

断片的な内容記述と映像の連続性にもとづいた ビデオ映像のシーン検索

是津耕司† 上原邦昭†† 田中克己†††

†通信・放送機構 神戸リサーチセンター

E-mail: zetsu@kobe-sc.tao.or.jp

††神戸大学都市安全研究センター

E-mail: uehara@kobe-u.ac.jp

†††神戸大学大学院自然科学研究科

E-mail: tanaka@in.kobe-u.ac.jp

時間的な連続性を持った映像に対する意味づけは、どの部分の映像をどの順序で見たかによって変わる。したがって、映像の内容記述をもとに、意味を持った一連の映像部分、すなわちシーンを検索するためには、欲しいシーンの意味的な内容を表現できる一連の連続した映像を見つけ出す必要がある。そこで本論文では、映画言語によって表される映像の構造的な連続性と、時刻印付オーサリンググラフによって表される内容の意味的な関連性を併せて用いることによって、意味的なまとまりを持った連続する映像としてシーンを検索する方法について述べる。

A Video Scene Retrieval Method based on Fragmentary Descriptions and Continuity of Video Contents

Kohji Zetsu† Kuniaki Uehara†† Katsumi Tanaka†††

†Kobe Research Center, Telecommunications Advancement Organization of Japan

E-mail: zetsu@kobe-sc.tao.or.jp

††Research Center for Urban Safety and Security, Kobe University

E-mail: uehara@kobe-u.ac.jp

†††Graduate School of Science and Technology, Kobe University

E-mail: tanaka@in.kobe-u.ac.jp

A video is a continuous image stream. Thus, the meanings for a part of a video depends on the contents and alignment of video images. It is important to identify the sequence of video images which has continuity of both semantics and syntax, in order to retrieve the scene. This paper describes how to retrieve a scene by semantic relationships and syntactic continuity of the video using a time-stamped authoring graph and cinematography or film editing.

1 はじめに

ビデオ映像を内容にもとづいて検索するためには、各々の映像の内容記述をインデックスとして付与する必要がある。従来の内容記述方式では、映像を事前に適当な時間区間に区切り、それぞれの区間の映像の内容を厳密に記述してた。[2][3][6]しかしこの方法では、内容に関する様々な検索要求に対応できるよう、時間区間を適切に区切り厳密な内容記

述を与えることが非常に困難である。これに対し、本研究で行なっている時刻印付オーサリンググラフ(=時刻印付ノードリンクグラフ) [1]による映像記述では、ある時点の映像に対する断片的な内容記述とそれらの間の意味的な関連性をもとに映像の内容を記述するため、従来の記述方式のような記述の厳密さととらわれることのない、柔軟な記述を行なうことができる。

時刻印付オーサリンググラフにもとづいた映像検

索では、映像に対する記述グラフの中から、検索要求としてあたえられた文を最も良く表現できる最小の部分グラフを見つけ出し、この部分グラフに対応する映像を選択し検索結果とする。これまでは、部分グラフに対応する映像として、部分グラフに含まれる各ノードに付けられた時刻をもとに、この時刻に上映される映像カットを全てのノードについて集めたカット集合を使用していた。しかし、こうして作られたカット集合の中の映像は互いに連続していないことが多く、検索結果として利用しにくいものであった。検索結果として意味のあるまとまった映像を得るため、適切な評価基準にもとづいて連続した映像を選択することが必要であった。

そこで本論文では、映像の構成に関わる特徴、すなわち映像操作や編集方法、画面構成やカットつながりから構造的な連続性を発見し、内容の意味的な関連性にもとづいてこれらの連続性を持った映像をつなぎ合わせることにより、意味的なまとまりをもった一連の映像、すなわちシーンを検索することを行なう。そのために、映画言語にもとづいた映像の構造的な連続性の発見と、時刻印付オーサリンググラフによる内容の意味的な関連性の表現を融合したシーンの検索方法について述べる。

2 連続性を持った映像への意味づけ

映像は連続性を持った情報メディアである。そのため、同じ映像でも、どの部分をどの順序で見たかによってその意味づけが変わってくる [7]。したがって、映像の内容記述を行なう際は、この映像の連続性による意味づけの変化を考慮に入れる必要がある。以下に、映像の連続性と意味づけの関係について述べる。

2.1 映像の構造的な連続性

カットは、映像を構成する時間的連続性を持った最小の単位である。カットは、連続して撮影された映像(ショット)から編集上の都合により切り出された映像部分であり、それ自身だけでは断片的な内容しか持たない。しかし、カットを注意深くつなぎ合わせることによって、あるまとまった意味づけを持つ一連の映像、すなわちシーンを作り出すことができる。あるシーンが、どのようなカット映像をど

のようにつなぎ合わせて構成されるか(映像のモニタージュ)は、そのシーンが表現しようとする内容によって異なる。映像を使って表現される様々な内容と映像の構成の間にどのような関係があるのかについては、映画言語 [8][9] の研究として行なわれている。

映画言語にもとづいて映像のモニタージュを分析することにより、個々のカット映像がどのようにつながっているのかが分かる。そこから、一連の映像として構成されている部分(シーケンス)を見つけ出すことができる。シーケンスは映像の構造的な連続性をもった映像部分であり、その部分の内容だけに関する意味づけを持つ。各々のシーケンスは、ある意味づけを持ったシーンに組み込まれることによって、初めてシーン全体のなかでの具体的な意味づけを持つことができる。

2.2 内容の意味的な関連性にもとづいたシーンの構成

シーンは、あるまとまった意味を持つように一連の映像をつなぎ合わせものである。したがって、ある意味づけを持ったシーンを得るためには、何らかの基準をもとに各々の部分映像が持つ意味をつなぎ合わせて、目的とする意味づけを表現できるような映像部分を全体の映像の中から見つけ出す必要がある。

時刻印付オーサリンググラフでは、ある時点の映像に対する断片的な内容記述をノードとし、それらの間の意味的な関連性をリンクとしたグラフによって映像の内容を表現する。ノードに付けられた断片的な内容記述は、その時点に上映されるあるまとまった映像部分に直接対応付けることができる。また、時刻印付オーサリンググラフでは、記述グラフに含まれる任意の部分グラフを使って、対応する映像部分の意味づけを行なうことができる。したがって、ある意味づけを持ったシーンを得るために、時刻印付オーサリンググラフの部分記述グラフを使って意味づけを表現し、この部分グラフの各ノードに対応する映像部分をリンクにしたがつなぎ合わせることにより、目的とする意味づけを持った一連の映像を得ることができる。すなわち、意味づけにしたがついてシーンを構成することができる。

3 シーンの検索

本節では、映画言語と時刻印付オーサリンググラフを使って実際にシーンを検索する方法について述べる。

3.1 連続した映像部分の発見

映画言語にもとづいて映像の連続性を判断し、連続する映像をまとめる。主な連続性の判断基準を以下に挙げる。

隣接するカットの連続性 隣接するカット間で以下の条件が満たされれば、それらを連続する一つの映像として見なすことができる。

構図の一致 画面内の人物の配置が一致している。

動きの一致 人物もしくはカメラの動きの方向が一致している。

平行編集 2つないし3つの映像が交互に現れる。これによって、それぞれの映像が表す出来事を相互に関連づけ、一つのまとまりのある映像を作り出す。

シーンの区切り フェイド、ディゾルブ、ワイプなどの編集効果は、それまで続いていききたシーンを終え、次のシーン平行することを示唆する。

こうした映画言語による判断基準を適用して映像の連続性を判断するために、以下のような映像の特徴量を使用する。これらの特徴量は、現在のイメージ処理技術などを駆使して自動的に取得することが可能である。[4]

人物の特定

人物の撮影距離 ロング／フル／ミディアム／クローズ・ショット、クローズ・アップ。

人物の画面内での位置 画面を2～3分割し、それぞれの人物がどの領域に属しているかを調べる。

人物の動き 動きの開始位置と終了位置（あるいは動きの範囲）、開始時刻と終了時刻、および動きの方向。

カメラの動き 動きの開始時刻と終了時刻、動きの方向、および動きの種類（パン、ティルト、ズーム、トラベリング、クレーン）。

実際の映像に対し、映画言語による判断基準にもとづいて連続した映像部分を見い出すことを行なった様子を図1に示す。図1の上部は、映像から得られる種々の特徴量を使って、映画言語によりどのような映像の連続性が見い出されたのかを示している。また、図1の下部は、全体の映像（約3分間）に対しどの程度の連続部分が見い出されたのかを示している。このように、ある程度ストーリー性のある映像では、映画言語によって効果的に映像の連続部分を見い出せることが分かる。

3.2 内容記述の検索

時刻印付オーサリンググラフによる映像記述の中から、検索要求に対応する記述部分を検索する。検索要求として欲しいシーンの内容を表現する自然言語文を与えると、この検索文中に含まれる単語（あるいはその類義語）がすべてリンクでつながれた最小の部分記述グラフを探し出す（リンクの重さは全て同じとする）。こうして最小の部分記述グラフが見つければ、このグラフが検索要求を最もよく表現していることになる。

3.3 シーンの抽出

時刻印付オーサリンググラフによる内容記述に対する検索の結果得られた最小部分記述グラフをもとに、対応するシーンを選択する。これは以下の手順で行なう。

1. 最小部分記述グラフに含まれるノードに付けられた時刻に上映される映像部分（カット、もしくは構造的な連続性を持ったシーケンス）を、すべてのノードに渡って選択する。
2. 選択された映像部分をつなぎ合わせて一つの映像にできる場合は、この映像を検索結果とする。もし、一つの映像につなぎ合わせる事ができなければ、この最小部分記述グラフによって意味づけできる一連の映像はないと判断する。

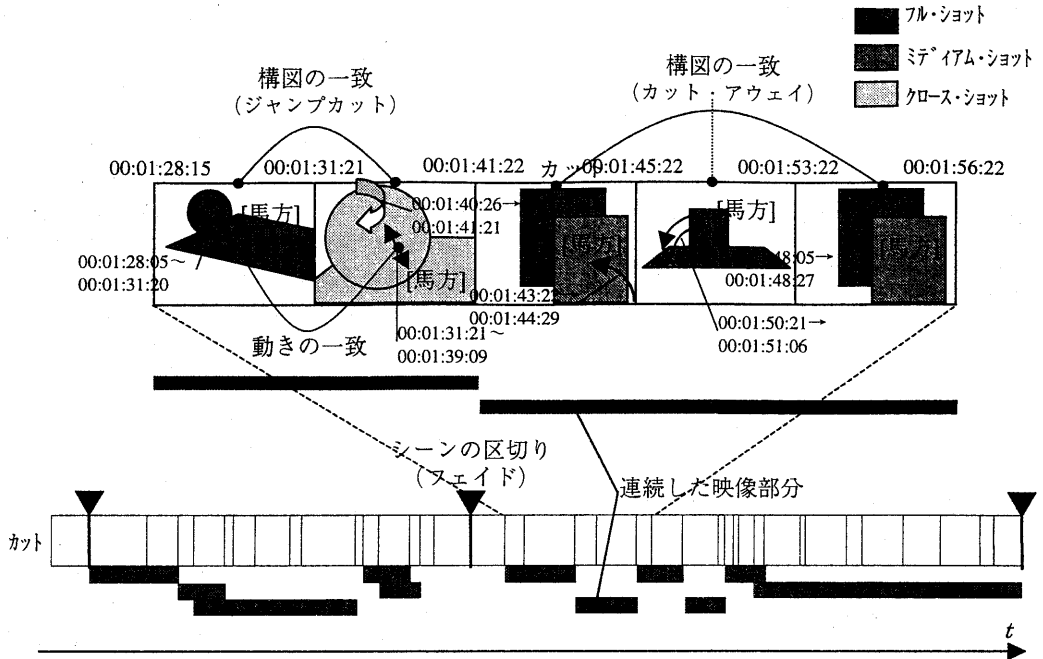


図 1: 映画言語にもとづいた映像の連続部分

3.4 検索例

上記の手順にしたがってシーンの検索を行なった様子を図 2 に示す。図 2 では、“馬方がタヌキに起こされる”シーンを検索している。まず、もとの映像に対し、映画言語にもとづいて構造的な連続性を持った映像部分がまとめられる (図 2 下部)。次に検索文“馬方がタヌキに起こされる”に対し、時刻印付オーサリンググラフによる内容記述の中から、検索文に含まれる単語間の関連性を表現する最小部分記述グラフを見つけ出す (図 2 中、太線で示されたグラフ)。最後にこの最小部分記述グラフに含まれるノードに対応する映像部分をつなぎ合わせる。最小部分記述グラフに含まれるノードに対応する映像部分は 4 つあるが、これらは 1 つにつなぎ合わせることが可能であり、この 1 つにまとめられた映像が検索結果となる (図 2 中の“検索結果の映像”)。

4 今後の課題

映画言語に基づいた映像の構造的な連続性の判断は、まだ明確な基準が十分に定義されておらず、主観的な判断基準に負うところが多い。したがって、

これらの判断基準を明確に定義しルール化することによって、連続性の判断を自動化することが今後の課題である。

5 おわりに

今回は、映像の構造的な連続性と内容記述の意味的な関連性をもとに、意味的なまとまりを持った一連の映像であるシーンを検索することについて述べた。映画言語にもとづき、映像操作や編集方法、画面構成や人物の動きなどの特徴から構造的な連続性をもった映像部分を見つけ出し、時刻印付オーサリンググラフによる内容の意味的な関連性にもとづいてこれらの映像をつなぎ合わせることにより、内容の意味づけに対応するシーンを検索する方法を示した。

謝辞

本研究において、貴重な映像資料の学術利用を許可して下さい、東映株式会社に感謝致します。なお、本研究は一部、文部省科学研究費重点領域研究

検索文：「馬方がタヌキに起こされる」

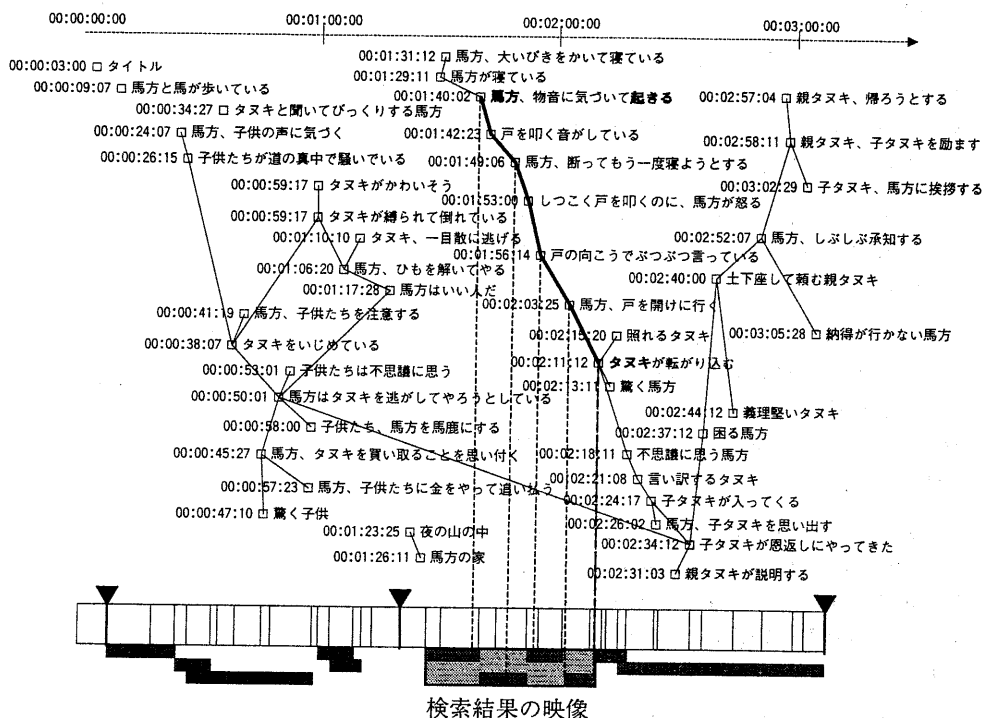


図 2: シーンの検索

(課題番号 08244103) による。

参考文献

- [1] 是津耕司, 上原邦昭, 田中克己, 木邑信夫: 時刻印付ノードリンクグラフによるビデオ映像のデータベース化, 情報処理学会研究会報告, 97-DBS111, pp 73-80 (1996).
- [2] Allen J.F: Maintaining Knowledge about Temporal Intervals, CACM, Vol. 26, pp. 832 - 843 (1983).
- [3] E. Oomoto, K. Tanaka: OVID: Design and Implementation of a Video-Object Database System, IEEE Transactions on Knowledge and Data Engineering, Vol. 5 No. 4, pp. 629 - 643 (1993).
- [4] H. D. Wactlar, T. Kanade, M. A. Smith, S. M. Stevens: Intelligent Access to Digital Video: Informedia Project, Computer theme issue on the US Digital Library Initiative (1996).
- [5] A. D. Bimbo, E. Vicario, D. Zingoni: Symbolic Description and Visual Querying of Image Sequences Using Spatio Temporal Logic, IEEE Transactions on Knowledge and Data Engineering (1996).
- [6] K. Uehara, M. Oe, K. Maehara: Knowledge Representation, Concept Acquisition and Retrieval of Video Data, Proc. of International Symposium on Cooperative Database Systems for Advanced Applications, pp. 527 - 534 (1996).
- [7] 中島義明: 映像の心理学, 新心理学ライブラリ = S1, サイエンス社 (1996).
- [8] ダニエル・アリホン: 映画の文法, 紀伊国屋書店 (1980).
- [9] クリスチャン・メッツ: 映画記号学の諸問題, 水声社 (1987)