# A Hybrid System for Interactive 3D Camera and Lighting Control

Mark Rudolph
*Institute for Innovation*
*Information Technology University Copenhagen*
*Rued Langgaards Vej 7*
*2300 Copenhagen*
*E-mail: mrudolph@itu.dk*

## Abstract

*With the accelerating convergence of cinema and computer games, the importance of the sequence of placements of virtual camera and lights has increased with the growing dramatic expressiveness and narrative complexity of the new interactive 3D medium. This paper explores a multiple source system for choosing camera and light placement. The first component is a set of special annotations to dialogue referred to as 'stage directions,' Next is a rule-base decision system reasoning about a dynamic fact base description of the scenes and the history of shots. Finally a live performance system allows skilled virtual cinematographers to make real-time inputs provided by a MIDI keyboard selecting shots from a constant constellation of camera placements and a computer keyboard mapping to pan, tilt and roll, and lighting selections. There is a large body of cinematic practice accumulated during the development of film art which may be captured in rule sets which express various styles of camerawork and montage. Traditionally the stylistic choices for cinema are made in post-production editing after much reflection and experimentation. The interactive 3D medium differs from film in that it is tightly constrained by time since it is fundamentally a dynamic live performance medium, and yet in another manner it enjoys greater freedom than film since virtual camera and lights are not subject to the limitations of physics placed on their material counterparts. The ideal scenario for interactive 3D cinematography suggested by the relative constraints and freedoms of the medium is the hybrid interactive system which we propose.*

## 1. Introduction

We create stories to share our experiences with each other and every age develops unique media to create representations of stories appropriate to the culture and technology of the period. The new medium of interactive 3D (i3D) is currently at that critical point of development when all the technical means now exist to enable the definition of a unique 'language' of the medium. It is at a similar point as cinema was about ninety years ago when directors discovered the freedom of time and space made possible by film editing, and the ability to merge interior thought and exterior action into a seamless whole. The language of cinema became defined by a grammar of 'montage' instead of a flow of continuous experience within the bounds of Aristotelean dramatic unities.

We can only imagine the full scope of the language of i3D since we are living in its historical infancy. However it is certain that the cinematic techniques of camera movement, montage and dramatic lighting will play a key role in its expressive capability. We can also assume that dialogue and narrative development in i3D need not be deterministic so that decisions about shot selection and lighting must be able to be made dynamically. Another consequence of non-deterministic dialogue and narrative is that the 'screenplay' for a production will not necessarily be linear. Therefore 'stage directions' associated with dialogue and action must be interpreted as suggestions instead of predictable events. They may be modified or even overruled by either a realtime artificially intelligent cinematographer and lighting director, or a human camera and lights performer, or both in cooperation. Non-deterministic dialogue and narrative have more general implications but they are beyond the scope of this paper.

The scenario under which we have developed our hybrid system of camera guidance and shot and lighting selection (see figure 1) is two person dialogue performed by virtual 3D characters able to modify their mouth positions in lip-synch with a synthesized text dialogue. Spoken text is supplemented by Java Speech Markup Language, JSML [1], metatags and associated stage directions expressed in a custom natural language via name-value pairs. JSML provides a means to modify and inflect the speech for expressive purposes, and to mark positions in dialogue so that events may be generated by the synthesizer and received by a 'director' able to execute the stage directions associated with the event.

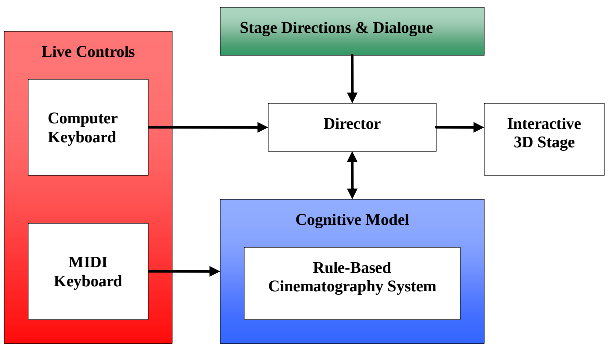


**Figure 1.(Hybrid interactive decision making system**)

Further, we keep a dynamic model of the scenes in the production and an inference engine capable of applying cinematic rules for camera shot selection, target framing and lighting modification, as well as object occlusion detection and appropriate camera adjustment for clear shots. Rule sets for shot selection and so on may be considered to be style sets. By introducing different rule sets it is possible to interpret a scene in varying manners and make quite different choices of shot framing, position and duration, and lighting position and color.

Finally, we believe that no automatic system can fully replace the skilled eye and instinct of a human camera and lighting expert. For this reason we provide a MIDI [2] keyboard interface for an initial camera placement selected from a discrete 'constellation' of camera marks and orientations. The key event is intercepted by the automatic cinematographer which checks for occlusion and makes fine adjustments of position for target framing and vertical orientation based on the previous shot history and other narrative state. The MIDI camera mark selection is assumed to be non-instantaneous due to system communication latency. A second live performance interface is provided via the computer keyboard which is unchecked by the automatic cinematographer and thus is essentially instantaneous. The range of camera modifications provided by the second live system is restricted to simple pan and tilt, and very rarely, roll and zoom. These camera moves do not affect camera position and therefore do not affect camera animation associated either with stage directions or the MIDI constellation selector since all animations start from whatever the current position and orientation of the camera and animate or cut to the prescribed final position and orientation. The computer keyboard interface also can make lighting changes and orientation adjustments to the actors. In all cases changes made via MIDI keyboard or the computer keyboard update the abstract model of the scene, lights and camera, so that the fact base of the inference engine of the cinematographer adapts to the present state of the scene for further decision making.

In order to reduce complexity our experimental production work has been in traditional two person dialogue with action by the characters confined to simple changes of orientation and minor motions. Our lighting has also been conventional and non-animated.

## 2. Related Work

The study of cinematography is a key to understanding the language of image, which in turn is a subtle but key element in the disclosure of narrative in the interactive 3D medium. The relationship between film cinematography and i3D originates in their common purpose of expressing meaningful artistic information through the camera. Cinematography and editing have been key to the narrative syntax of film, and a corresponding but unique language for i3D will be critical to the development of a more complex and expressive narrative genre in that medium also. The role of the camera and lighting in most computer games to the present has been little more than a swiveling gun sight and headlight, or occasional jumps back and forth from first person point of view to a third person view. The large body of expressive camera work and editing developed last century in film art has not yet been applied completely to computer games. A syntax for dialogue associated Stage Directions, a Rule-Based system to dynamically select and place camera and lights, and effective live interfaces for complementary shot insertion and camera navigation, will provide a powerful technique to enable the development of an unique i3D cinematography applicable especially to the case of non-deterministic dialogue and narrative.

Crucial to developing the technique and rule sets used in our application have been classic texts in film cinematography. There are many, but we relied mainly on the clear insightful texts by Katz [12] and Arijon [14].

Much experimental work has already been done in studying the application of cinematography idioms in i3D. For example, an interesting approach to camera control is shown in [3], but it is not applied at runtime. A system for generating camera specification in real-time is presented in [4]. A system that controls the cameras and lights in a virtual world inhabited by autonomous characters is presented in detail in [5], and a distributed approach to decision making is presented in [6], the decisions being delegated to separate modules, Director and Cinematographer. Further distributed decision making was explored in [10]. A very precise and insightful study of the formalization of rules of film cinematography can be found in [9], and some very interesting work on the dramatic and expressive elements of i3D cinematography can be found in [8].

Tangential to pure cinematography, but fundamental to the field of narrative and dramatic i3D is the OZ project from Carnegie Mellon University [7] which established many paths for the conception and implementation of i3D narrative and drama.

## 3. Basic tools of digital cinematography

The basic tools of digital cinematography consist of camera and light. The foundation of virtual camera and lighting techniques in our system are based on concepts described in "Digital Cinematography" manual, by Ben de Leeuw [11].

### 3.1. Camera

The camera plays the role of the eyes of the viewer, but also participates in the story telling through the language of image. The virtual camera is a moving point of view in the spacetime of the narrative. The sequence of choices of camera position corresponds to film montage but must be executed in realtime. However, because the camera is not restricted by physicality many positions and animations may be performed which are difficult or even impossible for a real camera. These are significant, especially for specialized camerawork such as modern dance, but the study of the full freedom of the virtual camera is outside the scope of our implementation which confines itself to traditional two person dialogue with minimal action.

**3.1.1. Camera positioning.** In order to create a meaningful shot, the camera needs to be placed in the right position in the right time, corresponding to what happens in the scene. A sequence of camera positions assists in creating a sense of emotional and narrative significance to dialogue and action.

The camera position sets the scope for the viewer's knowledge of a particular scene in relation to plot. If the camera is placed far away or high above the scene the effect tends to make the viewer an uninvolved observer. A close camera framing of faces as they speak and react tends to make the viewer an intimate viewer of emotionally significant dialogue. The relative strength of a character within the scene can also be expressed by the height of the camera. By setting the camera above a character the impression is that the character is smaller and weaker or under tension, whereas an upward camera view tends to suggest a more confident and powerful character role in the scene.

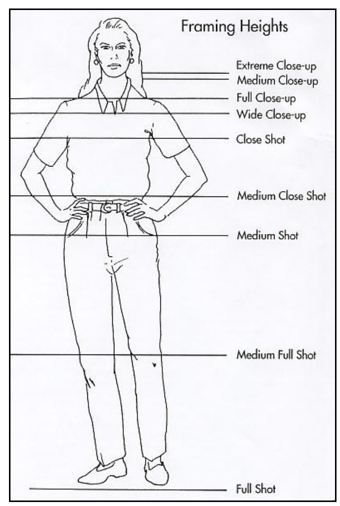


**Figure 2:(Basic Framing heights for human figure [12])**

**3.1.2. Camera framing and shot type**. If there is only one character on the screen, the shot can be one of ranges shown in figure 2. A similar categorization can be made for shots including two or more characters since usually one of the characters is the primary subject of the shot.

Camera range corresponds more importantly to the placement of the character(s) body in the visual frame. It is a decision related to still photography or even painting.

**3.1.3. Camera movement.** The camera has six degrees of freedom. It can move to any point in 3D space and in addition can change its orientation given any location. The additional three degrees of freedom are defined according to the possible axes of rotation. Pan is rotation around the vertical Y-axis, and tilt is rotation around the horizontal X-axis. An additional move used more rarely is 'roll' which a rotation around the Z-axis which creates a disturbing sense of imbalance and disorientation. Movement in space of the virtual camera is usually 'tracking' or following a moving character or 'dollying' in or out to dynamically alter the framing. A pseudo movement can also be achieved by 'zoom,' the altering of the field of view of the lens which gives an impression of dollying while remaining fixed in space.

**3.1.4. Camera punctuation.** Shots are joined together by two types of punctuation: smooth animation which we refer to as 'fly,' and abrupt changes traditionally termed 'cuts,' since they are associated with film splices of two visually discontinuous frames. An additional transition we have implemented is a 'dissolve,' which slowly dematerializes one shot to an opaque field and then slowly re-materializes to a new shot.

### 3.2. Lights

There are several basic types of light, which come from real-world cinematography and refer to the actual physical source and characteristics of the light. Common characteristics of all light are color and intensity.

Ambient light is the most basic light to be used in computer graphics and is not really light in the truest sense. It is essentially the maximum level of darkness in the scene. Ambient light has no physical devices associated with it, since it is all the environmental and reflected light in the scene. The higher the ambient light is, the brighter the darkest parts of the scene become. Ambient light should generally be kept very low, as it will flatten the look of the scene if it is too high.

Directional light consists of a vector field of parallel beams. The sun is an example of an idealized directional light which lights all objects in its scope.

A Point light is defined by a location in space which emits light of a defined intensity color and manner of attenuation.

Spotlight is a cone of light cut from a Point light and provides a very specific region of illumination.

Very specialized spotlight-like light can also be used to illuminate irregular regions such as eyes.

In lighting a scene there are several standard functions used in configuration for modeling faces and objects and giving dramatic effects.

**3.2.1. Key light.** Key light is the main source of illumination in the scene. Its function is to provide the primary illumination for the subject or focus of the scene. Often this light has a "source", that is, it has an onscreen or implied source that is easily identified.

The placement and intensity of the key light is a primary dramatic cue.

**3.2.2. Fill light.** Fill light is used to add a small amount of illumination in the dark areas left by the key light. The fill light is almost never more than 50% intensity of the key. The position of the fill light is usually a little less than 90° rotated from the key light on the ground plan. If the key light is set high, the fill light will generally be set low, and vice-versa.

**3.2.3 Back light.** Back light is used to separate the subject from the background. It is positioned a little less than 90° from the camera. The light is intended to add a thin line of illumination to the dark side of the subject that will delineate it from the background. The use of this dramatic lighting predates film since it can be found in even in late Renaissance painting.

## 4. System implementation

The system constructed for researching and producing virtual cinematography in the context of non-deterministic dialogue and narrative consists (simplified) of a Dialogue Generator, a Director, a Cognitive Model, two live Performance sources and the i3D Stage with Actors Camera and Lights.

Concentrating on camera and lights, the camera is a virtual viewpoint placed on the surface of a sphere centered at the origin (the target of view) and extending in the positive Z direction to a default position of ten meters. This primary sphere defines rotations around the target and scaling of the sphere produces dollying in and out. At the viewpoint location on the primary sphere is centered a second sphere to perform fixed location rotations corresponding to pan, tilt and roll. Thus the camera has six degrees of freedom as needed.

Further we define a letter pattern to describe the locations and orientations of the two actors in the dialogue. The pattern language extends to three and more characters but we consider just two for the present. The letter pattern is based primarily on a line of interest defined by the two center points of the actors speaking in the scene. In the case of two actors the only pattern is 'I' but this fundamental dialogue pattern is modulated by the orientation of the gaze of the actors.

The modulated letter pattern then defines a set of common positions of the camera based on the ‘Hollywood continuity‘ style of dialogue shooting. This discrete ‘constellation’ of shots is depicted in figure 3.

Another key feature of the spherical attachment of the camera is that three point lights are also attached at definable locations, colors and intensities – key, fill and back lights with default but parameterizable positions, orientations, intensities and colors, always carried wherever the camera moves or looks.

The actors are based on 3D heads (animating bodies will be added later.) The heads perform lip-synch visemes corresponding to the dialogue as well as blinking and occasionally looking aside or moving eyebrows and mouth all to give a sense of living presence rather than dead statues.

The stage is finite but unbounded since it has the topology of a torus and consists of a 5X5 grid which always reconfigures to maintain the current scene (defined by the location of the camera) in the central tile. In addition each tile can switch between an arbitrary array of scenes.

The key aspects of the system relative to the hybrid nature of the cinematic decision making are the Dialogue Generator (including Stage Directions,)

The Director and Cognitive Model, and the two live performance inputs, the MIDI constellation selector, and the computer keyboard pan, tilt, roll, zoom and lighting selector.

## 4.1. Dialogue and Stage Directions

Stage Directions include timed suggestions for camera shot selection and lighting as imagined by the writers of the dialogue. Stage Directions are directly bound to dialogue since JSML allows a word based event triggering of implied action such as shot selection and lighting choice.

**4.1.1. Stage Directions for Camera shot.** A complete Stage Direction for camera shot includes the following elements. They are the descriptor attribute of an XML message which defines:

- The name of shot in the shot constellation.
- The range of the shot, one of six different ranges:
  - ECU: extreme close up shot.
  - MCU: medium close up shot.
  - WS: waist shot.
  - MS: medium shot.
  - KS: knee shot.
  - FS: full shot.
- The type shot (TS): The camera animation can be cut shot (CS), the camera is translate abruptly from previous camera position to the new one or flying shot (FS), the camera fly from previous position to the new position with a certain speed.

An example of one Stage Direction for a camera shot is the following:

TS:FS S:FRONTMASTER R:MCU

**4.1.2. Stage Direction for lighting.** The Stage Direction can also give a suggestion about what parameters of lighting to use. As mentioned earlier there is a natural configuration of three related lights, key light, fill light, and back light. The Stage Direction can turn on and off any light during the conversation of characters, change the position, intensity or color of each light in the configuration. An example of one Stage Direction for lighting is the following:

L:KEY POSITION:DEFAULT COLOR:SEA INTENSITY:HIGH

**4.1.3. Stage Direction for actor animation.** During the conversation Stage Directions can direct how the character moves or orients himself via natural language suggestions.

An example of one Stage Direction for animation is the following:

A:MARY GO TO PAGODA

It means that the actor Mary should move to whatever object in the scene most closely corresponds to the ontological adjective PAGODA.

## 4.2. Director

The Director is the joining link among Dialogue Manager, Cognitive Model and the i3D stage. The role of the Director is mainly to ensure that speech, Stage Directions and animation are played at the right time. When the Director receives XML speech text it parses the text to isolate the Stage Directions from the actual speech. It then sends the Stage Direction to the Cognitive Model in order to be evaluated (parameterized,) and at the same time it forwards the JSML tagged speech to the synthesizer. The evaluated Stage Directions are then sent back to Director, which puts them in an event queue. When the synthesizer starts playing the speech it sends events to the Director such as speech start, word reached, and speech ended, so that an associated Stage Direction is executed at the right moment.

### 4.3. Cognitive Model

The Cognitive Model is a reflection of the stage and all scenes, all actors, and all instruments in the world including the camera and all lights. Its role is to parse human readable stage directions sent by the Director and calculate parameters for a Command design pattern sent back to the Director for event driven execution.

The Cognitive Model contains a map of the entire stage and all the current scenes in each tile. This geometric map is the information used by the Model to calculate animation paths in the world.

In addition, continuously updated facts about the position and orientation of the Camera and Actors permit the Model to maintain a working memory for cinematography rules used to calculate shot sequences.

The Cognitive Model also receives the live MIDI messages from the constellation selection piano keyboard controller and checks for occlusion and calculates the fine detail placement and orientation of the Camera actions. Both Stage Directions and live performance messages contribute to a flow of Commands to the Director.

By default the Director obeys live performance commands first, so that Stage Directions, automatic shot and lighting selection, and live performance commands can cooperate to good production effect. It is assumed that a skilled human makes better mixed performance decisions than either writers unaware of the live situation, or an expert system.

**4.3.1 Resource Bundle.** The Cognitive Model learns about and builds itself initially by reading a Resource Bundle of hierarchical descriptor objects representing stage, scenes and scene objects, as well as descriptors for Camera, Lights, and Actors.

The Bundle associates every Descriptor with a string. The 3D world is the sum of many items such as scenes, objects, actors, lights, camera marks, etc each of them described as a string. This type of representation allows the development of a tool that can automatically generate large 3D worlds.

For example an object (PAGODA) is defined with basic information such as the position, the size or bounding box, the looking or facing direction of the object. Each object also keeps track to which scene it belongs (HANOI) and other objects that it might contain (ROOMS).

The Bundle is loaded by the Cognitive Model at the startup when the i3D cinematography system is initialized. At that time the information of the items described in the Bundle is put into the working memory of the reasoning system as facts upon which the stylistic cinematography rules reason and assert shots in the dynamic montage.

These assertions cooperate with live performance events which together with the Stage Direction suggestions decide the next specific shot and lighting and update the Cognitive Model accordingly, affecting the working memory for the next automatic camera and lighting decision.

### 4.4. Automatic virtual Cinematographer

The reasoning and decision system behind the Cinematographer is a JESS rule-based inference system [13]. The rule sets associated with the inference system correspond to camera and montage styles. For our work we limited attention to one such style – the two person dialogue Hollywood continuity style. Its purpose is to advance the narrative through dialogue and acting and disclose dramatic and emotional significance through associated camera work and edits.

A rule-based inference system maintains a working memory of dynamic facts which in every decision cycle are matched against the left hand side of a set of shot selection rules. If there is a match that rule is placed on an agenda and the most appropriate or highest priority rule is executed. The right hand side of the rules are actions which may choose a shot, change the facts, or assert new or modified facts into the working memory.

The inference system is able to reason about facts representing the state of the stage, the story and characters, and the significance of the current dialogue. There is a close relation between the Cognitive Model and the inference system. The Cognitive Model relies on the reasoning system to make appropriate decisions about shot selection, for example, but may overrule if the shot chosen is blocked by a scene object, or if there is some other geometric problem. Each shot proposed by the reasoning system is then evaluated by the Model in terms of the position orientation and framing distance. These numerical arguments are then set in a Command design pattern object passed to the Director which queues the shot for execution at the appropriate instant, which usually corresponds to a change of speaker, but could also correspond to the speaking of a particular word.

Rules are constructed to reflect idiomatic film camera and editing good practice, and the grammar of the Hollywood dialogue continuity style. Elements of good practice are to preserve the same directional relationship between characters from shot to shot by never crossing the line connecting the speaking characters. Generally the camera targets the speaking character but framing should vary and close framing should be preferred for critical speech. Occasionally over the shoulder shots may be used to permit the observation of the listener as well as the speaker, and at key points a quick temporary cut to the listener may be used to catch an instantaneous reaction.

Lighting is generally set by the default positions and intensities of the configuration of key, fill, and back lights. However in some scenes or instants light may be arranged more perpendicularly to the character's forward direction in order to create more modeling and shadow. Many lighting practices derive originally from classical 17th and 18th century painting and then imitate implementations in later theatre and film. Since virtual lights have no physicality they can be moved arbitrarily close to a character without occluding a camera shot. In general lighting should suggest the light sources seen in the scene such as sun or room lamps, but if the camera is close light can be used to expressively model the face or highlight the eyes.

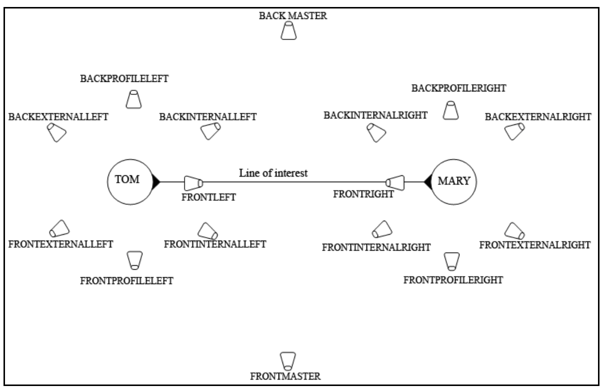


**Figure 3: (sixteen shot constellation)**

**4.4.1. Shot constellation.** Placing the camera in the right place is the essence of creating a meaningful shot. Where to place the camera determines the camera height and distance from objects. Each of these possibilities gives the scene a different meaning, and therefore requires careful consideration and planning.

In order to create a simple grammar for the set of shots used for presenting the conversation of two characters, sixteen different positions are maintained and recalculated with every movement. These sixteen positions make a constellation of camera shots for two characters. Based of the character positions and their respective forward vectors the Cinematography system dynamically recalculates the constellation. Each position has an absolute name in the constellation.

The following figure illustrates the location of each shot in the two person ('I' letter pattern) dialogue constellation.

**4.4.2. Absolute name of constellation shot.** The absolute name of camera shot is used by Stage Directions, Cinematographer and MIDI performer to suggest what shot to take. This way is very convenient because it works for every different position of character in the conversation. Below are listed the absolute name for camera shot in the shot constellation:

FRONTMASTER,
FRONTPROFILELEFT,
FRONTINTERNALLEFT,
FRONTLEFT,
FRONTEXTERNALLEFT,
FRONTPROFILERIGHT,
FRONTINTERNALRIGHT,
FRONTRIGHT,
FRONTEXTERNALRIGHT,
BACKMASTER,
BACKPROFILERIGHT,
BACKINTERNALRIGHT,
BACKEXTERNALRIGHT,
BACKPROFILELEFT,
BACKINTERNALLEFT,
BACKEXTERNALLEFT.

**4.4.3. Facts.** In order to generate a meaningful shot, the reasoning system uses much information such as character and object position, the state of the dialogue, the progression of the story, who is to be framed, obstacles, letter pattern of the characters, who is going to move and where, who was the previous speaker, the history of the previous shots, the current duration of the present shot, the position and type of lights and sounds. These pieces of information are facts stored in different modules in the working memory.

The Model maintains facts about the overall stage, each scene able to be staged, and all of the objects in any scene. Facts can be partitioned into namespaces and thus attention can focus on an arbitrarily specific range. The following is an example of attributes used to define a 3D object.

- Name.
- Size.
- Position.
- Forward vector.
- Ontological adjective.

**4.4.4. Ontological adjective.** Ontological adjective is a string that describes attributes associated with a scene object in natural language. The possibility of describing a scene object by an abstract string adjective and not simply a numerical value greatly increases the power of the query interface. In the most expansive application the query can bind a dialogue and its stage directions to an arbitrary scene if that scene has objects and locations with similar adjectival description. For example, a particular dialogue could take place in any scene with significant romantic significance or association with a particular location. Therefore a stage direction referring to *the blue pagoda near the lake could bind to any object with ontological adjective matching 'pagoda,' or 'lake,' or even to the scene location with ontological adjective with high 'romantic significance.'*

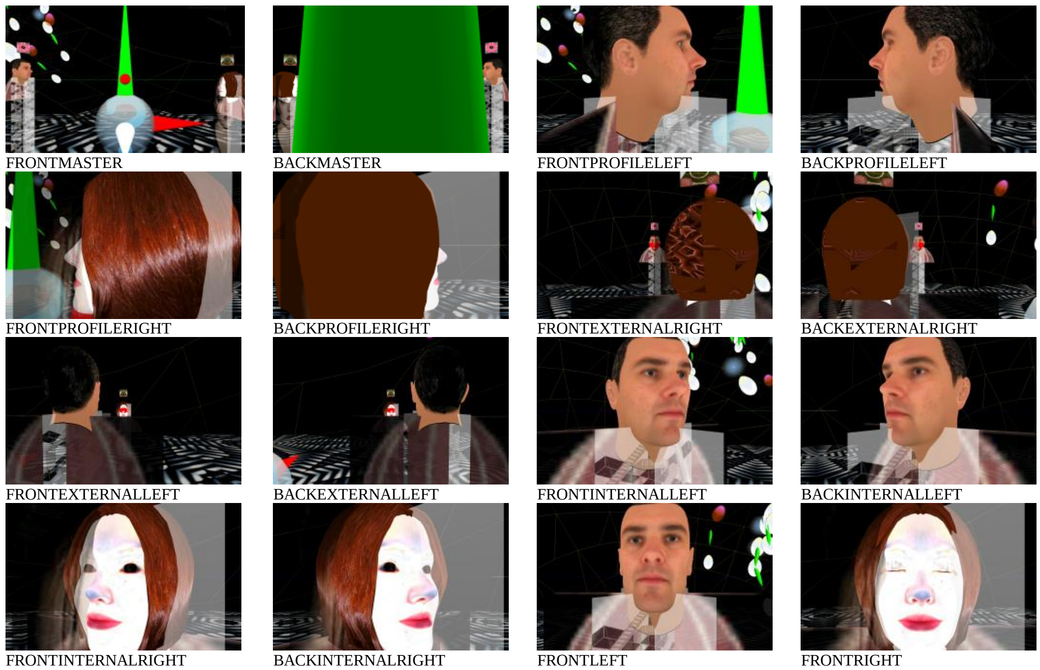


**Figure 4:(Sixteen shots for two people dialogue)**

**4.4.5. Commands.** Rules may modify facts as well as fire actions in the scene such as camera movement. However, namespace separations limit the rule firings to the module in which they are defined, and are triggered only by the fact base associated with the module. There are four types of rules: The first is concerned with the cinematography style, the second is to convert natural language Stage Directions into facts, the third is to query the working memory, and the fourth is to exchange data with the Cognitive Model.

**4.4.6. Cinematographer.** Cinematographer is the center of the reasoning system. This module of rules carries out a number of tasks related to the style of automatic camera movement and shot selection A shot may be a 'cut' as in a montage splice of film, or a 'fly,' a continuous animation.

Cinematographer is provided with a group of rules that may fire when it receives a request for a new shot, normally at a change of speaker. These idiomatic cinematography rules reason about the previous shot history and the state of the Letter Pattern, and finally stochastic means for variation, in order to decide the next shot.

Rules are applied as focus shifts from module to module, which is achieved via rules which may be termed meta-rules.

The default reference is the MAIN module which is global and the last reference, the last possible set of rules to fire.

**4.4.7. MAIN module.** This module stores information about facts that are not related to any specific scene. The actors exist during the entire story so must be visible from everywhere. Even though the actors move around different scenes, the actor facts stay in the MAIN module which apply by default after all other scene specific facts.

**4.4.8. Scene modules.** The scene is the current switched scenery of the central tile of the Stage. Each scene stores the objects that make up the scene in the form of associated hierarchical facts. A fact describes the position, size, orientation and ontological adjectives of an object. Keeping each individual scene in a separate module reduces the searching complexity and improves information management.

### 4.5. Live performance

Live performance complements the Stage Directions and automatic camera shot selection and control. Human skill and experience are superior to any expert system for cinematography. Live camera movement and constellation shot selection modify the shot selection facts and help the automatic system to stay in synchronization with the optimal stylistic sequence available. Currently the live performance is controlled from two sources. A MIDI keyboard is used to control constellation shot selection. In this manner fixed shot sequences can be notated and remembered as musical phrases, and mixed with other event information such as lighting choices to create a multi-channel MIDI score for playback or for editing in visual score-based applications. A key press is passed to the Cognitive Model which checks for object occlusion and decides fine detail such as framing relative to the shot history. In some cases the live shot can be overruled if it is highly contradictory to stylistic practice. The MIDI input and check by the Model is not absolutely realtime, but is acceptable for constellation shot selection. A second live performance control is the computer keyboard input which directly maps to the scenegraph so is instantaneous. The keyboard in this case is mapped to rapid pan, tilt, roll and zoom control, as well as light configuration settings. Because the computer keyboard control does not change camera position it need not be checked by the Model which permits minimum latency of application of effect. Thus the skilled live cameraman may be able to insert a selected constellation shot here and there, as well as rapidly modify the camera forward vector by pan and tilt in even a rapid ‘MTV’ style.

Every live constellation choice is sent to the Cognitive model which updates the Cinematography working memory for the next decision, and camera and light changes made by the computer keyboard are also sent to the Model to keep in synchronization with the live state of the camera and lights.

We expect that better interfaces than computer keyboard will enable more effective live camera control. This is a subject for further development. Independent of the controller we force camera animations to pass through spline curves defined by discrete points, and to obey automatic ‘ease-in’ and ease-out’ velocity rules. In addition we use quaternions to represent orientations so that interpolations do not create problems associated with axis-angle or Euler angle calculations.

## 5. Conclusion

The first stage of this research concentrates on the application of Hollywood continuity style and how to translate those cinematography idioms into information usable by the system, either in composed Stage Directions, stylistic rules, or live performance. A thorough understanding of cinematography idioms implies a need to have access to a large body of runtime facts in order to suggest stylistic camera shots.

In order to encapsulate camera and lighting idioms the system proposed implements a hybrid approach composed of three complementary components. First, a system of Stage Directions can convey by a relatively free specialized natural language syntax a suggestion for stylistic camera and lighting related tightly to dialogue. Second, the Cognitive Model system keeps track of the geometry and physical properties of the space of the narrative stage. This allows the system to have a complete picture of the world in which the camera works. By reflecting on a working memory of updated facts about the world, scenes, actors and camera, a rule-based cinematographer can reason about the physical and narrative world, and interpret shot selection based on stylistic camera and lighting practice. Finally, a live performance system consisting of a MIDI constellation shot selector and instantaneous computer keyboard pan, tilt, roll, zoom and lighting selection selector, permits skilled human intervention into the live montage.

## References


[1] JSML,http://java.sun.com/products/javamedia/speech/forDevelopers/JSML/.

[2] MIDI, http://en.wikipedia.org/wiki/MIDI.

[3] Christianson, D.B Anderson, S.E. He, L.W., Salesin, D.H., Weld, D.S and Cohen, M.F. 1996, “Declarative camera control for automatic cinematography.” Proceedings of the Thirteenth National Conference on Artificial Intelligence, 148-155.

[4] He, L. Cohen, M. and Salesin, D,“The Virtual cinematographer: A paradigm for automatic real-Time camera control and directing.”, 1996. Proceedings of the ACM SIGGRAPH ´96, 217-224.

[5] Bill Tomlinson, Bruce Blumberg, Delphine Nain from MIT Media Lab, “Expressive Autonomous Cinematography for Interactive Virtual Environment”, http://characters.media.mit.edu/Papers/aa2000badger.pdf

[6] Daniel Amerson and Shaun Kime, "Real-Time Cinematic Camera Control for Interactive Narratives.", http://liquidnarrative.csc.ncsu.edu/pubs/DAmerson01.pdf

[7] OZ Project Carnegie Melon University, http://www2.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/oz.html.

[8] K. Kennedy, R. E. Mercer, "Planning Animation Cinematography and Shot Structure to Communicate Theme and Mood," in Proceedings of the Second Intl Symp. On Smart Graphics, pp. 1-8, Hawthorne, New York, 2002.

[9] D. Friedman and Y. Feldman, "Knowledge-Based Formalization of Cinematic Expression and its Application to Animation," in Proceedings of Eurographics, pp. 163-168. Saarbrucken, Gemany, 2002.

[10] Tsai-Yen Li, Xiang-Yan Xiao, "An Interactive Camera Planning System for Automatic Cinematographer", in Proc. of the eleventh Intl. Multimedia Modelling Conference MMM 2005, pp. 310-315.

[11] Ben de Leeuw, "Digital Cinematography", Academic Press Limited, USA, 1997, ISBN 0-12-208875-1.

[12] Steven Katz, "Film Directing Shot by Shot" Michael Wiese Publications 1991 ISBN 0-941188-10-8.

[13] Ernest Friedman, "JESS in Action: Rule-Based Systems in Java", Hill ISBN 1930110898.

[14] Daniel Arijon, "Grammar of the Film Language", Silman-James Press, USA, Publications 1976 ISBN 1-879505-07-X.

[15] FuzzyJ ToolKit for the Java(tm) Platform & FuzzyJess, http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJToolkit2.html