Multimedia Presentations Using Augmented Transition Networks and Multimedia Input Strings

Shu-Ching Chen
School of Computer Science
Florida International University
Miami, FL 33199

Mei-Ling Shyu R. L. Kashyap
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907-1285

Abstract

An abstract semantic model called augmented transition network (ATN) to model multimedia presentations is proposed. An ATN is created based on a multimedia input string and it consists of a set of states and directed arcs. The advantages to using a multimedia input string to generate an ATN are its simplicity and ease of modification. The arc symbols of subnetworks represent the temporal and spatial relations of semantic objects. The separated condition/action table is used to control synchronization and quality of service of the multimedia presentation at both the coarse-grained and fine-grained levels. The formal proof of any multimedia input string that has the corresponding ATN is presented in this paper. Based on this proof, it shows that together ATNs and multimedia input string, multimedia presentations can be modeled.

Key words: multimedia presentation, augumented transition network (ATN)

1 Introduction

Many semantic models have been proposed to model temporal and spatial relations; however, there are four major disadvantages to most of these conceptual models. First, user interactions are not included in the conceptual models proposed by [1, 7, 11]. Second, the conceptual model does not have the mechanisms to handle the different delay situations. Some models handle a communication delay by adjusting the playout deadline schedule for the media streams; however, they do not provide the necessary actions for different communication delays [1, 5, 7]. Third, existing conceptual models are either too complex for the users to understand [1, 7, 10, 11] or too simple to let users see the whole view of the presentation schedule [5]. Fourth, few existing conceptual models model both temporal and spatial relations. They either develop a temporal model to capture synchronization information [1, 5, 7, 10, 11] or use image/computer vision techniques to get content-based information in the image or video. In our previous works, we have already shown that an augmented transition network (ATN) and multimedia input string can model the temporal and spatial relations, user interaction, user loop, graphical view [2, 3, 4, 8, 9]. In this paper, a formal proof is presented that given any multimedia input string there exists a corresponding ATN.

The organization of this paper is as follows. Section 2 defines the multimedia input string. In section 3, the augmented transition network semantic model is proposed. Conclusions are in section 4.

2 Multimedia Input String

Regular expressions [6] are useful descriptors of patterns such as tokens used in a programming language. Regular expressions are convenient ways of specifying certain set of strings. A multimedia input string is similar to a regular expression. It is used to represent the multimedia presentation sequence. In order to model the time sequence of a multimedia presentation, a multimedia input string is scanned from left to right to represent the time sequence.

Two notations $\mathcal{L}$ and $\mathcal{D}$ which are used to define multimedia input strings are defined as follows:
$\mathcal{L}$ = {A, I, T, V} is the set whose members represent the media type, where A, I, T, V denote audio, image, text, and video, respectively.
$\mathcal{D}$ = {0, 1, ..., 9} is the set consisting of the set of the ten decimal digits.

The following definitions will be based on the above notations.

Definition 2.1: Let $\mathcal{C}$ be a set that consists of the streams mi and cmi, where $m \in \mathcal{L}$, cm is the compressed version of m, and i is the index within a media type, m_i is the logical unit of the medium m and is a semantic event. This logical unit can be identified manually or detected automatically by the analysis of the content. For example, A_{100} means audio stream 100 and CA_{100} means compressed audio stream 100. For image, video, and audio, each m_i represents still image, video frame, and audio stream, respectively.

Definition 2.2: A multimedia input string can be created from $\mathcal{L}$ and D by applying the *union*, *concatenation*, *positive closure* and *Kleene closure* operations.

- **Concurrent:** The symbol "&" between two media streams indicates these two media streams are displayed concurrently.
- **Looping (positive closure):** $m^+ = \bigcup_{i=1}^{\infty} m^i$ is the multimedia input string of positive closure of m to denote m occurring one or more times. A multimedia input string uses "+" symbol to model loops in a multimedia presentation to let some part of the presentation be displayed more than once.
- **Optional (Kleene closure):** $m^* = \bigcup_{i=1}^{\infty} m^i$ is the multimedia input string of Kleene closure of m to denote m occurring zero or more times. In a multimedia input string, "*" symbol is used to represent media streams which can be dropped in the on-line presentation.
- **Contiguous:** Input symbols which are concatenated together are used to represent a multimedia presentation sequence and to form a multimedia input string. Input symbols are displayed from left to right across time sequentially. ab is the multimedia input string of a concatenates with b such that b will be displayed after a is displayed.
- **Alternative (Union):** A multimedia input string can model user selections by separating input symbols with the "|" symbol. So, $a|b$ is the multimedia input string of *a union (or) b*.
- Ending: The symbol "$" denotes the end of the presentation.

Users need to specify the tentative starting time and ending time of media streams. Based on the tentative starting time and ending time, an input ordered set that is sorted by the tentative starting time and ending time of these medias streams is constructed. A multimedia input string is constructed using the input ordered set and this multimedia input string is an input for an augmented transition network (ATN).

3 The Augmented Transition Network

A multimedia presentation consists of media streams displaying together or separately across time. The arcs in an ATN represent the time flow from one state to another. An ATN can be represented diagrammatically by a labeled directed graph, called a *transition graph*. The ATN grammar consists of a finite set of nodes (states) connected by labeled directed arcs. An arc represents an allowable transition from the state at its tail to the state at its head, and the labeled arc represents the transition function. An input string is accepted by the grammar if there is a path of transitions which corresponds to the sequence of symbols in the string and which leads from a specified initial state to one of a set of specified final states. Each nonterminal symbol consists of a subnetwork which can be used to model the temporal and spatial information of semantic objects for images and video frames and keywords for texts. In addition, a subnetwork can represent another existing presentation. Any change in one of the subnetworks will automatically change the presentation which includes these subnetworks. To design a multimedia presentation from scratch is a difficult process in today's authoring environment. The subnetworks in an ATN allow the designers to use the existing presentation sequence in the archives, which makes the ATN a powerful model for creating a new presentation. This is similar to the *class* in the object-oriented paradigm. Also, subnetworks can model keywords in a text media stream so that database queries relative to the keywords in the text can be answered.

States are represented by circles with the state name inside. The state name is used to indicate the presentation being displayed (to the left of the slash) and which media streams have just been displayed. The state name in each state can tell us all the events that have been accomplished so far. Based on the state name, we can know how much of the presentation has been displayed. When the control passes to a state, it means all the events before this state are finished. A state node is a breaking point for two different events. In ATNs, when any media stream begins or ends, a new state is created and an arc connects this new state to the previous state. Therefore, a state node is useful for separating different media stream combinations into different time intervals. The arc labels can tell us

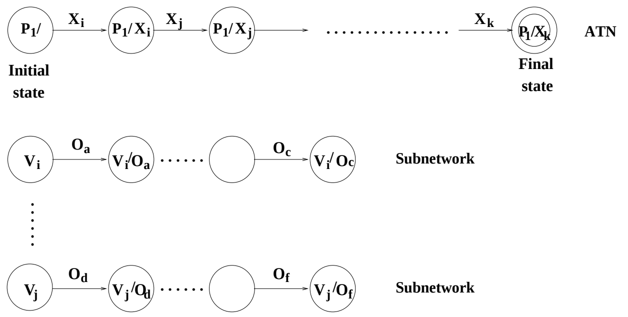

Figure 1: Augmented Transition Network Structure. Subnetworks such as V_i and V_j are constructed for video streams V_i and V_j, respectively. V_i and V_j appear in the arc symbols in the augmented transition network P_1.

which media streams or semantic objects are involved. Each arc represents a time interval.

In Figure 1, the initial state (state $P_1/$) has only one outgoing arc with symbol X_i. X_i is also the symbol associated with the arc pointing to state P_1/X_i. When certain conditions are satisfied for state P_1/X_i, the new input symbol X_j is read and advances to the next state (state P_1/X_j). This process continues until the final state P_1/X_k is met.

When the input string contains either images or video streams, subnetworks are constructed. As shown in Figure 1, P_1 is the augmented transition network. A subnetwork is constructed for each video stream. In this example, we assume that V_i and V_j appear in the arc symbols in the network P_1 so that V_i and V_j are two subnetworks constructed for video streams V_i and V_j. The input symbols of a subnetwork represent the relative spatial relationships among semantic objects [4]. A semantic object is an object in an image or video frame such as a "car". The state changes when there is any relative spatial location change of semantic objects or number of semantic object change. In multimedia database queries or browsings, if queries or browsings need to access information from videos or images, then the searching will go to the corresponding subnetworks. The searching will return to the original state when it reaches the final states. Therefore, this makes the ATN a *recursive* transition network.

Theorem 3.1: Let α be a multimedia input string. Then there exists an augmented transition network representing the language denoted by α.

Proof: Let $|\alpha|$ denote the length of α. The length of a multimedia input string α is the number of input

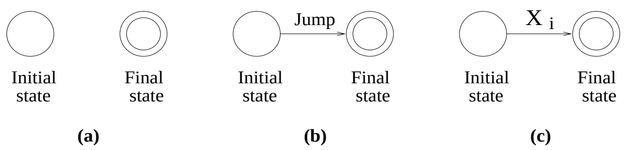

Figure 2: ATN representing elementary multimedia input strings of length 1:(a) for empty set ϕ, (b) for null input symbol and directly goes to next state, (c) for input symbol X_i.

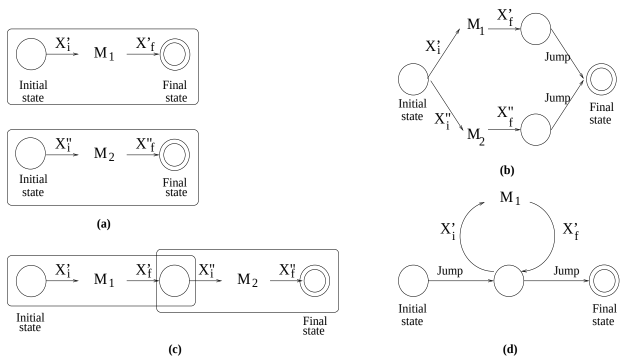

Figure 3: ATN representing increasingly longer multimedia input strings:(a) for β and γ, (b) for $\beta \mid \gamma$, (c) for $\beta\gamma$, (d) for β^*.

symbols in α. The proof is by induction on the length of α and we shall prove the theorem by constructing the required ATN.

BASIS: Let $|\alpha|$=1; α must be either an empty set ϕ, null input symbol (this means no input symbol is read and advances to the next state using JUMP notation), or an input symbol X_i. The ATN shown in Figure 2 represents the denoted multimedia input string.

INDUCTION STEP: Assume the theorem is true for multimedia input strings with n input symbols or less. We now show that it must also be true for any multimedia input string α having n+1 input symbols. α must be in one of the three forms:

$$(1)\ \alpha = \beta \mid \gamma \quad (2)\ \alpha = \beta\gamma \quad (3)\ \alpha = \beta^*$$

where β and γ are multimedia input strings, each having n or fewer characters. According to the induction hypothesis, the sets β and γ can be recognized by ATNs, which we shall denote M_1 and M_2, respectively, and have disjoint sets of states. Let their initial states be $X_i^{'}$ and $X_i^{''}$ and their final states be $X_f^{'}$ and $X_f^{''}$ as shown in Figure 3a. In Figure 3, each ATN contains just one starting node and one accepting node.

The set described by $\beta|\gamma$ can be recognized by an ATN composed of β and γ, as shown in Figure 3b. The set described by $\beta\gamma$ can be recognized by an ATN constructed in the following manner. Coalesce the final state of M_1 with the initial state of M_2 and regard the combined state as one that is neither initial nor final state. The resulting graph is shown in Figure 3c. The initial states of this graph are the initial state of M_1, while the final states are those of M_2. Clearly, this graph will represent a multimedia input string if and only if that multimedia input string belongs to $\alpha = \beta\gamma$. Finally, to represent the set β^*, construct the graph of Figure 3d.

Since every multimedia input string can be described by an expression obtained by a finite number of applications of the operations $|$, $\star$, and *concatenation* on the media streams, the theorem is proved. $\square$

4 Summary

In this paper, multimedia input strings and ATNs are proposed to model multimedia presentations. The ATN is a left to right model so that it represents the time flow from left to right. A condition/action table in the ATN is designed to separate the necessary conditions and actions from the graphical representation. In this regard, users are provided a clear view of the whole structure of the multimedia presentation. Moreover, a graphical representation can be used as the data structure inside the multimedia presentation, and the condition/action table is the only place that requires memory space.

Unlike a traditional abstract semantic model that only models the media streams, an ATN and its subnetworks can model the temporal and spatial relations of semantic objects based on multimedia input string. A formal proof is presented to show that given any multimedia input string there exists a corresponding ATN to represent it. Based on this proof, it tells us that multimedia presentations can be modeled by using an ATN and its multimedia input string.

Acknowledgements

This work has been partially supported by National Science Foundation under contract IRI 9619812 and the office of Naval Research under contract N00014-91-J-4126.

References

[1] H-J. Chang, T-Y. Hou, S-K. Chang, "The Management and Application of Teleaction Objects," *ACM Multimedia Systems Journal* vol. 3, pp 228-237, November 1995.

[2] S-C. Chen and R.L. Kashyap, "Temporal and spatial semantic models for multimedia presentations," 1997 International Symposium on Multimedia Information Processing, pp. 441-446, Dec. 11-13, 1997.

[3] S-C. Chen and R.L. Kashyap, "Empirical studies of multimedia semantic models for multimedia presentations," 13th International Conference on Computer and Their Applications, pp. 226-229, March 25-27, 1998.

[4] S-C. Chen and R.L. Kashyap, "A spatio-temporal semantic model for multimedia presentations and multimedia database systems," accepted for publication in *IEEE Transactions on Knowledge and Data Engineering*, 1999.

[5] N. Hirzalla, Ben Falchuk, and Ahmed Karmouch, "A Temporal Model for Interactive Multimedia Scenarios," *IEEE Multimedia*, pp. 24-31, Fall 1995.

[6] S. C. Kleene, "Representation of Events in Nerve Nets and Finite Automata, Automata Studies," Princeton University Press, Princeton, N.J., pp. 3-41, 1956.

[7] T.D.C. Little and A. Ghafoor, "Synchronization and Storage Models for Multimedia Objects," *IEEE J. Selected Areas in Commun.*, vol. 9, pp. 413-427, Apr. 1990.

[8] M-L. Shyu, S-C. Chen, and R. L. Kashyap, "Information Retrieval Using Markov Model Mediators in Multimedia Database Systems," 1998 International Symposium on Multimedia Information Processing, pp. 237-242, Dec. 14-16, 1998.

[9] M-L. Shyu and S-C. Chen, "Probabilistic Networks for Data Warehouses and Multimedia Information Systems," Submitted to *IEEE Trans. on Knowledge and Data Engineering.*

[10] T. Wahl and K. Rothermel, "Representing Time in Multimedia Systems," Proc. Int'l Conf. on Multimedia Computing and Systems, CS Press, Los Alamitos, Calif., pp. 538-543, 1994.

[11] Yahya Y. Al-salqan and Carl K. Chang, "Temporal Relations and Synchronization Agents," *IEEE Multimedia*, pp. 30-39, Summer 1996.