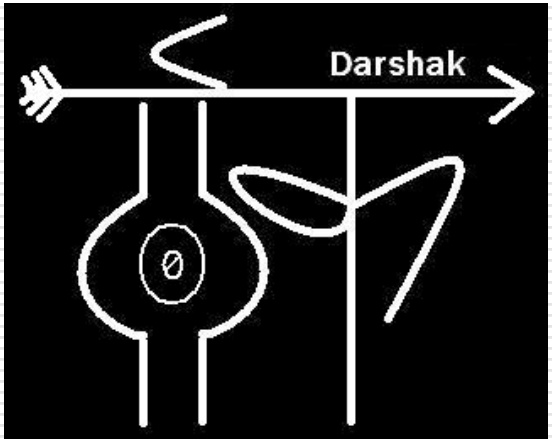
Darshak: An Intelligent Cinematic Camera Planning System



Darshak (Sanskrit) = Viewer

Arnav Jhala
Department of Computer Science
North Carolina State University

Advisor: **R Michael Young**

Committee:
James Lester
Jon Doyle
William Bares
Tim Buie

ahjhala@unity.ncsu.edu

Liquid Narrative Group

NC STATE UNIVERSITY

Introduction

- 3D Environments
  - Simulation and Training environments
  - Immersive 3D virtual worlds (video games, animated movies, virtual tours etc.)
- Virtual Camera: Tool for communicating information from within 3D environments
- Challenges
  - Dynamic environments
  - Communicating dynamic storylines
  - Navigation in complex environments

The Problem

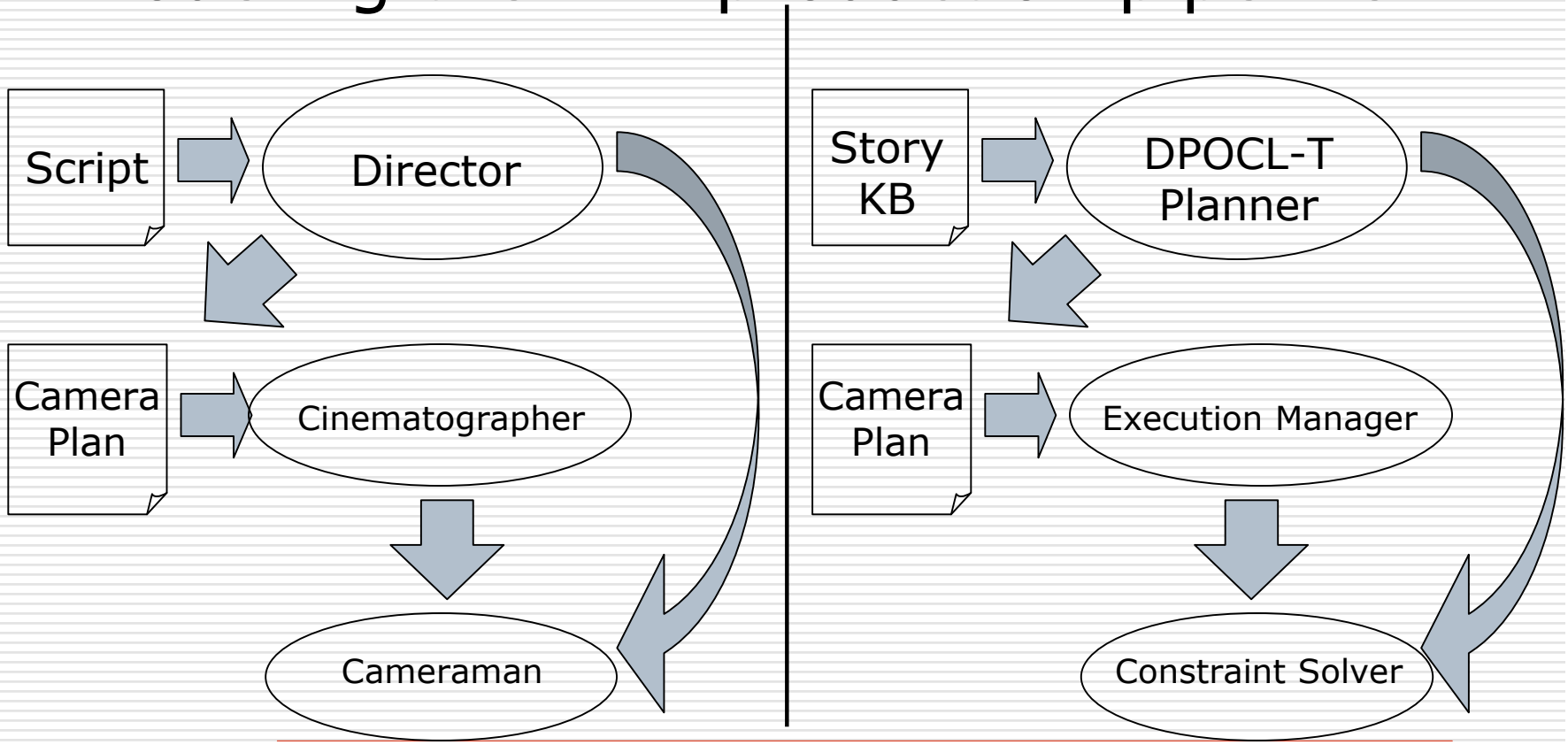
- Automatic placement of a virtual camera in a 3D environment with dynamic storylines
  - *Where* to place the camera?
    - Determined by the context and underlying 3D world layout
  - *How* to place the camera?
    - Solving equations to determine location, orientation, field of view and other camera parameters
  - *Why* to place the camera?
    - Satisfying the communicative goals to be conveyed to the viewer

Contributions

- Extending discourse planning techniques to a new medium
- Formalizing film idioms as plan operators and capturing the reasoning behind established techniques from cinema
- Improving navigation and storytelling in simulation and game environments

Approach

- Modeling the film production pipeline



Story Representation

- Knowledge base containing declarative description of the story
  - Story-world domain knowledge

```
(character Royce)(location living_room)
```

  - Story events and actions

```
(step s1)(act-type s1 speak)
(agent s1 Marguerite) (secondary s1 Royce)
(mood Marguerite angry (during s1))
```

  - Scene Boundary Markers

```
(Conversation c1) (Conv-Steps (s1 s2 s3))
(Conversation-Type Remote) (Conv-Start s1) (Conv-End s3)
```

Representing Cinematic Knowledge

- Camera shots are defined as *intentional* communicative acts *planned* by the camera planner to influence the *beliefs* and *mental states* of the *viewer (V)*
- Language of Film
  - Small set of well-defined primitive shots
  - Hierarchical nature
  - Causal relationships between shots
  - Composition and continuity rules

Primitive Operators (Shots)

Type: PanWithActor

Parameters: ?focus, ?shot-type, ?dir, $?T_{start}$, $?T_{end}$

Preconditions:
(infocus ?focus)@[$?T_{start}$)
(not (panning ?shot-type ?dir))@[$?T_1$, $?T_{start}$]

Constraints:
(> T_{end} T_{start}) (< (- T_{end} T_{start}) 10)

Effects:
(infocus ?focus)@[
(panning ?shot-type ?dir)

Look-At
- Extreme-Close-Up
- Close-Up
- Med-Shot
- Long-Shot
- Extreme-Long-Shot

Pan
- Pan-With-Actor
- Pan-ActorToActor

Track
- Track-Relative
- Track-Absolute

POV

Dolly

Tilt

Transitions
- Cut
- Fade In/Out
- Dissolve

Conversation Shots
- Internal
- External
- Reverse
- Master-Shot

Type: LookAt

Parameters: ?focus, ?shot-type, ?dir, $?T_{start}$, $?T_{end}$

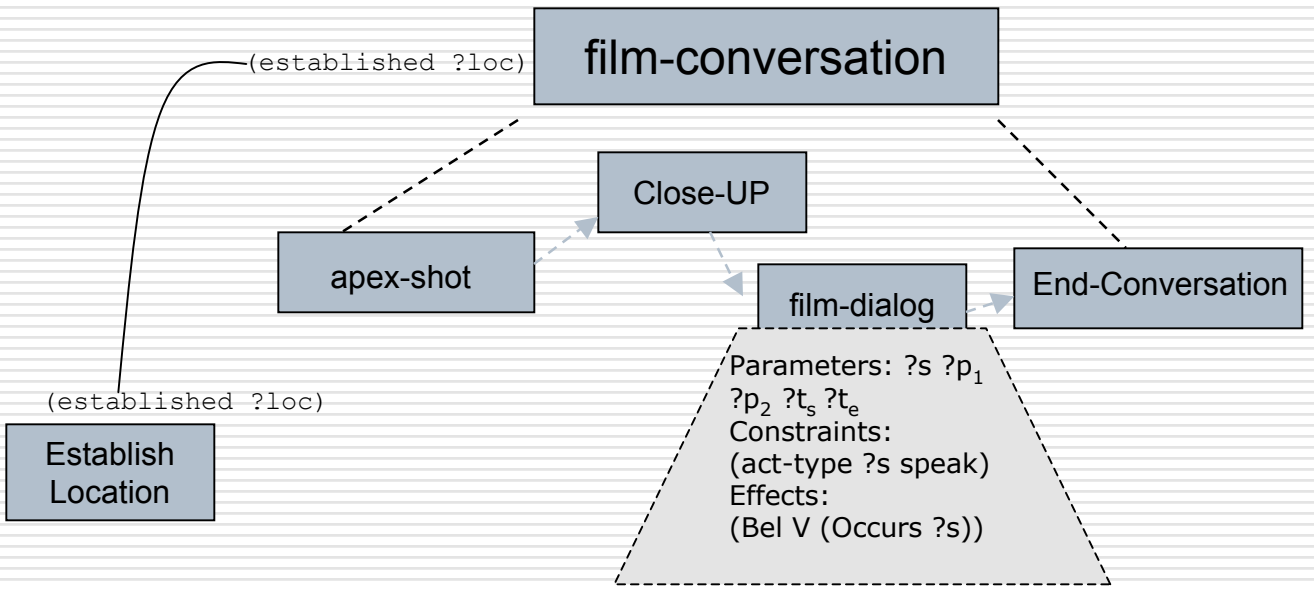
Preconditions:
(not (infocus ?focus))@[$?T_{start}$)

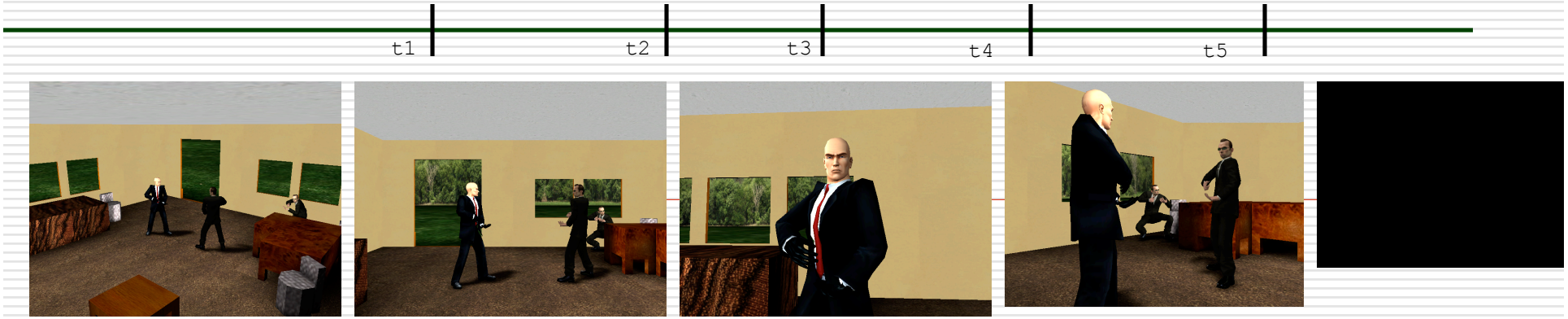
Constraints:
(> T_{end} T_{start})

Effects:
(infocus ?focus)@[$?T_{start,}$ $?T_{end}$)

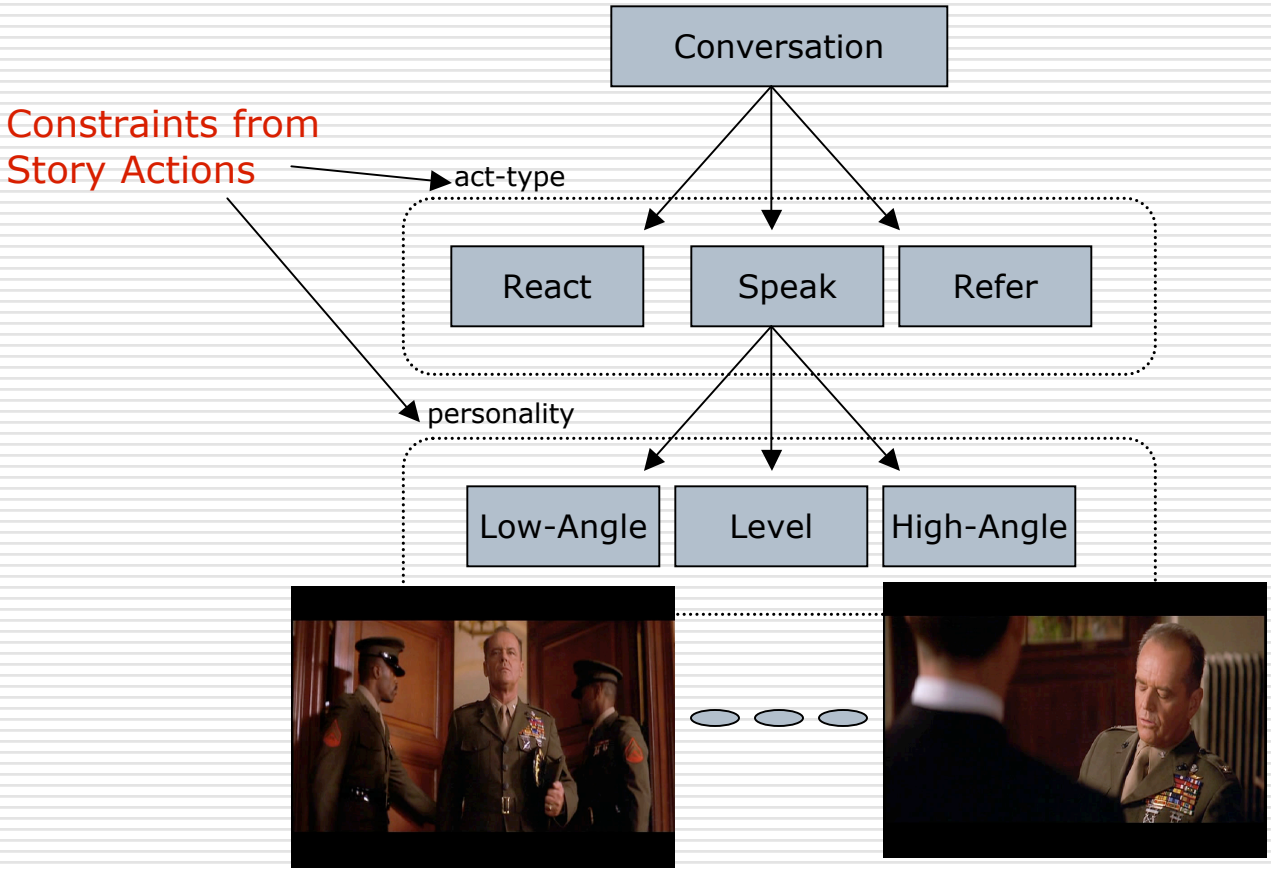
Abstract Operators (Scenes)

```
(define (decomposition film-conversation)
  :parameters (?c)
  :constraints ((conversation ?c) (conv-
        steps ?c ?slist))
  :tconstraints((starts-at ?st (start (first
        ?slist)))
  :steps (
        (step1 (apex-shot ?c))
          (forall ?step in ?slist
                (step2 (film-dialog ?step)))
        )
  :orderings ((step1 step2))
  :rewrites (((BEL V (Occurs ?c))
            ((forall ?step in ?slist
                (BEL V (Occurs ?step)
                  )))))
          )
```





Hierarchical Nature of Film Idioms



Planning Algorithm

- Decompositional Partial Order Causal-Link Temporal Planning (DPOCL-T)
- Camera actions are durative
  - e.g. shot-length affects tempo of the story
- Temporal Links between camera plan steps and story events
  - Camera actions are tied to time of execution of the event they film

Planning Process

Input Story: Bank Robbery (Annotated)

[S1] Guard leaves bank to go to starbucks
(effect S1 (not (at bank guard)))

(emotion guard neutral)
(mood s1 neutral)

[S2] Thief enters bank
(prec S2 (not (at bank guard)))
(effect S3 (at bank thief))

(knows thief (not (at bank guard))

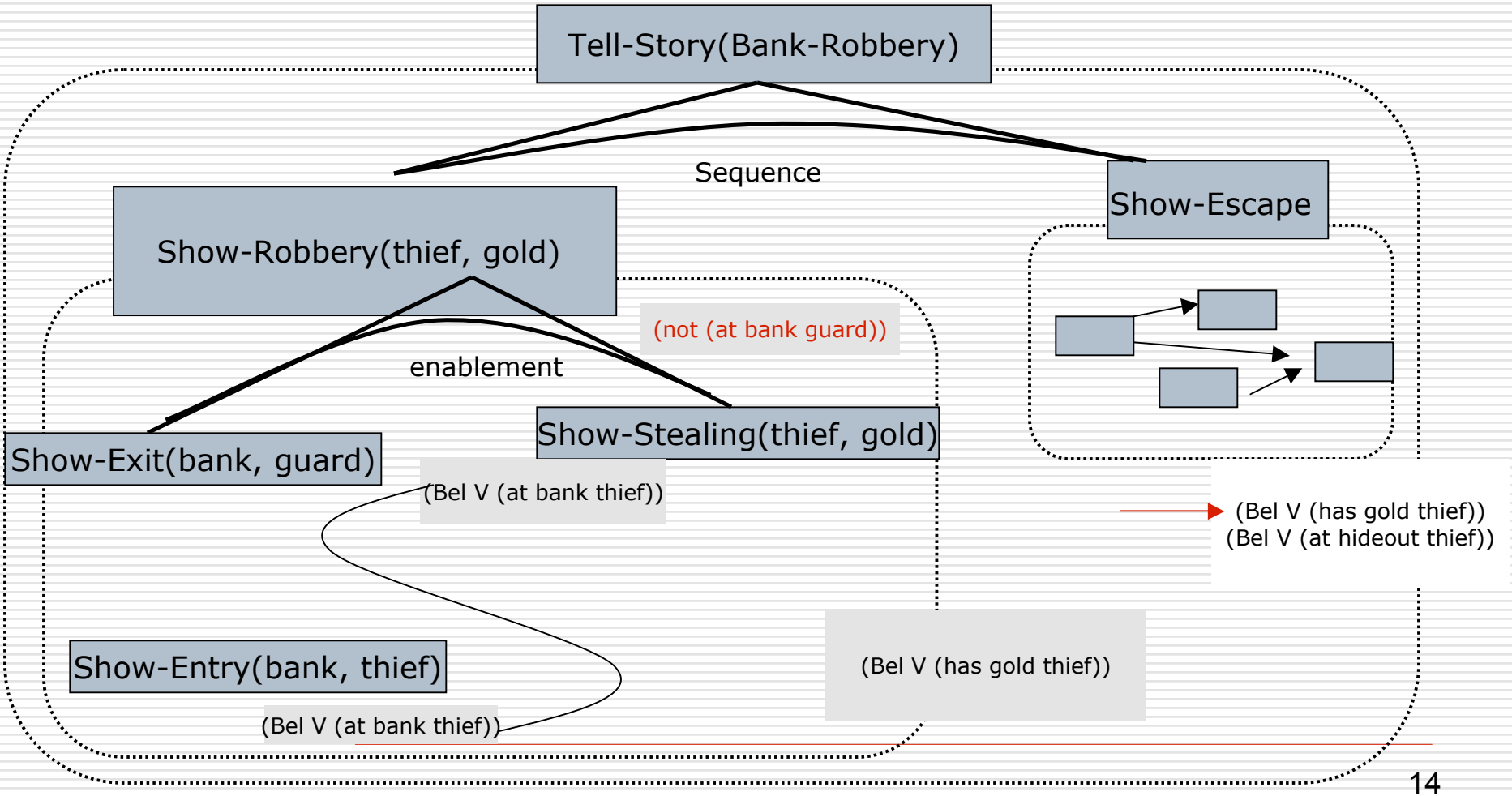
[S3] Thief steals gold
(prec S3 (at bank thief))
(effect S3 (has gold thief))

(mood s3 tense)

(emotion thief excited)
(mood s1 neutral)

[S4] Thief runs away
(effect S4 (at hideout thief))

Discourse Planning



Show-Exit(bank, guard)

Over-The-Shoulder (thief, s1)

(Bel V (Knows thief (not (at bank guard))))

S1

Over-The Shoulder

Parameters: ?p, ?a, ?e, ?ts, ?te
Constraints:
(character ?p) (step ?a) (effect ?a ?e)
(= ?ts (start ?a))
Effects:
(Bel V (Knows ?p ?e))

Cinematographer and Cameraman

- Cinematographer
  - Manages execution
  - Sets geometric constraints
  - Interface with the constraint solver
- Cameraman
  - Continuously maintains the camera at a solution that satisfies all the current constraints

Evaluation

- Viewer comprehension
  - The planner selects shots that convey intended information
  - Challenge: Removing voice-overs, facial expressions and other extra-diegetic communicative factors
- Can viewers pick out system generated movies from expert generated ones
  - Can the system generated movie be differentiated from expert generated ones
  - Experts use the same primitive operators that are available to the planner

Status

- Completed
  - Planning algorithm with interval temporal reasoning
  - Static idioms with limited moving camera shots
  - Toy domains
- In-Progress/Future Work
  - Point-based temporal reasoning
  - Testing more complex idioms
  - Applying to a larger domain
  - Evaluation

Open Issues

- Cinematic operators and heuristics
- Evaluation strategies
- Handling interactivity
- Incorporating feedback from the graphics engine

Related work: Geometric placement

- Visual primitives[Drucker et al. 94]
- Film Idioms[He li wei et al. 96]
- Constraint Solver[Bares, Lester 97]
- Genetic Algorithms [Halper, Olivier 01]
- Neural Networks [Hornung 03]

Drucker Steven *Intelligent Camera Control for Graphical Environments*, PhD Dissertation, MIT 1994

He Li-wei, Cohen Michael, Salesin David *The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing*, Computer Graphics Proceedings, 1996

Bares William, Lester James, *Cinematographic User Models for Automated Realtime Camera Control in Dynamic 3D Environments*, Proceedings of Sixth International Conference, UM-97

Nicolas Halper, Ralf Helbing, and Thomas Strothotte. *A Camera Engine for Computer Games: Managing the Trade-Off Between Constraint Satisfaction and Frame Coherence*. Computer Graphics Forum: Proceedings Eurographics 2001, 20(3):174-183, 2001.

Alexander Hornung, Gerhard Lakemeyer, Georg Trogemann: *An Autonomous Real-Time Camera Agent for Interactive Narratives and Games*. IVA 2003: 236-243

Related Work: Discourse Generation

- Theory of Discourse Structure[Grosz et al. 86]
- Rhetorical Structure Theory[Mann et al. 87]
- Planning approaches [Moore and Paris 89, Maybury 92, Hovy 93]
- Speech Acts for dialogue agents[Traum 99]

Grosz Barbara, Sidner Candace, *Attention, Intention and Structure of Discourse,* Proceedings of ACL 1986
W. C. Mann and S. A. Thompson. *Rhetorical Structure Theory: A Theory of Text Organization.* TR- ISI/RS-87- 190, USC ISI, Marina Del Rey, CA., June 1987
Moore, J.D. & Paris, C.L., *Planning text for advisory dialogues.* In Proceedings of the 27 th Annual Meeting of the ACL, pg. 203--211, Vancouver, B.C., Canada, 1989.
Maybury M, *Communicative acts for explanation generation,* IJMMS (1992) 37, 135-172
Hovy, E. *Automated Discourse Generation Using Discourse Structure Relations.* Artificial Intelligence 63, pp. 341-385, 1993
David R. Traum, *Speech Acts for Dialogue Agents*, in Michael Wooldridge and Anand Rao, editors, "Foundations And

References

- **Representational Requirements for a Plan Based Approach to Camera Control in Virtual Environments,** A Jhala, R M Young, to appear in Proceedings of *2nd Artificial Intelligence in Interactive Ditigal Entertainment Conference (AIIDE 2006),* Marina Del Rey, CA, 2006.

- **A Discourse Planning Approach to Cinematic Camera Control for Narratives in Virtual Environments,** A Jhala, R M Young, Proceedings of the *20th National Conference on Artificial Intelligence (AAAI-05),* Pittsburgh, PA, 2005.

- **An Intelligent Camera Planning System for Dynamic Narratives**, A Jhala, *Masters thesis*, Department of Computer Science, North Carolina State University, 2004.