

# シーン間関係に基づく動画オーサリング

## Video Authoring Based on the Relation between Scenes

田村和香子\*1 和泉憲明\*2 小林一郎\*1 橋田浩一\*2  
Wakako TAMURA Noriaki IZUMI Ichiro KOBAYASHI Kōiti HASIDA

\*1お茶の水女子大学大学院 人間文化創成科学研究科  
Graduate School of Humanities and Sciences, Ochanomizu University

\*2産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology (AIST)

Recently as movie cameras, digital cameras and low-priced video authoring tools have been getting commonly used, chances of authoring personal movies have increased. However, the quality of a created movie usually depends on creator's editing skill and knowledge. With this background, in this paper, we aim to make amateur possible to make high quality movies. As an approach to this, we propose a scenario-centered video authoring method in which we extract the scenario structure from a movie, construct its network model, and then edit the movie based on the network model.

### 1. はじめに

近年、デジタルカメラや携帯電話の動画撮影機能や Web 上の動画共有サイトなどの普及により、動画編集のニーズは、ますます高まっている。ここで、一般に、映画やドラマ、アニメ作品などの映像コンテンツに関する専門技術は、大別すると、構図や露出などの撮影技術と、物語構造の編纂技術とに分類できる。

現在、一般ユーザが利用可能な動画の編集環境は、主として、「動画を切り貼りする」や「動画の順序を入れ替える」といった操作をコアとしたもので、撮影技術を補完する役割のものが多く、このため、既存の映像コンテンツを再利用して新たなコンテンツを作成する場合は、動画の編集環境を使いこなせることが求められる。同時に、スキル不要のコンテンツ創造を可能にするためには、物語構造の操作のみにより、動画オーサリングが可能となることが不可欠である。

以上の観点から、本研究では、物語構造の操作による動画編集環境の実現を目的として、シナリオ中心の動画編纂手法を提案する。ここでは、セマンティックオーサリングの考え方 [1] に基づき、物語を構成するシーンとシーン間の関係に基づいてシナリオを構造化する。そして、この構造に基づいて映像コンテンツのシーン検索やシナリオ要約などの手法を与える。

### 2. シナリオ構造ネットワーク

一般に動画は、図 1 に示すように、時間の経過や場所の移動で区切られたシーンの集合として表すことができる。本研究では、このシーンを基本単位として動画を階層化し、さらに、シーン間の関係性やシナリオ全体におけるシーンの役割（ロール）を記述することで、シナリオ構造のネットワーク（以下、シナリオ構造ネットワークと呼ぶ）を構築することを提案する。

#### 2.1 シーンの定義と属性

ここでは、シナリオを構造化するために、まず、基本単位となるシーンを定義する。動画の分け方の粒度は、分割を行うユーザによって異なるが、本研究では、主に、動画の持つ時系列情報、または、場所情報の変化にて動画を区切り、区切られ

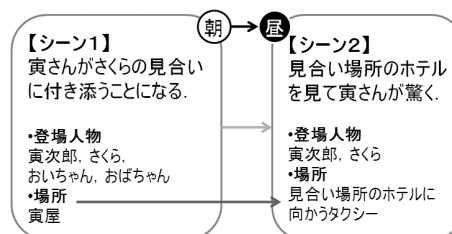


図 1: 動画を構成するシーンの一例

た単位をシーンとする。分割された各シーンの持つ情報は、次の属性により記述する。

- 登場人物  
シーンに登場する人物。  
人以外の動物なども含むが、ただ映り込んでいるだけの通行人など、シーンへの関連度が低いものは除く。
- 場所  
登場人物たちが行動している場所。
- 内容  
登場人物の行動など、そのシーンで起こっている出来事の内容。

例えば、「場所 P で登場人物 A と B が勉強をしている」という内容のシーンの場合、シーンの持つ属性は表 1 のように記述する。

表 1: シーン属性の記述例

シーン：場所 P で登場人物 A と B が勉強をしている

属性	属性値
登場人物	A,B
場所	P
内容	P で A と B が勉強している

#### 2.2 シーンの階層化とシーン間関係

次に、シーンを内容の関連によって分類し、シーンの内容的なグループをイベントとして定義する。ここでは、シーンとイベントの構造を「包含関係」により与える。そして、シーンとイベント間の関係として、「話題転換関係」と「因果関係」、「前後関係」を導入することにより、シナリオのネットワーク構造を与える。ここで、図 2 に具体的なシナリオのネットワーク構造を示すが、その構成は次のとおりである。

連絡先: 田村和香子, お茶の水女子大学大学院人間文化創成学科学専攻, 東京都文京区大塚 2-1-1 小林研究室, 03-5978-5709, E-mail:wakako.t@koba.is.ocha.ac.jp

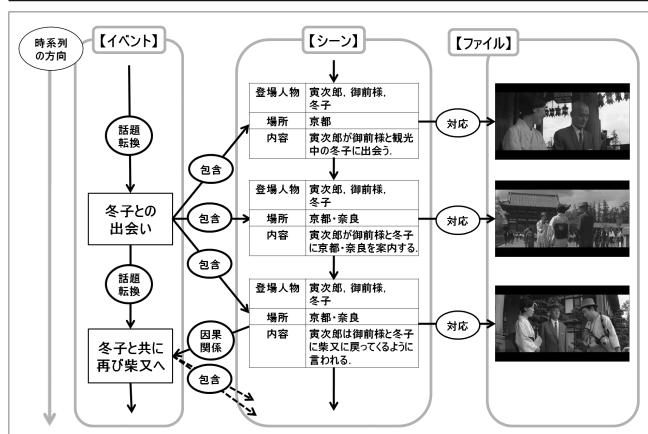


図 2: シーンの階層化とシーン間関係

- イベント  
内容の関連によってシーンを分類したもの。  
「イベント 包含 シーン」の形式で記述することとし、図 2 のような木構造を構成する。例えば、「寅次郎が御前様と観光中の冬子に会う」、「寅次郎が御前様と冬子に京都・奈良を案内する」、「寅次郎は御前様と冬子に柴又に戻ってくるように言われる」というシーンがある場合、このシーン集合を「冬子との出会い」というイベントとして扱う。シーンを階層化することは、扱うシーンの数が多くなればなるほど有効である。
- 動画ファイル  
各シーンに対応した実際の動画ファイル。  
「シーン 対応 動画ファイル」の形式で記述する。図 2 では、動画ファイルのサムネイルを表示した。この動画ファイルは、必ずしも一つのファイルである必要はなく、ファイルの集合である場合もある。
- 話題転換  
イベント イベント間をつなぐ関係。  
前後で話題が変わっている場合に記述する。図 2 の場合、寅次郎が冬子と出会い、数日を京都・奈良で過ごした「冬子との出会い」というイベントと、冬子と共に柴又へ戻ってきた寅次郎と寅屋の人々が再会する「冬子と共に再び柴又へ」というイベントの前後で、話題が変わっている。これを、イベント間の関係として、話題転換として記述する。
- 因果関係  
シーン、イベント間をつなぐ原因・結果の関係。  
シーン間に原因・結果の関係がある場合、「原因のシーン 因果関係 結果のシーン」という形で記述する。シーン シーン間以外に、イベント イベント間、イベント-シーン間にも必要に応じて与える。図 2 では、「寅次郎は御前様と冬子に柴又に戻ってくるように言われる」というシーンが原因で、結果として「冬子と共に再び柴又へ」というイベントが発生している。これを因果関係として記述する。
- 前後関係  
ストーリーの構造を与える時系列の前後関係は、シーンの並びで表現する。図 2 の場合、シーンの順序は上から下に向かっている。但し、回想シーン等、プロットと呼ばれる時系列の前後関係とは異なるシーンの記述は、2.3 節で説明するルールを用いて与える。

### 2.3 ルールの導入

本研究では、イベントまたはシーンがシナリオ中でどのような役割を担っているかを示すものをルールと呼ぶ。

例えば、「寅次郎が御前様と観光中の冬子に会う」という内容のシーンが、シナリオ中で「登場人物: 冬子の紹介」の役割を担っていると判断した場合、図 3 のようにルールを記述する。イベントも同様で、「冬子との出会い」というイベントが、「寅さんがマドンナに恋をする」ということを表す役割を担っていると判断した場合、図 3 のようにルールを記述する。

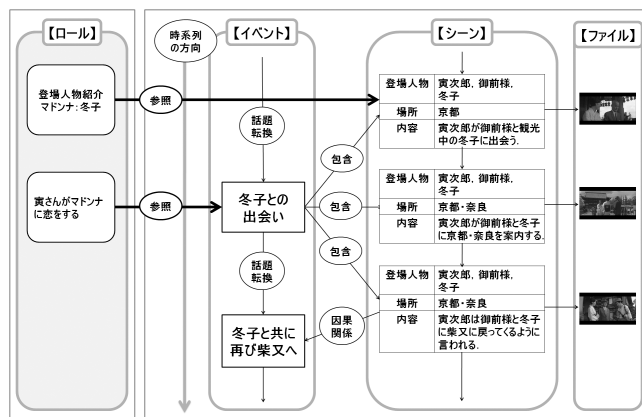


図 3: ルールの記述例

ルールの一例としては、表 2 のようなものが挙げられる。これはあくまで一例であり、シナリオ構造を持つ作品であれば必ずこのルールが含まれるわけではない。

表 2: ルールの一例

ルールの列により、プロットが記述可能となる

ルール	ルールの内容
登場人物紹介	名前や性格等、登場人物の情報を持つ
状況説明	場所や天気等、状況の情報を持つ
問題発生	問題が起こる
問題解決	問題を解決する
回想	過去に起こった出来事を思い出す

また、図 4 のようにルールを用いれば、回想シーン等、時系列の前後関係とは異なるシーンを記述することができる。すなわち、物語のストーリーとは異なった順序でコンテンツのシナリオを定義するプロットの構造を与えることが可能となる。

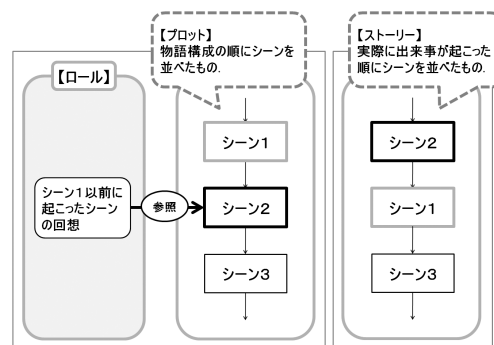


図 4: プロットとストーリー

### 3. シナリオ構造ネットワークに基づく動画編集手法の提案

以上の枠組みにより構築したシナリオ構造ネットワークに基づいて、次の三つの動画のオーサリング手法を定義する。

● 属性を用いたシーン検索

シーンの検索は、イベントの記述を文字列検索したり、登場人物や場所、内容の属性を検索することにより、関連する動画を検索結果とする。結果として、特定の条件に一致したシーンのみを選択することが可能となる。

● シーン間関係を用いたシーンの取捨選択

使用したいイベントまたはシーンを選択するとともに、選択したイベントやシーンから因果関係でつながるイベント・シーンを選択することで意味的につながったシーンを抽出し、取捨選択することができる。

● ロールを用いたダイジェスト作成

ロールを利用することにより、シナリオのテンプレートを用意することが可能になる。例えば、「起・承・転・結」というロールセットを用意して、そこにシーンを割り当てれば、起承転結の成立した作品を作ることができる。同じテンプレートを用いて作品を作ったとしても、ユーザによってロールが参照するシーンの選択は異なるので、シナリオ構造を維持したまま、ユーザの個性を反映した作品を作ることができる。また、シリーズ作品を作る際に利用すれば、シリーズ共通のパターンをロールセットとして用意することも可能である。

表 3: シリーズ作品のロールセット例

ヒロインが敵に襲われる
ヒロインのピンチにヒーローが駆け付ける
敵の攻撃にヒーローがピンチになる
ピンチを脱したヒーローが敵を倒す

4. 動画オーサリング環境の構築および実験

本研究では、シナリオ構造ネットワークの構築例として、映画「男はつらいよ」(第一作: 1969年8月公開)を用いた。

本シリーズは、テキ屋稼業を生業とする”フーテンの寅”こと寅寅次郎が主人公の人情喜劇作品である。第一作は、二十年振りに故郷・柴又に帰ってきた寅が、御前様の娘・冬子に惚れ直す話であり、また、寅の妹・さくらが博とめでたく結婚するものこの第一作である。

「男はつらいよ」シリーズの第一作を元に構築したシナリオ構造ネットワークの一部とその説明を図5に示す。シリーズ共通のロールセットは、表4の四つを用意した。

表 4: 「男はつらいよ」シリーズのロールセット

寅さんが柴又に戻ってくる
寅さんが柴又から旅に出る
寅さんがマドンナに恋をする
寅さんがマドンナにふられる

実験では、第一作の本編全体をシナリオ構造ネットワークの形で記述し、第3章で提案した手法を適用した。実際に属性検索を行った結果を図6に示す。条件を指定して検索を実行した結果が、シナリオ構造ネットワークへのリンク関係として表示される。

表4に示したロールセットを利用して、「寅さんがマドンナに恋をする」と「寅さんがマドンナにふられる」の二つのロールからシーンを抽出し、さらに抽出されたシーンから因果関係でつながっているシーンを取捨選択することで、第一作における寅次郎の恋模様のダイジェストを作ることができた。

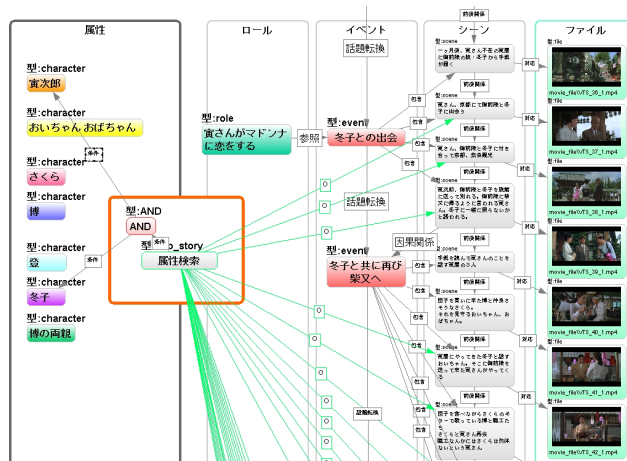


図 6: 編集ツールの実装

5. 考察

動画編集手法としては、テキストや音声などの各種メディアを利用した動画のセグメント自動分割に基づく編集が数多く提案されている。伊藤ら [2] は、音声と映像を分離し、音声テキストを要約することから動画の要約区間を決める手法を提案している。Girgensohn らは、Hitchcock システム [3, 4] にて、全体の動画から重要なムービークリップを抽出するためのスコア計算を導入し、動画の要約を行っている。また、Hua ら [5, 6] は、動画の編集を編集層と表現層の二つのレイヤーで捉え、それぞれのレイヤーにおいて動画編集のためのテンプレートを導入し、質の高い動画編集を行っている。動画をセグメント分割する基礎的な手法として、Li ら [7] は、テキスト情報に基づき動画のセグメント分割を行っている。

それらの手法に対して、我々の提案する手法は、物語構造の操作による動画編集を行うために、シナリオの構造をネットワークで可視化したシナリオ構造ネットワークを構築し、それに対して、提案した編集手法を組み合わせることで、特定の属性を持つシーンや、シーン間関係でつながっているシーンの抽出、ロールを用いたダイジェストの作成など、様々な目的に合わせた編集を行った。また、シナリオ構造をネットワークの形で記述したことで、視覚的にシナリオの構造をとることが容易になった。さらに、ロールを用いてシーンの役割を記述することで、ユーザがそのシーンの役割をどのように捉えているのかを、明確に表すことができた。これは、例えば、複数の人間が関わって作品を作る際に、それぞれがシーンに対するロールを記述することで、お互いのシーンの役割の捉え方を理解し合いながら、編集を進められるという点で非常に有用であると考えられる。

6. おわりに

本研究では、物語構造中心の動画編集環境が重要との観点から、シーン間の関係に基づいた、ネットワーク型のシナリオ構造を提案した。ここでは、構造的に動画の内容を表現することが可能となり、意味的なシーンの検索やシナリオの要約に関する動画オーサリングの枠組みを与えた。シナリオにおけるイベントやシーンの役割をロールとしてモデル化することで、プロットやストーリーを明示的に取り扱うことが可能となった。結果として、ユーザの目的とするシナリオ構造に基づいて動画を編集することができた。

今回、実験では映像コンテンツを構造化し、シナリオに基づく検索や要約を行ったが、一般に、動画の編集プロセスやアプローチは多様である。本研究で提案したシナリオ構造ネット

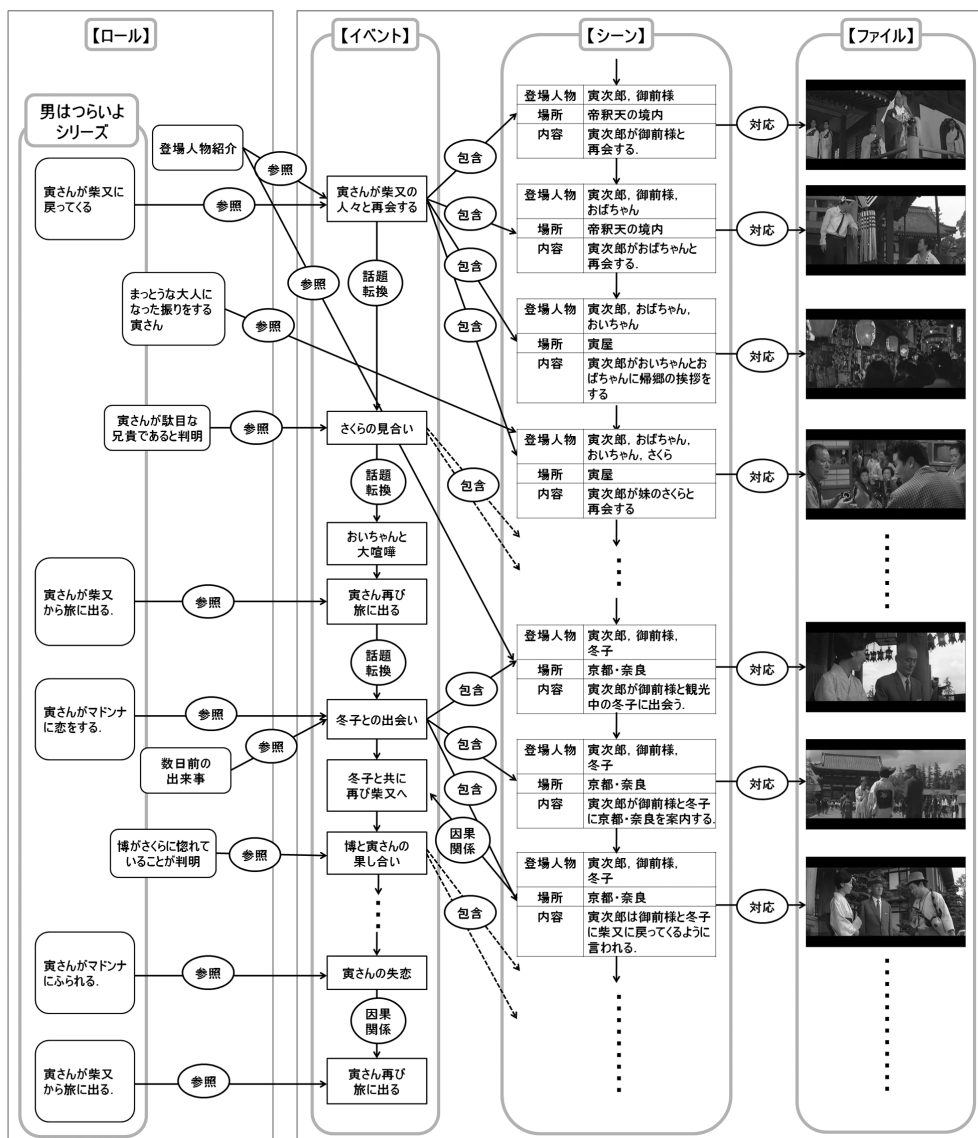


図 5: シナリオ構造ネットワーク構築例「男はつらいよ」

シーン定義によって分割されたシーンが実際の動画ファイル (図 5 ではサムネイルを表示) に対応付けられ、さらに、イベントと、内容の関連によって分類されたシーン集合が包含関係となり階層化されている。また、イベントとイベントの間は「話題転換」でつながれ、原因・結果の関係にあるシーンとイベントの間が「因果関係」でつながれている。ルールには、シリーズ共通のルールと各シナリオに依存したルールが存在し、それぞれの役割を持つイベントやシーンを参照している。

ワークを用いた編集手法は、物語中心の編集プロセスである。今回、本研究の対象は映画などの映像コンテンツであったが、物語構造の編纂手法は、漫画や小説などにも適用可能と考えられる。このためには、表現の違いについての深い考察などが不可欠であるが、これらは今後の課題である。

参考文献

[1] 橋田 浩一, 和泉 憲明: オントロジーに基づく知識の構造化と活用. 情報処理, 48(8),843-848 ,2007.  
 [2] 伊藤 一成, 酒井 康旭, 斎藤 博昭: メタデータ解析と自然言語処理を併用した動画要約, IPSJ SIG Notes, Vol.2004, No.3(20040115) pp. 41-48 2004-DBS-132-(6).  
 [3] Girgensohn, A., et al. Home Video Made Easy . Balancing Automation and Use Control. In Proceedings of INTERACT, pp. 464-471, 2001.  
 [4] Girgensohn, A., et al. A Semi-automatic Approach to Home Video Editing. In Proceedings of ACM Symposium on User

Interface Software and Technology, Vol. 2, pp. 81-89, 2000.  
 [5] Xian-Sheng Hua and Shipeng Li: Interactive video authoring and sharing based on two-layer templates, HCM '06: Proceedings of the 1st ACM international workshop on Human-centered multimedia, pp.65-74, 2006.  
 [6] Xian-Sheng Hua and Shipeng Li: Personal media sharing and authoring on the web, MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia, pp.375-378, 2005.  
 [7] Ying Li and Youngja Park and Chitra Dorai: Atomic topical segments detection for instructional videos, MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia, pp.53-56, 2006.  
 [8] Andreas Girgensohn and Frank Shipman and Lynn Wilcox: Hyper-hitchcock: authoring interactive videos and generating interactive summaries, MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia, pp.92-93, 2003.