Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2010)
M. Otaduy and Z. Popovic (Editors)

A Real-time Cinematography System for Interactive 3D Environments

C. Lino[1], M. Christie[1], F. Lamarche[1], G. Schofield[2] and P. Olivier[2]

[1] IRISA/INRIA Rennes Bretagne Atlantique, France
[2] School of Computing Science, Newcastle University, UK

Abstract

Developers of interactive 3D applications, such as computer games, are expending increasing levels of effort on the challenge of creating more narrative experiences in virtual worlds. As a result, there is a pressing requirement to automate an essential component of a narrative – the cinematography – and develop camera control techniques that can be utilized within the context of interactive environments in which actions are not known in advance. Such camera control algorithms should be capable of enforcing both low-level geometric constraints, such as the visibility of key subjects, and more elaborate properties related to cinematic conventions such as characteristic viewpoints and continuity editing. In this paper, we present a fully automated real-time cinematography system that constructs a movie from a sequence of low-level narrative elements (events, key subjects actions and key subject motions). Our system computes appropriate viewpoints on these narrative elements, plans paths between viewpoints and performs cuts following cinematic conventions. Additionally, it offers an expressive framework which delivers notable variations in directorial style.

Our process relies on a viewpoint space partitioning technique in 2D that identifies characteristic viewpoints of relevant actions for which we compute the partial and full visibility. These partitions, to which we refer as Director Volumes, provide a full characterization over the space of viewpoints. We build upon this spatial characterization to select the most appropriate director volumes, reason over the volumes to perform appropriate camera cuts and rely on traditional path-planning techniques to perform transitions. Our system represents a novel and expressive approach to cinematic camera control which stands in contrast to existing techniques that are mostly procedural, only concentrate on isolated aspects (visibility, transitions, editing, framing) or do not encounter for variations in directorial style.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.6]: Methodology and Techniques—Interaction Techniques

1. Introduction

With the advent of increasingly realistic 3D graphical environments, a requirement has arisen for the developers of many interactive applications to create more immersive narrative experiences for the user (eg. games, machinima, storytelling). Additionaly, there is a particular demand to recreate the cinematic experience of conventional movies. This not only requires an underlying narrative model to drive action and behavior, but the ability to specify and control virtual cameras such that they obey the conventions of traditional film making [Ari76]. This integration of narrative and cinematography in interactive 3D graphics applications requires camera control techniques that can address both low-level issues such as the visibility of scene elements, and high-level issues such as composition on the screen and enforcement of continuity rules between the shots. Though a number of systems have been proposed to support camera staging (*i.e.* to decide where to position and move the camera, see [CON08]), the integration of such work within an appropriately expressive camera editing process (*i.e.* how, when and where to cut between viewpoints) remains an open problem that has received limited attention from the computer graphics research community.

© The Eurographics Association 2010.

The provision of fully automated camera control for interactive environments, including viewpoint computation, viewpoint planning and editing, is a complex problem that raises three important issues:

1. Such a system should provide the user with well-composed viewpoints of key subjects in the scene following established conventions in the literature, while maintaining the visibility of these elements in fully interactive 3D environments. For example, the simultaneous computation of the visibility of multiple targets in a dynamic environment requires a pragmatic solution to provide some guarantee as to the efficacy of results, as well as a quantified estimation of visibility for a large range of viewpoints.
2. A second issue is related to the fundamental challenge of operationalizing editing rules and conventions in a sufficiently expressive model of automated editing. In relation to camera control, these rules and conventions are often expressed as a set of idioms – a stereotypical sequence of camera setups that allows the viewer to understand the sequence of actions occurring. For example, such idioms would describe conventional camera placements and movements for variety of settings involving two or more key subjects in dialogue.
3. The last issue is related to the possible variations in directorial style that a tool needs to offer. Style in cinematography remains a dimension that is difficult to define and is generally characterized in terms of empirical evidence. Nonetheless, an essential task consists in identifying some of the paramters and devices that underpin directorial style, such as pacing, dynamicity, preferred views and enforcement of specific cinematic conventions. As a consequence, the challenge consists in the design of a model expressive enough to control such paramters and devices and to yield a range of variations in results.

Contributions We propose a fully automated system that constructs a cinematically expressive movie in real-time from a flow of low-level narrative elements generated as the environment evolves. In essence, our approach contributes to the domain on the following points:

Real-time integrated system Our model proposes an integrated approach that characterizes full and partial visibility of key subjects, performs camera planning between viewpoints, and enforces continuity editing rules. Current contributions in the field generally focus on individual aspects of camera control (visibility and path-planning [OSTG09], editing [CAwH*96, ER07, AWD10], screen composition [BMBT00]).

Expressive cinematographic engine Our cinematographic engine encodes the empirical knowledge from practice and literature on camera staging (*e.g.* Internal, External, Parallel) as well as classical shot sizes (close, medium, long) through the notion of *Director Volumes* (spatial characterization of viewpoint regions around key subjects). Additionally, knowledge related to continuity editing and style is encoded by filtering operators over the Director volumes (line-of-interest, line-of-action, thirty-degree angle, ...). At last, our engine enforces on-the-screen composition that selects the best view from the resulting regions using a local search optimisation technique and following well-known composition rules.

Variation in directorial style Our system provides means to express notable variations in directorial style in a real-time context by integrating a filtering process over the Director Volumes, and controlling indicators such as pacing, camera dynamicity and composition to enforce high-level narrative dimensions such as isolation, dominance and affinity between key subjects.

Our paper is organized as follows. We first present a state of the art spanning over the multiple aspects of camera control (viewpoint computation, path-planning, editing). We then propose an overview of our approach in Section 3, before detailing the construction of Director Volumes (in Section 4) and the mechanics of the reasoning process for editing (in Section 5). We finally present results on a full 3D model of the canteen scene in Michael Radford's 1984 movie, in which we demonstrate variations in the application of directorial styles (pacing and dynamicity), together with variations in narrative dimensions.

2. Related work

Camera control in computer graphics is receiving an increasing attention from the research community (see [CON08]). Contributions span a range of subproblems, from viewpoint computation and the specification of screen-space properties, to camera motion control and the planning of visually felicitous paths. To contextualize our formulation of automated cinematography we consider a number of significant contributions across the broad spectrum of problems to be solved.

Viewpoint Computation Viewpoint computation refers to the process of positioning the camera to furnish an image with the desired set of cinematographic properties and which conveys the appropriate information wrt. the key subjects. The ability to compute viewpoints with specifiable properties is of particular value in a range of applications, including visual servoing, medical and scientific visualization, and image-based rendering. For such applications, viewpoints typically maximize the visibility of salient features and highlight the spatial relations between objects with a view to enhancing a viewer's perception and understanding of a scene. Vasquez *et al.* [VFSH03] proposed that a good view should maximize the amount of information about a scene for a viewer. Drawing on Information Theory they use the notion viewpoint entropy (defined in terms of the probability distribution the relative area of the projected faces over the sphere of directions centered in the viewpoint) to compute a minimal set of N good views.

© The Eurographics Association 2010.

Others have considered the problem of viewpoint selection as one of information composition. For example, Bares and colleagues [BMBT00] automatically generated camera shots using the composition heuristics of expert photographers. Christie & Normand [CN05] proposed a semantic partitioning of space into volumes that capture the regions of space around scene elements for which a camera will have a consistent set of visual properties. However such approaches to viewpoint computation mostly concentrate on static scenes in offline contexts (with the notable exception of [HHS01] and [CM01] who handle some elements of composition in real-time). By contrast, we rely on the real-time computation of characteristic views while propagating full and partial visibility of multiple key subjects and ensuring composition of elements on the screen using a local search optimization process.

Camera Planning Existing methods for camera motion planning augment classical planning approaches with features such as frame coherency and visibility along the path. Planning techniques can be divided into two main classes: local approaches, that constrain the search to a reduced set of candidate solutions (*e.g.* [HHS01]), and global approaches, which require the construction of, and search over, an a priori representation of the whole environment (*e.g.* [OSTG09]). Local approaches consider the use of techniques such (dynamic) potential fields [Bec02], camera-space gradients using visual servoing [CM01] or probabilistic roadmap techniques (PRM) using lazy evaluators to update the accessibility/visibility relations in the graph [LC08]. Although such techniques are responsive to local changes and dynamic elements of the scene (*i.e.* occluders) they typically fail to produce camera paths with properties that account for the global properties of a scene. Integrating global knowledge of the environment significantly improves the quality of the path but generally relies on an offline analysis of static environments (*e.g.* using PRMs [NO03]). Recently, Oskam *et al.* [OSTG09] presented a single target tracking real-time system that uses a global visibility-aware roadmap which locally adapts at run-time to dynamic occluders.

By contrast, our approach to camera planning computes paths along which we characterize the visibility of multiple key subjects. Global visibility is computed for static occluders through a cell-and-portal representation. We however do not encounter for dynamic occluders.

Editing Incorporating the editing process (where, when and how to cut between viewpoints) in computer graphics applications has been relatively under-addressed [HCS96, CAwH*96, BL97, AK01, FF04]. Typically this is realized as a set of idioms (series of reference viewpoints linked to a given configuration of key subjects) using a procedural language (*e.g.* a finite state machine): when a given action occurs, the associated viewpoint and transition is applied. Such approaches fail to recognize the importance of narrative discourse level [You07]. Integration of these narrative aspects has only recently been considered in [ER07, Jha09]. However, the realization has again been procedural in character, a script (*i.e.* a series of actions) is taken as input and an automated visualization of the scenario *w.r.t.* procedural cinematic rules is produced with an offline process. More recently, Assa *et al.* have proposed to study the correlation between the motion of key subjects in the 3D environment and the motion of the key subjects on the screen for a number of candidate cameras [AWD10]. This correlation metric is utilized to select the most appropriate sequence of views while enforcing coherency between shots. However key subject motion is only one of the key elements in editing.

In the application contexts we envisage, narrative elements are not known in advance and are continuously fed into our system at run-time. The automated cinematography system we propose maintains rhetorical consistency with past shots as in real movies, performs at interactive framerates, interrupts the current shot to handle new narrative elements as they become more relevant and offer means to enforce an expressive range of directorial styles.

3. Overview

Our system takes as input a 3D interactive environment (a geometry in which motions of key subjects are not known beforehand) together with a flow of *narrative elements* that provide a description of the actions performed in the environment. This notion of *narrative element* is a central notion in our approach and we look at its coherent translation into cinematographic viewpoints/idioms. A narrative element is a component of the discourse and conveys relevant information of an action from the story. Narrative elements carry in accordance with the nature of the story, prototypical actions (a key subject stands up, walks, talks to another key subject) to more subtle notions such as the display of a relationship between key subjects (*e.g.* show dominance, conflict or isolation). The output of our system is a movie which conveys these narrative elements according to some cinematic rules and directorial style parameters.

Our system follows a four-step process to compute viewpoints and transitions (see Fig. 1):

Selecting Narrative elements Whenever a new cut/transition is enabled, our system selects at run-time the most relevant narrative element among all events occuring during the same time step. The relevance is computed by an external process (*e.g.* a narrative engine) that distinguishes the relative importance of actions running in parallel.

Computing Director Volumes This stage turns selected narrative elements into Director volumes by composing areas of characteristic viewpoints that enable the portray of the element (the Semantic Volumes) with an analysis the full and partial visibility of key targets (Visibility Volumes).

© The Eurographics Association 2010.

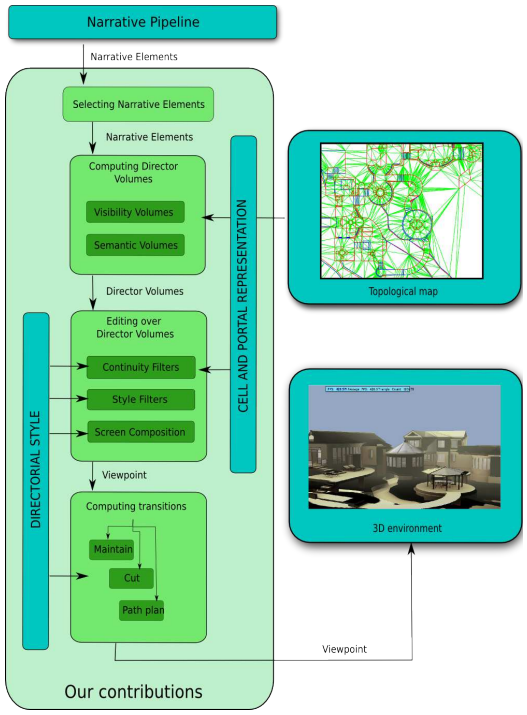


Figure 1: *Overview of our real-time cinematography system. First, relevant narrative elements are selected from a flow of events. Narrative elements are then turned into Director Volumes (all regions of space from which the narrative element can be portrayed). Editing is performed by reasoning over Director Volumes to select and compute an appropriate shot, given continuity rules, style rules and composition rules. Finally a transition between the previous and new shot is computed (as a cut or as a continuous transition).*

Editing over Director Volumes The selection of appropriate Director Volumes among all available is performed by applying multiple *filtering* processes. *Continuity filters* typically enforce classical continuity rules between shots by removing or pruning inconsistent regions in the set of Director Volumes (for example those that would make the camera cross a line-of-action). A second filtering process selects the most appropriate shots with relation to the narrative event by enforcing elements of directorial style (*Style Filters*). In a last step, a numerical optimization process is performed in each resulting Director Volume to select the best shot in terms of composition (exact locations of key subjects in the screen).

Computing transitions Whenever it appears necessary to switch between Director Volumes (end of narrative event, new narrative event, target(s) occluded, pacing in cuts), our system selects an appropriate transition (a cut or a continuous transition). For continuous transitions, we rely on a Dijkstra search process to plan an appropriate path between the current camera and a set of target Director Volumes.

In the sequel we detail the major steps of our computational process, namely computing Director volumes, reasoning over Director volumes and performing editing cuts and transitions.

4. Computing Director Volumes

At the core of our approach is a spatial representation, the *Director Volume*, that aggregates viewpoints that give rise to qualitatively equivalent shots in terms of information conveyed to the user, and in terms of visibility of key subjects.

4.1. Computing Semantic Volumes

The first step consists in translating the selected narrative element into semantic volumes. In our approach, each narrative element is associated with a collection of stereotypical viewpoints that convey the element according to established cinematographic conventions (see Fig. 3). A narrative element is modelled by a semantic tag (the performed action), a duration, and the key subjects involved in the action. Typically an action such as *Symes speaks to Smith* can be portrayed through a large collection of viewpoints described in the literature (Internal, External, Apex, Subjective) together with a range of distances (Close-Up, Medium Shot, Long shot, Extreme Long Shot) [Ari76]. Do note that for such an action, the utterance of a key subject may be portrayed either by framing the talking key subject, the listening key subject, or both (the choice is a matter of directorial style and continuity). By default each narrative element is portrayed by all possible viewpoints, and some specific elements may require a subset of shots (*e.g. Symes looks at Smith* obviously requires that the camera should not be located behind Syme). In the task of precisely designing the semantic volumes that encodes cinematographic knowledge, we followed an experimental process that consists in positioning and evaluating many possible shots in a 3D modeler for key configurations of subjects (single subject, two subjects facing, two subjects not facing, ...). Such as process was guided by empirical evidence in literature and a professional experienced modeler. As a result we characterized the spatial extents of viewpoint regions for key configurations of subjects. These semantic volumes are displayed in Figure 3. Additionally, we studied the evolution of these semantic volumes with relation to the distance between the key subjects, and the orientations of the key subjects. To represent such volumes, we rely on annotated Binary Space Partitions (a-BSPs), which are classical BSPs augmented with semantic tags on the branches. This structure is dynamically and procedurally generated with respect to the configuration of the key subjects, for each frame, and allows us to efficiently characterize and partition the environment into such sets of characteristic viewpoints (see accompanying videos).

© The Eurographics Association 2010.

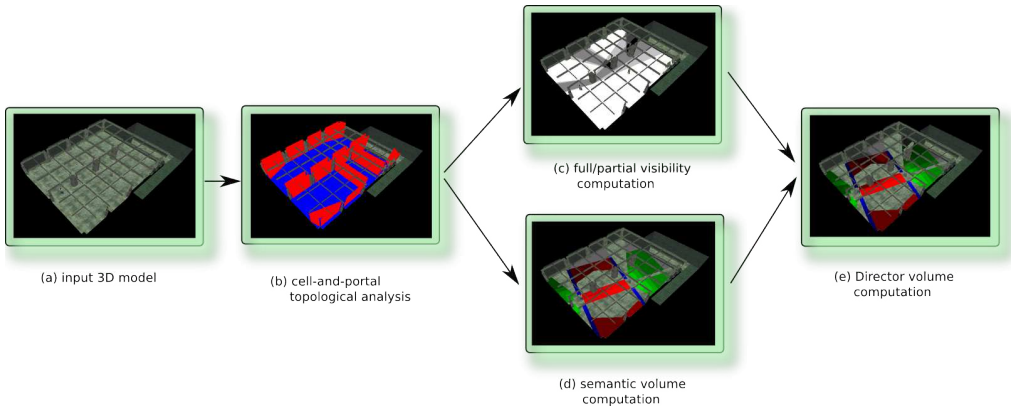


Figure 2: *Computation of Director Volumes (a) the input 3D model with two key subjects, (b) the pre-computed cell-and-portal representation (cells in blue, portals in red), (c) the visibility computation for both key subjects (white is visible, black is occluded and grey is partially visible), (d) the semantic volume computation which identifies characteristic viewpoints (internal/external/apex) and shots sizes (close shot to extreme long shot), (e) intersecting semantic and visibility volumes to generate Director Volumes.*

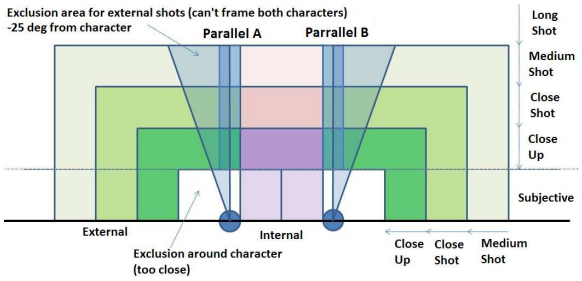


Figure 3: *Design of semantic volumes for a classical interaction between two key subjects (adapted from [Ari76]). Each circle represents a key subject.*

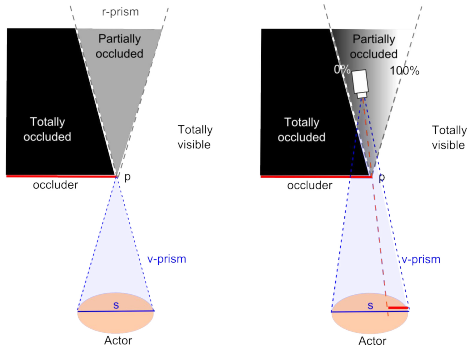


Figure 4: *From a vertex p, two stabbing lines are defined with respect to a key subject. By combining both stabbing lines, a categorization into three regions (full occlusion, partial visibility and full visibility) is obtained. 2D visibility of a key subject corresponds to a segment s linking its extreme points. From a viewpoint included in the region of partial visibility, the visible portion of the key subject is evaluated by using relative side of the occluder* w.r.t. *the stabbing line passing by the viewpoint and the extremity p of the occluder).*

4.2. Computing Visibility Volumes

The second step consists in computing and characterizing the visibility of key subjects involved in a narrative action. We address the problem of computational cost in visibility determination by reducing the practical complexity through two assumptions. Firstly, we abstract the 3D scene using a 2D cell-and-portal representation together with 2D convex hulls to represent key subjects. We use visibility propagation between adjacent cells to characterise areas of full or partial visibility over key subjects. The cell-and-portal representation is based on a topological analysis of the environment [Lam09]. Secondly, we ignore visibility issues due to dynamic occluders. The visibility process is therefore restricted to dynamic targets in static environments.

To characterize partial visibility, we extend classical cell-and-portal visibility propagation techniques [TS91] to handle from-region visibility where subjects are approximated as convex cells [COCSD00]. We base our visibility analysis on the use of stabbing lines to compute a dynamic categorization of the full and partial visibility of a key subject. For this purpose, we first identify the current convex cell in which the key subject is located, and list all adjacent cells which can be reached through a portal. For a point p (a vertex of a portal), two stabbing lines are defined such that each line is tangential to, but on opposite sides of the convex hull of the key subject (fig. 4).

© The Eurographics Association 2010.

The stabbing line linking an extremal point e of a key subject to a point p (vertex of a portal) separates regions where e is visible and e is occluded. Since the key subject is fully included between stabbing lines, the visibility categorization of the whole key subject is then performed by combining visibility categorization of each stabbing line: (1) the visibility prism (v-prism) is defined as the triangle between p and the extremal points, and the segment s (joining both extremal points) is seen as the abstraction of the 2D visible part of the actor from p. (2) the reversed prism (r-prism) is computed and defines the partial visibility region for point p. Three regions are then obtained: a region where the key subject is fully occluded, a region where the key subject is fully visible, and an intermediate region where the key subject is partially visible. Moreover, for any viewpoint v with partial visibility, an estimation of the visibility degree of the actor at v is processed by computing the visible portion of segment s.

The process is repeated by computing the visibility information for all portals connected to the current cell (see Fig. 5). Areas of full visibility, partial visibility and no visibility are then further propagated in the neighbor cells through the cell-and-portal representation in a way similar to [TS91] (for an illustration see Fig. 2c). An a-BSP representation is used inside each topological cell. A node of this a-BSP represents a stabbing line, and a leaf corresponds to a sub-cell (sub-region of a cell) for which visibility is characterized. This a-BSP representation enables (1) a cumulative cell subdivision process for different key subjects; and, (2) a reduction of the search complexity when characterizing a given viewpoint in the topological cell.

We have computed semantic volumes corresponding to characteristic viewpoints, and visibility volumes with relation to key subjects. We build Director Volumes by combining visibility and semantic volumes – through the fusion of our a-BSP representations– to obtain a full characterization of the possible viewpoints portraying a narrative element (see Fig. 2c and d). This process lays the groundwork for the camera editing process by proposing a range of possible viewpoint regions where the camera can move to or cut to.

5. Reasoning over Director Volumes

Once the space of viewpoints have been characterized in terms of Director Volumes, the next step consists in selecting the best candidate volume following an encoding of continuity rules and directorial style.

In such a reactive environment, our system constantly reasons as to whether to maintain the current location of the camera, perform a cut, make a transition to another viewpoint, or react to a new narrative element that has occurred. These choices depends on: (1) the current and incoming narrative elements (and their likely duration), (2) the current viewpoint, (3) the past viewpoint and (4) continuity and style parameters.

Our editing model is parameterized in terms of *continuity* and *style*.

5.1. Continuity editing

Continuity affords compliance with cinematographic conventions and enforces the maintenance of the spatio-temporal context when the camera changes viewpoint. Conventions are well established in relation to the key subjects' motions, spatial locations, line-of-interest and actions. For example, when switching between shots in a dialogue, relative screen locations of key subjects should be maintained. Though some conventions are implicitly encoded within idioms, it is necessary to maintain continuity when the camera changes from one narrative element to another, when visibility fails in the middle of an idiom, or in case of interleaved idioms (parallel editing). As a result, the continuity process acts as a filter over the viewpoints and idioms.

Continuity rules are encoded as filtering operators which process a Director Volume and returns a new Director Volume in which viewpoint regions may have been pruned. Our cinematography engine encodes the following continuity rules:

Line-of-action continuity Between two shots relative to the same key subject, coherency must be maintained in the apparent direction of motion of this key subject. Therefore, viewpoints located on the opposite side with relation to the direction of motion are discarded by inserting a new plane into the a-BSP and computing the resulting intersection of this half space with Director Volumes.

Line-of-interest continuity In narrative elements that involve two or more key subjects, once the side of the line-of-interest (imaginary line linking two key subjects) has been chosen, it should not be crossed, unless using an extreme long shot (that re-establishes the key subjects in relation to the environment) or a continuous transition. The line-of-interest is recomputed at each frame and all Director Volumes on the opposite side, but extreme long shots, are discarded. In a similar way, the process injects a new plane in the a-BSP representation and computes the resulting intersection.

Change in angle or size Between two shots related to the same narrative element, there should be at least a thirty-degree difference in orientation, or a notable difference in size of portrayed elements. The rule is implemented by first removing all viewpoints from director volumes that subtend an angle lower and 30 degree to the subject (computed by injecting a new half plane in the a-BSP), and then building the union of the resulting director volumes with all director volumes which are at least two units different in size (in a graduation that ranges in close-up, medium close-up, long shot and extreme long shot).

© The Eurographics Association 2010.

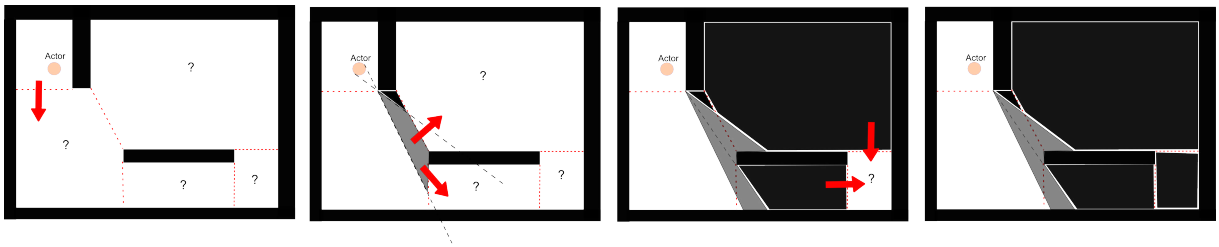


Figure 5: *Visibility propagation using the cell-and-portal representation. Each visibility prism issued from an edge of a portal is propagated through the portals of the adjacent cells. The process is repeated inside each cell until all cells are explored.*

5.2. Style in Director Volumes

Style in Director volumes is defined as a number of preferences as to which idioms/viewpoints should be considered first when available (or visible). Style is encoded as a filtering process that ranks the Director Volumes according to different style parameters (variation in the choice of successive volumes, preferred viewpoints, narrative dimensions to enforce such as affinity, dominance or isolation – see Results in Section 6).

Unless cases of failure (all Director Volumes have been filtered), the order in which the filters are applied have no impact on the result but has an impact on the computational cost.

5.3. Failures in available Director Volumes

A number of cases occur in which filtering operators may remove all possible Director Volumes (before applying Style filters). We handle such failures in two ways. First, in our design, extreme long-shots are not impacted by continuity rules. Such shots are generally used to establish the relation between a key subject in his environement, and film makers classicaly use this device. Second, continuous transtions can be applied to violate continuity rules (crossing the line of interest) as long as the spatial continuity is maintained.

5.4. Enforcing Screen Composition

The last step consists in selecting an appropriate viewpoint inside a Director Volumes (or a set of Director Volumes in case of multiple candidates) and enforcing some composition rules (*e.g.* rule of the thirds and balance in the image). By definition each Semantic Volume encompasses some implicit elements of composition (number of key subjects and relative locations of the subjects). For example, an apex shot should portray two key subjects respectively on the left and on the right on the screen. In our approach we express the composition through the exact location of key subjects on the screen. When key subjects are key subjects, we actually consider key subjects eye location or key subjects heads location (depending on the shot size). Thereby, each Semantic Volume directly contains values of composition for one, two or more key subjects on the screen. Since the volumes already encompass a notion of size (close-shot to extreme long-shot), we do not model the size of the key targets on the screen. These default values can be modified to enforce some specific dimensions (*e.g.* dominance of a key subject can be expressed by constraining his eye-line to be higher on screen than the eye-line of the dominated key subject – see Results in section 6).

In order to select the appropriate shot, a local search optimization process is performed in the set of Director volumes. Starting from the barycenter of the director volume, a number of neighbour candidates are generated and evaluated with regard to composition. The best neighbour is chosen as the current configuration and the process iterates on a new set of neighbors. Iteration stops whenever no improvement is made or when a threshold in iterations is reached. At this step, the handling of dynamic occluders could be included by adding a cost related to the visibility of key subjects for dynamic occluders. Simple techniques such ray-casts or depth rendering can be integrated with little overhead.

5.5. Performing Cuts and Transitions

In our approach we introduce two indicators of editing that control cuts and transitions *pacing* and *dynamicity*:

Pacing

As described in literature, cuts are motivated either by style or by necessity. When a matter of necessity, cuts may be performed at any moment in the movie. Such cuts are motivated by the occlusion of a key target, or the fact that a key target leaves the frame. When a matter of style, cuts are driven in our model by a specific indicator: the *pacing* which represents the rate at which cuts are performed. Our pacing is modelled with two boundary values: a minimum (p_{min}) and a maximum (p_{max}) shot duration. The probability of cutting within these bounds is driven by a Gaussian (the closer to p_{min} the less probable, the closer to p_{max} the more probable). Cuts will occur either due to the arrival of narrative elements of higher relevance, or due to the modelling of a narrative element as a complex idiom (sequence of semantic volumes);

© The Eurographics Association 2010.

Dynamicity

Changes in viewpoints may be realized through continuous transitions (*i.e.* camera paths). Such changes are typically a matter of cinematographic style, which we model through the *dynamicity* indicator. In our implementation, dynamicity ranges from static shots, panoramics, travellings and free camera motions. Interestingly, continuous transitions may allow the director to break continuity rules, for example by crossing the line-of-interest or the line-of-action. Thus in cases where no Director Volumes are applicable after filtering on continuity, we allow the camera to continuously move to any visible Director Volume.

Building a Planning Graph To enforce continuous transitions between Director Volumes, we build a roadmap from both the cell-and-portal connectivity graph and the a-BSP partitions contained in each cell. By construction, the cell-and-portal decomposition offers a connectivity graph that avoids obstacles. A finer roadmap is constructed by: (1) locally sampling the portal segments and BSP boundaries to generate more way-point nodes; and (2) creating edges that connect all pairs of way-points nodes together.

Computing Transitions The task of planning a path between Director volumes is expressed as a classical multi-criteria optimization process that includes criteria such as the visibility of key subjects along the path and the length of the path. The search is performed with a Dijkstra process that handles multiple target cells and stops as soon as a solution is found. The cost function is expressed as a weighted sum of the path length and visibility of key subjects. The cost c of a path t composed by s segments is therefore defined as

$$c(t) = \sum_{s \in t} \phi(s)\Delta(s)$$

where $\phi(s)$ stands for the visibility v_s along the segment s ($\phi(s) = \alpha * d_s + 1$, $\alpha > 0$, where d_s is the distance between v_s and $[v_{min}, v_{max}]$, and $d_s = 0$ when $v_s \in [v_{min}, v_{max}]$). $\Delta(s) = |s| > 0$ represents the length of the segment s. Such a representation allows to favor paths with little, no or many occlusions on the way.

6. Results

We illustrate the key features of our automated cinematography system by exploring the possible variations in terms of camera dynamicity (from fixed shots to complete dynamic shots) and high-level narrative dimensions of dominance, affinity and isolation. For the materials, we rely on a fully animated 3D environment of the Canteen scene in Michael Radford's 1984 movie, built for the need of comparing edits. A full annotation of all actions in the movie (more than one hundred) provide the narrative element inputs to our system. All 3D models, XML representations of narrative elements and cinematic idioms, together with resulting videos are available at http:\\www.cameracontrol.org/1984-Canteen-scene .

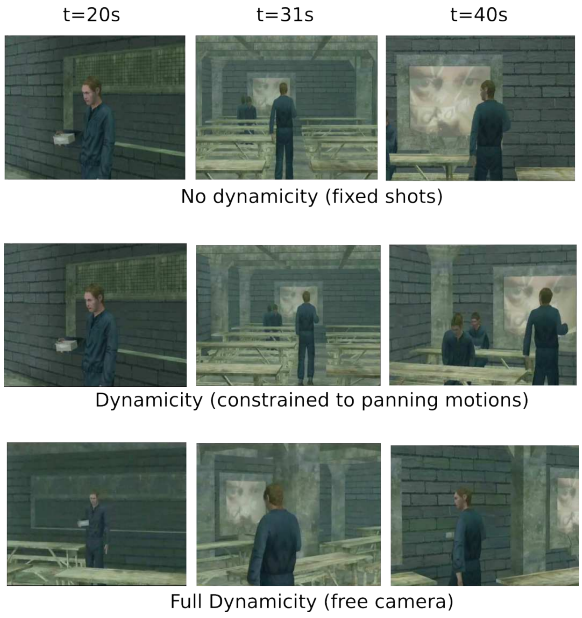


Figure 6: *Variations in dynamicity for the narrative action Parsons walks to the table at time frames $t = 20, t = 31, t = 40$ for the same geometry. On the top row, the camera setups are fixed (in location and in orientation). In the middle row, the camera is allowed some panning motions, and in the last row, the camera is fully free in the environment.*

6.1. Variations in Dynamicity

Our second example illustrates the possibility our system offers to control the camera dynamicity, which encompasses both the motion camera performs inside a Director Volume (none, panoramic, travelling) and the transition between shots (cut or continuous transitions). Here we display and compare four levels of dynamicity: no dynamicity (static shots, cuts between shots), panning dynamicity (allows panning motions, cuts between shots), panning+travelling dynamicity and full dynamicity (free camera motions, continuous transitions between shots).

Figure 6 illustrates the results of variations in dynamicity for the narrative action *Parsons walks to the table*, at three different time steps.

As a result, our camera control system efficiently plans sequences of camera shots while enforcing cinematic continuity. Once a side of the dialogue between Smith and Syme is chosen, the line-of-interest rule is maintained. Similarly when key subject Parsons walks to the table, the line-of-action rule is maintained. Please refer to videos for a detailed view of results. Table 6.1 details the performances for all values of dynamicity. This computation includes the cost of creating and intersecting the visibility and semantic volumes, filtering the director volumes, planning paths (when necessary) and computing the screen composition. The cost

© The Eurographics Association 2010.

	None	Pan.	Trav.+Pan.	Full
Frame rate	244.3	245.6	222.4	202.0

Table 1: *Average frame rates for variations in dynamicity. Though full dynamicity requires the re-creation of the roadmap and the computation of paths at frequent intervals (every 0.5s), the process remains fully real-time (all tests performed on a Intel i-7 core with 4Gb of RAM.*

of the Full dynamicity remains important due to the necessity to re-create the roadmap and perform the queries at frequent intervals (in this benchmark, every 0.5s). Cases where source and target volumes are far away impact strongly the framerate (Table 6.1 only displays an average of the whole scene).

6.2. Variations in Narrative Dimensions

To illustrate the expressiveness of our cinematography engine, we propose to explore three narrative dimensions (Dominance, Affinity, Isolation) through the implementation of a specific style filter. We first describe the devices to express such dimensions in cinematography before detailing the means to enforce them in our system.

Affinity. Affinity between key subjects is classically accomplished by preferring balanced and symmetric shots where similarity between key subjects can easily be established. In such, apex shots are preferred to external shots, symmetry is enforced in the framing and eye-levels are constrained at the same height inside shots and between shots. Furthermore reaction shots generally follow key subject utterances.

Dominance. Dominance is accomplished by offering asymmetric views, in which dominant key subjects appear larger, more often and for longer durations than dominated key subjects. As a consequence, external shots and subjective views of the dominant key subject are preferred over internal and apex shots. In terms of eye-levels, the dominant key subject gaze is generally higher than the dominated key subject. Another device that is considered is the use of high-angle shots of the dominated key subject with low-level shots of the dominant.

Isolation. Isolation of a key subject is enforced by preferring shots displaying only that key subject (over apex and external shots), spending more time on him than other key subjects. In terms of composition, the key subject is generally framed within his environment (medium shot or long shot) with large empty space around or in front of him.

In our system, the enforcement of these dimensions impact both the Style filter as well as the composition process. For each dimension a specific Style filter has been implemented. This style filter typically ranks the preferred views first and undesired views last. In case of dominance for example, all Director Volumes that represent apex shots are ranked negatively and all external and subjective views of the dominant key subject are ranked positively. Within the composition process, the default neutral screen compositions are replaced by specific values. Dominance, once again, is enforced by changing the height of the dominant key subject eye-lines in external shots, and using low angles in subjective view of the dominant key subject (see Figure 7, bottom left shot).



Figure 7: *Exploration of narrative dimensions in the dialogue between Symes and Smith, at four different moments. In bottom left shots, Syme is considered the dominant key subject. We see that the composition changes (higher eye-line on screen) and that an external shot is selected. In bottom right shots Syme is the isolated key subject; Syme is the only key subject framed, and composition reserves important space ahead of the key subject.*

Figure 7 displays a comparison of results for neutral, affinity, dominance and isolation dimensions in the 1984 canteen scene. This example displays the expressiveness of our framework in which new filtering operators can easily be added to enforce specific camera behaviors (while maintaining previous ones active). Refer to videos for more details. All four results present similar performances (around 240 fps) since the cost of the new filtering process is low.

6.3. Discussion

A limitation of our approach stands in that the visibility computation for static occluders is performed in 2D. Though the 2D cells are extruded to 2D$\frac{1}{2}$ (by considering the respective ceiling heights) and composition is computed in full 3D, the propagation of visibility between cells fundamentally relies on a 2D process. Therefore, in environments with multiple levels, inter-level visibility will not be computed although this could potentially be addressed by pre-computing inter-cell visibility for cells on different levels using ray casting techniques (in a way similar to [OSTG09]). Additionally vis-

© The Eurographics Association 2010.

bility of key targets wrt dynamic occluders can be handled by including ray-casting or depth rendering techniques in the screen-composition process.

7. Conclusion

In this paper we have presented simple and efficient solution mechanisms to the problem of interactive camera control. We address the innate complexity of well understood problems such as visibility determination and path planning required in real-time camera control, while tackling higher-level issues related to continuity between successive shots in an expressive editing model. Our real-time cinematography system encodes cinematic idioms and coherency rules to produce appropriate edits and camera paths from a set of narrative actions. The model relies on a spatial partitioning providing a characterization into visibility cells (fully visible, partially visible, or fully occluded) and characteristic viewpoint cells (the semantic volumes). We reason on these cells to identify how, when and where shot transitions should be performed, utilizing a filtering-based encoding of cinematic conventions together with the possibility to implement different directorial styles. The expressiveness of our system stands in stark contrast to existing approaches that are either procedural in character, non-interactive or do not account for proper visibility of key subjects.

8. Acknowledgements

This work was supported in part by the EU FP7 network of excellence IRIS (ICT-231824) Integrating Research in Interactive Storytelling.

References

[AK01] AMERSON D., KIME S.: Real-time Cinematic Camera Control for Interactive Narratives. In *The Wotking Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment* (Stanford, CA, 2001).

[Ari76] ARIJON D.: *Grammar of the Film Language*. Hastings House Publishers, 1976.

[AWD10] ASSA J., WOLF L., DANIEL C.-O.: The Virtual Director: a Correlation-Based Online Viewing of Human Motion. In *Proceedings of the Eurographics Conference* (2010).

[Bec02] BECKHAUS S.: *Dynamic Potential Fields for Guided Exploration in Virtual Environments*. PhD thesis, Fakultät für Informatik, University of Magdeburg, 2002.

[BL97] BARES W. H., LESTER J. C.: Cinematographic user models for automated realtime camera control in dynamic 3D environments. In *Proceedings of the sixth International Conference on User Modeling* (Vien New York, 1997), Springer-Verlag.

[BMBT00] BARES W., MCDERMOTT S., BOUDREAUX C., THAINIMIT S.: Virtual 3D camera composition from frame constraints. In *Proceedings of the ACM international conference on Multimedia (MULTIMEDIA '00)* (2000), ACM Press.

[CAwH*96] CHRISTIANSON D. B., ANDERSON S. E., WEI HE L., SALESIN D. H., WELD D. S., COHEN M. F.: Declarative Camera Control for Automatic Cinematography. In *Proceedings of the 13th National Conference on Artificial Intelligence and the 8th Innovative Applications of Artificial Intelligence Conference* (1996), AAAI Press / MIT Press.

[CM01] COURTY N., MARCHAND E.: Computer animation: A new application for image-based visual servoing. In *Proceedings of International Conference on Robotics and Automation* (2001).

[CN05] CHRISTIE M., NORMAND J.-M.: A semantic space partitioning approach to virtual camera control. In *Proceedings of the Eurographics Conference* (2005), vol. 24, Computer Graphics Forum.

[COCSD00] COHEN-OR D., CHRYSANTHOU Y., SILVA C., DURAND F.: A survey of visibility techniques for walkthrough applications. *Transations on Visualization and Computer Graphics* (2000).

[CON08] CHRISTIE M., OLIVIER P., NORMAND J.-M.: Camera control in computer graphics. *Computer Graphics Forum 27*, 8 (2008).

[ER07] ELSON D. K., RIEDL M. O.: A lightweight intelligent virtual cinematography system for machinima production. In *Proceedings of the 3rd Conf. on AI for Interactive Entertainment* (2007).

[FF04] FRIEDMAN D. A., FELDMAN Y. A.: Knowledge-Based Cinematography and Its Applications. In *Proceedings of the 16th European Conference on Artificial Intelligence* (2004).

[HCS96] HE L., COHEN M. F., SALESIN D. H.: The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In *Proceedings of SIGGRAPH* (Aug. 1996), ACM Computer Graphics.

[HHS01] HALPER N., HELBING R., STROTHOTTE T.: A camera engine for computer games: Managing the trade-off between constraint satisfaction and frame coherence. In *Proceedings of the Eurographics Conference* (2001), vol. 20, Computer Graphics Forum.

[Jha09] JHALA A.: *Cinematic Discourse Generation*. PhD thesis, Faculty of North Carolina State University, 2009.

[Lam09] LAMARCHE F.: Topoplan: a topological path planner for real time human navigation under floor and ceiling constraints. *Computer Graphics Forum 28*, 2 (2009).

[LC08] LI T.-Y., CHENG C.-C.: Real-time camera planning for navigation in virtual environments. In *SG '08: Proceedings of the 9th international symposium on Smart Graphics* (2008), Springer-Verlag.

[NO03] NIEUWENHUISEN D., OVERMARS M. H.: *Motion Planning for Camera Movements in Virtual Environments*. Tech. Rep. UU-CS-2003-004, Institute of Information and Computing Sciences, Utrecht University, 2003.

[OSTG09] OSKAM T., SUMNER R. W., THUEREY N., GROSS M.: Visibility transition planning for dynamic camera control. In *Proceedings of Symposium on Computer animation (SCA)* (2009).

[TS91] TELLER S. J., SÉQUIN C. H.: Visibility preprocessing for interactive walkthroughs. In *SIGGRAPH '91: Proceedings of the 18th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1991), ACM.

[VFSH03] VÀZQUEZ P.-P., FEIXAS M., SBERT M., HEIDRICH W.: Automatic view selection using viewpoint entropy and its application to image-based modelling. *Computer Graphics Forum 22*, 4 (2003).

[You07] YOUNG R. M.: Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies 8*, 2 (2007).

© The Eurographics Association 2010.