

意味内容に基づくビデオ・データベース・ブラウジングのための ハイパービデオ・モデル*

牛尼 剛聡†

九州芸術工科大学 芸術工学部‡

渡邊 豊英†

名古屋大学大学院 工学研究科§

1 はじめに

ビデオは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。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るという特長がある。ビデオの内容を観測するためのプレゼンテーションに方式としては再生が行われるのが一般的である。しかし、再生では、ビデオを構成する静止画像（フレーム）をサンプリング周期毎に1画面ずつ表示する必要があるために、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開発が期待されている。

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。シーン検索では、利用者が観測したいビデオ中の場面があらかじめ明確であり、検索要求としてそうした場面を同定する条件を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、シーン検索での内容確認が効率的でない場合がある。一方、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一覧を利用者に提示する。しかし、この方法では、利用者の視点が考慮されていないため、利用者の興味とは異なった冗長な部分が多く含まれているために興味のある場面を発見するのが困難となる場合や、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である場合がある。

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。

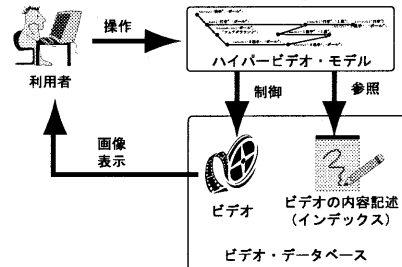


図1: ハイパービデオ・モデルの位置づけ

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 内容記述

ビデオの内容記述はイベント-アクティビティ・モデル [1] に基づいて行う。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。イベントはビデオ中の物理的 / 概念的な実体の特性の変化、および実体の特性を変化させる実体間の相互作用を表す。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 hit("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。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 イベントの構造化

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、同一の実体を引数を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引き起こす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】任意の2イベント ev_i, ev_j に対して、 $t(ev_i) < t(ev_j)$ かつ $\arg(e_i) \cap \arg(e_j) \neq \emptyset$ ならば、 $\text{depend}(ev_j, ev_i)$ である。なお、ここで、 $t(ev_i)$ はイベント 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 e_i の引数（実体）からなる集合を表す。

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 $G = (E, \text{depend})$ 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2に示す。この例では、野球の試合においてサードゴロ

* A Hyper-Video Model for Semantic Content-based Browsing on Video Database

† Taketoshi USHIAMA and Toyohide WATANABE

‡ Faculty of Design, Kyushu Institute of Design

§ Graduate School of Engineering, Nagoya University

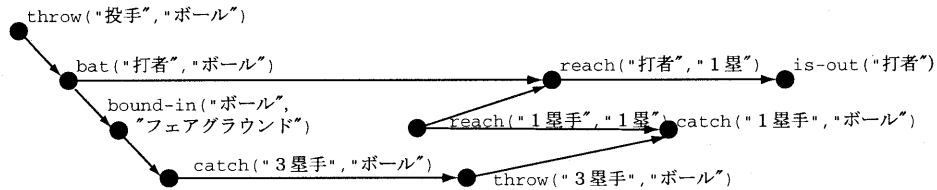


図 2: イベント依存グラフの例

で打者がアウトになる部分を表している。

2.3 ビデオのプレゼンテーション

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過程を表しているが、過程を観測しなかったとしても大雑把な内容は把握することができる。したがって、基本的に利用者に対してはイベントが発生した場面を提示し、利用者が必要に応じてイベントが発生する過程を観測可能とすることにより、効率的なブラウジングが実現できる。

また、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、遷移的に依存するイベントを確認することもできる。

3 ハイパービデオ・モデル

3.1 データ構造

我々が提案するハイパービデオ・モデルのデータ構造を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 は因果グラフ、 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。

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。

3.2 基本的な操作

ブラウジング操作を導入に必要となる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 $\phi(ev)$ を利用者に提示する。

【シーンの再生:play(ev_1, ev_2)】 2 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。

3.3 ブラウジング操作

上記のデータ構造に対して以下の操作を定義する。

【前方向移動:forward】 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。この操作が呼び出された際には次の操作を実行する：1) $pop(S_{ev})$, 2) $push(S_{ev}, cur)$, 3) $cur = child(top(S_{ev}), top(S_{view}))$, 4) $display(cur)$ 。

【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。この操作が呼び出された際には次の操作を実行する：1) $pop(S_{ev})$, 2) $push(S_{ev}, cur)$, 3) $cur = parent(top(S_{ev}), top(S_{view}))$, 4) $display(cur)$ 。

【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。この操作が呼び出された際には次の操作を実行する：1) $pop(S_{ev})$, 2) $push(S_{ev}, cur)$, 3) $cur = child(top(S_{ev}), top(S_{view}))$, 4) $play(top(S_{ev}), cur)$, 5) $display(cur)$ 。

【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。この操作が呼び出された際には次の操作を実行する：1) $play(top(S_{ev}), cur)$, 2) $display(cur)$ 。

【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。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、 $en \in args(cur)$ 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) $push(S_{view}, en)$, 2) $push(S_{ev}, cur)$ 。

【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。この操作が呼び出された際には次の操作を実行する：1) $pop(S_{view})$, 2) $pop(S_{ev})$, 3) $cur = top(S_{ev})$, 5) $display(cur)$ 。

4 おわりに

今後、イベントの出現確率を利用した意味関係の導入と、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、本モデルを実装し、有効性を検証する。

参考文献

- [1] 牛尾剛聡, 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).