Dynamic Context Adaptation in Multimedia Documents

Paola Bertolotti
Università di Torino
Torino, Italy
bertolot@di.unito.it

Ombretta Gaggi
Università Ca' Foscari
Venezia, Italy
ogaggi@dsi.unive.it

Maria Luisa Sapino
Università di Torino
Torino, Italy
mlsapino@di.unito.it

ABSTRACT

Multimedia documents are collections of media objects, synchronized by means of sets of temporal and spatial constraints. Any multimedia document definition is valid as long as the referred media objects are available and the constraints are satisfiable. Document validity depends on the context in which the document has to be presented. In this paper, we introduce a framework to characterize context adaptation, in the presence of both physical and user oriented context requirements. We define *semantically equivalent* presentation fragments as alternative to undeliverable ones. In the absence of equivalence, undeliverable media are replaced with candidates that minimize the loss of information/quality in the presentation.

1. INTRODUCTION

Multimedia documents are collections of media objects, synchronized by means of sets of temporal and spatial constraints. Any multimedia document definition is valid as long as the referred media objects are available and the constraints are satisfiable.

To ensure that a document is presentable several aspects should be taken into account: media and resource requirements have to be compatible with the resource availability (e.g., network bandwidth, CPU time) and with the presentation device type (e.g., desktop, laptop, PDA, or cell-phone). In addition, users' preferences and the environment in which the document is being presented might need to be considered. These aspects that govern how a document is presented to the user are collectively referred to as the *context* in which the documents have to be rendered [6].

Adaptive systems provide different versions for a document, each taking into account different features of the distinct rendering contexts. The *process of adaptation* first detects the cause of the problem with the document, i.e., which set of media cannot be presented and which are the violated constraints. Secondly, it identifies alternative media that can replace the undeliverable ones and that are suitable for the given context. To do so, physical aspects of the media, semantic information associated with the document components, and the temporal and spatial constraints of the presentation are taken into account.

In this paper, we introduce a framework to characterize hypermedia context adaptation, in the presence of both physical and user oriented contexts. We describe the context by means of a *set of properties*, which are stored in a *database*. Context properties are represented as logical assertions. They include (i) users' profiles (e.g., language, preferred media, physical capabilities, etc.); (ii) users' device characteristics (e.g., screen size, supported file formats, etc.); and (iii) the environment description (e.g., the situation in which the user is accessing the document).

For the *first phase* of the adaptation process (i.e., the phase that checks the feasibility of the presentation and, if any problems occur, identifies the critical points in the presentation specification) we rely on a resolver to compare the features of the media occurring in the presentation with the context in which the presentation has to be delivered. For the *second phase* of the adaptation process (i.e., the phase that replaces undeliverable (sets of) media with deliverable ones), we introduce a method to identify *semantically equivalent* media (or presentation fragments) as candidates for the replacement. The notion of *semantic equivalence* will be based on users' explicit statements as well as on inferred relations. In the absence of equivalence, we will choose to replace the undeliverable media with a candidate that minimizes the loss of information/quality in the presentation.

The paper is organized as follows: Section 2 defines the presentation automaton, an intermediate reference model that captures the relevant aspects of any presentation. Section 3 characterizes closed fragments as candidates for replacement during the process of adaptation. Section 4 is devoted to fragment equivalence, to the preference relation defined over fragments and to the adaptation rules. Section 5 presents some related work and concludes the paper.

SAC'06 April 23-27, 2006, Dijon, France

2. A UNIFORM REPRESENTATION FOR DOCUMENT DYNAMICS

To avoid being limited to any specific model for multimedia presentations, we consider an intermediate reference model, that abstracts the relevant aspects of any presentation.

DEFINITION 2.1. *A* multimedia document *is 4-tuple* $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC}, \mathcal{SC} \rangle$*, where* $\mathcal{MI}$ *is the set of media objects to be presented to the users,* $\mathcal{E}$ *is the set of possible events (start/end of media items playback),* $\mathcal{TC}$ *is the set of temporal constraints, and* $\mathcal{SC}$ *is the set of spatial constraints.*

We assume that events are of the form *start(m)*, *end(m)*, or *stop(m)*, which denote the start, the natural termination, and the user induced stop of the presentation of medium m, respectively.

Without loss of generality, we will assume that *temporal constraints* are expressed as bounds on difference[1]. Thus, we will assume that temporal constraints have the form $c_1 \leq inst(e_1) - inst(e_2) \leq c_2$, where $inst(e)$ represents the time instant in which the event e will occur, and c_1 and c_2 are constants defining a lower and an upper bound on the difference between the two mentioned instants. We will conceptually distinguish between the locator of the actual medium file (loc) and the instance of the medium that has to be presented. Consequently, each presentation of the associated medium will have exactly one starting time instant and one ending time instant. With a slight abuse of notation, in the following we will drop "$inst$" from the specification of the temporal constraints, and we will directly express the relationships on events. We refer to [1, 3] for a discussion on different alternative ways to express temporal constraints in multimedia presentations.

In this paper we will mostly consider temporal/dynamic aspects, and will not explicitly deal with the spatial constraints[2]. Therefore, from now on we will refer to multimedia documents as 3-tuples $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC} \rangle$, where the spatial component has been dropped, since it is out of the scope of the discussion.

2.1 The presentation automaton

The general characterization given so far for any multimedia document implicitly defines possible evolutions of the presentation - one distinct evolution for each solution to the constraints. In this section, we introduce an automaton, to characterize the behavior of a given presentation in terms of the paths along the graph of the automaton.

[1] We don't make any explicit assumption on the presence or the absence of disjunction, since we do not explicitly deal with constraint solving: we rely on the availability of the appropriate constraint solver which gives, when needed, a solution to the current set of temporal constraints.

[2] This is not a limitation, since the approach we use for temporal constraints could be suitably adapted to spatial constraints of the form, $c_1 \leq pos(m_1) - pos(m_2) \leq c_2$, where $pos(m)$ indicates a coordinate of the screen area on which the media m has to be delivered, and c_1 and c_2 are constants defining a lower and an upper bound on the difference between such coordinates.

DEFINITION 2.2. *Any presentation of a given document* $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC} \rangle$*, can be represented by means of an automaton* $AUT(D) = \langle S, s_0, TEV, s_f, \delta \rangle$ *where (i)* S *is the set of the states of the presentation. Each state is a pair* $s = \langle \mathcal{AM}, \mathcal{TC}' \rangle$*, where* $\mathcal{AM}$ *denotes the set of media items that are active (i.e., playing, or being displayed), while* $\mathcal{TC}'$ *represents the current set of temporal constraints, i.e., the union of the initial set of temporal constraints* $\mathcal{TC}$ *and the set of equalities* $e = t$*, for every event* e *occurred so far at the corresponding time instant* t*; (ii)* $s_0 = \langle \emptyset, \mathcal{TC} \rangle$ *is the initial state; (iii)* TEV *is the alphabet, i.e., the set of symbols that label possible transitions. It represents the set of possible timed events:* $TEV = 2^{EV} \times R$*, where* $EV = \{start(m) \mid m \in MI\} \cup \{end(m) \mid m \in MI\} \cup \{stop(m) \mid m \in MI\}$ *is the set of possible events, and* R *is the domain for the time instants; (iv)* s_f *denotes the final state, i.e., the state in which the presentation is terminated - no item is being played and nothing remains to be played, according to the temporal constraints; (v)* $\delta : S \times TEV \rightarrow S$*, is the transition function. Given a state* $s = \langle \mathcal{AM}, \mathcal{TC}' \rangle$*, a set of contemporary events* ev*, and a time instant* t*,* $\delta(s, ev, t) = s' = \langle \mathcal{AM}', \mathcal{TC}'' \rangle$ *where* $\mathcal{AM}' = \mathcal{AM} \setminus \{m \in \mathcal{MI} | stop(m) \in ev \vee end(m) \in ev\} \cup \{m \in \mathcal{MI} | start(m) \in ev\}$*, and* $\mathcal{TC}'' = \mathcal{TC}' \cup \{stop(m) = t | m \in \mathcal{AM} \wedge stop(m) \in ev\} \cup \{end(m) = t | m \in \mathcal{AM} \wedge end(m) \in ev\} \cup \{start(m) = t | m \in \mathcal{MI} \wedge start(m) \in ev)\}$.

The states of the automaton are implicitly associated with the temporal dimension: each state is entered at a specific time instant associated with the events that cause the transition from the previous state to it. It is exited when the next set of events (if more than one, then simultaneously) occurs. Obviously, the initial state models the situation in which no media is active and the set of occurred events is empty. Without loss of generality we assume that the initial state is entered at time 0. When the presentation ends the set of media items is empty and the set $\mathcal{TC}^*$ contains the solution chosen for the given set of temporal constraints $\mathcal{TC}$. The set $\mathcal{TC}^*$, in some sense, registers the history of the presentation.

REMARK 2.1. *Assuming infinite time,* the alphabet of the automaton is potentially infinite. *Accordingly, the set of states is also potentially infinite.* We restrict our attention to the cases in which both the alphabet and the set of states are finite. *This is a minor restriction, as infinite presentations are only the ones in which infinite loops exist, and the sequence of time instants associated to events has periodicity. We refer to [2, 7, 12] for a discussion on a compact (finite) representation for these presentations.*

Next, we present an example of presentation automaton. Consider a virtual tourism application, which returns a presentation containing a set of locations, holiday camps, hotels or camping, which adhere to the user's query parameters. Each location is described by an audio file ($audio_i$) in parallel with a sequence of pictures ($pic_{i,j}$). All the locations returned by the query are displayed in sequence.

This synchronization is described by the constraints $0 \leq start(audio_i) - start(pic_{i,1}) \leq 2$, which introduces a tolerance of 2 time units in the simultaneous start of the audio description and of the first picture of each location, $0 \leq stop(pic_{i,j}) - start(pic_{i,j}) \leq 5$ and $0 \leq start(pic_{i,j+1}) - stop(pic_{i,j}) \leq 1$ which display each picture for no longer than 5 time units and introduce a tolerance of 1 time unit between the visualization of a picture and the following one, and $0 \leq stop(audio_i) - stop(pic_{i,n}) \leq 0$ which states the simultaneous stop of the audio file and the last picture illustrating that location. Finally, the temporal constraint $0 \leq start(audio_{i+1}) - stop(audio_i) \leq 2$ allows a tolerance of 2 time units between the rendering of two locations. Figure 1 shows an automaton illustrating the temporal evolution of a possible solution of the set of constraints.

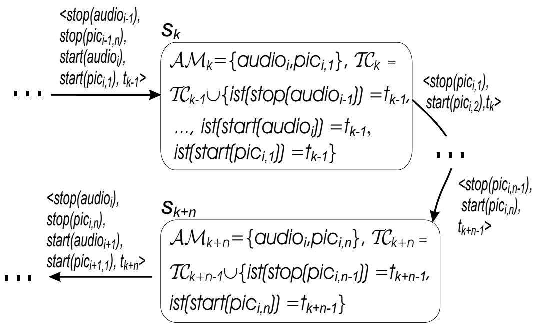

Figure 1: A fragment of the automaton representing the tourism application.

In the next section, we introduce the concept of *significant fragments*, that play the role of "information units" that will be considered as candidates for replacement.

3. STRUCTURAL CHARACTERIZATION THROUGH FRAGMENTS

A fragment of a presentation D is a "sub-presentation" of D, that is, a restriction of D to a given set of objects $\mathcal{MI}'$ which can be replaced without affecting the presentation constraints of the other objects.

In order to isolate fragments that are suitable candidates for replacement, it is important to take into account the tight temporal connections existing among objects in the given document. Specifically, the temporal synchronization constraints can involve several media items and it is not always possible to isolate a portion of the presentation to be replaced, without affecting the remaining presentation items. To take into account mutual temporal relationships, we need to specify the set of fragments that can be a candidate for replacement without affecting the remaining part of the presentation. This leads to the notion of *closed fragment*.

Definition 3.1. *Let $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC} \rangle$ be a multimedia document, and $F = \langle \mathcal{MI}_F, \mathcal{E}_F, \mathcal{TC}_F \rangle \subseteq D$ be a fragment of D. F is a* closed fragment *if: (i) media items in the fragment are completely presented within a delimited time interval, and their removal does not leave unaccomplished media presentations in the rest of the document, and (ii) for every media item $m_F \in \mathcal{MI}_F$ for which a temporal constraint which relates m_F to any media item in $\mathcal{MI} \setminus \mathcal{MI}_F$ exists, at least one of the following conditions holds: (i) $start(m_F)$ is the event occurring in the temporal constraint, and m_F starts at the time instant in which F starts or (ii) $end(m_F)$ is the event occurring in the temporal constraint, and m_F ends at the time instant in which F ends.*

If we consider the tourism application example introduced in Section 2.1, each location is completely described by the audio narration and the sequence of pictures that can therefore be considered a closed fragment of the entire document. The same statement applies to a set of sequential locations. At a deeper level of details, each state represents a closed fragment if we consider a single picture.

An *undeliverable* closed fragment with respect to a given context is any closed fragment that is not presentable in, i.e., compatible with, that context (resources, etc).

4. DYNAMIC ADAPTATION

In this section, we discuss the detection and replacement of undeliverable fragment; we define a notion of *media equivalence* and a partial order on semantically equivalent items and fragments on which context adaptation is based.

The semantic equivalence between media items, or between presentation fragments, is based on the information content they provide. Content can be expressed by means of different formalisms: MPEG 7, metadata descriptions, or logic assertions are examples of possible formalisms.

We assume the system has access to a *database* in which metadata associated to each media item, including its level of detail, is stored. Description of the physical resources that are needed to present the item is also available through the database. Formally, the semantics of media items is represented as follows.

Definition 4.1. *The semantics of a media item m is a pair $Sem(m) = \langle Cont_m, Lv_m \rangle$, where (i) $Cont = \{c_1, c_2, \ldots, c_n\}$ is the specification of the content, $c_i \in Dom$[3] can be keywords explicitly listed by the author, or extracted from the metadata associated to the item, or they can be concepts extracted by means of any tool for automatic extraction of semantics; (ii) $Lv_m : Dom \rightarrow R$ is a function that, given a concept c, assigns a level of detail value for the object m with respect to c. The function Lv_m is a combination of $l_m \in R$ which is the level of detail of the object m (for instance, the resolution of an image), and $W_m : Dom \rightarrow R$, a function which assigns to every concept, the weight for that concept in the considered media item*[4]

The notion of semantics of a media item can be extended to define the semantics of a document.

Definition 4.2. *The semantics of a document $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC} \rangle$ is a pair $Sem(D) = \langle Cont_D, Lv_D \rangle$, where*

[3]We do not make any strong hypothesis on the domain *Dom* of elements.

[4]The function W_m may express different notions depending on the type of media item. For example it can express term frequency for text media items, or the percentage of the image representing the concept, for images.

(i) $Cont_D = \bigcup_{m \in \mathcal{MI}} Cont_m$, *and (ii)* $Lv_D : Dom \to R$ *is a function such that* $Lv_D(c) = max_{m \in \mathcal{MI}} Lv_m(c)$.

Based on the above definition of document semantics, we can state when two or more documents or document fragments can be seen as equivalent; i.e., they can be used as candidates for replacement during adaptation.

DEFINITION 4.3. *Let* D_1 *and* D_2 *be two multimedia documents,* $Sem(D_1)$ *and* $Sem(D_2)$ *be their semantics expressed with the same domain* Dom*, and* $\equiv$ *be any equivalence relation on* Dom*.* D_1 *and* D_2 *are semantically equivalent if (i)* $Cont_{D_1} \equiv Cont_{D_2}$*, that is, if for all* $c_i \in Cont_{D_1}$ *there exists* $c_j \in Cont_{D_2}$ *such that* $c_i \equiv c_j$*, and viceversa, and (ii) for every pair* c_i, c_j*,* $c_i \in Cont_{D_1}$ *and* $c_j \in Cont_{D_2}$ *such that* $c_i \equiv c_j$*, it holds that* $Lv_{D_1}(c_i) = Lv_{D_2}(c_j)$.

A simple example of equivalence relation on concepts is string equality. Different notions of equivalence can be used in different applications, depending on the available knowledge representation system, such as available ontologies which determine concept equivalence.

Equivalent documents, or equivalent document fragments, are the natural candidates for replacement, when a given document D cannot be presented and adaptation is needed. Unfortunately, equivalent alternatives are not always available and in most cases suboptimal candidates have to be used instead. Suboptimal candidates are found by relaxing the constraint about the equivalence of the detail of the objects, so that simpler (i.e., less detailed) versions for the same content can be presented. The level of details of the documents is compared by means of a distance function on documents, defined as follows.

DEFINITION 4.4. *Let* $\mathcal{D}$ *be the domain of the documents,* D *and* D' *be two documents,* $Sem(D) = \langle Cont_D, Lv_D \rangle$*, and* $Sem(D') = \langle Cont_{D'}, Lv_{D'} \rangle$*, be their semantics, where* $Cont_D \equiv Cont_{D'}$.
$dist_{lev} : \mathcal{D} \times \mathcal{D} \to R$ *is a function which returns the distance between the detail level of its arguments.*

Examples of distance functions are the following:

- $dist_{lev}(D, D') = max\{(Lv_D(c) - Lv_{D'}(c)) \mid c \in Cont_D\}$;
- $dist_{lev}(D, D') = \Sigma_{c \in Cont_D}(Lv_D(c) - Lv_{D'}(c))$;
- $dist_{lev}(D, D') = \Sigma_{c \in Cont_D}(\mid Lv_D(c) - Lv_{D'}(c) \mid)$.

The distance function defined on documents is the basis for the definition of the *preference relation* defined over documents.

DEFINITION 4.5. *Let* D*,* D'*, and* D'' *be three documents,* $Sem(D) = \langle Cont_D, Lv_D \rangle$*,* $Sem(D') = \langle Cont_{D'}, Lv_{D'} \rangle$ *and* $Sem(D'') = \langle Cont_{D''}, Lv_{D''} \rangle$*, be their semantics, where* $Cont_D \equiv Cont_{D'} \equiv Cont_{D''}$*, and let* $dist_{lev}$ *be any distance function on the detail of the documents. Both* D' *and* D'' *are semantic alternatives to* D*.* D' *is* preferred *to* D'' *wrt.* D*,* $D' \succeq_D D''$ *if* $dist_{lev}(D, D') \leq dist_{lev}(D, D'')$.

Since the adaptation process looks for self-contained fragments as candidates for replacement, the above definitions are usually applied to document fragments, as modules to be replaced, instead of to entire documents.

4.1 Detecting alternatives for replacement

Some systems allow an author to explicitly express replacement candidates for a single media or for part of a document. In general authors can express their alternatives by means of ad hoc constructs provided by the underlying model (like the SMIL [11] `switch` tag) or using appropriate metadata specification stored in the database. The metadata is used with the media items' features, semantic content annotations, and the device requirements.

In the first case, for context adaptation the system does not need to consider the semantic properties of possible alternatives: the system simply checks which one, among the listed alternatives, is the most suitable one for the current resource constraints.

In the second case, whenever a set of media $\mathcal{M} \subseteq \mathcal{MI}$ from the original document D cannot be delivered in the current context, the *adaptation module* uses metadata to choose the most appropriate adaptation approach. Without loss of generality, we assume that the author defines her alternatives in terms of logical assertions $Alt(F, F')$, where *both* F *and* F' *are closed fragments.* Given these fragments, the system (i) checks whether $\mathcal{M}$ is the set of media of a *closed fragment*[5], and if this is the case (ii) *computes* alternatives for the fragment to be replaced. When adaptation is needed, with the goal of returning the optimal alternative, the database is first queried, to find a *semantical equivalent* fragment P to the undeliverable one F which needs adaptation[6]. The query is iteratively repeated until an alternative presentable in the current context is found, if any. If this first phase does not succeed (no presentable equivalent alternative exists), the system checks if any *weaker version* can be found for the undeliverable fragment and returns the top one (according to the $\preceq_F$ order relation)[7].

In case of failure, the constraint on the *semantic content equivalence* is relaxed, and the subsets of $Cont_F$ will be considered in decreasing order (according to the $\subseteq$ order relation) to find the maximal $K \subseteq Cont_F$ for which a presentable alternative exists for the given context[8].

Note that in the last step of the fragment selection method, we do not make any specific hypothesis on the order according to which maximal subsets are looked for. For example, we might first order the subsets according to their resource consumptions and try to choose the optimal al-

[5]If $\mathcal{M}$ does not define a closed fragment there can be some temporal constraints between "inner" media items of the fragment (i.e., media items whose start and end does not coincide with the start and the end of the fragment) and objects of the presentation. In this case it would not be possible to relate such media, external to the fragment, to media in the replacing candidate.

[6]The query is $(Alt(F, P) \vee Alt(P, F)) \wedge presentable(P) \wedge (P \equiv F)$ where $presentable(P)$ is a predicate which checks if P can be delivered in the current context.

[7]The query $(Alt(F, P) \vee Alt(P, F)) \wedge presentable(P) \wedge (\forall P', (Alt(F, P') \wedge presentable(P')) \to P \succeq_F P')$ binds P to the best presentable alternative for F.

[8]The query is $(Alt(F, P) \vee Alt(P, F)) \wedge presentable(P) \wedge (\forall P', (Alt(F, P') \wedge presentable(P')) \to P \succeq_{K,C} P')$., where $\succeq_{K,C}$ is the *preferred alternative* of the document C with respect to the content K.

ternative in terms of resource usage/saving.

We remark that the same method can be used to enhance the quality of a multimedia presentation: we can substitute a fragment F with a presentation $P \succeq F$ to increase the level of detail of the original document.

4.2 Fragment substitution

If the system is able to find out a suitable candidate P for the replacement of the closed fragment F not compatible with the current user context, a set of *adaptation rules* must be applied to tune the original document $D = \langle \mathcal{MI}, \mathcal{E}, \mathcal{TC} \rangle$ into a new deliverable $D' = D[F/P] = \langle \mathcal{MI}', \mathcal{E}', \mathcal{TC}' \rangle$.

While integrating the new fragment P in the "deliverable part" (D_F) of the original presentation, temporal constraints from the two documents have to be combined. If the same medium occurs in both P and D_F, its name clearly refers to distinct presentation instances: the documents to be integrated have disjoint media sets, and disjoint constraint and event sets, accordingly, thus no conflict can arise due to the integration.

The adaptation of the document affects the set of media items and the constraints that will temporally relate the new candidate fragment to the remaining part of the original document. If the closed fragment $F = \langle \mathcal{MI}_F, \mathcal{E}_F, \mathcal{TC}_F \rangle$ is replaced by $P = \langle \mathcal{MI}_P, \mathcal{E}_P, \mathcal{TC}_P \rangle$, the resulting document contains $\mathcal{MI}' = \mathcal{MI} \setminus \mathcal{MI}_F \cup \mathcal{MI}_P$ and $\mathcal{E}' = \mathcal{E} \setminus \mathcal{E}_F \cup \mathcal{E}_P$, i.e., all items and events from F have been removed, while items from P have been added.

The management of temporal constraints is more complex since the replaced closed fragment was temporally related to the remaining part of the presentation.

From the closeness hypothesis for replaced fragments, it is known that in the presence of any temporal constraint which binds a media item $m_F \in \mathcal{MI}_F$ with another media item in $\mathcal{MI}_D \setminus \mathcal{MI}_F$, either m_F starts, or it ends with F. Therefore, each constraint on event $start(m_F)$, where m_F is a media items starting with F, (or on event $end(m'_F)$, where m'_F is a media item ending with F), is replaced with a reference to the start (the end, respectively) of the fragment P. We denote the obtained set of constraints by $\mathcal{TC}[F/P]$. Then $\mathcal{TC}'$ is equal to $\mathcal{TC}[F/P] \setminus \mathcal{TC}_F \cup \mathcal{TC}_P$ which is obtained by removing the temporal information related to the closed section F from $\mathcal{TC}[F/P]$ and by adding the set $\mathcal{TC}_P$ of synchronization constraints of the presentation P.

Although it preserves all the explicitly constrained mutual relationships between the events, this solution does not ensure that the resulting set of constraints $\mathcal{TC}'$ is solvable. As an example, it might be the case that the actual duration of the fragment P is longer than the duration of the replaced fragment F, and a constraint binds two events, one before F and one after F, e.g., $c_1 \leq end(D) - start(D) \leq c_2$ narrows the overall duration of the presentation.

To deal with these cases, we introduce a final *translation phase*, which relaxes the strict unsatisfiable constraints, under the hypothesis that the duration of the two fragments (the replaced F and the replacing one P) are partially known, i.e., we have $min_F \leq dur_F \leq max_F$ and $min_P \leq dur_P \leq max_P$.

Since during the adaptation process $inst(ev)$ is known, if ev occurs before the beginning of F, we can translate each critical constraint in the form $c_k \leq inst(ev_k) \leq c'_k$. Then, we replace it with the constraint $c_k + \delta \leq inst(ev_k) \leq c'_k + \delta'$ where $\delta = min_P - min_F$ and $\delta' = max_P - max_F$ could also be negative numbers.

If this translation is not possible, it is the author's responsibility to ensure that the alternatives he has defined are presentable, also in terms of temporal constraints.

To clarify the complete document adaptation process, we consider again the multimedia document presenting a tourism application of Section 2.1. The user can play the presentation through different devices like a desktop computer, a PDA or a cell phone. Using a desktop computer, the answer to the user query is rendered as described in Section 2.1. Otherwise, the system provides a better solution according to the current context.

Let us assume that the user can query the database through the desktop computer and decides to switch to her cell phone, which can not play media $audio_i$, due to their file format, and can not visualize the associated sequence of pictures. Therefore, given the location i, the *undeliverable* fragment F_i contains the set of media items $\{audio_i, pic_{i,1}, \ldots, pic_{i,n}\}$. The database contains the predicate $Alt(F_i, F'_i)$ for each location, F'_i being a sequence of text descriptions, named $text_{i,j}$, each one associated to a smaller image, $im_{i,j}$, like in a MMS. We can also assume that $Cont(F'_i) \equiv Cont(F_i)$ (i.e., the two fragments have the same semantic content), but F'_i has a lower level of details.

The result is shown in Figure 2. The original set $\mathcal{TC}$ is modified by removing temporal constraints involving media items in F_i, and adding the temporal constraints describing the synchronization between text descriptions and images, i.e., $0 \leq start(text_{i,j}) - start(im_{i,j}) \leq 0$, $0 \leq stop(text_{i,j}) - stop(im_{i,j}) \leq 0$, and $0 \leq stop(text_{i,j}) - start(text_{i,j}) \leq 5$. Moreover, the constraint $0 \leq start(audio_{i+1}) - stop(audio_i) \leq 2$ is translated into $0 \leq start(audio_{i+1}) - stop(text_{i,n}) \leq 2$.

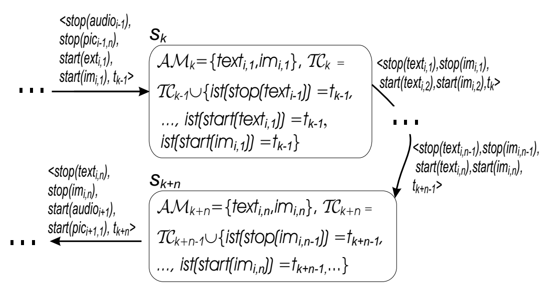

Figure 2: A fragment of the automaton representing the history case after the adaptation process.

The same elaboration applies to the next location, unless the user turns back to the desktop computer. As already discussed in Section 4.1, the same approach can be used also to enhance the quality of the presentation, if some change in the user context allows it. As an example, if the network connection supports delivery of video files, some audio narrations can be substituted with a video file with an higher level of detail. In this case, the audio file $audio_i$ can be considered a closed fragment, so the adaptation process is trivial.

5. DISCUSSION AND CONCLUSION

In this paper we have introduced a framework for the dynamic adaptation of a multimedia document to a given context, which is described in terms of available resources (i.e, the device), user profile and environment. The evolution of a multimedia document is described by means of an automaton that records, at each step, the set of active media, and the occurred events.

The adaptation process allows the system to replace a set of media items with an alternative set defined by the author through the underlying model or using metadata. To be "replaceable", therefore *adaptable*, a set of media items must build a closed fragment of the original document of which the system can evaluate the set of keywords.

The possible alternatives for replacement are calculated taking into account the semantic content and the level of detail of the *undeliverable* fragment: the system first looks for *semantic equivalent* alternatives; if they do not exist, the system checks if any *weaker version* can be found.

The notion of equivalence is also proposed in [4]: the authors propose an adaptation system in which media elements or fragments of a multimedia document can be replaced by different media elements of different quality and type, but that are semantic equivalent alternatives. To select potential alternatives, first maintenance of the semantic of the presentation is verified, then the preservation of the information flow is checked. The adaptation is done during presentation playback taking into account changes in the context. In contrast to our approach, this model defines the equivalence based on some discriminating aspects, like subject and duration, while in our approach we impose also an order between fragments relating the same information.

In [5] the authors propose an architecture of a context-aware document adaptation system. A *resolver* defines which media items can be used in different contexts. Media features and context properties are expressed through logical assertions. In contrast to our approach, different alternatives, explicitly listed by the author in the multimodal document, are assumed to be semantically equivalent without specifying media contents and their level of detail. Our framework provides a mechanism to automatically detect candidates for replacement, guaranteeing the same information, at the highest level of detail.

Our adaptation algorithm also "adapts" temporal constraints defined on the replaced objects, to properly tight them to the context in which they are inserted. This is an innovative aspect of our approach. Synchronization between components of a multimedia presentation is not taken into account in [10], which presents a system adapting multimedia Web content to optimally match the resources and capabilities of diverse client devices.

[9] presents a general framework for multidimensional adaptation, which defines a graph-based representation abstraction embedded within a multidimensional utility space. The representation graph is the basic structure on which dimensional tradeoffs, cost metrics and utility metrics are defined. The evaluation of cost and utility metrics guides adaptation. Adaptation is mostly seen as adjusting access to data to reflect current system conditions. Similarly to our approach, the authors define multidimensional ordering on which their adaptation is based, but they do not consider the impact of adaptation on the temporal aspects of the presentation.

In [8] the authors propose a semantic approach, which transforms a multimedia document into one compatible with the context. A model is a potential execution of a document and a context defines a particular class of models; the adaptation process looks for the models of the original document that belong to the class defined by the context: if such models do not exist, it produces a document whose models belong to the class and are close to those of the original document. In this approach the distance between two objects is obtained taking into account temporal relations between them and not their content.

In the future, we plan to investigate in different directions: in particular we plan to analyze the possibility of, given a database of multimedia documents, automatically extract closed fragments of a presentation; moreover we better investigate the characterization of the user situation considering its activity, in order to infer an appropriate level of detail of the presentation.

As a further extension, we plan to associate a temporal interval of validity to each adaptation rule: this means that some rules are always valid, but others are evaluated only during some specific time intervals. Similarly, adaptation rules can be expressed in causal forms, instead of being just logical assertions, with mutual dependency relations.

6. REFERENCES

[1] E. Bertino and E. Ferrari, Conceptual Temporal Models for Multimedia Data. In *IEEE TKDE*, 10(4), 1998.

[2] C. Bettini and R. De Sibi. Symbolic Representation of User-defined Time Granularities. In *Proc. TIME'99* 1999.

[3] V. Brusoni, L. Console, and P. Terenziani. On the computational complexity of querying bounds on differences constraints. In *Artificial Intelligence*, 74(2), 1995.

[4] S. Boll, W. Klas, and J. Wandel. A Cross-Media Adaptation Strategy for Multimedia Presentations. In *ACM Multimedia*, p. 37–46, 1999.

[5] A. Celentano and O. Gaggi. Context-aware Design of Adaptable Multimodal Documents. In *Proc. MWD04 at IEEE MSE04*, p. 427–434, 2004.

[6] A. K. Dey. Understanding and Using Context. In *Personal and Ubiquitous Computing Journal* 5(1), p. 4–7, 2001

[7] L. Egidi and P. Terenziani. A lattice of classes of user-defined symbolic periodicities. In *Proc. Int. Symposium on Temporal Representation and Reasoning*, 2004.

[8] J. Euzenat, N. Layaïda, and V. Dias. A Semantic Framework for Multimedia Document Adaptation. In *Proc. 18th IJCAI*, 2003.

[9] D. Gotz and K. Mayer-Patel. A General Framework for Multidimensional Adaptation. In *ACM Multimedia*,2004.

[10] R. Mohan, J.R. Smith, and C.-S. Li. Adapting Multimedia Internet Content for Universal Access. In *IEEE Transaction on Multimedia* , 1(1), 1999.

[11] Synchronized Multimedia Working Group of W3C. Synchronized Multimedia Integration Language (SMIL) 2.0 Specification. http://www.w3.org/TR/smil20, August 2001.

[12] A. Tuzhilin and J. Clifford. On Periodicity in Temporal Databases In *Information Systems*, 20(8), 1995.