Template-based Adaptive Video Documentaries

Cesare Rocchi, Massimo Zancanaro
ITC-irst
Italy
{rocchi,zancana}@itc.it

Abstract
In this paper, we introduce an approach to the adaptive composition of video documentaries. The adaptation is based on templates that encode rules for the dynamic selection, sequencing and composition of video shots. We introduce a formalism to represent 2D video documentaries together with a template language to define adaptation rules. Finally, we discuss rhetorical patterns, strategies that we abstracted out during the realization of a museum mobile guide.

Keywords
Adaptivity, Cinematic Techniques, Multimedia

INTRODUCTION
Research in Adaptive Hypermedia (AH) has primarily focused on outputs which are based on static media, mainly the combined use of images and texts [5]. Yet recent works have tried to establish computational frameworks to include also time based resources, both audio and video (see respectively [14] and [11]). The introduction of temporal media affects the key stages of the composition processing, namely content selection, content organization and presentation rendering.

Duration over time represents an additional dimension that the selection process has to take into account. Unlike classical hypertexts, the organization and rendering of temporal resources requires synchronization, e.g. the mention of an object in an audio commentary has to match the moment in which it is displayed on the screen. On a more 'aesthetic' viewpoint, multimedia presentations and video in particular should also adhere to some principles of cinematography, which encompass 'good' selection of camera movements and 'good' intertwine of transitions. For example, a panning camera highlights the spatiality of a picture, whereas a long cross fade better conveys a change of topic, etc. (see [3]).

The composition of video documentaries is usually performed by selecting a set of prerecorded clips from an underlying database and 'fuse' them together in a sort of movie ([11]). In this paper, we present a different approach. The building blocks of our presentations are not prerecorded and annotated movie clips, but shots that can be defined by means of a scripting language. A shot is made of camera movements, ap-plied to 2D images, and a sequence of audio files. By means of this finer granularity, presentation designers can focus on both the structure of the final presentation and the content of single clips at the same time. We chose to focus on videos describing 2D images because they represent a simple, yet realistic, testbed to study content adapta-tion in video documentaries. This kind of videos is usually employed in art documen-taries and, recently, they have been used by National Geographic to produce light (non-adaptive) web documentaries .

In the field of Adaptive Hypermedia ([5], [7]) there have been attempts to integrate time based media into systems that dynamically generate or compose multimedia presentations tailored to particular users (for an overview see [1]). In [2] André presents an interesting comparison of two systems, WIP and PPP. The Cuypers system ([15]) generates web-based multimedia presentations from an underlying database using constraint satisfaction techniques. Not and Zancanaro ([14]) have implemented an adaptive hypermedia system, for a mobile museum guide, which generates presen-tations including audio files. Lindley and colleagues ([11]) adopt RST ([12]) as a basis for a system that generates news from a digital archive of movie clips. RST served as a starting point also for the model of communicative devices presented in [17].

In this paper, we introduce a framework for the dynamic composition of this kind of documentaries to produce user-tailored presentations. In section 5, we introduce the notion of 'Rhetorical Patterns', schemata of solutions for recurrent problems in tem-plate authoring, drawing on the analogy with 'Design Patterns' in the programming field.

TERMINOLOGY
The shot is the building block of a video clip. In the field of cinematography a shot is defined as a continuous view from single camera without interruption. In this work, since we deal with still images, we refer to a shot as a sequence of camera movements applied to the same image. The camera can move along three dimensions x, y, and z axis, (respectively pan, tilt and zoom). The main transition effects are: display - the first frame of the shot to be displayed immediately replaces the last frame of the shot currently on display; cut - a white space is inserted between two shots; fade - a shot is gradually replaced by (fade out) or gradually replaces (fade in) a black screen or an-other shot; cross fade (or dissolve) which is the composition of a fade out on the dis-played shot and a fade in applied to the shot to be shown.

Starting from these notions we defined XSCRIPT ([18]), an xml-based markup lan-guage that represents a description of a Video Documentary. A script for a Video Documen-

tary is made of two main parts, a list of media items which are part of the video and the specification of the order in which the shots have to be played and the transitions between them. Each shot encompasses an audio track that contains a list of audio files - the audio commentary - and a video track that specifies both the image to display and the camera movements that will be applied on it.

XCRIPT is specifically targeted at representing videos that describes 2D images. It can be easily transformed into SMIL for rendering. In the context of PEACH, we are using an XSCRIPT player implemented in Macromedia Flash to simplify the integra-tion in the present user interface of the mobile guide.

XASCRIPT - ADAPTIVE TEMPLATE LANGUAGE

XASCRIPT extends XSCRIPT and allows defining adaptation rules and constraints. While XSCRIPT describes the structure of an actual documentary, XASCRIPT is a language for the definition of templates, intensional descriptions of a set of potential documentaries, with multiple-choice points on user-dependent parameters. Once a video documentary is requested, the adaptation engine elaborates the templates ac-cording to the current user model and returns a script for a video documentary in XSCRIPT form. Adaptation rules are the tool that enables authors to state constraints and strategies to select shots and apply transition effects. An adaptation rule is a <condition, action> pair, where the condition tests the requirements and the action composes the pieces of the documentary.

```
  <rule>
    <UM-expression>
      $user.hasSeen('shot01')
    </UM-expression>
    <fade-in shot="shot02" duration="2"/>
    <display shot="shot03"/>
</rule>
```

Figure 1. An example of rule in XAScript.

For example, referring to Figure 1, if the condition holds (the user has seen shot01), then shot02 fades in, otherwise shot03 is displayed. Rules can also be embedded to match more complex situations.

Our framework allows defining many dimensions along which adaptivity can be re-alized. While designing templates, the author can always refer to two key variables: the user for which the documentary is going to be presented, and the current docu-mentary under processing. In this way, the author explicitly inserts conditions related to user's features (e.g. preferences, previous seen presentations, device dimensions) and also conditions related to the current composition process (e.g. previous selected effect, list of already selected shots).

By referring to the current composition process, templates' authors can define:

- dependencies among content units: e.g. the selection of a unit requires or pre-supposes another shot to be chosen;
- dependencies among presentation forms: e.g. the selection of a fade requires the previous effect not to be a cut.

THE ADAPTIVE VIDEO ENGINE

The system is implemented in Java. The adaptive video engine has been implemented on top of the Jakarta Velocity. First, the selected XASCRIPT template is translated in VTL (the template language of Velocity) then the User Model is queried to update the Velocity Context and finally the Velocity Engine is employed to instantiate the template. The result is an XSCRIPT document.

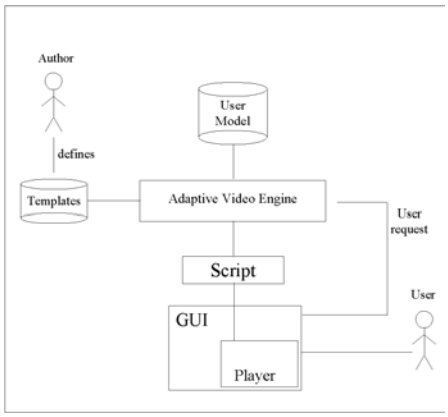

Figure 2. Architecture of the system.

The communication among components is implemented trough TCP/IP sockets, over which XML based messages are exchanged. The XSCRIPT player is implemented in Macromedia Flash MX.

At the moment the engine is used in a bigger architecture, in the context of the PEACH project. It is meant to compose presentations on the fly, whenever the visitor asks for a video documentary. Presentations can be played both on PDA, if the user is actually visiting the museum, and on a "virtual window"[16].

The goal of the overall system is to engage the user while is visiting a museum and help her in learning new information and relate it to her previous knowledge. Currently the engine supports the following user's features to perform adaptation: Dis-course history, a model of what the visitor has already seen and heard during the current session and previous ones. This information can be useful to draw compari-sons among exhibit or prompt suggestions; Spatial Position. If the user is visiting the museum with a PDA, a system tracks her current position and the history of previous movements. This kind of data can be exploited to select audio that contains explicit spatial or temporal references like "this is the January fresco", "as

you just saw", and so on; Background Knowledge. Information about user's previous knowledge about the exhibit she is visiting. Through this information the author can define rules for selecting content tailored to the knowledge level of a particular user; Device. The device that she has at the moment, to browse the material related to the museum (e.g. PDA, desktop PC). For example, if presentations running of PDA are streamed over a wireless connection, the author might prefer to select images and audio with less defi-nition, to reduce user's wait; Interests. A model of user's interests in the topics of the domain. This is particularly useful to generate suggestions and lead the user to visit exhibits related to her interests.

RHETORICAL PATTERNS FOR TEMPLATE AUTHORING

Besides the 'classic' issues in AH, there is an increasing interest on the authoring of AH systems, as pointed out in [6]. Hypermedia authors usually exploit graphical interfaces to prepare the content and define the adaptation mechanism of the system, but these tools are not enough, since they do not help focusing the adaptation strategies, which are the core of every adaptive system. Authors, in fact, can be distracted by low-level issues, like how to not miswrite a rule or where to locate a source file. An interesting approach is relying on the notion of pattern.

Although the implementation of templates for video documentaries can not be com-pared with a programming task, they have some commonalities. Presentation designers have to take into account many features like coherence of the selected content, cohesion of the presentation form, choice of transition effects, preferences of the user, etc. In the programming field, Design Patterns encode 'scheme of solutions' to recur-rent problems. During the writing of templates for adaptive presentations, designers have to face recurrent problems. For example, they often have to address the following issues: Deepening: if the user has already been exposed to a topic t, how to select and present material related to t (e.g. highlight its features)? Comparison: Comparisons maximize the extent to which a visitor's understanding of an exhibit coheres with her other knowledge, and help to prevent the hearer from forming misconceptions ([13]). How to refer to previously mentioned material m, related to the current topic? Sugges-tion: suggestions are complementary to comparisons. If the user has not visited an exhibit e and the designer thinks that e - considering the current context - is worth a visit, how to lead the user to visit e? Hide: if the user has already heard about a topic t, or she is not 'ready' to watch a presentation about t, how to skip it and keep the discourse coherent? Sequence: if the user can see a then she can also see b and c; how to suggest such a continuity? Exemplification: if the topic t is generic, say 'a painting techniques', how to provide visual and aural explanations, so that the user can more effectively understand *t*?

In analogy with the programming field, we will call Rhetorical Patterns (RPs) the scheme of solutions to these problems. Each of the previous questions, in fact, relates to the structure of the discourse and how such a structure affects the message underly-ing the documentary. In our context RPs encode adaptation strategies, namely ways to find a solution given a context, a particular combination of user dependent features.

Table 1. Classification schema for Rhetorical Patterns

Name:	A name identifying the pattern.
Problem:	Concise description of the problem to be solved.
Conditions:	The list of conditions which have to be true, in order to apply the pattern.
Motivation(s):	Explanation(s) of why and how this pattern solves the problem stated above.
Solution:	Description of the solution. Can contain a specific example to better explain the idea.
Related Patterns:	A list of patterns that is worth considering if the current one applies.

We propose to classify RPs in a way similar to the description of Design Patterns. According to the schema in table 1, each pattern has a name, a statement of the problem it is meant to solve, a list of applicability conditions, a list of motivations stating how and why the problem is solved by applying such a pattern, the solution, or a sketch of solution if the problem is generic and, finally, possible related patterns. In the examples presented below we show how a pattern can be mapped onto a schema of template in a straightforward way. Our approach is inspired by the Rhetorical Structure Theory (RST) presented by Mann and Thompson [12]. While the nature of RST is descriptive, in that rhetorical relations are meant to functionally relate pieces of text one another, our approach is rather focused on the proposal of stereotypical solutions to recurrent problems in defining templates for video documentaries. The nature of rhetorical patterns proposed here is thus prescriptive, for they are a sort of recipes to prepare 'good' documentaries. We then foresee the possibility to couple recurrent problems to particular configurations of discourse structures that provide a solution to those problems. A Suggestion pattern, for instance, can be described as in table 2.

The purpose of a suggestion is to persuade the visitor to see an exhibit. This is a suitable pattern if the currently presented exhibit or topic is related to another. This way the user is helped in contextualizing and relating information by means of simi-larities or even differences between two artworks. The solution is to insert a piece of video at the end of the presentation, which contains an explicit invitation to visit an-other exhibit and, if needed, the path to reach it.

Table 2. Classification schema for Rhetorical Patterns

Name:	Suggestion
Problem:	Lead the user to visit an exhibit.
Conditions:	1. The current exhibit *e* is related to another one *o* (e.g. they share the same style, the depict the same scene) 2. The user has not visited *o*.
Motivation(s):	The user is explicitly told (via audio or video) to go visiting *o*. This helps her to understand similarities and/or differences between the two exhibits and contextualize her information absorption.
Solution:	At the end of the video presenting *e*, insert a shot which represents *e* and an audio file which points the user to *o*. If needed insert also spatial references to help the visitor finding *o* location.
Related Patterns:	Comparison. If condition 2 does not hold maybe a comparison applies.

Comparisons provide a framework in which the visitor can contextualize her knowledge. Besides preventing the visitor from misconceptions, comparisons allow also to relate items each others, thus enabling descriptions which are more effective, since they refer to things the visitors are familiar with. This pattern applies when there are two or more exhibits that share similarities. The solution is to insert media items that make explicit reference to concepts well known to the visitor, or pieces of art she has already visited.

Table 3. The Comparison Pattern

Name:	Comparison
Problem:	Show similarities between two exhibits.
Conditions:	1. The current exhibit *e* is related to another one *o*. 2. The user has already visited *o*.
Motivation(s):	The comparison among two items can help the user to find the relation among them (e.g. slightly different painting style, same subject depicted by two different authors)
Solution:	Insert one or more shots which explicitly refer to the features of *o* and show their similarities. Screen can be split into two part to help the user recalling *o*.
Related Patterns:	Contrast. Besides similarities also differences can be presented.

Since they are related, Comparison and Suggestion patterns can be easily mapped onto the schema of template in Figure 3.

```
...
<rule>
  <UM-
expression>$user.hasVisited(<exhibit-id>)
  </UM-expression>
  <!– here shots(s) presenting comparisons
-->
  <!– here a shot containing the sugges-
tion -->
</rule>
```

Figure 2. Architecture of the system.

We have also identified patterns for providing follow-ups, hiding media items, creating simple sequences (e.g. to enumerate similar items) and providing exemplifications. At the moment we are implementing a graphical user interface which allows authors to exploit our pattern approach during the writing of templates, by proposing schemata of templates like the one showed above.

CONCLUSION

In this paper, we presented a framework for the dynamic composition of video documentaries, where media items are dynamically, selected, sequenced and rendered according to a set of contextual parameters which include user preferences, topic, perspective, type of device, etc. The core of this framework is XASCRIPT, a flexible mark up language that we developed to define templates in a declarative way. When a presentation is requested the adaptation engine selects the appropriate template and composes a presentation by applying rules and constraints defined in the templates. The resulting output can be played on both mobile devices and desktop machines.

We then presented an approach to facilitate the implementation of templates by discussing the notion of 'Rhetorical Pattern', schemata of solutions for recurrent problems that arise during the preparation of the templates. In our view these patterns help authors focusing on the interaction with the information space they organize and user's needs; and also to cope with the usually big amount of content to be organized during the authoring of a hypermedia system.

ACKNOWLEDGMENTS

This work has been supported by the PEACH and TICCA projects, funded by the autonomous Province of Trento.

REFERENCES

[1] André E.: The Generation of Multimedia Documents. In: R. Dale, H. Moisl and H. Somers: A Handbook of Natural Language Processing: Techniques and Applications for the Process-ing of Language as Text, Marcel Dekker (2000) 305-327

[2] André E.: WIP and PPP: A Comparison of two Multimedia Presentation Systems in Terms of the Standard Reference Model. In: Computer Standards and Interfaces, Volume 18, No: 6-7, pp. 555-564, 1997

[3] Arijon D.: Grammar of the Film Language. Silman-James Press, Los Angeles, CA (1976)

[4] Borchers J.: A Pattern Approach to Interaction Design, John Wiley & Sons (2001)

[5] Brusilovsky P.: Adaptive Hypermedia. User Modeling and User-Adapted Interaction 11 (2001) 87-110

[6] Cristea A., deMooij A., Designer Adaptation in Adaptive Hypermedia Authoring. In: Pro-ceedings of International Conference on Information Technology, Las Vegas (2003) 444-448

[7] De Bra P., Aerts A., Smits D., Stash N.: AHA! Version 2.0, More Adaptation Flexibility for Authors. In: Proceedings of the AACE ELearn conference, Montreal, (2002) 240-246

[8] Garzotto F., Retalis S., Cantoni I., Papasalouros A.: Patterns for designing adap-tive/adaptable e-learning experiences. In: Proceedings of the First International Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia at the IASTED Interna-tional Conference on Web-based Education. Innsbruck (2004)

[9] Germán D. M., Cowan D. D.: Towards a Unified Catalog of Hypermedia Design Patterns. In: Proceedings of the 33rd Hawaii International Conference on System Sciences, Hawaii (2000)

[10] Graham I.: A Pattern Language for Web Usability. Addison-Wesley (2003)

[11] Lindley C., Davis J., Nack F., Rutledge L.: The application of rhetorical structure theory to interactive news program generation from digital archives. Technical Report INS-R0101, CWI Centrum voor Wiskunde en Informatica, (2001)

[12] Mann W. C., Thompson S.: Rhetorical Structure Theory: a Theory of Text Organization. In: The Structure of Discourse. Ablex Publishing Corporation (1987)

[13] Milosavljevic M.: The Automatic Generation of Comparisons in Descriptions of Entities. PhD Thesis. Department of Computing, Macquarie University (1999)

[14] Not, E., Zancanaro, M.: The MacroNode Approach: Mediating Between Adaptive and Dynamic Hypermedia. In: Proceedings of the International Conference on Adaptive Hyper-media and Adaptive Web-Based Systems. Lecture Notes in Computer Science. Springer-Verlag, UK (2000) 167-178

[15] van Ossenbruggen J., Geurts J., Cornelissen F., Rutledge L., Hardman L.: Towards Second and Third Generation Web-Based Multimedia. In: The Tenth International World Wide Web Conference, Hong Kong, (2001) 479-488

[16] Rocchi C., Stock O., Zancanaro M., Kruppa M., and Krüger A. The Museum Visit: Gener-ating Seamless Personalized Presentations on Multiple Devices. In Proceedings of IUI2004, January 2004.

[17] Rutledge L., Davis J., van Ossenbruggen J., Hardman L.: Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure. In: Proceedings of International Confer-ence on Multimedia Modeling, Nagano, Japan (2000)

[18] Zancanaro M., Rocchi C., Stock O.: Automatic Video Composition. In: Proceedings of 3rd International Symposium on Smart Graphics Heidelberg, Germany (2003) 192-201