Appeared in Proceedings of 9th International Conference Computer Graphics, Imaging and Visualization, 2012.

Context-aware Camera Planning for Interactive Storytelling

Chia-Hou Chen
Computer Science Department,
National Chengchi University
Taipei, Taiwan
e-mail: g9711@cs.nccu.edu.tw

Tsai-Yen Li
Computer Science Department,
National Chengchi University
Taipei, Taiwan
e-mail: li@nccu.edu.tw

***Abstract*--In the application of 3D interactive narratives, virtual camera is a crucial element for appropriate presentation of scenarios happening dynamically. In this work, we have designed a virtual cinematographic system to generate appropriate camera plans automatically in a 3D virtual environment for interactive storytelling. In such a system, not only the story line but also the relative positions between the actors may be changed at run time due to interaction with the user. Therefore, a camera plan that can reflect the changes in dynamic scenarios needs to be generated automatically at run time. We have also designed a scripting language to describe the directive details of the story and implemented a camera planning system to generate camera configurations according to the scenario descriptions. We will use an example to illustrate the idea of interactive narratives in 3D environment and how appropriate camera motions can be generated according to the result of interaction.**

Keywords - Virtual Camera, Real-Time Camera Planning, Interactive Storytelling, Context-Aware

I. Introduction

Interactive Storytelling (IS) is a new research topic attracting much attention in recent years. In an interaction storytelling system, a user can not only be a consumer of the story but also participate in the creation process of the final story that cannot be determined ahead of time. Despite its great potential in enhancing user experience for entertainment, it also presents a great challenge to the author as well as the presentation system due to the uncertainly resulting from user interactions at run time. This is especially true for 3D virtual environments with real-time animation since it means that 3D contents need to be generated at run time.

Most 3D virtual environments, such as 3D online games, use pre-programed rules and canned animations to interact with the users at real time. Due to the constraints of using canned motions for character animation, the ways that users can interaction with the system are also limited. In addition, the camera motions adopted in such a system are also limited because of the lack of description about story context. As a result, most of the camera motions are either arranged in advance or operate with simple rules to avoid occlusion.

New forms of interactive entertainment are under development, and machinima is one of them. Machinima refers to the use of real-time 3D computer graphics rendering engines to create a cinematic production. It is a new form of interactive entertainment allowing users to participate in the creation of the story flow while keeping the story plot designed by the original story author. Although rooted at gaming, machinima systems require more detail scripts and camera control to create cinematic effects. Therefore, it often leverages the techniques from Artificial Intelligence to generate the motions of the characters as well as camera control.

In this work, we aimed at designing a Machinima system that allows authors to write scripts with Morphology to express the structure of a story. With the help of common functions in Morphology, the system can use the story context specified by the author and the interactions with the user in a virtual cinematographic planner to generate appropriate camera motions automatically. In other words, we hope that story line can change dynamically at run time as a result of various types of interactions with the user, and the camera planning module can arrange camera positions according to the story contexts at run time as well.

In order to realize such a novel system, we have developed the following four components in this work:

- Using Morphology to analyze and build story structure: decomposing a story into story elements, each of which serves as a distinct function in the story context;
- Mapping each story context into a corresponding camera control module;
- Designing camera control modules to generate camera motions that follow the idioms in Cinematography;
- Designing appropriate interaction modules allowing the user to participate in the development of the story.

II. Related Work

Previous research pertaining to this work falls into two categories: *camera planning* and *interactive narratives* as described in more details below.

A. Automatic camera control

Automatic camera control has long been a design goal for computer animation systems. The earliest planning system for camera control was proposed by Blinn [1] and focused on solving it as a geometry problem. Since then, much research has been conducted for solving this problem. Recently, Christie has conducted a literature survey [2] on camera control and classified the approaches into three categories: *algebraic system*, *constraint satisfaction system*, and *motion planning system*. In [3][4], the authors partitioned the space into semantic volumes with distinct meanings in Cinematography and then used local search techniques to compute good representatives of each volume. Due to the computation complexity of most approaches, real-time methods were not available until recently. In [5], several camera control modules have been developed for the user to select and switch in real time.

B. Camera planning in interactive narratives

The concept of interactive narratives was proposed long time ago and put into simple web-based games in its early stage. Due to the rapid development of Artificial Intelligence and Computer Animation in recent years, interactive narratives start to be realized with richer contents in immersive 3D virtual environment. In an interactive virtual environment, perception of a user is created through virtual cinematographer. He et al. [6] was the first to propose using hierarchical finite state machine to realize film idioms in virtual cinematography. Bares et al. [7] used a constraint-based approach to solve the camera planning problem in 3D space. However, these approaches usually assume that the type of camera controls have been determined beforehand and ignore how it is related to the story context.

Amerson and Kime [8] proposed the FILM (Film Idiom Language and Model) system to construct idioms for Cinematography in a story. However, it does not provide automatic reasoning functions to compute camera positions. Jhala and Young [9] used the DPOCL algorithm to compute casual links between event nodes in a story and then perform dynamic camera planning. In addition, Hornung et al. [10] attempted to build the relations between story elements and camera controls in order to implement a real-time camera planning system. Story context was introduced in this work for camera planning but the story plot is assumed to be linear and determined beforehand.

III. MORPHOLOGY FOR DIGITAL STORYTELLING

Positioning of a camera is closely related to the context of a story as well as the director. For example, the story context determines who and what should be included in the camera window while the director's style will determine how the camera is positioned such as viewing angle, viewing distance, and intercuts. Therefore, it is important to be able to script a story with appropriate structure and context annotation for camera control. In this section, we will describe Propp's Morphological model for describing a story, the context parameters that we have designed for camera control, and the interactive storytelling model that we have used for script design.

TABLE 1. CONTEXT PARAMETERS AND THEIR RANGES

Context Parameters	Variable Range	Description
Function	[Absentation, Interdiction, Violation, …]	Story function
Object Might	$\in [-1.0 ... 1.0]$	Importance of target
Tension	$\in [-1.0 ... 1.0]$	Tension in a given context
Scene Type	[Action, Dialogue, Both]	Type of scene
Action Type	[Physical, Mental, Predicate]	Type of action
Event Coherence	$\in [-1.0 ... 1.0]$	Coherence of an event in the context.
Character Emotion	$\in [-1.0 ... 1.0]$	Emotion of the character to be stressed

A. Propp's story model

Propp [11] is a Soviet formalist scholar who analyzed the basic plot components of Russian folk tales to identify their simplest irreducible narrative elements, which are called *functions*. A story is comprised of a sequence of functions in pairs or in certain patterns. For example, "interdiction" and "violation of interdiction" are paired functions that occur in sequence. "Pursue" and "rescue" are another example. There are 31 functions in total as concluded by Propp. In addition to function, Propp also proposed that a character in a story could play one of the eight different roles such as "hero," "villain," and "donor." Since these functions and roles were induced from folk stories, it is not complete for describing general stories of all types. Nevertheless, this model is widely used in the literature to express story structures. For example, Grasbon [13] used Propp's Morphology model to propose a high-level scripting system allowing the story designers to control the details of a story in various levels. Hartmann [14] extended Propp's formalism with a new "Motivation" category to provide non-linear story branches.

B. Context parameters

A good coverage of cinematographic skills, including the challenging ones, has been analyzed and enumerated in [9]. We hope to use these cinematographic skills, commonly used in film production, to present a story with theatricality. In [10], a set of dramatic parameters have been defined to describe a story event for its contextual meaning, which determines the view angle and control type of a camera. We have used the definitions as a basis to design our own context parameters as shown in Table I. These parameters will be mapped into control parameters of a camera in a later section.

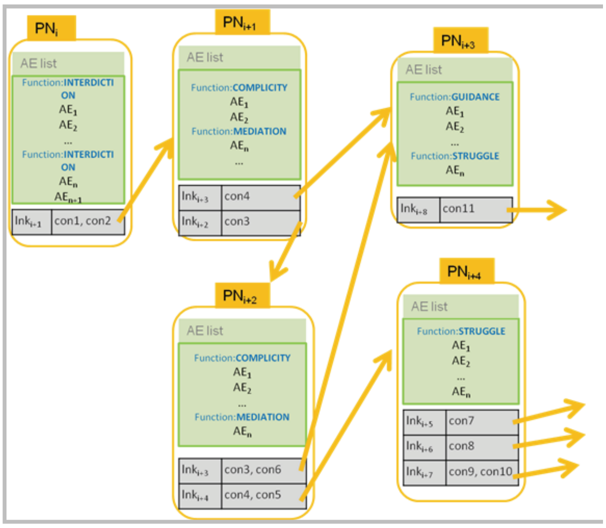

Figure 1. Story structure and the relations between plot nodes

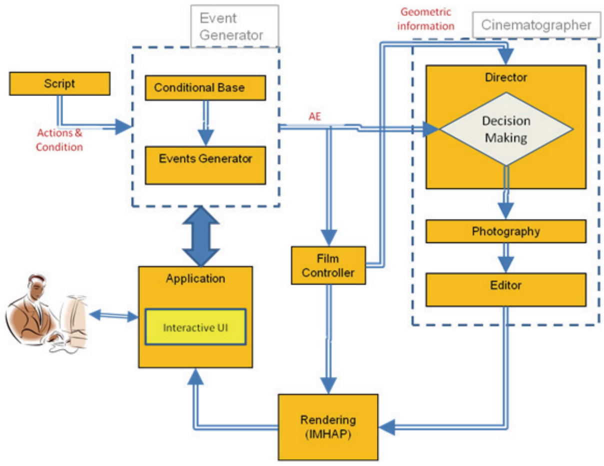

Figure 2. System architecture

C. *Story play for interactive narratives*

We have used the Morphology structure proposed by Propp to design our screenplay. *Action Event* is the most basic element in our script. Each action event contains information about description of an action, the scene context (through context parameters such as scene type, character emotion, and function defined in Section III.B), and the preconditions for triggering this action.

According to Propp's theory, a story can be decomposed into a sequence of functions, which are conceptual elements. We think a function is realized in a story through its corresponding action event. Several related action events can be grouped into a *Function Group* (denoted by *Fn*), and a sequence of *Fn*'s can be used to compose a *Plot Node* (denoted by *PN*) while several *PN*'s comprise a *Scene*. The story structure we have proposed is shown in Figure 1. In this work, the story structure with contextual information is used to guide the selection of camera controls and appropriate camera parameters to generate the best views to express the story.

IV. DESIGN OF VIRTUAL CINEMATOGRAPHY SYSTEM

A. *Problem description*

In a 3D interactive storytelling system, a user can use multiple modes, such as first-person navigation and third-person view to interact with the scene. In either case, we hope that the user can concentrate on the story itself or moving his/her avatar instead of controlling the camera. In other words, we hope that the camera can be the best storyteller and its motion can be generated automatically with the best effort according to the story plot and scene contexts. In order to provide this service, the system should be able to define the camera planning problem (tracking, panning, intercuts, etc.) according to the story plot and contexts and then compute the best camera configuration according to the scene contexts such as choosing the best view aiming at the target object while avoiding obstacle obtrusion. More specifically, two layers of information need to be defined clearly. One is the relation between objects and the scene as the story develops. The other is the mapping between scene contexts and camera controls through Morphology and Cinematographic rules to create appropriate camera configuration arrangement in real time.

B. *System Architecture*

Our 3D interactive narrative system consists of two key modules at run time to drive the screenplay as shown in Figure 2. The first key module is the Event Generator, which prepares Action Events with dialogs, captions, as well as character animations on the screen based on the input script. The other key module is the Cinematographer module, which takes the action events passed from the Event Generator module and maps them into camera control parameters. As the story progresses, the Film Controller module issues corresponding commands to the Cinematographer module and the Rendering module to update the scene with correct character animation and camera configurations. In the next two subsections, we will describe the Event Generator module and Cinematographer module in more details.

C. *Event Generator*

As described in the previous subsection, a story play consists of plot nodes (*PN*) connected through links. Each plot node may branch to one of several *PN*'s if the conditions for the *PN* all become true. In other words, an interactive story can be considered as a graph consisting of *PN*'s, whose traversal is determined at run time as a result of user interaction. When the system enters a *PN*, the Event Generator sends the *AE*'s in sequence to the Cinematographer module for camera planning and to the Film Controller to generate motions for the animated characters.

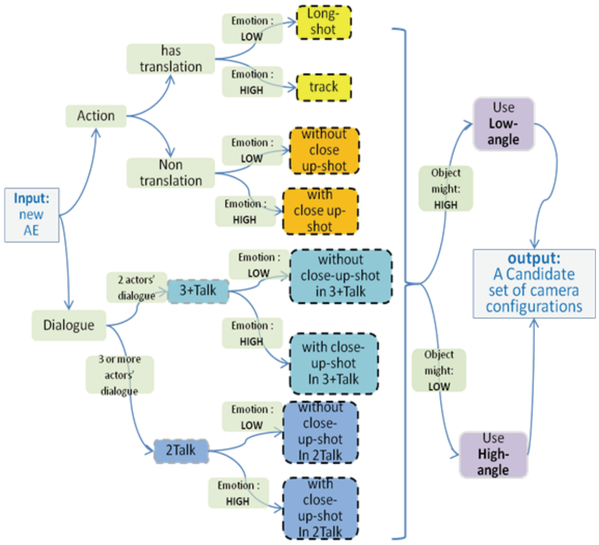

Figure 3. Decision tree for the camera module

In every *PN*, there exist one or more links (denoted by *lnk*) to other *PN*'s. Each link acts as a branch to another *PN* if the conditions (denoted by *con*) for the link all become true. Every time when an AE is triggered, the system will update the conditions that may be changed. When a *PN* (for example, PN_i) comes to the end, the system may branch to one of the connected links [lnk_{i+1}, lnk_{i+2}, lnk_{i+3}] if all the conditions (for example, con_a, con_b, etc.) for the link are satisfied. In addition, in some exceptional cases, in order to ensure the tempo and flow of a story, the system may continue to develop the story after some timeout when the user cannot issue effective commands to the system to move the story to the next *PN*.

D. *Cinematographer*

The Cinematographer module determines how to take shots, given a story context. The key function of the Cinematographer module is a decision making procedure that maps a given story context into camera parameters including appropriate selection of control methods. A story context mainly consists of a function in Propp's terminology for an Action Event, one or several characters or objects acting in the scene, which affect how a camera shot should be taken. In other words, a camera configuration should be affected by factors such as the actor, its action, emotion, and tension in the scene.

We use a hierarchical decision tree to determine how to map the story context into camera configurations as shown in Figure 3. The system will first determine the type of scene as either Action or Dialog. If it is an Action scene, we will further consider the motion of the acting character to be static or moving. If it is moving (has transition) and the emotion value is high, a tracking control is adopted. Otherwise, a long-shot is used. On the other hand, if a dialog scene is assumed, we will check the number of actors involved in the dialog. When the emotion (arousal) is high, a close-up shot may be preferred.

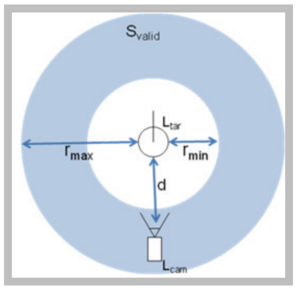

Figure 4. Relative position and legal camera region in the navigation mode

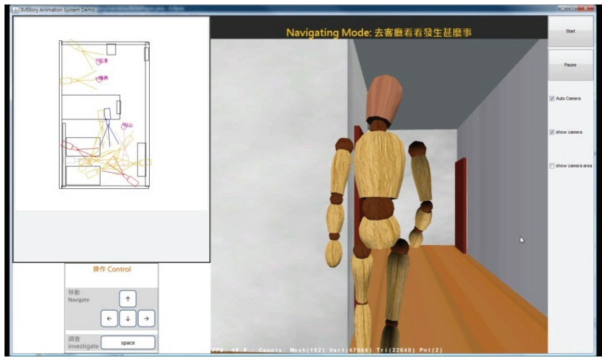

Figure 5. The navigation mode

Finally, the Might parameter in Table 1 affects the view angle (pitch). When the value of Might is high, the camera will take a depression angle (looking down) and, otherwise, an elevation angle (looking up).

Through the decision making process, the attributes of the final camera control are determined by the following tuple $<M, d, h, \theta>$, where *M* is the control mode, *d* is the distance of the camera from the target, *h* is the camera height, and θ is the angle from the vertical axis. We have designed three camera control modes: *static*, *tracking*, and *dialog*, according to the literature in the field of Cinematography [15][16]. Note that beside the constraints imposed by the environment such as occlusion, how the context parameters affect the camera configurations is a subjective matter that usually reflects the artistic style of the film director.

V. IMPLEMENTATION AND EXAMPLE

In order to demonstrate that the abovementioned design provides automatic camera planning and brings novel user experience with Machinima, we have implemented a 3D interactive virtual environment system for interactive narratives. We have designed a scene as part of a short story called *Strange Box*, originally written by Akagawa Jiro. The story is about two detectives (Katayama and Daisuke) and Katayama's sister, Harumi, investigating a murder case happened in a closed room. The main actor that the user can play in this example is Katayama. There are two camera modes that the user can use in the story-

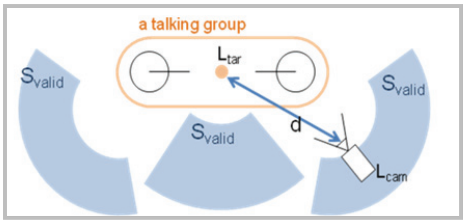

Figure 6. Relative position and legal camera regions in the storytelling mode

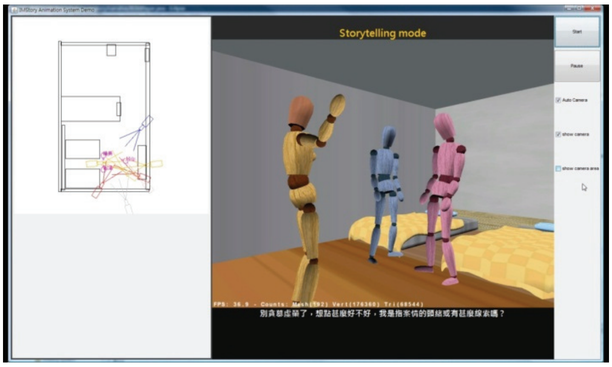

Figure 7. The storytelling mode

telling process: *Navigation* and *Storytelling*, as described in more details below.

A. Navigation mode

In the avatar navigation mode, the camera uses an out-of-body third-person shot behind the avatar that is manipulated by the user. The relative position and shooting parameters between the avatar (L_{tar}) and the camera (L_{cam}) are shown in Figure 4. Given a set of parameters $<M, d, h, \theta>$, we first define a legal region of a ring shape (see Figure 4) with nominal shooting distance as d and a radius range (r_{min}, r_{max}). A legal configuration for the camera is defined as a non-occluded location inside this ring region (S_{valid}). When the initial configuration of the camera is illegal, a search procedure is evoked to find a non-occluded one. In addition, the new camera position should be as close to the previous location as possible to avoid unnecessary jumps. Once the location is determined, the height and shooting angle are set to h and θ, respectively.

When the system enters such a mode, the story usually is in a situation where the user is asked to move to a destination, such as another room, to continue the story with another scene. The system switches to the storytelling mode when the destination region is reached or the time is out. A snapshot of the scene under the navigation mode is shown in Figure 5.

B. Storytelling mode

The storytelling mode is used in an interactive narrative system to tell a scripted story to the user without intervention. In such a mode, the Cinematographer module receives an *AE* from the Event Generator and output a set of parameters $<M, d, h, \theta>$. According to M, a set of possible shooting regions (S_{valid}) are computed as the potential positions for the camera as shown in Figure 6. Before a camera location is selected, we need to ensure that the view is not occluded as in the case of navigation mode.

When the scene is about a dialog in the storytelling mode, the camera arrangement with intercuts is used. Four factors are considered to determine if an intercut is necessary: duration of the current shot, target visibility, similarity between two actions, and tension value in the scene context. The longer the current shot, the less visible on the target, the less similar of two actions, and the more tension for the new scene, the more likely the camera will perform an intercut. A snapshot of scene under the storytelling mode is shown in Figure 7.

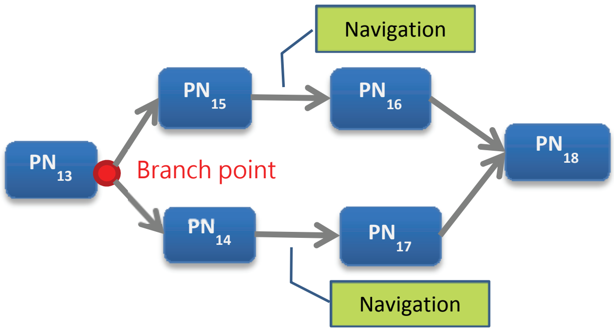

Figure 8. Example story graph with branches

C. Example scene of interactive storytelling

An example of story graph for interactive storytelling with branches based on user interaction is shown in Figure 8. In the branch point of the story fragment, the user is prompted with a question of selecting one of the two branches about willingness to help out a character or not at end of PN_{13} in Figure 8. Depending on that the user avatar chooses to help or not, the story will go through the upper branch (PN_{15} and PN_{16}) and the lower branch (PN_{14} and PN_{17}), respectively. At the end of PN_{15} and PN_{14}, the story enters the navigation mode and the actors move to PN_{15} and PN_{17}. Because of this interaction with the user, the final and relative position of the user avatar compared to the other two actors might be different. Therefore, although both branches enter PN_{18} at the end, the camera shots could be totally different as shown in Figure 9. This example demonstrates that the interactive storytelling system provides an interesting user experience allowing a user to participate in the development of a story. Because of the dynamics resulting from the interaction, automatic camera planning and control is crucial to present the story to the user.

VI. Conclusions

In this paper, we have adopted Morphology proposed by Propp to design an interactive storytelling system. The system allows a user to participate in the development of

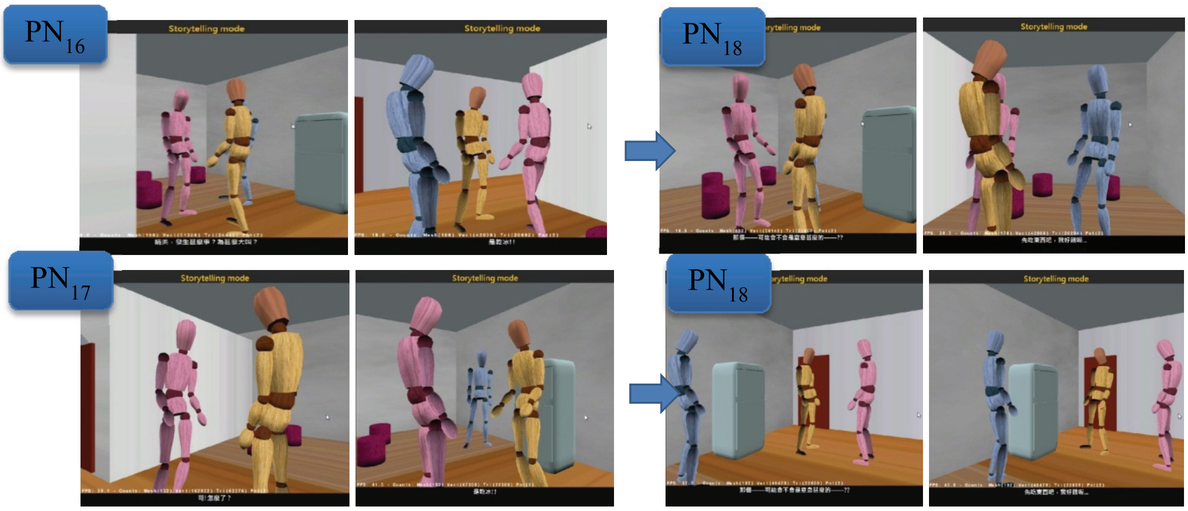

Figure 9. Comparison of camera shots through different branches of the story plot

a story through navigation and selection of branches for different story plots. A key component in this kind of system is the corresponding camera planning and control modules that generate appropriate camera configurations to present the scene contexts based on the objects in the environment and the Morphological function of the story element. We believe that we are the first to realize such a 3D interactive storytelling system with virtual cinematography. Currently, the authoring of interactive story is still carried out manually. In the future, it is highly desirable to design authoring tools to design and compose interactive scripts for this kind of system.

ACKNOWLEDGEMENTS

This work was partially supported by National Science Council under contract number NSC 99-2221-E-004 -008 -MY2.

REFERENCES

[1] J. Blinn, "Where am i? What am i looking at," Computer Graphics and Applications, IEEE, vol. 8, 1988, pp. 76-81.

[2] M. Christie, P. Olivier, J.-M. Normand, "Camera Control in Computer Graphics," Computer Graphics Forum, 8(27), 2008.

[3] M. Christie and J. M. Normand, "A semantic space partitioning approach to virtual camera composition," Computer Graphics Forum, 24(3): 247-256, 2005.

[4] M. Christie, F. Lamarche and F. Benhamou, "A Spatio-temporal Reasoning System for Virtual Camera Planning," Proc. of Smart Graphics, 2009, pp. 119-127.

[5] C. Lino, M. Christie, F. Lamarche, G. Schofield, and P. Olivier, "A real-time cinematography system for interactive 3D environments," Proc. of the 2010 ACM SIGGRAPH/Eurographics Symp. on Computer Animation, 2010, pp. 139-148.

[6] L.-W. He, M. F. Cohen, and D. H. Salesin, "The virtual cinematographer: a paradigm for automatic real-time camera control and directing," SIGGRAPH '96 Proc. of the 23rd annual conf. on Computer graphics and interactive techniques, 1996, pp. 217-224.

[7] William H. Bares, J. P. Gregoire, and J. C. Lester, "Realtime constraint-based cinematography for complex interactive 3D worlds," '98 Proc. of the fifteenth national/tenth conf. on Artificial intelligence/Innovative applications of artificial intelligence, 1998, pp. 1101-1106.

[8] D. Amerson, S. Kime, and R. M. Young, "Real-time cinematic camera control for interactive narratives," Proc. of the 2005 ACM SIGCHI Intl. Conf. on Advances in computer entertainment technology, 2005, pp. 369-369.

[9] A. Jhala and R. M. Young, "A discourse planning approach to cinematic camera control for narratives in virtual environments," Proc. of the 20th national conf. on Artificial intelligence, 2005, pp. 307.

[10] A. Hornung , G. Lakemeyer, and G. Trogemann, "Autonomous real-time camera agents in interactive narratives and games," Proc. of the IVA2003: 4th Intl. Working Conf. on Intelligent Virtual Agents, 2003, pp.236.

[11] V. Propp, Morphology of the Folktale, 1958.

[12] U. Spierling, D. Grasbon, N. Braun, and I. Iurgel, "Setting the scene: playing digital director in interactive storytelling and creation," Computers & Graphics, vol. 26, 2002, pp. 31-44.

[13] D. Grasbon and N. Braun. "A morphological approach to interactive storytelling," Proc. of the second international conf. on Entertainment computing, 2001.

[14] K. Hartmann, S. Hartmann, and M. Feustel, "Motif definition and classification to structure non-linear plots and to control the narrative flow in interactive dramas," Virtual Storytelling, 2005, pp. 158-167.

[15] D. Arijon, Grammar of the film language. London ; New York: Focal Press, 1976.

[16] J. V. Mascelli, The five C's of cinematography: Motion Picture Filming Techniques, Silman-James Press, 1998.