medium Long Shot (MLS)**: Depicts the subject from slightly above the knees to the top of the head, integrating the subject within the environment.
- **Medium Shot (MS)**: Shows the subject from waist-level to the top of the head.
- **Medium Close-Up (MCU)**: Frames the subject from chest-level to the top of the head.
- **Close-Up (CU)**: The subject's head or object fills most of the frame.
- **Extreme Close-Up (ECU)**: A small detail or part of the face fills the entire frame.

Obviously, different shot sizes are primarily distinguished by the relative positions of key human joints, rather than merely by the occupied frame area as used in previous works [22, 50, 51, 61, 66]. Inspired by the definition of shot size in cinematographic visual language theory, we propose a novel key-joint-chain-based shot size descriptor. Specifically, we focus on landmarks most relevant to shot size perception in filmmaking—head top, chest, waist, knees, and feet—denoted as $\mathcal{J} = \{j_{\text{headtop}}, j_{\text{chest}}, \ldots, j_{\text{feet}}\}$. This cinematographic definition captures the semantic essence of shot size while remaining robust to minor variations in pose, viewpoint, and body shape, thereby ensuring stylistically faithful reproduction of classic cinematic shots. As shown in Fig. 6, the baseline model[62] fails to match the shot size: the character appears smaller than in the reference image. In contrast, our method maintains consistent shot size even when the character poses differ, demonstrating robustness to appearance variations.

Our task specifically targets the stylistic replication of classic cinematic shots, requiring high-level visual style consistency while tolerating minor differences in actor poses and physique. In practice, significant pose variations rarely occur between the reference and rendered scenes, ensuring that the spatial layout of the key-joint-chain, which represents high-level pose feature, remains pretty similar.

(a) Visualization of shot size and framing in jaws[62].

(b) Visualization of shot size and framing in our method.

Figure 6. Comparison of shot size and framing. The baseline method (top) produces misaligned in shot size and framing, while our method (bottom) achieves more consistent alignment with the reference image.

8.2. Framing Loss

As a core component of film style, framing determines where subjects—especially human characters—are positioned to convey narrative emphasis, emotional tone, and visual harmony. In practical filmmaking, particularly within classical Hollywood cinema, framing is closely intertwined with shot size: the subject's spatial placement typically follows conventions that also define how much space they occupy in the frame. As we have already addressed subject size through our *Shot Size Loss*, the *Framing Loss* is specifically designed to capture spatial placement. In cinematography, the spatial position of such elements in the 2D image plane is determined jointly by scene composition and camerawork. Given inevitable differences in scene content between reference and target shots, perfect spatial matching of all elements via camera adjustment is generally infeasible. Among these elements, human characters are most frequently the primary narrative focus and the dominant compositional anchor in the frame. Motivated by this cinematic principle, we concentrate on the spatial placement of characters as visual anchors for framing alignment, similar to previous work[20, 23, 62]. A straightforward strategy, used in JAWS [62], is to perform pose-level alignment by matching the 2D positions of all detected skeletal joints. However, full-pose alignment is highly sensitive to variations in body pose, orientation, and camera viewpoint. As shown in Fig. 6, the baseline model fails to match the framing: the character is centered in the image, whereas in the reference frame, the character is positioned on the right. In contrast, our method preserves consistent framing even under pose variations, demonstrating robustness to changes in appearance and subject positioning.

To improve robustness against these variations, we represent character placement in the frame using the centroid of a subset of reliably key joints (same with key joints in Section 8.1), rather than the full set of joint coordinates. This cen-troid acts as a compact descriptor of the character's overall spatial location in the frame, effectively capturing the framing intent while discarding high-frequency pose variations.

By focusing on centroid alignment, this formulation captures the cinematographic intention of subject placement while maintaining robustness to pose, orientation, and articulation differences, thus providing a stable framing constraint across heterogeneous scenes.

8.3. Filmic Space Loss

Filmic space is the spatial construct perceived within the film frame [28] (see Section 11.5). Its key attributes include depth cues and juxtaposition. Since we only focus on single shot rather than shot sequences editing, we focus on depth cues, which convey the perceived three-dimensionality of a scene. As noted in [28] (see Section 11.5), filmic space can be characterized by the depth, proximity, size, and proportions of objects and places within the image. To encode these properties in a way that matches human perception and is robust to monocular-scale ambiguity, we adopt perceptual distance, rather than metric distance, as the filmic space feature.

A long-standing convention in film theory and practice organizes the frame into three planes—foreground, midground, and background—as a mise-en-scène strategy to create the illusion of depth and to structure attention [7]. This tripartite scheme is both canonical in classical film language and cognitively natural: observers coarsely "chunk" depth into near/mid/far zones, supporting stable judgments of spatial layout and narrative salience. Motivated by this, we segment each frame into three coarse depth layers by thresholding the perceptual depth value of each pixel: (i) foreground ($\mathcal{F}$), (ii) character ($\mathcal{C}$), and (iii) background ($\mathcal{B}$). This representation aligns with cinematographic semantics (framing elements in the foreground, principal actors at mid distances, setting in the background) and encourages the model to allocate features to the same coarse regions that directors intentionally compose.

However, directly matching per-pixel depth between reference and rendered frames is fragile when scene content differs. Thus, for each layer we compute a representative depth value given by the kernel density estimation (KDE) of that layer's depth distribution to serve as a stable estimate of its distance, as defined in Eq.10.

To obtain a scale-robust perceptual descriptor, we then derive two relative depth ratios as defined in Eq.9: (1) the ratio of the character layer's depth feature to the background layer's depth feature, and (2) the ratio of the character layer's depth feature to the foreground layer's depth feature. These ratios effectively encode the perceived depth separation between the main character and the background (or foreground), respectively.

This loss encourages the generated scene to maintain

Figure 7. Camera Angle loss.

similar depth relationships as the reference. By matching relative depth cues in this way, we ensure that the perceived spatial composition – how "deep" or "flat" the scene looks – remains consistent with the reference style.

8.4. Camera Movement Loss

Previous cinematic transfer approaches[62] employs an global optical flow estimator to measure and match inter-frame motion between the reference and rendering videos. However, while optical flow effectively captures camera-induced inter-frame motion, it neglects the critical motion parallax phenomenon [18, 60]: identical camera motions produce depth-dependent apparent motion, with nearer objects exhibiting larger displacements than distant ones. Thus, directly aligning global optical flows between reference and rendered scenes with distinct depth distributions can introduce misleading. Moreover, a global endpoint-error (EPE) objective is implicitly weighted by pixel counts; large background regions dominate the loss and obscure character dynamics—particularly when foreground and background exhibit large depth disparities and distinct motion regimes(e.g., in dolly zoom or bullet-time effects).

To address this, we propose a depth-layered optical flow decomposition strategy. We first compute a dense depth map for each frame using a state-of-the-art depth estimation network[69]. Subsequently, we compute optical flow separately within each depth layer, as defined in Eq.11.

8.5. Camera Angle Loss

For each frame, we infer three angles $\mathbf{a} = (\psi, \theta, \phi)$ (yaw, pitch, roll), as shown in Fig. 7.

9. Implementation Details

We represent the input 3D scene S using 3D Gaussian Splatting (3DGS) [26] from its high quality and efficiency. To extract their cinematic features, we leverage several state-of-the-art models:

- For shot size loss $L_{\text{shortsize}}$ and framing loss L_{framing}, we adopt HRNet [56] (via the MMPose library) trained on the CrowdPose dataset [31] to infer the joint position of

the character. Nevertheless, the 14 annotated joints in CrowdPose are designed for the pose estimation task in computer vision, which does not align with the 5 key joints required for camera language transfer. Thus, we construct these 5 key joints as follows: (i) the *head* position is directly obtained from the annotated head top keypoint. (ii) The *chest* is estimated as the midpoint between the left and right shoulders. (iii) The *waist* is estimated as the midpoint between the left and right hip. (iv) The *knees* and (v) *feet* are defined as the midpoints of the respective left-right joint pairs (knees and ankles). The set of visible key joints $\mathcal{J}_{\text{ref}}^{\text{vis}}$ is obtained by applying a confidence threshold of 0.5 to the joint predictions from HRNet, where joints with confidence greater than 0.5 are considered visible.

- For filmic space loss L_{space} and camera movement loss $L_{\text{cam-move}}$, we employ DepthAnything [68] to estimate the perceptual depth of each pixel in a frame. We use YOLOv8 for character mask generation (*i.e.*, character layer $\mathcal{C}$).

- For camera movement loss $L_{\text{cam-move}}$, we use RAFT [57] to estimate the optical flow in each depth layer. The optical flow is computed in a backward manner, *i.e.*, for the i-th frame, it is estimated using the i-th and $(i-1)$-th frames. The camera movement loss is not applied to the first frame.

- For camera angle loss L_{angle}, we use OrientAnything [64] to estimate characters' orientation.

10. Ablation Study

Shot Size Loss Fig. 4 (second row, fourth column) shows the effect of removing the shot size loss. Without shot size loss, the character appears significantly smaller compared to the reference, indicating a mismatch in perceived subject scale. In contrast, our full model accurately preserves the character's size, closely matching the reference shot and maintaining intended composition.

Framing Loss Fig. 4 (second row, third column) shows the effect of removing the framing loss. In this case, the character appears on the left side of the frame, deviating from the reference composition in which the character is located on the right. Incorporating the framing loss in our full model corrects this spatial misalignment, resulting in a composition that accurately matches the subject placement in the reference image.

Filmic Space Loss Filmic space loss is designed to resolve the inherent ambiguity in shot size: the same shot size can arise either from placing the camera physically closer to the subject or from using a longer focal length. However, the two ways result in markedly different filmic spaces due to their distinct perspective properties. Fig. 4 (first row, fourth column) shows our ablation study for filmic space loss. The left image in second line is rendered without ap-

plying the filmic space loss, resulting in a deep space with a wide-angle appearance: foreground expansion, strong line convergence, and a diminished background relative to the subject. In contrast, the right image in second line is rendered using our full model, demonstrating a shallower space with flattened perspective and stronger depth compression, consistent with the reference image.

Camera Movement Loss Fig. 5 shows the ablation study for camera movement loss. Camera movement loss encourages consecutive frames to exhibit similar visual motion patterns to those in the reference clip, promoting perceptual coherence across frames. Moreover, its formulation facilitates efficient optimization.

Camera Angle Loss Fig. 4 (first row, third column) shows the effect of removing the camera angle loss. Without this term, the subject's orientation deviates noticeably from the reference, and the optimization becomes less stable, often requiring more iterations to converge. With the loss included, the subject's pose aligns more consistently with the reference view.

11. Related Definition from Professional Film Book

The following definitions are quoted from A Dictionary of Film Studies by Kuhn and Westwell [28] for reference in our discussion of cinematic loss. All excerpts are quoted for academic reference purposes.

11.1. Shot

"Continuous action on the cinema screen resulting from what appears to be a single run of the camera. The shot is the basic building block of all films—which normally consist of a series of shots edited together (see editing; medium specificity). Shots are generally characterized by 1. The apparent distance between camera and subject (see framing; shot size). 2. The angle of the camera in viewing the subject (see camera angle). 3. The movement of the camera during the shot (see camera movement). 4. The number of characters within the frame (e.g two-shot, three-shot)."

11.2. Shot Size (shot scale, shot type)

*"An informally agreed and widely accepted set of conventions which describe and define different framings of a film image, or apparent distances between camera and subject. In the extreme long shot (ELS, XLS), often used as an establishing shot to set up the location for the scene, the frame is **dominated** by a landscape or a setting. The long shot (LS) shows **the subject (usually a character in the film)** in its entirety, along with the background. The medium long shot, three-quarter shot, or American shot (MLS) shows the subject from above the **knees** to above the **head**, but still as part of the setting. The medium shot or mid shot shows the subject from **waist-level** to **the top of the head**. The medium closeup or medium close shot (MCU, MCS) shows a character from **chest** level to the top of the head. In the closeup or close shot (CU, CS) the head **takes up more than half of the frame**; while in the extreme/tight shot/close up (ECU, XCU) a portion of the face, or a small object, fills the frame. In classical Hollywood cinema, these definitions were carefully maintained in order to match studios' standardized production methods; but over time they have become less precise, and framing guides for filmmakers are often inconsistent in their descriptions of shot sizes. For the contemporary cinematographer, director, and camera operator working together on a production, the solution to this imprecision is to agree before filming precisely how they wish to define each type of shot size, bearing in mind that framing can be an important expressive tool. They might agree, for example, to use carefully composed framings reminiscent of classical cinema, with neat uncrossed edges, a clear space about the head, and solid centring in the frame. This choice could suggest a sense of order and stability that might enhance character or story in, say, a heritage film. Alternatively, choosing framings that look 'grabbed', awkward, uneven, and inconsistent might be a stylistic support for a storyline involving hurry or unease. Shot size, then, is not a set system but an integral part of a film's style. In Watchmen (Zack Snyder, US, 2009), for example, the decision was made for the film to stay with careful, graphically composed, comic-book style images. This choice recreates the look of the graphic novel and creates a sense of period, because the shot sizes are similar to those of classical Hollywood cinema. However, because the visual style for contemporary action films has shifted to unsteady, uneven framings, with odd and inconsistent shot sizes (as, for example, in 127 Hours (Danny Boyle, US/UK/France, 2010)), Watchmen might look somewhat static and out-of-date in its style by comparison. Given such flexibility in shot sizes one might ask whether any random framing and shot size will be acceptable to audiences. It is not, because an aesthetically controlled film will be consistent in its use of shot sizes, and there is a history of framing that will be familiar to viewers from paintings, pho-*

tographs, films, and television programmes, and which will be drawn on in interpreting the style of a particular film. The variations in shot size and framing—and above all the disparities in the scale and the fragmentation of the human body on the screen—that are possible in cinema constitute a key point in the medium's distinctiveness. In the early years of cinema, viewers and commentators alike were astonished above all by the closeup, with its capacity to convey detail and emotion. In film studies, shot sizes can be treated as a component of film style, and attention to patterns and variations in their use in films and groups of films can be illuminating, for example, in analyses of authorship in film and of national cinemas, for example, as well as in histories of film form. See also shot."

11.3. Scale

"If the same object were filmed at different shot scales it would often signify quite differently. Shot scale can foster intimacy with a character, or conversely, it can swallow the character in its environment. Orson Welles exploited divergent shot scales in Citizen Kane (1941) to demonstrate the changing power relationship between Charles Foster Kane and his lawyer. As a boy, his figure is lost in the snow at the back of the shot as the lawyer arranges for his adoption. As a young man he rebels against Bernstein's oversight, rising in the frame as he asserts himself."

"**Extreme Long Shot** A framing in which **the scale of the object** shown is **very small**; a building, landscape, or crowd of people will fill the screen. Usually the first or last shots of a sequence, that can also function as establishing shots. The following examples of framing from Eyes Wide Shut (Stanley Kubrick, 1999) and A Summer Tale (Conte d'Été, Eric Rohmer, 1996) well illustrate the range of uses for this particular shot scale." "These two extreme long shots are also establishing shots. However, their primary function is different. Whereas Rohmer give us a standard establishing shot that introduces the locale where the main characters are about to meet, Kubrick uses the ballroom shot mainly as a brief transition between two more important scenes. While the two shots above have similar sizes, some extreme long shots can be significantly larger, particularly if shot from the air with the help of cranes or helicopters. This kind of extreme long shot is also called bird's eye view shot, since it gives an aerial perspective of the scene."

"**Long Shot** A framing in which the scale of the object shown is small; a standing human figure would appear nearly the height of the screen. It makes for a relatively stable shot that can accomodate movement without reframing. It is therefore commonly used in genres where a full body action is to be seen in its entirety, for instance Hollywood Musicals or 1970s Martial Arts films. Another advantage of the long shot is that it allows to show a character and her/his surroundings in a single frame, as in these two images from Eyes Wide Shut (Stanley Kubrick, 1999) and A Summer Tale (Conte d'Été, Eric Rohmer, 1996)."

"**Medium Long Shot** Framing such than an object four or five feet high would fill most of the screen vertically. Also called plain américain, given its recurrence in the Western genre, where it was important to keep a cowboy's weapon in the image."

"**Medium Close-Up** A framing in which the scale of the object shown is fairly large; a human figure seen from the chest up would fill most of the screen. Another common shot scale."

"**Close-Up** A framing in which the scale of the object shown is relatively large. In a close-up a person's head, or some other similarly sized object, would fill the frame. Framing scales are not universal, but rather established in relationship with other frames from the same film. These two shots from Eyes Wide Shut and A Summer Tale can be described as close-ups, even if one starts at the neck and the second at the upper chest.. Framing scales are usually drawn in relationship to the human figure but this can be misleading since a frame need not include people. Accordingly, this shot from The Color of Paradise (Rang-e Khoda, Majid Majidi, Iran,1999) is also a close-up."

"**Extreme Close-Up** A framing in which the scale of the object shown is very large; most commonly, a small object or a part of the body usually shot with a zoom lens. Again, faces are the most recurrent images in extreme close-ups, as these images fromThe Color of Paradise (Rang-e Khoda,Majid Majidi, 1999), The Stendhal Syndrome (La Sindrome di Stendhal, Dario Argento, 1996), and My Neighbor Totoro (Tonari No Totoro, Miyazaki Hayao, 1988) demonstrate. With regard to the latter, it should be noted that while all of these film terms equally applies to animation, the technical procedure to achieve a particular effect can be very different. For instance this last frame is a drawing of Totoro's teeth, not a zoom on his face, as it would have been the case

in a live-action film.”

11.4. Framing

*“The **arrangement and composition** of elements in a film frame, i.e. the entire rectangular area of a film image as projected or as visible on the screen. In the Hollywood studio era shot sizes were standardized to ensure continuity for editing, and these standards still dominate framing today. To this extent framing for film is pragmatic: it follows a set of rules which can be seen in the vocabulary of shot size. A standard closeup, for example, crops just below the shoulders and puts the eyes of the actor along an imaginary line that cuts across the top third of the screen horizontally: this placing of the eyes provides for a small amount of space above the actor’s head. To enhance consistency of framing, lighting and focus visually isolate the actor from the background, ensuring his or her dominance in the image. Even performance is subsidiary to framing: screen actors stay relatively still, except as their role requires them to move—in which case in order to ensure correct framing they keep in position by following marks on the floor.” “In film studies, framing is treated as a component of film style, and as such is widely referenced in textual analysis and in studies of mise-en-scene, of authorship, and of genre. Framing is also a key element in film reception, and here the significance of the standardization of film framings cannot be overstated: audiences the world over are familiar with these conventions. **However, standardized framings may be modified by practices associated with artistic and dramatic forms such as photography, painting, and drama; and this can produce striking stylistic effects.**”*

*“**Rule of Thirds** A flexible compositional ‘rule’ taught as part of painting and photographic practice and which may be extended to the framing of shots in filmmaking. Its aim is to **indicate where significant elements may be placed in the frame** in order to attract the viewer’s attention, and also produce a well composed—visually coherent and harmonious—image. This idea of composition, based on geometrical principles, stems from ideas developed from the ancient Greek and Roman periods which still hold sway in Western culture today, the argument being that geometrical ‘rules’ follow the ‘rules’ of nature. The rule of thirds ordains that the frame be divided into thirds both vertically and horizontally: if lines were drawn to mark these thirds they would look like the grid used to play noughts and crosses, but with flatter rectangular spaces. The intersections of the four gridlines represent the approximate points where objects in the frame would be placed. In the case of filmed closeups, for example, the subject’s eyes would be lined up to match the upper horizontal third. However, conventions of composition change and develop, and in filmmaking centred framings are more common than those using rule of thirds.”*

11.5. Filmic Space (Cinematic Space, Cinematographic Space, Film Space)

*“The space created within the film frame as opposed to the space of the real world or of the profilmic event. Filmic space is a wholly distinct type of space, one that can only be created on the cinema screen through the techniques and language of cinema—one of the distinctive attributes of film as a medium being that it creates its own patterns of spatiality (and temporality). The key attributes of filmic space are firstly, that it is two-dimensional but assumes the appearance of three-dimensionality because of **depth cues** provided through mise-en-scene: composition and framing, lighting, deep focus cinematography, camera movement, and other formal and stylistic elements; and secondly, that from the juxtaposition, by means of editing, of shots recorded in different locations a coherent and intelligible topography can be created in and for the world of the film. Filmic space can be treated as an element of film form; and it also figures in the processes by which viewers become drawn into the world on the screen and follow various cues in mentally navigating that world. In other words, filmic space is a factor in film spectatorship and a component of the cinematic apparatus. In film studies, analysis of filmic space as a formal element readily generates insights into cinema’s capacity to engage the viewer: this is undoubtedly a factor in a ‘spatial turn’ noted in recent film studies. And indeed systematic scrutiny of spaces and their organization within individual films and groups of films can be a productive strategy in textual analysis, and can also provide a basis for phenomenological inquiry into filmic space and the cinematic experience. ”*

11.6. Space

“The representation of space affects the reading of a film. Depth, proximity, size and proportions of the places and objects in a film can be ma-

nipulated through camera placement and lenses, lighting, decor, effectively determining mood or relationships between elements in the diegetic world."

"**Deep Space** *A film utilizes deep space when significant elements of an image are positioned both near to and distant from the camera. For deep space these objects do not have to be in focus, a defining characteristic of deep focus. Staging in deep space is the opposite of staging in shallow space. Deep space is used throughout many Iranian films such as The Color of Paradise (Rang-e Khoda,1999). Director Majid Majidi likes to integrate the characters into their natural surroundings, to map out the actual distances involved between one location and another in order to emphasize just exactly how hard it is for a particular character (especially children) to move from one place to another.*"

"**Shallow Space** *The opposite of deep space, in shallow space the image is staged with very little depth. The figures in the image occupy the same or closely positioned planes. While the resulting image loses realistic appeal, its flatness enhances its pictorial qualities. Striking graphic patters can be achieved through shallow space. In these frames from My Neighbor Totoro (Tonari No Totoro, Japan, 1988) Miyazaki fills the entire background with a lamp-eyed, grinning cat-bus. Shallow space creates ambiguity: is the cat brimming with joy at the sisters' encounter, or is he about to eat them? Shallow space can be staged, or it can also be achieved optically, with the use of a telephoto lens.This is particularly useful for creating claustrophic images, since it makes the characters look like they are being crushed against the background.*"

11.7. Camera Movement

"*In film studies, camera movement is examined as part of the wider study of film form and film style, both past and present; and as a stylistic 'signature' of certain directors, genres, or film movements. Camera movement was rare in early cinema, when films were usually composed of static wide shots, or tableau shots, giving the impression of action taking place on a stage set.*"

11.8. Zoom Shot

"*A shot taken with a zoom (varifocal, or variable focal length) lens in which focal length is changed from wide-angle to telephoto, or vice versa, in the course of recording the shot.* **A zoom shot creates the impression of the camera moving towards (zoom in) or away from (zoom out) the subject, though no camera movement is involved.** *However, although a zoom shot may create mobile framing as, say, a tracking shot does,* **perspectival relations and depth of field are different** *in each case. A zoom in enlarges elements in the image and flattens its planes together, while a zoom out does the opposite. A zoom in can be effective in rapidly and dramatically drawing the viewer into a scene or bringing the viewer's attention to a detail; and a zoom out in revealing the background and the surroundings of a character or activity. Zoom shots are a staple of forms of filmmaking such as documentary and news gathering, where mobility and speed of reaction are paramount and/or where the filmmaker prefers subjects to be unaware of the camera. Originally designed for aerial and reconnaissance photography, zoom lenses became a standard tool in news filming around 1950, and the practice of zooming in or zooming out during the course of a shot began in the late 1950s.*"

11.9. Lens

"*There are two main types of lens: prime lenses which have a fixed focal length and zoom lenses, which have a variable focal length. Focal length is the distance (measured in millimetres) from the optical centre of the lens to the focal point on the film stock where the image is sharp and clear (i.e. in focus). Prime lenses are diverse, but are often split into a number of types. Firstly, normal—or standard, or middle-focal-length—lenses which are taken to correspond to, and reproduce the sense of perspective seen by, the human eye; these lenses have a focal length of between 35 mm and 50 mm. Secondly, wide angle, or short-focal-length, lenses: with a focal length beginning at 12.5 mm these lenses can create a sense of distortion; figures appear to loom, objects look unnaturally large in the foreground and the background diminishes with steepened perspective. Wide-angle lenses are often used by filmmakers wishing to convey something out of the ordinary. Thirdly, telephoto, or long-focal-length, lenses, with focal lengths ranging from 85 mm to as high as 500 mm, bring distant objects close but flatten space and depth in the process.*"

11.10. Camera Angle (Angle of Framing, Angle of View)

*"**The placement, or implied placement, of the film camera in relation to the subject.** The normal **camera height** is about eye level, producing the most common view, a straight-on angle on the subject. When the camera is**placed above** the subject, the result is a **high-angle** or extreme high-angle shot. A camera placed below eye level produces a low-angle or extreme low-angle shot. A canted angle or Dutch angle is produced where the camera is tilted so that the shot is composed with vertical lines at an angle to the side of the frame or the horizon line of the shot is not parallel with the bottom of the frame. Camera angle is not a set system but an integral part of a film's style, and there are no hard-and-fast rules about the meanings of different angles on a subject (a character shown in low angle does not necessarily suggest that he or she is powerful and overbearing, for example; nor does a view from a high angle always indicate vulnerability). In film studies, camera angle is treated as a component of film form and film style, and it is assumed that, as with all aspects of film form, a particular camera angle will derive its meaning or meanings from its place and function within the film as a whole and in combination with other formal elements (such as framing, lighting, shot size, etc.). In the point-of-view shot, for example, an angled shot can simply represent the direction of a character's look, and as such would be fully integrated into the film's narrative system. On the other hand, angle of view can be deployed expression is tically and even abstractly, as in the extreme high-angle studio shots in Busby Berkeley's production numbers for 1930s Hollywood musicals such as Gold Diggers of 1933 (Mervyn LeRoy, US, 1933); Footlight Parade (Lloyd Bacon, US, 1933); and Dames (Ray Enright, US, 1934), in which the kaleidoscopic spectacle of symmetry and pattern in movement becomes the focus of attention, engagement, and pleasure. The same device is used with very different meanings in Ratcatcher (Lynne Ramsay, UK, 1999), where high angle shots signal a move from a space of reality to one of interiority and imagination. In some avant-garde films, extreme high and low angles are used as a defamiliarization device (see russian formalism): for example, Fernand Léger's quirky angles on familiar objects in Ballet mécanique (France, 1924) offer fresh ways of looking at ordinary and everyday objects."*

11.11. Mise-en-scène

The following is from Chapter 4 of Film Art: An Introduction [7].

"Depth cues also pick out planes within the image. Planes are the layers of space occupied by persons or objects. Planes are described according to how close to or far away from the camera they are: foreground, middle ground, background. "