Outline

- Organisation
- Last lecture
- AUTEUR – automatic generation of slapstick video sequences
- Additional applications
Video – summary

Investigated
- Temporal visual medium
- Effect of time on interaction and creativity
- AI techniques for the representation of time, action, plan

Example
Video, Frames, Scripts, Conceptional Dependencies, Plans, Goals, Story Grammar, Semantic Links

Findings
- Video, though based on common human content and thematic structures, provides its own realities of time and space which are interwoven in the narrative structure.
- A story is a representational system based on two main layers, structure and content, each serving two distinct purposes (form and substance).
- A primarily structure-oriented approach (grammar) to the temporal aspects of video with respect to generation and interaction is not appropriate.
- A planning approach (planner or agent) seems more workable for automatic video generation.
AUTEUR - Aim

- Automatic composition of visual slapstick sequences.
- Use an existing, arbitrary database.
- Investigate machine creativity.
AUTEUR – Generation Example I

Motivation

Narrative rules (joke structure)
Access via content representation
Editing rules (continuity)

Narrative rules (continuity)
Access via content representation
Editing rules (continuity)
AUTEUR – Generation Example II

Realisation

Narrative rules (joke structure)
Access via content representation
Editing rules (continuity)
AUTEUR – Generation Example III

Resolution

Narrative rules (joke structure)
Access via content representation
Editing rules (continuity)
AUTEUR – Generation Example IV

Result
AUTEUR - Architecture

Frank Nack
Kennisgebaseerde media
AUTEUR - Architecture

Retrieval System

Editor
  Structure Planner
  Content Planner
  Visual Designer
  Visual Constructor

Development Tools

Conceptual link

Interface

Knowledge Base
  Semantic net
  Filmic and editing knowledge

DB Video material

DB Video representation

Information flow
AUTEUR - Architecture

- 145 Conceptual Dependencies
- 26 humour rules
- 37 editing rules
- 85 clips (5 sec – 9 min)
AUTEUR – Structure Planner
The conflict between the expected and what actually occurs.

The implicit upgrading of the mood of the viewer.
AUTEUR – Plot Model II

Plot

Episode
- Mot
- Event
- Action
- Subaction

Episode
- Real
- Event
- Action
- Subaction

Episode
- Reso
- Event
- Action
- Subaction
H-Strategy 1
An action forms the most suitable subject for a joke, then an actor, then an object, and finally a location.

H-Strategy 2
If the action portrays an intention [goal], interrupt the action in a way that is unexpected by the character, so that the goal cannot be fulfilled and the character’s mood is downgraded or he or she suffers in some way. (Mischief + Schadenfreude)

H-Strategy 16
A sequence of actions that is meaningful is more preferable for the construction of jokes than a sequence of unrelated actions.

H-Strategy 24
A relationship between two oppositional characters should be established in such a way that the goal of one character is to interrupt the goal of the other in such a way, that is unexpected by the second character. The reaction of the second character must then be influenced by the first so that the second character's mood is downgraded or he suffers in some way. (Mischief + Schadenfreude).
AUTEUR – Content Planner

![Diagram of the AUTEUR framework with nodes and connections representing retrieval system, video material, video representation, knowledge base, semantic net, filmic and editing knowledge, development tools, conceptual link, and information flow.]
AUTEUR – Content Planner

Motivation  =>  Realisation  =>  Resolution

<table>
<thead>
<tr>
<th>Motivation</th>
<th>Realisation</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>[approach]</td>
<td>[[search_money, insert_money+], [wait]]</td>
<td>[look_change, take_cup+, leave]</td>
</tr>
<tr>
<td>[]</td>
<td>[[process],[provide_cup+]]</td>
<td>[]</td>
</tr>
</tbody>
</table>

available video material

scene relevant material
### Frame

<table>
<thead>
<tr>
<th>Name</th>
<th>slip</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain</strong></td>
<td>motion</td>
</tr>
<tr>
<td><strong>Nature of location</strong></td>
<td>outdoors</td>
</tr>
<tr>
<td><strong>Set of objects</strong></td>
<td>[banana_peel, dog_shit, soap, ice]</td>
</tr>
<tr>
<td><strong>Body part / related object</strong></td>
<td>[shoe]</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>[road]</td>
</tr>
<tr>
<td><strong>Relation Location -&gt; Object</strong></td>
<td>under</td>
</tr>
<tr>
<td><strong>Relation Object -&gt; Body part</strong></td>
<td>under</td>
</tr>
<tr>
<td><strong>Intention</strong></td>
<td>[unintentional]</td>
</tr>
<tr>
<td><strong>Result actions</strong></td>
<td>[sit, lie, kneel, shake, look_back]</td>
</tr>
<tr>
<td><strong>Result mood</strong></td>
<td>[anger, rage, astonishment]</td>
</tr>
</tbody>
</table>
AUTEUR – Semantic Description – Event and Concept

Script (to meet)

<table>
<thead>
<tr>
<th>Name</th>
<th>Actor number</th>
<th>Gender</th>
<th>Intention</th>
<th>Motivation</th>
<th>Realisation</th>
<th>Resolution</th>
<th>Episode</th>
</tr>
</thead>
<tbody>
<tr>
<td>meeting</td>
<td>2</td>
<td>any</td>
<td>meet</td>
<td>[walk]</td>
<td>[look at]</td>
<td>[shake_hand]</td>
<td>date</td>
</tr>
</tbody>
</table>

Conceptual dependency

<table>
<thead>
<tr>
<th>Emotional class name</th>
<th>Body part</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>pleasure</td>
<td>Head</td>
<td>[lip, up]</td>
</tr>
<tr>
<td>--</td>
<td>--</td>
<td>whistle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abstract concept name</th>
<th>Representation structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>[[shadows], [passing]]</td>
</tr>
</tbody>
</table>
AUTEUR – Semantic Description – Semantic Graph

- walk
- meet
- look
- move
- listen
- glide
- collide
- Road
- Shoe
- Pleasure
- Motion

Links:
- Opposition link
- Synonym link
- Subaction link
- Association link
- Conceptual link
- Causal link
- Conceptual structure action
- Other conceptual structures

Frank Nack
Kennisgebaseerde media
AUTEUR – Visual Designer
### Description Scheme (Character)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shot ID</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Startframe</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Endframe</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Identifier</strong></td>
<td>Identifier for a character, e.g. a name or a number</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>male, female, hermaphrodite, artificial</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>e.g. young, old, 25, etc.</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td>e.g. black, white, Asian, etc.</td>
</tr>
<tr>
<td><strong>Appearance</strong></td>
<td>a structure including:</td>
</tr>
<tr>
<td></td>
<td><strong>role</strong> e.g. lawyer, plumber, stewardess, etc.</td>
</tr>
<tr>
<td></td>
<td><strong>Costume</strong></td>
</tr>
<tr>
<td></td>
<td><strong>.kind</strong> e.g. business suit, apron dress, overall, etc.</td>
</tr>
<tr>
<td></td>
<td><strong>.colour</strong> a doublet list providing the major colour for the top and bottom part</td>
</tr>
<tr>
<td></td>
<td>e.g. [black, white]</td>
</tr>
<tr>
<td></td>
<td><strong>appeal</strong> e.g. casual, formal, etc.</td>
</tr>
</tbody>
</table>
## Description Scheme (Denotative Video Aspects)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shot ID</td>
<td></td>
</tr>
<tr>
<td>Startframe</td>
<td></td>
</tr>
<tr>
<td>Endframe</td>
<td></td>
</tr>
<tr>
<td>Identifier</td>
<td>Identifier for a character, e.g. name or number</td>
</tr>
<tr>
<td>Relative Position</td>
<td>(Screen position first frame, screen position last frame), e.g. (left, right), (left, middle), (right, right), etc.</td>
</tr>
<tr>
<td>Action</td>
<td>e.g. eat, drink, walk, read, etc.</td>
</tr>
<tr>
<td>Speed of action</td>
<td>e.g. slow, medium, fast</td>
</tr>
<tr>
<td>Direction of action</td>
<td>left, up-left, up, up-right, right, down-right, down, down-left, front, back, circular</td>
</tr>
<tr>
<td>Body gesture</td>
<td>a structure containing:</td>
</tr>
<tr>
<td></td>
<td><strong>full body</strong></td>
</tr>
<tr>
<td></td>
<td>horizontal, vertical, left-diagonal, right-diagonal</td>
</tr>
<tr>
<td>Head</td>
<td></td>
</tr>
<tr>
<td>profile</td>
<td>right, left, half-left, half-right</td>
</tr>
<tr>
<td>movement</td>
<td>up-down, left-right, up, down, left right, circle</td>
</tr>
<tr>
<td>eyebrows</td>
<td>up, down, straight, etc.</td>
</tr>
<tr>
<td>line of sight</td>
<td>left, right, straight, up, down, etc.</td>
</tr>
<tr>
<td>mouth</td>
<td>up, down, straight, open</td>
</tr>
<tr>
<td>Hand</td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>action/related object e.g. (tap/table)</td>
</tr>
<tr>
<td>right</td>
<td>action/related object e.g. (holding/head)</td>
</tr>
<tr>
<td>Foot</td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>e.g. tap, lift, etc.</td>
</tr>
<tr>
<td>right</td>
<td>e.g. tap, lift, etc.</td>
</tr>
</tbody>
</table>
### Description Scheme (Media Dependent Aspects)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shot ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>Shot length</td>
<td>in frames (25 frames for a second)</td>
</tr>
<tr>
<td>Startframe</td>
<td></td>
</tr>
<tr>
<td>Endframe</td>
<td></td>
</tr>
<tr>
<td>Shot kind</td>
<td>a structure including:</td>
</tr>
<tr>
<td></td>
<td><strong>lens movement</strong></td>
</tr>
<tr>
<td></td>
<td>zoom-in [start camera dist., end camera dist.]</td>
</tr>
<tr>
<td></td>
<td>zoom-out [start camera dist., end camera dist.]</td>
</tr>
<tr>
<td></td>
<td>masking left, middle, right</td>
</tr>
<tr>
<td></td>
<td><strong>lens state</strong> (deep focus, foreground-focus,</td>
</tr>
<tr>
<td></td>
<td>background-focus)</td>
</tr>
<tr>
<td></td>
<td><strong>camera distance</strong> (extreme close-up, close-up,</td>
</tr>
<tr>
<td></td>
<td>medium, long, long, extreme long)</td>
</tr>
<tr>
<td></td>
<td><strong>camera movement</strong> (pan_left, pan_right, tilt_</td>
</tr>
<tr>
<td></td>
<td>up, tilt_down, roll_left, roll_right)</td>
</tr>
<tr>
<td></td>
<td><strong>camera position</strong> (left, middle, right)</td>
</tr>
<tr>
<td></td>
<td><strong>camera angle</strong> (overhead, high-angle, eye-level,</td>
</tr>
<tr>
<td></td>
<td>low-angle)</td>
</tr>
<tr>
<td></td>
<td><strong>film speed</strong> (slow motion, normal, fast motion)</td>
</tr>
<tr>
<td>Shot colour</td>
<td><strong>colour</strong> list of the dominant colours</td>
</tr>
<tr>
<td></td>
<td><strong>black &amp; white</strong></td>
</tr>
<tr>
<td>Shot granularity</td>
<td>fine, medium, strong</td>
</tr>
<tr>
<td>Shot contrast</td>
<td>high, medium, low</td>
</tr>
</tbody>
</table>
AUTEUR – Visual Designer

E-Strategy 1
If sequence.kind = Motivation then
  Camera distance of Shot to be chosen is long
  => location.function = outdoor
  medium long or medium
  => location.function = indoor
AUTEUR – Video spatial relations between frames

<table>
<thead>
<tr>
<th>Shot A</th>
<th>Shot B</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) extreme close-up</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) close-up</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>(3) medium close-up</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) medium</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>(5) medium long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) extreme long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: X indicates spatial relation; O indicates no spatial relation.

Frank Nack
Kennisgebaseerde media
AUTEUR – Visual Designer

Value of camera device | Level of content detailness
---|---
extreme close-up | object:subparts[form, colour]
character: subparts
one detail of either head, hand, feet

close-up | object:instance shape
character: Instance
or parts of either Head+Id, Hand or Foot

medium close-up | object:instance shape
character: Instance Appearance, Head

medium | object:instance shape
character: Instance Appearance

long | object:instance shape
setting: Time, Location, Lighting

extreme long | object: instance shape
setting: Time, Location, Lighting
number and size of characters to be portrayed and their spatial relationships involved.

break up into sub-groups of the kind described to the left and then use their configuration.

Legend:
Symbols:  □ shot,  | cut,  ◦ subject

Type of spatial relations:
- line,
- △ triangle or half circle,
- □ square or circle,
- × over cross

Spatial relationships in the shot and created via a cut:

---

Based on the decomposition (1):
- hierarchical knowledge representation of subjects, the decomposition starts on parts level:
  - Class
  - Subclass
  - Instance
  - Parts
  - Subparts

Decomposition (2 - n):
- the content of each shot should present the same hierarchical level within the knowledge representation, e.g: Instance - Instance, or Parts - Parts.
E-Strategy 6
If a sequence is to be established
where location of shot A ≠ location of shot B
or the sequence is the first sequence to be established
then create a memory structure of the spatial relations between all characters of Shot B

Location-Memory-Structure
Start Shot-id
End Shot-id
List of structures
List_of_content_relations
List_of_used_shots

E-Strategy 9
If number of character = 3 and
camera distance of both shots ≥ medium long
then
shot A ([left | right]) with shot B ([middle]) => circle / triangle
shot A ([left | middle]) with shot B ([right]) => circle / triangle
shot A ([middle | right]) with shot B ([left]) => circle / triangle
shot A ([left]) with shot B ([middle]) with shot C ([left]) => line or its permutations.
E-Strategy 12
If an action for a character is required and there is no shot available to portray that action then isolate the character in a shot retrieve the body part related to the action retrieve a suitable shot where a body part performs the required action build a bridge into or out of this sequence if necessary continue with the sequence which was interrupted by this subsequence.
AUTEUR – Video shaping

Frames

Stratamodel

TIME (H:Min:Sec:Frames)

walk

talk

eat

sit

Shot

Annotations
E-Strategy 29
If sequence.action.tempform = contraction and action is a single action then
favour decomposed forms of presentation where the camera distance of shot A ≥ camera distance shot B

E-Strategy 31
If camera distance of a shot ≤ close-up and then
clip it to a length ≤ 60 Frames.

E-Strategy 32
If close-up < camera distance of a shot < long and sequence. kind = motivation then
clip it to a length ≤ 108 Frames.
AUTEUR – Summary

- Planner approach (backward chaining – goal driven)
- Addresses continuity problem BUT incompletely
- Closed system
- Requires lots of knowledge
- Editing strategies are explicit BUT not final
- Story generation is flexible BUT presentation should be generated rather than using existing material
Video – Additional Applications
Video – Additional Application

History is in your hands! Through an audience response-measuring device (Polling meter) connected to a computer, viewing audiences respond to periodic questions regarding their satisfaction with marketing polls. These questions occur every 6 minutes during the story. The loudest applause determines the winning answer.

Your answers to these questions allow the computer program to create historical narratives that mirror and even exaggerate your biases and desires. Just drop, watch and enjoy. At long last, Terminal Time gives you the history you deserve!

The Terminal Time engine uses the past 1,000 years of world history as "fuel" for creating these custom-made historical documentaries. Each program generated by the machine can be either projected on a screen or broadcast on television monitors. (Although the video and sound are constructed in the computer, the signal is compatible with standard video technology.) Each program lasts approximately 30 minutes.

Audiences experience the project in a similar manner to a movie. A short breakdown of how the audience experiences the work follows:

1. Audience members find their seat and the "movie" begins with a short title sequence and introduction.
2. After introduction, the story begins with a 2-min introductory video.
3. Then, there is a 5-min history lesson.
4. After the lesson, there is a 5-min discussion with the audience.
5. Finally, there is a 5-min exit.

(The answers to these questions changes the story that will follow.)
Introduction

This page presents information about VOX POPULI, a system for automatic generation of biased Video Sequences. In few words, VOX POPULI uses rhetorical annotations to generate video sequences. Our annotation schema encodes the verbal information contained in the audio channel, identifying the claims the interviewees make and the argumentation structures they use to make those claims. Based on this schema, we construct a semantic graph which is traversed by rhetoric-based strategies selecting video segments.

On-line SMIL demo

The demo presents a Web interface to generate a short video documentary, as described in the paper. Please report problems to us.

As an example, a prerendered video sequence is here. The Toulmin analysis of the first part of this video sequence (as described in the above-mentioned paper) is the following:

Other projects using Vox Populi

VJ Cultuur is the result of a one-year long research about the culture of VJs in the Netherlands and consists of video interviews with 12 Dutch VJs. From the start of this project it was decided to present this video material in an interactive way using Vox Populi.

VOX POPULI in the media

Radio Interviews

Interview (in Dutch) for the Dutch radio program Hoordenlicht over scientific developments, broadcast on national radio. A local copy of the interview is here.

Interview on the Amsterdam radio Onda Italiani in Dutch, unfortunately it was not recorded.

Press Interviews

Interview with Elke van Cassel, Cursor (weekly information magazine of the University of Eindhoven), nr. 13, 2006.

Press Releases

There has been a press release about Vox Populi which has appeared on the following sites:

- Fact Sheet (English) from CWI, the research center I worked for.
- Alphagalileo (English), independent resource for European research news.
Video – Additional Application III

Agent Stories  
(1994-1999)

**Kevin Brooks**  
Writer / Designer

**Glorianna Davenport**  
Research Supervisor

It is possible to create non-linear stories which allow many different variations of the tale to be interactively played out. Agent Stories is a thinking tool useful for building and ultimately displaying these computationally-based audio/visual narratives.

The writer builds a web of small story “chunks” which are tied together via predefined relationship links. Once this webbed story world or “storybase” is created, the user is able to request a story as constructed by any one of a number of “story agents.”

Story agents are software agents with behavioral traits that describe the type of stories they are prone to construct. Although this research is not tied to a particular delivery system, this technique of story definition and construction suggests a model for interactive TV.

The limitation of the program is that story agents can only be as distinct and “interesting” as the complexity of the storybase allows.

---

**Crossing the Street** is an evolving and expanding collection of first-person short stories about the effects of taking journeys, large and small.

---

Davenport G; Bradley B; Agamanolis S; Barry B; Brooks KM (2000).  
**Synergistic storyscapes and constructionist cinematic sharing**  

**Metalinear Cinematic Narrative: Theory, Process, and Tool**  
MIT Ph.D. Thesis.

**Programming Narrative**  
Video – Additional Application VI

Elastic Catalogue
(1996)

Michael Murtaugh
Designer/Programmer

Glorianna Davenport
Research Supervisor

Davenport G; Murtaugh M (1997).
**Automatist storyteller systems and the shifting sands of story**
*IBM Systems Journal*, vol.36, no.3, 1997, pp.446-56. Publisher: IBM, USA.


**ConText: Towards the Evolving Documentary**
*ACM Multimedia '95, November 1995.*

**Movies of the Future: Storytelling with Computers**
*American Cinematographer, April, cover, pp 4-12.*
Summary

- The content determines the application
- Content description
  - Application dependent
  - Complex
  - Recourse demanding
  - Time critical
  - Incomplete
- Modular Schemata
- Description environment
  - Production supportive
  - Archive supportive