Capturing Triadic Conversations
A Visual Director System for Dynamic Interactive Narratives

Bingjie Xue and Stefan Rank
Drexel University, Philadelphia PA, USA
stefan.rank@drexel.edu

Abstract
Film cinematography has been developed and applied for more than a century to involve and engage the viewer in visual storytelling. Interactive storytelling games can benefit from these cinematic conventions to enhance visual experience. However, even conversation scenes in games are highly dynamic, and pre-authoring camera parameters using cinematography principles is often insufficient. This paper proposes an automatic Visual Director System focused on dynamic conversation scenes involving three characters and reports on work in progress on a prototype applied to the recreation of a movie scene. Based on principles of cinematography and the study of film scenes, cinematic conventions for triadic conversations are encoded modularly as an artificial intelligence game component that selects suitable shots for dynamic scenes.

Introduction
Interactive storytelling games transform passive story recipients into active users (Gansing 2003). IS games provide the players with opportunities to participate in stories in real time. Game designer Chris Crawford states that interactive storytelling (IS) does not simply translate to "games with stories" (Crawford 2004). Instead of just watching, the player can control the story plot and direct story development. Compared to other types of games, which focus on gameplay rather than story plot, IS games are more like films, in that their major focus is on narrative. As Eku Wand points out, 3D interactive storytelling games present “excellent opportunities for integrating and combining new possibilities with traditional film viewpoints” (Wand 2002).

Virtual cameras play a vital role in 3D IS games because they directly affect player experience and game enjoyment (Pinelle, Wong, and Stach 2008). Virtual cameras provide the player's points-of-view. The current work considers in particular games that do not employ a first or third-person camera but that rather use other means of camera positioning, sometimes only in conversation scenes, to achieve cinematic aesthetics. The current work reports on work-in-progress towards an intelligent camera system, the Visual Director System, focusing on triadic conversations, i.e. scenes that involve three characters, and easy integration into a game engine.

Related Work
Hawkins (2005) conceptually integrates framing, editing, and artificial intelligence to present how traditional film cinematography techniques can be adapted for common use in games. Newman (2008) tries to explore the middle ground between film and games by pointing out the most relevant cinematic techniques that can be applied in games. He implies that important film theories such as the “Five C’s of cinematography” (Mascelli 1976), and the “Rule of Thirds” are also considered essential in game cinematography. Haigh-Hutchinson (2009) reviews existing game camera systems by introducing camera theory, camera system design and camera implementation methods. Haigh-Hutchinson examines some of the most common game genres and discusses the camera solutions that are often applied, analyzing their advantages and disadvantages respectively. Oliveros (2004) applies film idioms to game cinematics for 3D games. Similarly, Flaherty (2007) adapts digital cinematography techniques for game cinematics.

Many other scholars (Christianson et al. 1996) (He, Cohen, and Salesin 1996) (Charles et al. 2002) (Kardan and Casanova 2008) (Kardan and Casanova 2009) (de Lima et al. 2009) (de Lima, Feijó, Furtado 2012) examined the field of automated camera planning, which combines expertise in Computer Graphics and Artificial Intelligence to emulate the composition and editing decisions found in traditional film to reduce the burden of manually placing and moving virtual cameras. The Declarative Camera Control Language (DCCL) provides a camera control system constrained by film principles (Christianson et al. 1996). The Virtual Cinematographer system (He, Cohen, and Salesin 1996) explored a cinematography model in a real-time environment. Although it was created to be able to be
adopted in different kinds of scenes, the Virtual Cinematographer has two weaknesses (Amerson, Kime, and Young 2005). Firstly, the system is limited to creating effective shots for scenes that it is familiar with; secondly, each transition between two idioms will break the continuity of the scene, creating a rupture in the narrative, hence it is not suitable for real-time use in highly dynamic scenes. Amerson, Kime, and Young (2005) have proposed a system for real-time camera control in interactive narratives named Film Idiom Language and Model (FILM). An outstanding contribution of FILM is that information about common film idioms is encoded in a scene tree using the FILM Language. Based on the conceptual foundation of FILM, Charles et al. (2002) created a similar system that uses abstract film idioms as basic rules to choose the best camera position and angle at any time using action recognition applied on a plan-based plot representation. Kardan and Casanova (2008) present an approach to generate shots for 3D computer graphics cinematic sequences for Maya animation from event-based descriptions of scenes of conversations between groups of actors. Kardan and Casanova (2009) also proposed a set of heuristics for editing of film clips that would attempt to mimic the thought processes of a human editor, using editing heuristics informed by a set of stylistic rules, but only for static animation scenes. De Lima et al. (2009) created a Virtual Cinematography Director System, extended in (de Lima, Feijó, Furtado 2012) towards a real-time editing method for interactive storytelling systems. The system can automatically generate adequate shot transitions, swap video segments to avoid jump cuts, and create adequate looping scenes in real-time. Burelli and Jhala (2009) present an autonomous camera control system for real-time 3D games using a constraint-based camera employing artificial potential fields; moreover, the system is able to produce high quality visual results and smooth camera movements. However, this system can only handle dynamic characters in a navigating base scene.

Just as de Lima states, “the visual result of any current interactive storytelling system that uses 3D worlds to dramatize the stories is far from the excellent quality of a computer animated feature film” (de Lima, Feijó, Furtado 2012). Part of the reason for this lies in the limitations of automatic camera systems, in particular regarding the parameterization of these systems that would allow for directorial style and regarding their range of application. Further, those systems are often dependent on a particular representation of the narrative. This directly inspires our design approach towards a modular system that makes minimal assumptions about the narrative structure and uses a focus on a particular type of scene, triadic conversations, to work towards a reusable system that allows the parameterization to reproduce the aesthetic qualities found in film. This includes the consideration of character emotions and the mood and style of a scene, not considered by previous work.

**Approach**

We propose an automatic camera system that is able to evoke a film aesthetic and enhance the visual experience for players. Rather than positing suitable AI methods, we seek to find out how cinematic conventions and styles are used in three-people conversations in film in order to iteratively encode identified conventions as game AI components useful for dynamic scenes in interactive storytelling games. In addition to the study of cinematography literature, a study of examples of film dialogue has been conducted. By analyzing conversation scenes directed by different accomplished directors that involve three characters, we intend to catalog how rules are translated in successful films, providing insights into how this can be encoded into dynamic conversation in games using general rules while allowing for capturing the variability of styles.

Several triadic conversation scenes were studied, twenty of them analyzed in detail and we present one here as an example. It is taken from the film “The Lady Vanishes”, directed by Alfred Hitchcock. In the film, a young girl, named Iris, realizes that a lady seems to have disappeared from the train; then she begins looking for the woman with Gilbert, a young man she met during the journey. In this scene, Iris and Gilbert have had some discovery, and they are about to tell Dr. Hartz, who shows particular interest in the case. But Dr. Hartz turns out to be the one behind the whole incident. This scene is the final revelatory moment in the film and develops from a happy to a scary mood as the story unfolds. The director divides three characters into two groups, and shows the story using a range of framing from a medium long shot (Apex) to close-ups showing the facial expression on characters.

![Image](image_url)

**Figure 1: The Lady Vanishes Film Scene**

This scene exemplifies several rules that are present in most of the analyzed scenes and that can be translated into game cinematography establishing a baseline for camera setups.

In a conversation scene, personal cameras usually follow the rule of thirds: important elements are placed on imaginary lines dissecting the camera view into thirds horizontally and vertically.

Over-the-shoulder shots are realized using information on the skeleton of characters targeting a character’s shoulder and using characters’ eye position as head target.

All algorithms are coded modularly in Unity 3D, a cross-platform game engine used to develop videogames for desktop platforms, consoles and mobile devices. A third-person interactive story has been created as an animation using existing 3D assets modeled in Maya and imported into Unity to demonstrate the VDS. This scene recreates
the scene from the movie The Lady Vanishes mentioned above. In this demonstration scene, different moods are evoked and different camera idioms can be applied. The protagonist in this scene, Dr. Hartz, can be freely directed by the player through the scene, and interact with other NPCs.

**Preliminary Results**

A Unity prototype has been created for automatic static real-time camera position setups, camera angle setups, and editing functions in three-person conversation scenes. This section provides an overview of the set of functionality achieved so far.

**Single character framing and angles**

Functions to set camera framing that ranges from close-ups to long shots, as well as functions to set camera horizontal and vertical angles are encoded in the VDS.

**Line of action calculation and apex camera**

VDS can define the line of action in the conversation and capture multiple characters in the camera in any position or size.

**Personal cameras for 3-people conversations**

The cameras for individuals are always staying on the same side of the apex camera based on the 180 rules; the individual camera can also function as over the shoulder shot. Rule of Thirds has been encoded in the system so that the character on left is displayed on the left one third and the character on right is displayed on the right one third.

**Camera switching function**

Algorithms for switching between apex cameras and cameras for one of the characters have been implemented. The camera can switch based on which character is talking.

**Over the Shoulder Shot**

If the characters are basically positioned face-to-face, this function can provide a shot of a target character or characters from another character’s shoulder. Over the shoulder shots can be set to use the left or right shoulder based on the scene.

**Summary and Future Work**

A Unity 3D prototype for real-time camera setups regarding automatic camera positioning, camera angles, and editing has been created for three-person conversation scenes. Further extension to more aspects and cinematographic styles of dynamic triadic conversations is underway. The goal of this research is a Visual Director System that is useful and easily reusable for dynamic three-character scenes in interactive storytelling games. The VDS is expected to result in computer-generated camera work that conforms to the anticipations associated with film-like scenarios.

Despite the spatial relations, dramatic tension and character actions that can change unpredictably in a dynamic interactive scene, the VDS can act as a director to ensure that the story is conveyed to players effectively using film techniques. Such a Visual Director System can extend the traditional way of storytelling in games, and enhance the visual experience for players in interactive storytelling games.
References


