

ハイパービデオを利用したビデオ・データベースの意味構造に基づく ブラウジング手法

牛尼 剛聡*

渡邊 豊英†

ushiama@kyushu-id.ac.jp watanabe@nuie.nagoya-u.ac.jp

近年、長時間に渡る活動が記録されたビデオをデータベース化し、利用者が必要とする情報を簡単な操作で効率的に取得可能なアクセス機構に対する要求が高まっている。本論文では、イベント-アクティビティ・モデルを用いて内容記述がなされたビデオを自動的にハイパービデオ化することにより、ビデオの意味内容に基づいて構造的なブラウジングを実現する手法を提案する。イベント-アクティビティ・モデルでは、ビデオの内容をイベント系列として表現するが、本手法では、イベント間に定義される依存関係に基づいてビデオ・フレーム間にリンクを設定する。利用者は、設定されたリンクを選択することにより、観測中のビデオの文脈にもとづいて興味を持った内容を効率的に確認することができる。

A Method for Semantic Content-based Video Database Browsing with Hyper-Video

Taketoshi Ushiana* Toyohide Watanabe†

We have studied on the mechanism that support users to browse video database efficiently based on their interests. In this paper, we introduced an approach for the browsing video databases with hyper-videos that are generated automatically from content descriptions of videos. In our hyper-videos, video frames are linked each other based on their semantic relationships, and end users can traverse these links. This mechanism supports non-linear semantic content-based accesses of video, and users are able to observe interesting frames and scenes in video interactively according to the contexts.

1 はじめに

ビデオは動的な事象を人間が直感的にわかりやすい形式で表現可能であり、対象世界の動的な事象を記録・伝達するために様々な分野で利用されている。近年のビデオ配信技術の進歩により、一般の利用者がインターネットを介して大量のビデオにアクセス可能な環境が整備されつつある。こうした中で、大量のビデオから必要な情報を簡単な操作で効率的に取得するためのアクセス機構に対する期待が高まっている。

ビデオはフレームと呼ばれる静止画像の時系列である。ビデオに記録された活動を利用者が観測するためには、何らかの内容提示操作を実行する必要がある。

内容提示操作としては再生を利用するのが一般的である。しかし、再生による内容提示では、サンプリング周期毎にビデオ・フレームを切り替えて表示する必要があるため、ビデオの内容を効率的に把握することが困難である。この困難さは、大量のビデオを対象とするビデオ・データベースにおいては致命的な欠点となるため、効率的な内容確認方式の実現が重要となる。

従来、効率的な内容確認を実現するための方式として、シーン検索と概略表現を用いたブラウジングが提案されてきた [1]。

シーン検索は利用者が検索要求を検索式として記述し、検索式に指定された条件を満足するシーンを検索結果として返す [2-7]。利用者は検索結果となったシーンを再生することにより、シーンの内容を確認する。シーン検索は、利用者が観測を要求するビデオ中

*九州芸術工科大学芸術工学部
Faculty of Design, Kyushu Institute of Design

†名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

の場面があらかじめ明確であり、そうした場面を同定する条件を検索要求として記述可能である場合には有効である。しかし、場面の意味がビデオの文脈に依存する場合には、検索する以前に利用者がビデオの内容を把握する必要があるため、シーン検索による内容確認が非効率的となる場合がある。また、高度なシーン検索を実現する検索式を記述するためには、検索式の構文やデータ構造等の専門的な知識を必要とするため、専門的な知識を持たない利用者の利用が困難である。

概略表現によるブラウジングでは、ビデオ中の特徴的なフレームのサムネイル画像を一覧性の高い形式で利用者に提示したり、ビデオを再生する際の表示速度を動的に変化させることにより、利用者の効率的な内容確認を実現する [8-14]。ブラウジングによる内容確認はシーン検索に比べて操作が簡単であり、文脈を把握するのにも適している。しかし、これらの方法では、利用者の視点を十分に反映することが困難であるため、利用者の興味をもたない冗長な部分が含まれてしまうことが多い。したがって、利用者が興味のある場面を発見するのが困難となることや、利用者が観測したい内容が省略されてしまうため対象とする場面が観測不可能となることがある。

現在、我々は内容記述（インデキシング）がなされたビデオを自動的にハイパービデオ化することにより、利用者の興味を反映した効率的な内容確認を実現するための研究を行っている。本研究におけるハイパービデオは、ビデオ中の場面間の意味的な関係に基づいてリンクが存在し、ビデオを非線型的にアクセスする機構を提供する。ハイパービデオでは、あらかじめ利用者がビデオの具体的な内容を知らない場合であっても、観測中の文脈に応じて、利用者は各自の興味に応じた場面を選択して対話的に内容を観測することができる。本稿ではビデオの内容記述に基づいてデータベース内のビデオをハイパービデオ化し、構造的なアクセスを提供するハイパービデオ・モデルを提案する。本モデルの位置づけを図1に示す。

なお、本研究ではスポーツ中継などのように、類似した事象が繰り返し発生する対象を記録したビデオを対象とし、本論文では野球の試合を記録したビデオを例として説明する。

2 アプローチ

2.1 ビデオの意味内容とモデル化

ビデオをブラウジングして内容を効率的に把握するためには、ビデオの意味内容を反映したアクセスが必

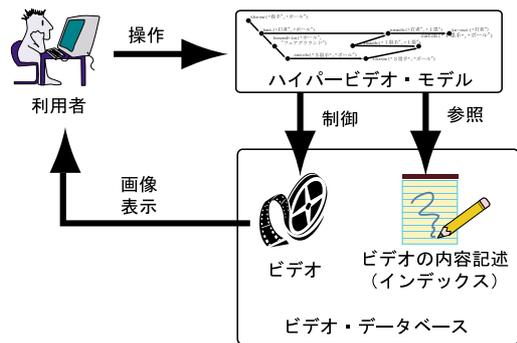


図 1: ハイパービデオ・モデルの位置づけ

要である。こうしたアクセスを提供する機構を実現するためには、ビデオの意味内容を適切にモデル化しなければならない。従来のブラウジング手法の多くが利用している意味内容は、画像処理や音声処理により抽出した特徴量から自動的に導出可能なものに限定されている。すなわち、これらの手法ではボトムアップ的にビデオの意味内容モデルを設計している。人手によってビデオの意味内容を記述するためには多大な労力を必要とすることから、このようなボトムアップ的なアプローチは実用に向けた現実的なアプローチである。しかし、現在の画像・音声処理技術では高次の意味内容を抽出することができないため、従来のビデオの意味内容モデルでは利用者の意図を十分に反映することができない。今後、ビデオ・データベースに対するブラウジングの要求はますます高度化することが予想され、そうした要求を満足するためには高次の意味内容を適切に表現可能な一般化された内容記述モデルが必要である。

ビデオは対象世界の多種多様な活動を記録している。我々はビデオに記録された対象世界の活動をモデル化するためのイベント-アクティビティ・モデル [7] を提案している。本モデルでは、対象世界のスナップショットを物理的または概念的な実体の集合とし、動的な活動を実体の特性の特徴的な変化および変化の過程としてとらえるアプローチに基づいている。イベント-アクティビティ・モデルでは、実体の特性の変化、および実体の特性を変化させる実体間の相互作用をイベントとして表す。また、実体の特性が変化する過程をイベント系列として表し、これをアクティビティと呼ぶ。イベントとアクティビティは直交する概念ではなく、アクティビティとはイベントの並びを解釈した抽象的な概念である。ビデオの内容全体を表すアクティビティをコンテキストと呼ぶ。

表 1: コンテキストの例

発生順	イベント
1	play("試合 1")
2	is-set("ストライク・カウント", "0")
3	is-set("ボール・カウント", "0")
4	is-set("アウト・カウント", "0")
5	is-set("イニング", "1")
6	is-batter("打者")
7	throw("投手", "ボール")
8	catch("捕手", "ボール")
9	is-set("ボール・カウント", "1")
10	throw("捕手", "ボール")
11	catch("投手", "ボール")
12	throw("投手", "ボール")
13	catch("捕手", "ボール")
14	is-set("ストライク・カウント", "1")
15	throw("捕手", "ボール")
16	catch("投手", "ボール")
17	throw("投手", "ボール")
18	bit("打者", "ボール")
19	bound-in("ボール", "フェアグラウンド")
20	reach("1 塁手", "1 塁")
21	catch("3 塁手", "ボール")
22	throw("3 塁手", "ボール")
23	catch("1 塁手", "ボール")
24	reach("打者", "1 塁")
25	is-out("打者")

本論文ではイベントを

イベントの種類(実体1, 実体2, ..., 実体*n*)

と表記する。たとえば、打者 A がボールを打ったことを表すイベントを bat("A", "ball") と表記する。ここで、"A", "ball" はそれぞれ実体「選手 A」「ボール」を表す。野球の試合におけるコンテキストの例を表 1 に示す。表 1 に示したコンテキストは試合開始後先頭打者が 3 塁ゴロでアウトになる場面を表している。

ビデオの内容記述を行う際には、イベントは発生した順に順序付けられ、発生した時刻を記録したフレームに対応づけられる。

2.2 ハイパービデオとブラウジング

ブラウジングとは、提示情報の部分的な省略、提示情報の詳細度の低減、構造的な情報提示等により効率的に情報を獲得する操作である。

現在、テキストに対する構造的な情報提示によって効率的な内容確認実現するための手法として、ハイパーテキストが広く利用されている。ハイパーテキストでは文書を構成する要素間の意味的な関係にもとづいてリンクが設定され、利用者は目的や興味に応じて非線形的に文書要素の内容を確認することができる。

本論文では、ハイパーテキストと同様に、ビデオを構成する要素間にリンクを設定することによりハイパービデオ化し、ビデオを非線形的にアクセスするための機構を提供する。これにより、利用者の目的や興味に応じてビデオの内容を効率的に把握可能なブラウジングを実現する。

ビデオをハイパービデオ化するにあたって、ビデオを構成する要素としてフレームとシーンを想定する。シーンとはビデオ中で意味的なまとまりを表すフレーム部分系列である。このとき、ビデオの構成要素間に設定可能なリンクは以下の 4 種類である。

1. フレーム フレーム
2. フレーム シーン
3. シーン フレーム
4. シーン シーン

なお、本論文で提案する手法の目的は、長時間にわたる活動を記録したビデオを効率的に把握することであると、異なるビデオの要素間のリンクは考えない。

2.3 意味構造に基づくブラウジング

イベント・アクティビティ・モデルではイベントがフレームに対するインデックスとなり、アクティビティ（イベント系列）がシーンに対するインデックスとなる。したがって、利用者にフレームを提示することにより、提示したフレームに対応付けられたイベントが表す実体の特性の変化や実体間の相互作用に関する情報を提供可能である。また、利用者に対してシーンを提示（再生）することにより、提示したシーンが表す実体の特性が変化する過程に関する情報を提供可能である。なお、ビデオに記録されたイベントは対応付けられたフレームを観測することにより一意に決定できると仮定する。実際に利用者が内容を確認する際には一つのフレームだけでは発生したイベントの種類の特定に曖昧さが生じる場合もあるが、実装においてフレームを表示すると同時にイベントの種類を表す文字列をフレームに付帯して提示することにより上記の

仮定を満足させることができる。

ハイパービデオを用いて効率的なビデオの内容確認を実現するためのアプローチを以下に示す。

- 利用者がビデオの内容を観測するために表示されるフレーム系列をプレゼンテーションと呼ぶ。プレゼンテーションには必ずしもすべてのフレームが含まれている必要はない。たとえば、イベントが発生していないフレームは直前に発生したイベントから次にイベントが発生するまでの活動の過程を表しているが、過程を観測しなかったとしても活動の概略は把握することができる。マンガは過程が省略されていても観測者に動的な事象を伝達可能であることを示す例である。したがって、利用者に対してはイベントが発生した場面を提示し、イベントが発生する過程は利用者が必要に応じて選択的に提示することにより、効率的なブラウジングが実現できる。過程を省略するためには、過程の先頭に位置するイベントと末尾に位置するイベントとをリンク付け、利用者がそれらのリンクにしたがって表示フレームを変更すれば良い。
- ビデオには複数の活動が平行して発生する様子が記録されることが多い。また、一つの活動は複数の部分的な活動を含むのが一般的である。しかし、ビデオに記録された活動に関する情報を取得するために、利用者はビデオに記録された特定の活動にしか興味がない場合がある。したがって、ビデオ中で関連のある部分活動をリンクで結び付けることにより、興味のある一連の部分活動に関する情報のみを取得可能とする。

本論文で提案する手法では、上記のリンクをイベントの依存関係にもとづいて設定する。

3 ハイパービデオ・モデル

3.1 イベント依存関係

イベントは実体間の相互作用及び実体の特性の変化を表し、対象となる実体を引数として持つ。したがって、イベントは引数となる実体間の関係として捕らえることができる。このとき、コンテキストにおける実体間の関係はイベント・シーケンス図を用いて直観的に表現することができる。イベント・シーケンス図の例を図2に示す。イベント・シーケンス図においては、縦方向に実体が配置され、横軸が時間の流れを表す。また、横軸に垂直な直線が関係（イベント）を表し、直線上の黒丸が関係付けられた実体を表す。この例では、野球の試合において三塁ゴロで打者がア

ウトになる部分を表しており、1に示したコンテキストの17から25までに対応する。

いま、実体は自発的に特性を変化させないと仮定する。このとき、同一の実体を引数として持つ複数のイベントに関して、時間的に前に発生したイベントが、時間的に後に発生したイベントを生起させるきっかけとなると考えることができる。こうした性質に基づいてイベントを構造化するために、イベント依存関係を導入する。

【イベント依存関係】 $t(ev_i)$ がイベント ev_i が発生したビデオ上の相対時刻を表し、 $arg(ev_i)$ はイベント ev_i の引数（実体）からなる集合を表すとき、任意の2イベント ev_i, ev_j に対して以下の2条件とともに満足するならば、 ev_j は ev_i に依存するといひ $depend(ev_j, ev_i)$ と表記する。

1. $t(ev_i) < t(ev_j)$ かつ $arg(ev_i) \cap arg(ev_j) \neq \emptyset$ である。
2. $t(ev_i) < t(ev_k) < t(ev_j)$ かつ $arg(ev_i) \cap arg(ev_j) \cap arg(ev_k) \neq \emptyset$ を満足するイベント ev_k が存在しない。

イベント集合 E 上のイベント依存関係はノードがイベントを表し、エッジがイベント依存関係を表す有向グラフ $G = (E, depend)$ として捉えることができる。このグラフをイベント依存グラフと呼ぶ。イベント依存グラフの例を図3に示す。図3に示したイベント依存グラフは図2で示したイベントの依存関係を表している。

3.2 データ構造

本論文で提案するハイパービデオ・モデルでは、イベントの依存関係をリンクとして利用する。イベント依存関係をリンクとして表現したハイパービデオの例を図4に示す。利用者は全てのイベントに興味があるわけではない。注目する実体を限定することにより、観測するイベントの数を減少させることができる。プレゼンテーション中で利用者が注目する実体を視点と呼ぶ。視点が指定されている場合には、イベント依存グラフ上で前後のイベントが一意に決定できる。また、グラフ中の分岐を持つノード上では視点を変更することにより、推移的に依存するイベントを確認することもできる。図中の矩形がフレームを表し、矩形内の文字列が対応づけられたイベントを表す。矩形間の矢印がリンクを表し、矢印に対応づけられたラベルが視点を表す。

我々が提案するハイパービデオ・モデルのデータ構

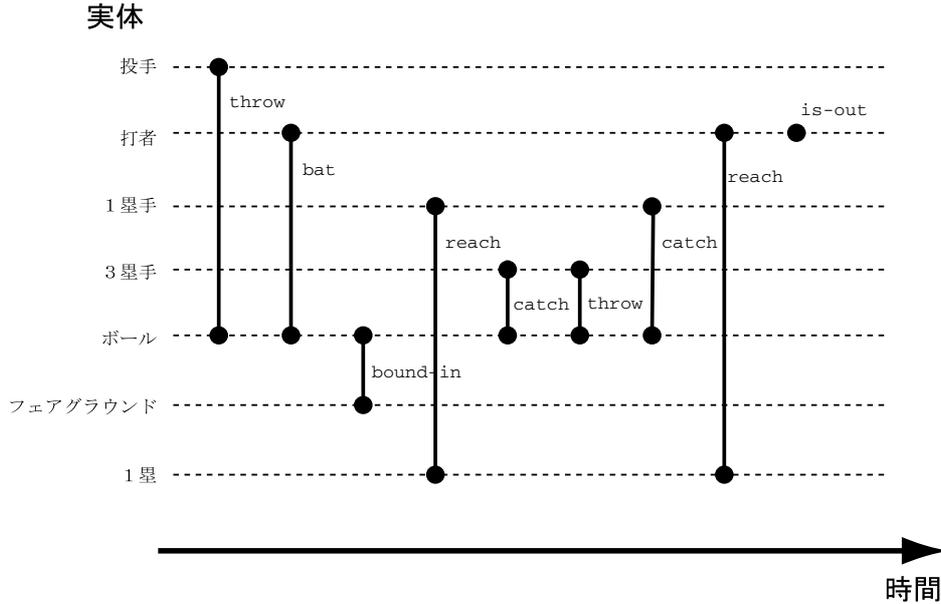


図 2: イベント・シーケンス図の例

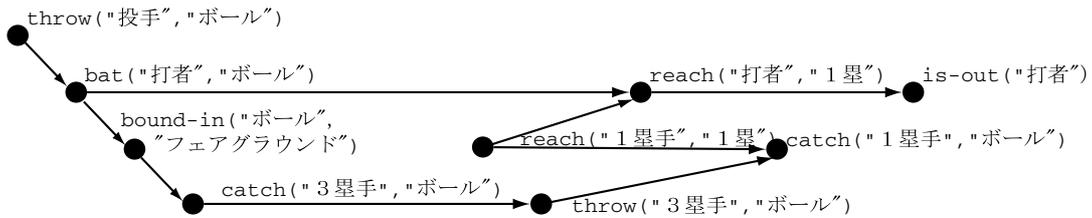


図 3: イベント依存グラフの例

造を6ヶ組 $D = (F, G, \phi, S_{ev}, S_{view}, cur)$ として定義する。ここで、 F はビデオ・フレームの全順序集合、 G はイベント依存グラフ、 ϕ はイベントからフレームへの写像、 S_{ev} はイベントを要素とし観測したイベントの履歴を保持するスタック、 S_{view} は実体を要素とし視点変化の履歴を保持するスタック、 cur は現在のイベントを指す変数である。

スタック S_{ev} と変数 cur は、シーンを再生するために利用される。スタック S_{view} は視点を変更して内容確認した後、元の視点に復帰して内容確認を継続する際に利用する。

なお、イベント依存グラフ上のノード ev に対する親ノードのうち、引数に実体 en を含むものを $parent(ev, en)$ として表記し、 ev に対する子ノードのうち、引数に実体 en を含むものを $child(ev, en)$ と表記する。

3.3 基本的な操作

ブラウジング操作を導入に必要な基本的な操作を定義する。

【フレームの表示:display(ev)】 イベント ev に対応づけられたフレーム $\phi(ev)$ の内容を利用者に提示する。

【シーンの再生:play(ev_1, ev_2)】 2個のイベント ev_1, ev_2 のそれぞれが対応づけられるフレーム $\phi(ev_1), \phi(ev_2)$ で挟まれた部分フレーム系列(シーン)を再生する。

【スタックに対する操作】 スタック S に対して、先頭の要素を参照する $top(S)$ 、先頭に要素 e を挿入する $push(S, e)$ 、先頭の要素を削除する $pop(S)$ の操作が定義されている。

3.4 ブラウジング操作

上記のデータ構造に対してブラウジングに利用する操作を導入する。本モデルで提供する操作は全部で6

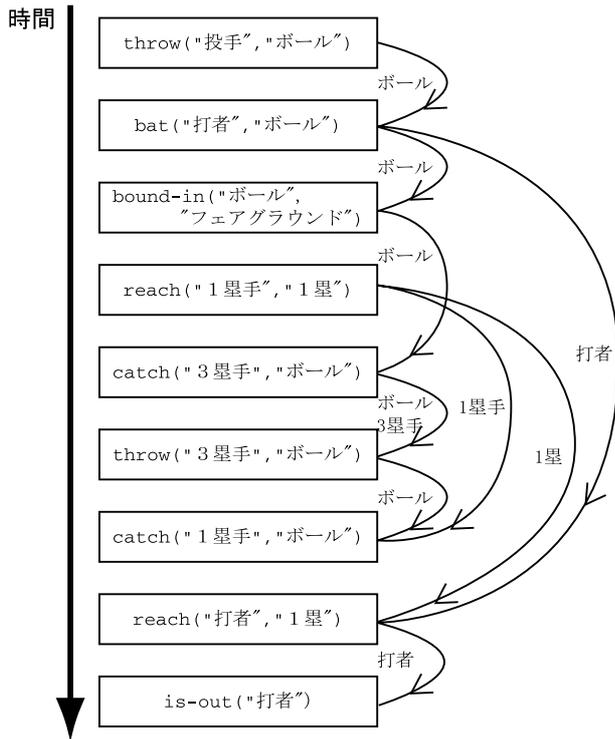


図 4: ハイパービデオの例

種類である。それらは、視点に関する操作、リンクをたどる操作、シーンを表示する操作の3カテゴリに分類できる。以下に、それぞれの操作の定義を示す。

3.4.1 視点に関する操作

【視点の変更:change(*en*)】現在の視点から別の実体 *en* に視点を変更する操作である。なお、指定可能な実体は現在観測中のイベントの引数のいずれかである。すなわち、 $en \in \text{args}(cur)$ を満足しなければならない。この操作が呼び出された際には次の操作を実行する：

1. $\text{push}(S_{\text{view}}, en)$
2. $\text{push}(S_{\text{ev}}, cur)$

【視点の復帰:return】現在の視点に変更する直前の視点に移動する操作である。この操作が呼び出された際には次の操作を実行する：

1. $\text{pop}(S_{\text{view}})$
2. $\text{pop}(S_{\text{ev}})$
3. $cur = \text{top}(S_{\text{ev}})$
4. $\text{display}(cur)$

3.4.2 リンクをたどる操作

【前方向移動:foward】現在観測中のイベントの子となるイベントのうち、現在の視点に一致するイベントに移動する操作である。現在観測中のイベントの結果を確認するために利用する。この操作が呼び出された際には次の操作を実行する：

1. $\text{pop}(S_{\text{ev}})$
2. $\text{push}(S_{\text{ev}}, cur)$
3. $cur = \text{child}(\text{top}(S_{\text{ev}}), \text{top}(S_{\text{view}}))$
4. $\text{display}(cur)$

【後方向移動:backward】現在観測中のイベントの親となるイベントのうち現在の視点に一致するイベントに移動する操作である。現在観測中のイベントの原因を確認するために利用する。この操作が呼び出された際には次の操作を実行する：

1. $\text{pop}(S_{\text{ev}})$
2. $\text{push}(S_{\text{ev}}, cur)$
3. $cur = \text{parent}(\text{top}(S_{\text{ev}}), \text{top}(S_{\text{view}}))$
4. $\text{display}(cur)$

3.4.3 シーンの内容を提示する操作

【通常生成:n-play】現在のイベントから、現在の視点に基づく次のイベントまでを再生する操作である。一般的なシーン再生と同様な機能である。この操作が呼び出された際には次の操作を実行する：

1. $\text{pop}(S_{\text{ev}})$
2. $\text{push}(S_{\text{ev}}, cur)$
3. $cur = \text{child}(\text{top}(S_{\text{ev}}), \text{top}(S_{\text{view}}))$
4. $\text{play}(\text{top}(S_{\text{ev}}), cur)$
5. $\text{display}(cur)$

【確認再生:c-play】直前に観測したイベントと現在のイベントに挟まれた区間を再生する操作である。概略を確認したシーンの内容を具体的に観測するために利用する。この操作が呼び出された際には次の操作を実行する：

1. $\text{play}(\text{top}(S_{\text{ev}}), cur)$
2. $\text{display}(cur)$

4 ユーザ・インタフェース

前説で導入したハイパービデオ・モデルに基づいてビデオ・データベースをブラウジングする際のユー

ザ・インタフェースについて述べる。

オブジェクト間にリンクが設定されたハイパー・メディアの特徴はオブジェクト間のリンクにもとづいて利用者の興味に応じた非線型的なアクセスを提供することにある。したがって、ハイパー・メディアをブラウジングするためのユーザ・インタフェースにおいては、個々の要素の内容を提示するばかりでなく、以下に示す機能を提供することが必要とされる。

1. 利用者が注目するオブジェクトに対して設定されたリンクとその種類を利用者に提示する
2. 提示されたリンクから、利用者が特定のリンクを選択する

たとえば、HTMLで記述されたハイパーテキストでは、文書、単語、画像などの要素から、他の文書、文書中のパラグラフ、画像等にリンクを設定できる。HTMLにおいては基本的に同一の要素に対して複数のリンクを設定することができないため、リンクづけされた要素を指定するだけで一意にリンク先を決定することができる。このようなハイパーテキストをブラウジングするために現在一般的に利用されているWebブラウザでは、リンクが設定されたオブジェクトに対して、アンダーラインを施す、特定の色にする、マウス・ポインタのロールオーバーによりハイライトさせる等の処理を行い、利用者にリンクの存在を提示する。

本論文で提案するハイパービデオ・モデルでは、同一のフレームに対して複数のリンクが存在する。したがって、フレームを指定しただけではリンク先を一意に決定することができない。しかし、視点が指定された場合には順方向および逆方向にたどることができるリンクは一意に決定することができる。このことから、本モデルに基づいたブラウジングを実現するユーザ・インタフェースが利用者に対して提供しなければならない最低限の構成要素は以下ようになる。

1. 特定のフレームやシーンの内容を利用者に提示するための機能
2. 特定のフレームにおいて指定可能な視点を提示し選択するための機能
3. 6種類のブラウジング操作のそれぞれを実行するための機能

特定のフレームにおいて指定可能な視点を提示し選択するためには、視点として指定可能な実体をメニュー形式で提示する手法や、指定可能な実体に対応するフレーム中の領域をマウス・ポインタでクリックする等の手法が考えられる。ただし、後者の手法で



図 5: ブラウジング・ウィンドウの実行例

は、それぞれのフレームにおける実体の領域をあらかじめ取得しておく必要があるため、実現が困難である。

5 システム・プロトタイプの実装

本手法に基づいてビデオ・ブラウジングを実現するシステム・プロトタイプを実装した。本システムは我々が開発中のビデオ・データベース・システム STRIKE(STream data Retrieval system based on Indexing with Key Events)[7]の一部を構成する。開発環境はシリコン・グラフィックス社製ワークステーション O2 であり、IRIX 上で動作する。開発言語としては C++ を用いた。なお、現時点の実装ではシーン再生操作は提供しておらず、今後の課題である。

図 5 に野球の試合を記録したビデオに対するシステムのブラウジング・ウィンドウの表示例を示す。ウィンドウの左側はフレームとシーンを提示する部分である。スライド・バーは提示中のフレームを表している。上向きの三角形は視点の復帰を表し、下向きの三角形は視点の変更を表す。変更する実体は右下のメニューにより指定する。右向き矢印は前方向移動を表し、左向き矢印は後方向移動を表す。

6 おわりに

本論文ではイベント・アクティビティ・モデルに基づいて内容記述がなされたビデオをイベント依存関係に基づいてハイパービデオ化し、簡単な操作でビデオに記録された活動に対する情報を効率的に取得するための機構を提案した。

本論文で提案したイベントの依存関係に基づくリンクは、粒度が細かいために利用者の興味によっては効率的な内容確認が実現できない場合がある。この問題を解決するために、イベント依存関係の他にブラウジングの際に利用可能なリンクを導入する必要がある。今後、イベントの出現確率を利用した意味関係、

および時間概念の包含関係を導入し、それらを利用してさらに効率的なブラウジングの実現を目指す。また、実験により本論文で提案した機構の有効性を検証する予定である。

謝辞

本研究の一部は科学研究費奨励研究(A) (課題番号12780235)、科学研究費基盤研究(B)(2) (課題番号10558053)による。ここに記して謝意を表します。

参考文献

- [1] R. M. Bolle, B. Yeo and M. M. Yeung: “Video Query: Beyond the keywords”, Technical Report 20586, IBM (1996).
- [2] E. Oomoto and K. Tanaka: “OVID: Design and Implementation of a Video-Object Database System”, *IEEE Trans. on Knowledge and Data Engineering*, Vol. 5, No. 4, pp. 626–643 (1993).
- [3] A. Duda, R. Weiss and D. K. Gifford: “Content-Based Access to Algebraic Video”, *Proc. of IEEE Int’l conf. on Multimedia Computing and Systems*, pp. 140–151 (1994).
- [4] 田淵仁浩, 村岡洋一: “動画像データベース中の系列データを指定する条件の不完全さを許容できる問い合わせ処理と MeSOD モデル”, 信学論 (D-I), Vol. J76-D-I, No. 6, pp. 288–299 (1993).
- [5] C. Declair and M. Hacid: “A Database Approach for Modeling and Querying Video Data”, *Proc. of Int’l conf. on Data Engineering*, pp. 6–13 (1999).
- [6] J. Koh, C. Lee and A. Chen: “Semantic Video Model for Content-based Retrieval”, *Proc. of IEEE Int’l conf. on Multimedia Computing and Systems*, Vol. II, pp. 158–162 (1999).
- [7] 牛尼剛聡, 広部一弥, 渡邊豊英: “利用者の視点に基づくシーン検索のためのイベント-アクティビティ・モデル”, 信学論 (D-I), Vol. J82-D-I, No. 1, pp. 256–267 (1999).
- [8] S. Srinivasan, D. Ponceleon and D. Petkovic A. Amir: ““What is in that video anyway?”: In Search of Better Browsing”, *Proc. of IEEE Int’l conf. on Multimedia Computing and Systems*, Vol. I, pp. 388–393 (1999).
- [9] M. Yeung, B. Yeo and B. Liu: “Extracting Story Units form Long Program for Video Browsing and Navigation”, *Proc. of IEEE Int’l conf. on Multimedia Computing and Systems*, pp. 296–305 (1996).
- [10] R. Lienhart, S. Pfeiffer and W. Effelsberg: “Video Abstracting”, *Comm. of ACM*, Vol. 40, No. 12, pp. 55–62 (1997).
- [11] 是津耕司, 上原邦彦, 田中克己: “映像の意味的構造の発見と動的 skimming への応用”, 第9回データ工学ワークショップ論文集 (1998).
- [12] X. Sun, M.S. Kankanhalli, Y. Zhu and J. Wu: “Content-Based Representative Frame Extraction for Digital Video”, *Proc. of IEEE Int’l conf. on Multimedia Computing and Systems*, pp. 190–193 (1998).
- [13] 橋本隆子, 白田由香里, 真野博子, 飯沢篤志: “TV受信端末におけるダイジェスト視聴システム”, 情報処理学会論文誌, Vol. 41, No. SIG3(TOD6), pp. 71–84 (2000).
- [14] M. G. Christel, M. A. Smith and C. R. Taylor D. B. Winkler: “Evolving Video Skims into Useful Multimedia Abstracts”, *Proc. of ACM CHI*, pp. 171–178 (1998).