

論理式によるマルチメディア同期表現モデル

佐藤 龍雄 林 正薫 岡田 謙一 松下 温

慶應義塾大学理工学研究科

異機種・分散環境でマルチメディアデータの共有・交換を行うためには、単一のメディアの情報と各メディア間の関係の情報とに分ける必要がある。メディア間の関係には大きく分けて空間的な関係と時間的な関係が考えられるが、本稿ではメディア間の時間的な関係を標準的に記述するための同期モデルを考えた。このモデルでは、メディアをプレゼンテーションするための条件をすべて論理式で表すことができる。また、ユーザからの入力を1つのメディアとして扱うことにより、ユーザとのインタラクションを含む同期関係も同様に論理式で表現できる。実際に、この同期モデルに基づいたアプリケーション(電子動物図鑑)を作成した。

A Presentation Model for Multimedia Synchronization by Boolean Algebra

Tatsuo SATO Joung-hoon LIM Kenichi OKADA Yutaka MATSUSHITA

Faculty of Science & Technology
Keio University

3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi 223, Japan

For treating multimedia information in distributed environment, it is an important issue how describing the synchronization relationships among media which contain not only text, images, but also audio and video. This paper proposes a describing method for specifying synchronization relationships among media by boolean expression. Using this method we implement prototype system, named, "Multimedia Encyclopedic" in the UNIX environment.

1 はじめに

近年のコンピュータ技術の発達により、コンピュータは様々なメディア（テキスト、静止画、動画など）を扱えるようになった。また、様々なメディアを統合することにより、従来にはない新たな計算機環境が実現できる。

異機種・分散環境におけるマルチメディア情報の共有や交換のためには、各メディアデータの標準的な記述方法が必要である [3, 4, 5]。現在、単一のメディアに関するフォーマットや符号化には、様々な種類が存在する。このような状況下でマルチメディア情報のフォーマットを定めると、莫大な数のフォーマットが考えられる。たとえば、音声、静止画、テキストの符号化がそれぞれ a, b, c 通りあったとすると、その組み合わせは $a \times b \times c$ 通り存在する。

このような状況を打開するためには、マルチメディアの情報を各メディアごとの情報に分け、各メディア間の関係を標準的な方法で記述すればよい [1, 2]。メディア間の関係は、空間的な関係と時間的な関係に分類できる。空間的な関係とは、どのメディアをディスプレイ上のどの位置にレイアウトするかという情報である。一方、時間的な関係とは、どのメディアをどのようなタイミングでプレゼンテーションしたらよいかという情報である [7]。

その中で我々は、とくにメディア間の時間的な同期関係に着目し、その記述方法を提案する。

動画や音声といった時間依存のメディアを扱うようになると、メディア間の時間的な同期関係、すなわち、いつ、どのメディアを表示もしくはプレゼンテーションするかという情報は、すべてのメディア間に普遍的に存在する。言い換えれば、時間的な同期関係はすべてのメディアを結び付ける重要な役割を果たす情報である。

本稿では、まずメディアにはどのようなものが考えられるか、また、どのようにそれらが分類されるか論じる。次にメディアの同期モデルについて説明し、その同期モデルに基づいてメディア間の同期関係を論理式を用いて表現する方法について述べる。最後に、この同期モデルに基づいたマルチメディアアプリケーション（マルチメディア動物図鑑）のインプリメントに関して説明を行う。

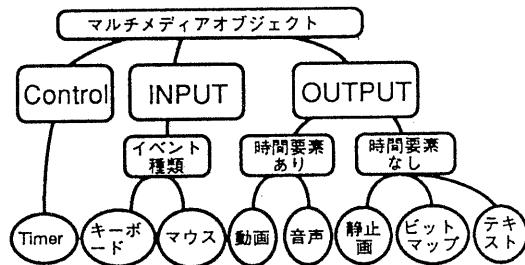


図 1: メディアの分類

2 メディアおよびメディアオブジェクト

2.1 メディアの意味

メディアという用語は、漢字では‘媒体’と書くことができる。現代社会では、‘メディア’も‘媒体’も非常に多くの意味を持つに至った。

日常では、‘マスメディア’という言葉に代表されるように、情報の伝達媒体もしくは手段として広く使われている。また、磁気ディスク、光磁気ディスク、CD-ROM などのデジタル情報の蓄積媒体の種類を‘メディア’と呼ぶこともある。

ここでは‘メディア’とは、テキスト（文字）、音声、静止画、動画といった、情報の表現形式のことを意味する。

2.2 メディアの分類

我々は、Fig.1 に示すように、すべてのメディアを INPUT, OUTPUT, そしてそのどちらにも分類されない Timer メディアの 3 種類に分類した。

OUTPUT は人間が知覚することができるメディアである。OUTPUT メディアはさらに TDM (Time Dependent Media: 時間依存メディア) と TIM (Time Independent Media: 時間非依存メディア) に分類される。TDM は‘時間軸をもつメディア’とも呼ばれ、音声、動画などがある。TIM は時間軸とは無関係のメディアであり、テキスト、静止画、などがある。

我々が提案するモデルでは、ユーザの入力も 1 つのメディアとして扱うことができる。これを INPUT メディアと呼ぶ。ユーザの入力には、例えば、マウスによるボタンのクリック、キーボードによる入力

等が考えられる。基本的な入力操作を組み合わせることで、メニュー選択などの入力方法を実現できる。

Timer は、時間の経過を表現するメディアである。ユーザの五感で知覚されず、また、ユーザの入力を反映するものでもないで、OUTPUT, INPUT のどちらにも分類できない。Timer には、次のような使い方が挙げられる。

- プレゼンテーションのディレイ

たとえば動画の表示開始から 1 秒遅れてナレーションを開始したい場合、1 秒間の長さを持つ Timer を使い、Timer と動画を同時に開始、Timer の後にナレーションを開始、という具合に考える。

- 一定時間の TIM プレゼンテーション

テキスト、静止画といった TIM は、時間要素を持たないため、オートマチックスライド（一定時間経つと次に移るスライド）を実現するためには Timer メディアを用いなければならない。一般に TIM は、TDM, INPUT メディアまたは Timer メディアのうちのどれか 1 つ以上と組み合わせなければならない。

- 期限付きの入力

Timer メディアと INPUT メディアを組み合わせると、期限付きの入力が実現できる。すなわち、ある一定期間に入力がないと自動的に次のシナリオに進むことができる。

2.3 メディアオブジェクト

1 つのプレゼンテーションシナリオは、いくつかのメディアから構成されている。また、各メディアはそれぞれいくつかの基本的な単位に分割できる。たとえば、一連の動画はシーンやカットごとに分割することができる。

この分割されたシーンやカットごとに音声やサブタイトルなどと同期をとることを考える。これら分割されたデータをメディアオブジェクトと呼ぶ。すなわち、マルチメディア情報はメディアオブジェクトの組み合わせであるといえることができる。

メディアオブジェクトという用語を使うと、本研究の目的は、“メディアオブジェクトどうしのプレゼンテーションタイミングの標準的な記述を定めることである”と言い換えることができる。

3 同期モデル

3.1 メディアオブジェクトの機能

マルチメディア情報とはメディアオブジェクトの組み合わせであると捉えることにより、メディア間の時間的な同期関係はメディアオブジェクト間の同期関係であると考えられることができる。

ここでは、メディアオブジェクトがどのような機能を持っていれば互いに同期をとってプレゼンテーションを行うことができるか、ということについて述べる。

すべてのメディアオブジェクトには 3 つの機能がある。

まず 1 つ目はカウンタである。このカウンタの初期値は 0 で、カウントするごとに 1 づつ値が増えゆき、あらかじめ設定された値に達するとカウントをやめる。このカウンタはリアルタイムのカウントを行う。したがって、1 秒間に n 回カウントする場合、 $1/n$ の精度の時計が内蔵されていると考えることができる。

2 つ目はメディア値の変化である。メディア値の初期値は 1 であり、カウンタの値があらかじめ設定された値に達すると 0 に変化する。

3 つ目はプレゼンテーション条件である。プレゼンテーション条件は、各メディアオブジェクトのメディア値を変数とする論理式（ブール代数式）で表現される。この論理式をプレゼンテーション式と呼び、式の計算結果をプレゼンテーション値と呼ぶ。条件が真であればプレゼンテーション値は 1 であり、偽ならば値は 0 である。プレゼンテーション条件が真である、すなわちプレゼンテーション値が 1 である間、カウンタの動作が許される。カウンタの値が設定値に達すると、カウントは中断される。設定値に達していなくてもプレゼンテーション条件が偽であれば、カウントが中断される。

カウンタの動作はプレゼンテーションが行われることに相当する。すなわち、カウンタが動作している期間とプレゼンテーションが行われる期間は等しい。カウンタが設定値に達するということは、もうこれ以上プレゼンテーションすべきものがないことを意味する。

以上 3 つの機能により各メディアオブジェクトは互いに同期をとってプレゼンテーションを行うことが可能になる。この中でプレゼンテーションの同期

関係と最も密接に関わっているのがプレゼンテーション条件である。この条件式の設定がメディアオブジェクト間の同期関係を決定する。ある同期関係を実現するためにはどのように条件式を設定すればよいか、ということに関しては、5節で述べる。

3.2 メディアオブジェクトの構造

前節に挙げた3つの機能、すなわち、カウンタ、メディア値、プレゼンテーション条件は、メディアオブジェクト間で同期をとりながらプレゼンテーションを行うときに用いられる。実際には、メディアオブジェクトは次の2種類の情報から構成される。

- プレゼンテーション条件
- メディアの情報

プレゼンテーション条件は各メディアオブジェクトのメディア値を論理変数とする論理式（ブール代数）で表現される。メディア値の変化によりプレゼンテーション条件の結果も変化する。この条件式の結果をプレゼンテーション値と呼び、条件式が真なら1、偽なら0の値をとる。また、条件式中の論理変数はメディア値を表していることから、“メディア変数”と呼ぶことにする。

プレゼンテーション値が1であればそのメディアオブジェクトはプレゼンテーションを行うことができる。しかし、値が0であればプレゼンテーションを行うことができず、したがって、人間が目や耳で見たり聴いたりできない状態であるといえる。

メディア値は1もしくは0の値をとる論理変数で、プレゼンテーションが行われるまでは1の値をとり、プレゼンテーションの終了により0の値に変化する。たとえば30秒の音声情報が含まれているメディアオブジェクトの場合、プレゼンテーション条件が1である間は音声を再生し続けるが、もう再生すべき音声が無くなった（データの終わりに達した）場合は、メディア値が1から0に変化する。ただし、音声の再生中であってもプレゼンテーション条件が0になると再生は中断される。

メディアの情報とは、各メディアに固有の属性、たとえば、先の音声メディアの場合では、音量、もしステレオ音声なら音像の定位などのプレゼンテーションの際の属性、そして、サンプリング周波数、音声の再生時間、圧縮の有無や圧縮方法などの音声

の符号化・復号化に関する情報が含まれる。さらに、符号化された音声情報、またはその音声情報が格納されている場所（ファイル上のどの位置か、またはデータベースのどの位置にあるか、など）に関する情報が含まれる。

4 同期の表現

メディアオブジェクトはカウンタ、メディア値、プレゼンテーション条件の3つの機能を含んでいる。ここでは、この3つがINPUT,OUTPUT,Timerの各メディアではどのような意味になるのか説明する。

4.1 TDMにおけるメディアオブジェクトの意味

音声や動画などの時間に依存するメディアでは、プレゼンテーションに要する時間（再生時間）は、符号化の際に決定される。したがって、カウンタの設定値は再生時間を意味する。

メディア値が1から0に変化するという事は、再生が終わったことを意味する。また、プレゼンテーション値が1なら再生のための処理を行い、0ならその時点で再生を中断する。再びプレゼンテーションが1になると、中断された状態から再生を再開する。

4.2 TIMにおけるメディアオブジェクトの意味

テキストや静止画などの時間に依存しないメディアは、いったんプレゼンテーションを行う（表示を行う）と、いくら時間が経過しても表示内容に変化は起きない。したがって、メディア値は常に1であり、0へは変化しない。

プレゼンテーション値が1であれば表示を開始し、0になれば表示を終了する。また、TIMのメディアオブジェクトには、カウンタは意味をなさない。強いていえば、設定値が無限大のカウンタで、プレゼンテーション値が1であれば必ずカウントを行う、すなわちプレゼンテーションを行う。

4.3 INPUT メディアにおけるメディアオブジェクトの意味

INPUT メディアでは、OUTPUT メディアのような“表示”や“再生”の概念はない。そのかわり、メディア値の変化はユーザからの入力に対応させる。たとえば、マウスがクリックされるとメディア値は1から0に変化する。

プレゼンテーション値が1の時は入力を受け付けるが、0のときはユーザの入力操作は無視される。

入力が行われるまでにかかる時間は予測不可能なので、カウンタの設定値は、設定不能かつその値を得ることができない値であると考えられる。したがって、TIM の場合と同様、INPUT メディアの場合もカウンタは意味をなさない。

4.4 Timer メディアにおけるメディアオブジェクトの意味

Timer メディアは時間のみを提供するメディアである。したがって、人間にとっては見たり聴いたりといった五感で感じることができるメディアではない。Timer メディアはまさにカウンタを持っており、プレゼンテーション値が1であるかぎりカウントを行う。0のときはそこで一時中断し、再び1となれば中断したときの値からカウントを再開する。Timer メディアで設定される時間はカウンタの設定値を意味する。例えば、 $1/n$ 秒ごとに1つつ増えるカウンタの場合、1秒の間隔を実現するためには設定値を n にすればよい。 n 回のカウントを終えるとメディア値は1から0に変化する。

5 論理式による同期関係の表現方法

メディア間の同期には様々なパターンが考えられるが、実際にはいくつかの基本的なパターンを組み合わせることにより表現できる。すなわち、複雑な同期関係は基本的な同期関係の組み合わせである。ここでは、基本的な同期の種類とそれに対応するプレゼンテーション式について説明を行う。

5.1 時間的同期関係の基本要素

メディア間の関係は次の2種類の同期関係の組み合わせであらわすことができる [9]。

1. 連続開始型 (sequential synchronization)

これは、1つのメディアが終了した直後に別のメディアが開始する。すなわち、2つ以上のメディアが順次プレゼンテーションされる。

2. 同時開始型 (simultaneous synchronization)

これは、2つ以上のメディアが同時に開始または終了する。すなわち、2つ以上のメディアが同時にプレゼンテーションされる。パラレル同期 (parallel synchronization) ともいう。

5.2 基本的な同期に関する論理式の記述例

それぞれの同期について、本方式ではどのように表現されるかを以下に示す。

1. 連続開始型をあらわすプレゼンテーション式

2つのメディアオブジェクト M_1, M_2 があり、各々のメディア値を m_1, m_2 、プレゼンテーション値を p_1, p_2 とすると、以下のように定めれば良い。

$$\begin{aligned} p_1 &= m_1 \\ p_2 &= \text{not } m_1 \text{ and } m_2 \end{aligned}$$

はじめ m_1 の値は1であるから、 $p_1 = 1$ となり、 M_1 は表示を始める。一方、 M_2 のプレゼンテーション値 p_2 は0となるため、 M_2 は表示されない。時間が経ち、 M_1 の表示が終了すると m_1 は1から0に変わるため、 $p_1 = 0$ となり、 M_1 は表示をやめる。一方、 p_2 の値は1となるので、 M_2 は表示を開始する。このようにして、1つのメディアの終了とともに次のメディアが開始することができる。

n 個のメディアオブジェクトが順にプレゼンテーションされる場合、 i 番目のメディアオブジェクト M_i のプレゼンテーション式 p_i は、一般的に次のようにあらわせる。

$$p_i = \text{not } m_1 \text{ and } \dots \text{not } m_{i-1} \text{ and } m_i$$

2. 同時開始型をあらわすプレゼンテーション式

連続開始型の例と同じく、2つのメディアオブジェクトを M_1, M_2 とする。

同時開始型には、その終了方法により、2通りの条件式が存在する。

1. プレゼンテーション条件が

$$p_1 = m_1 \text{ or } m_2$$

$$p_2 = m_1 \text{ or } m_2$$

のとき。

プレゼンテーション値 p_1 が 1 となり、 p_2 も同時に 1 をとるので、'M1' と 'M2' は同時に表示を開始する。時間が経ち、両方の表示が終わった時点で両者の表示処理を終了する。このタイプの同期では、一方の表示処理が行われているときは必ず他方も表示処理が行われている。すなわち、同時に複数のメディアのプレゼンテーションを行える。

メディア値が 0 だがプレゼンテーション値が 1 である場合は、メディアによって異なる解釈が与えられる。例えば、動画像なら最後のシーンを表示し続ける処理が考えられる。

2. プレゼンテーション条件が

$$p_1 = m_1 \text{ and } m_2$$

$$p_2 = m_1 \text{ and } m_2$$

のとき。

'M1' の表示が開始されるとプレゼンテーション値 p_1 は 1 となり、 p_2 も同時に 1 をとる。したがって、同時に表示を開始する。しかし、一方のメディアの表示が終了し、メディア値が 1 から 0 に変化すると、両方のメディアのプレゼンテーション値が 0 となり、同時に表示を終了する。

5.3 ユーザとのインタラクションの記述

次に、ユーザ入力に起因した同期に関して述べる。ここでは、例として、入力メディアをマウスのクリックと想定して説明する。

マウスのクリックを行うためには、人の目に見える部分が必要である。これをボタンと呼ぶ。ボタンの役割を果たすメディアとしてビットマップ図形を考え、このビットマップ図形をアイコンと呼ぶ。ユーザの入力を受け付ける時期を、アイコンがディスプレイ上に表示されることでユーザに認識させ、入力不可能なときはディスプレイ上からアイコンが

消えるという方法でユーザとのインタラクションが行われる。

最初にアイコンが表示されていて、それをクリックするとアイコンが消え、ビデオやナレーションなどが開始する場合について考える。

ユーザからの入力がない場合、マウスのメディア値は 1 であり、マウスのクリックにより値は 0 に変化する。

マウス入力をメディア 'M1'、アイコン (ビットマップ図形) をメディア 'M2' とすると、次のような条件式にすればよい。

$$p_1 = m_1$$

$$p_2 = m_1 \text{ and } m_2$$

m_1 は最初、1 であるため、 p_1 の値は 1 となる。このとき、入力可能な状態だがユーザの入力はまだない。また、 m_2 は最初、1 であるため、 p_2 の値は 1 となる。したがって、アイコンが表示される。

ユーザがマウスをクリックすると、メディア値 m_1 が 1 から 0 に変わる。 p_1 は 1 から 0 に変化し、'M1' は入力不可能な状態となる。よって、マウスの機能が働かなくなり、ユーザの入力が許されなくなる。それにともない、 m_1 の値の変化により p_2 の値は 1 から 0 に変わり、アイコンの表示は行われなくなる。

5.4 Timer メディアを用いた同期

Timer メディアを T とあらわすことにする。このメディアは、ある定められた時間 τ を経過するとメディア値 t が 1 から 0 に変化する。しかし、何の表示も行われぬ。Timer メディアは次のような使い方が考えられる。

1. ある時間 τ を過ぎると t が 1 から 0 に変わることを利用して、時間要素のないメディアに表示期間を与えることができる。

たとえば、静止画メディア M のプレゼンテーション条件 p は、以下のようにあらわせる。

$$p = t$$

2. 同期のタイミングのある短い時間ずらすことができる。

たとえば、動画 M_1 と音声 M_2 (メディア値をそれぞれ m_1, m_2 とする) は同時にプレゼンテ-

シオンされるが、実際は音声は動画の表示より τ 秒遅らせたい場合、Timer メディアを用いて以下のようにプレゼンテーション条件を定めればよい。

$$\begin{aligned}
 p_t &= t \\
 p_1 &= m_1 \\
 p_2 &= \text{not } t \text{ and } m_2
 \end{aligned}$$

p_1, p_2, p_t はそれぞれ M_1, M_2, T のプレゼンテーション値である。

6 マルチメディア動物図鑑の作成

我々が提案した同期モデルに基づいてプレゼンテーションを行うプロトタイプシステムを作成した。このシステムでは、テキスト、静止画、動画、音声メディアを使用することができ、マウスを用いたユーザとの対話的な処理も可能である。また、そのアプリケーションとしてマルチメディア動物図鑑を作成した。

本システムは SUN ワークステーションにビデオボードを付加し、NTSC 信号の動画表示を可能としている。動画はレーザーディスクを RS232C を用いてコントロールしている。音声は標準で装備されている音声デバイスを使用した。

動物図鑑のシナリオは、メディアオブジェクトのプレゼンテーション条件式とメディアの属性（表示位置、音量など）から成る。このプロトタイプシステムの基本動作は非常に単純である。まず、作成されたシナリオを読み込み、そのシナリオ中のすべての論理式を評価し、その結果に基づいて各メディアのプレゼンテーションを行っていく。メディア値が変化するたびに論理式の評価を繰り返し、テキストや動画の表示タイミングを制御する。

図2は動物図鑑のシナリオの一部を図示したものである。まず、シーンIでは静止画とテキストが表示され、ユーザがマウスをクリックすると次のシーンに進むことができる。次のシーン（シーン4）では動画とテキストが表示され、ナレーションが流れる。動画とナレーションの両方が終わった時点で次のシーン（シーン5）に移り、シーン4と同様に動画、テキストが表示され、音声流れ出す。

これらのシーンは

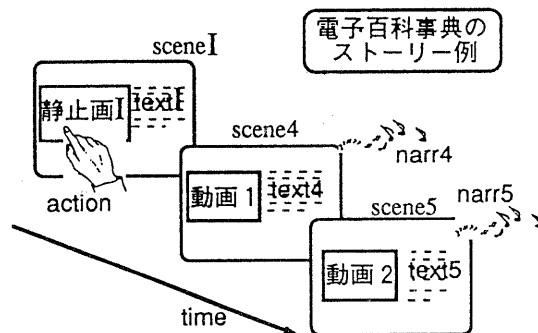


図2: シナリオの例

メディアの名前.メディアの種類 { プレゼンテーション条件式 }

という形式で以下のようにシナリオを表現できる。

```
picI.picture{action-lion and not (voice1
and voice2 and voice3) and action }
```

```
action.mouse{action-lion and not (voice1
and voice2 and voice3) and action }
```

```
textI.text{action-lion and not (voice1
and voice2 and voice3) and action }
```

```
video1.video{not action and (video1
or voice4)}
voice4.voice{not action and (video1
or voice4)}
text4.text{not action and (video1 or
voice4)}
```

```
video2.video{not action and not (video1
or voice4) and (video2 or voice5)}
voice5.voice{not action and not (video1
or voice4) and (video2 or voice5)}
text5.text{not action and not (video1
or voice4) and (video2 or voice5)}
```

7 まとめ

本稿ではマルチメディアデータの共有・交換を行うためにはメディアとメディア間の情報とを分離する必要があるという主張を行った。我々はメディア間の関係の中でも普遍性が最も高い時間的な同期関係に着目し、メディア間の時間的な同期関係のモデルおよびその表現方法を提案した。我々が提案したモデルでは、メディア間の表示タイミングとユーザとのインタラクションを論理式で表現することができる。論理式を導入することにより、いくつかの

基本的な同期関係を簡潔に表現することが可能となり、それらを組み合わせることでもっと複雑な同期関係を記述することができる。

また、この同期表現および同期モデルに基づいたマルチメディアアプリケーションを作成した。

今後は論理式を自動生成することができるオーサリングツールの開発や、同期モデルに基づいたマルチメディアコンピュータの実現が課題となる。

参考文献

- [1] T. D. C. Little, A. Ghafoor, Synchronization and Storage Models for Multimedia Objects, IEEE Journal on Areas in Communications, Vol. 8, No. 3, April 1990
- [2] T. D. C. Little, A. Ghafoor, Network Considerations for Distributed Multimedia Object Composition and Communication, IEEE Network Magazine, November 1990
- [3] International Organization for Standardization, ISO Document No. 8613, ISO, Geneva, Switzerland, Mar. 1988.
- [4] ISO/IEC JTC1/SC2/WG12, Multimedia and Hypermedia information coding Expert Group — Multimedia and Hypermedia Information coded representation, Version 1, June, 1990
- [5] ISO/IEC JTC1/SC18/WG8, Hypermedia Time-based Structuring Language(HyTime), Sixth Draft, 1990
- [6] R. Steinmets, Synchronization Properties in Multimedia Systems, IEEE Journal on Areas in Communications, Vol. 8, No. 3, April 1990
- [7] T. Yoneda, Y. Matsushita, A New Communication Tool: Time Dependent Multimedia Document, IEEE Phoenix 92 to be published.
- [8] C. Nicolaou, An Architecture for Real-Time Multimedia Communication Systems, IEEE Journal on Selected Areas in Communications, Vol.8, No.3, April 1990
- [9] A. Poggio, J. J. Garcia Luna Aceves, E. J. Craighill, D. Moran, L. Aguilar, D. Worthington, J. Hight, CCWS: A computer-based multimedia information system, IEEE Comput., Oct. 1985