

## 映像文法のためのカット先読み機構を備えた自動ダイジェスト生成システム

西澤 尚宏<sup>†</sup> 鎌原 淳三<sup>‡</sup> 春藤 憲司<sup>§</sup> 塚田 清志<sup>§</sup>  
有木 康雄<sup>\*</sup> 上原 邦昭<sup>◇</sup> 下條 真司<sup>†</sup> 宮原 秀夫<sup>†</sup>

<sup>†</sup> 大阪大学大学院 基礎工学研究科  
〒 560-8531 大阪府豊中市待兼山町 1-3  
nisizawa@ics.es.osaka-u.ac.jp  
simojo@cmc.osaka-u.ac.jp  
miyahara@ics.es.osaka-u.ac.jp

<sup>‡</sup> 神戸商船大学 情報処理センター  
kamahara@cc.kshosen.ac.jp

<sup>§</sup> 毎日放送株式会社  
shunto@mbs.co.jp  
tsukada@mbs.co.jp

<sup>\*</sup> 龍谷大学 理工学部  
ariki@rins.ryukoku.ac.jp

<sup>◇</sup> 神戸大学 都市安全研究センター  
uehara@kobe-u.ac.jp

### 概要

短時間で映像情報の概要を把握するには、要約映像であるダイジェストが有用である。従来のダイジェスト手法ではシーンの重要度のみを用いてダイジェストを生成するため、カット間の整合性が考慮されず、アップの映像が続くなどユーザが視聴する際に見づらい映像になる可能性がある。そこで我々はダイジェスト生成に映像文法を導入することにより、自然なダイジェスト生成を可能とする手法を提案する。本手法ではシーン毎に各カットが次選択候補カット集合を持ち、現在のカットから映像の文法に基づいて適切なカットを選択する。この際、適切なカットを発見するために次選択候補のリストを先読みするのが特徴である。本稿では、この手法を映像素材に適用した自動ダイジェスト生成システムを紹介する。このシステムでは、映像のインデックス情報に従って選択するシーンとダイジェスト再生時間を動的に選択できる。このシステムを使って映像の文法を利用しない場合と比較し、映像の専門家である放送局の人にとって自然な映像であることを示した。

キーワード: ダイジェスト, 自動生成, 映像文法, 個人化, 映像インデックス情報, カット先読み

## Automatic Digest Generation System with Look-Ahead Mechanism for Video Syntax

Takahiro Nishizawa Junzou Kamahara Kenji Shunto Kiyoshi Tsukada  
Yasuo Ariki Kuniaki Uehara Shinji Shimojo Hideo Miyahara

### Abstract

To grasp an outline of films in short time, it is useful to view the digest movie as an summary of it. In the conventional techniques to generate digest movies, they skim only importance of scenes, therefore, a consistency between scenes is not taken into consideration, and produced digest movies may have trouble watching for users. For example, continuous close-up scenes into the digest movie. Then, we propose the method to produce smooth digest by using video syntax. In this method, each scene have group of cut which can be chosen next, and appropriate next cut is chosen using video syntax from them. The feature of this method is looking-ahead the list of cuts which can be chosen as the next cut, in order to find a suitable one. In this article, we introduce the automatic digest generation system which applied this method to the video material. In this system, specifying scene type and playout-time of a digest can be dynamically chosen according to video index information. The digest using video syntax is compared with the not used video syntax, we show that the digest movie is more smooth than conventional digest movie for video specialists in broadcast company.

**key words:** Digest, Automatic Generation, Video Syntax, Personalization, Video Index, Cut Look-ahead

# 1 はじめに

近年の計算機技術の進展により、マルチメディア情報を扱うことが容易になってきている。これまで、我々は映像の要約であるダイジェストを自動で生成する研究を行ってきた [1]。ダイジェストを自動で生成することにより、ユーザが見たい時間に自由にダイジェストを見ることができ、またユーザの興味を反映したダイジェストを提供することが可能になった。さらに、制作側もダイジェスト生成のコストを削減することが可能になった。

従来の自動でダイジェストを生成する手法は、シーンに映像の意味的情報やユーザの興味から優先度を与え、その情報のみを用いてダイジェストの生成を行ってきた [2]。しかし、この場合ではカット間の整合性は考慮されておらず、ユーザが視聴する際に

- アップの映像が続く
- カメラが動きっぱなしで安定しない
- 各カット時間が不自然である

などの見づらい映像になる可能性があるといった問題があった。そこでダイジェストの生成に、視聴者が違和感なく理解しやすいようにプロが考えた映像文法を導入することを考える。映像文法を用いることによりカット間の整合性をとり、より自然なダイジェストの生成が可能となる。ここでの映像文法とは、映像と映像を繋ぐ際に、前後の映像間の違和感を無くするための規則の集合である。

映像文法を用いるためには、映像がどのように撮影した映像であるか、といった細かい情報が必要である。そのうち、いくつかの属性情報は自動で与えることが可能である [3]。しかし、解析による属性情報だけでは映像文法を用いるには十分でないため、現状は意味的な属性情報を別に付与する必要がある。これらの属性情報を用いて重要なシーンの選択を行い、ダイジェストを生成する。

ダイジェストの生成は優先度の高いカットを選択することによって行われる。この際の選択方法として、全カットの中から優先度の高いものを一度に選ぶ手法がある [4]。しかし、映像文法を用いるためこの手法を用いることはできず、前カットと次候補カットを一つずつ比較する必要がある。

さらに映像文法を適用したダイジェストを生成する上で問題となるのは、既存の方法が優先度のみでシーンを選択するだけでよかったのに対して、映像文法を適用することによって組み合わせることが可能なシーンが制限されることである。これ

によりあるシーンに接続可能な次のシーンをどのように選択するかという問題が生じる。

そこで、本稿では映像文法を適用するにあたって必要なインデックス構造の定義を行い、ダイジェストを生成するシステムにおいて映像文法を適用するために必要な先読み機構の提案を行う。先読み数を調整することにより、生成されたダイジェストの時間を目標としていた時間にほぼ近づけることが可能となる。また、この先読み機構を備えたシステムを実装し、実際のダイジェスト生成に適用することによって評価を行った。

## 2 自動ダイジェスト生成

本研究における自動ダイジェスト生成とは、シーンやカットといった断片に分解された映像を、それぞれの断片の優先度やダイジェスト自身のストーリー性に従って再構成することにより行われる。これらの作業を自動化することにより、従来制作にかかっていたコストの軽減になる。また、各ユーザ毎にダイジェストを生成することが可能になるため、ユーザの興味を反映したダイジェストを作ることができる。

### 2.1 シーンとカット

本研究においては、映像の断片としてシーンとカットという概念を用いる。シーンとは意味的に一致する映像の集合を意味する。同一シーン内における映像は、同じ状況下で同じ対象を撮影しているものと考えることができる。一方カットとは、映像の性質まですべて一致する映像のまとまりを表す。同一シーン内においても、カメラワークやショットサイズが変われば、違うカットであると考えられる。本研究ではカットを再構成することにより、ダイジェストを生成する。

### 2.2 映像の優先度

ダイジェストを生成する際に、どのカットが重要であるかを判断するパラメータが必要である。そこで各カットに対して優先度のパラメータを用意する。ダイジェストを生成するときには各カットの優先度を参照し、高いものを順に組み合わせればよい。本稿において優先度の値は、カットの重要度と映像文法の適合度から決定される。このカットの重要度とは、各カットの情報量や映像の

テーマとの関連性, またコンテンツの提供者自身の意図から, 提供者によってあらかじめ定められているものである。

### 2.3 ダイジェストにおけるシナリオ

複数の映像を組み合わせてダイジェストを制作する際には, ダイジェストにストーリー性を持たせることが重要である。ストーリー性の無い, 優先度の高いシーンを順に組み合わせただけのダイジェストでは視聴者が正しく情報を得ることは難しい。そこで, ダイジェストをどのような流れで構成するかを記述したシナリオを用意することにより, ストーリー性を与えることができる。

## 3 映像文法

映像の編集を行う際に, 編集者の意図を視聴者に正確に伝えようとした場合には, 繋ぎ合わせ方にある規則が存在する。すなわち, 映像の概念や事実を正確に伝えるためには, その規則に従う必要がある。この規則の集合を映像文法と呼ぶ。

映像文法に関係する映像要素としては, ショットサイズ, リズム, カメラワーク等がある。これらの映像要素は 4.2 節のようにインデックス化し, 映像文法の評価関数により評価する。

### 3.1 ショットサイズ

ショットサイズとは, 被写体とカメラの距離に対して決定される要素である。ショットサイズはタイトショット (TS), ミドルショット (MS), ルーズショット (LS) に分類できる。ルーズショットは比較の対象から引いて撮ったショットであり, タイトショットはより対象に近づいて撮ったショットである。TS は映像のインパクトは大きい LS のほうが映像の理解をしやすい, といった特徴がある。ショットサイズにおける編集規則としては,

- 同一の対象を撮っているとき, ショットサイズが同じものを接続しない
- ショットサイズが急激に変化するものは接続できない

などがある。

### 3.2 リズム

リズムは各カットの時間から決定される要素である。カットの時間の変化が乏しい場合や, カットの時間が映像内容に合っていない場合には視聴者はどのカットが重要であるのかが判断しづらく, 制作者の意図を映像から得ることが難しくなる。したがって, 個々のカットの長さを考えて, リズム感のある映像にする必要がある。リズムにおける編集規則は,

- 動きのないカットは長すぎではない
- LS は 6 秒, MS は 4 秒, TS は 2.5 秒程度とする

などがある。

### 3.3 カメラワーク

カメラワークにはフィックス, パン, チルト, フォロー, ズーム等がある。映像の内容に合わないカメラワークを用いた場合, 視聴者にとって見づらく, 制作者の意図を理解しづらい映像になる。また視聴者が飽きやすい映像になることも考えられる。そこで, バランス良く違和感の無いカメラワークの接続をしなければならない。カメラワークにおける編集規則は

- カメラが動くカットでは, 開始点と終了点を固定したものしか使えない
- カメラが動くカットでは, 一定の速度で移動するものしか使えない
- 同じカメラワークのカット同士や, カメラが動くカット同士を一定数以上連続して繋げない

などがある。

## 4 実験システムの実装

本章では, 3 章において説明した映像文法を導入した自動ダイジェスト生成システムの説明を行う。本実験では映像ソースとして「食べ物屋の紹介番組」用に撮影した編集前の映像を用いた。未編集の映像をデータベースとして保存しておき, システムはユーザの要求を受けたときにそのユーザの希望を反映したダイジェストを生成し, 提供する (図 1)。

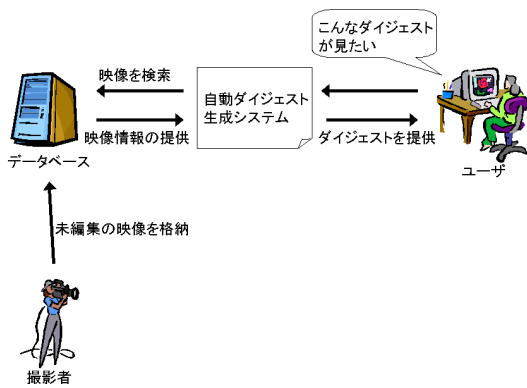


図 1: 自動ダイジェスト生成システム

#### 4.1 システムの概要

ユーザは、自分の嗜好情報と見たいダイジェストの長さをシステムに入力する。すると、システムはその情報を元に、データベース上の映像を用いてダイジェストの生成を行う。生成システムでは、入力であるカット群から映像文法を用いて出力カットを選択し、それらを結合してダイジェストとして出力する(図 2)。

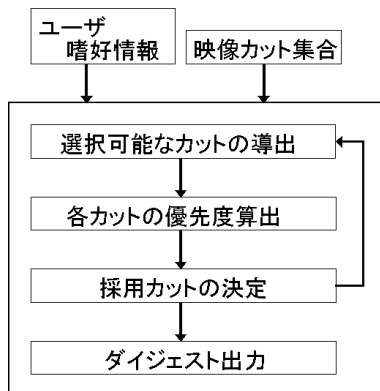


図 2: ダイジェスト生成の流れ

ダイジェストを生成するには、まず初めに先頭カットとして使用できるカットの候補を探し、それらに対してカット自身の重要度から出力カットの先頭を決定する。出力カットに対して、映像自身の構成とユーザの嗜好情報から次に選択可能なカットを決定できるので、それら全てに対してカットの重要度と映像文法の適合度からそのカットの優先度を算出し、一番優先度の高いものを次の出力カットとして選ぶ。これを全ての出力カットが決定するまで繰り返すことによりダイジェストの生成を行う。

#### 4.2 映像インデックス情報

映像文法を用いるために、ソースとなる映像には映像インデックス情報があらかじめ与えられている。これは、ソース映像の各カットがどのような意味、情報を持っているのかを記述しているものであり、表 1 に記述した属性から定義される。カットの開始時間、終了時間は対象のカットがソースの映像の中のどこに位置しているかを表すものである。

映像インデックス情報のうち、カット開始時間や終了時間、カメラワークやショットサイズ等は画像解析を用いることによって自動で与えることが可能であるため、映像にインデックス情報を与える作業量はそれほど多くないと考えられる。しかし、撮影対象の状況の分類や撮影対象の ID、さらにカットの重要度といった属性情報は画像解析から与えることは難しいので、人間の手で与えてやらねばならない。

#### 4.3 ダイジェストの個人化

本システムでは、撮影対象の状況を表 2 のように分類した。ダイジェストの生成を行う際に、ユーザが興味に応じて状況の割合を指定することにより、ダイジェスト中に使われる各状況の数が増減する。すなわち、ユーザが見たい情報を多く含むダイジェストが生成される。また、それに加えてダイジェストの時間もユーザが指定できる(図 3)。従って、各ユーザに対してより個人の興味に近いダイジェストを提供することが可能である。例えば、ユーザが「料理をしている場面をより多く見たい」と思い、ユーザ入力画面で料理のパラメータを大きく設定すれば、他の状況に対して料理をしているシーンの多いダイジェストが生成される。このユーザの嗜好情報とダイジェストの時間から、ダイジェストに使用される各状況のカット数は決定される。

外観	お店の外観を映している場面
内装	お店の内装を映している場面
材料	お店で出される料理に使われる材料を映している場面
調理	料理を調理しているところを映している場面
料理	完成した料理を映している場面

表 2: 状況属性

情報の属性	情報の例
カット ID	37
シーン ID	13
カット開始時間 (秒)	367.4
カット終了時間 (秒)	381.2
撮影対象の状況の分類	お店の内装
撮影対象の ID (同じ ID なら映っている対象が同じ)	8
カメラワーク	Fix
ショットサイズ	Middle Shot
カットの重要度 (5 段階)	4
直前カットとカメラが連続しているか?(yes/no)	no
カットの途中切断が可能か?(yes/no)	yes

表 1: 映像インデックス情報



図 3: ユーザ入力画面

#### 4.4 カットの先読み

本システムでは映像文法を用いるため、現在のカットと次の選択候補のカットの映像情報が必要となる。そこで、1つ1つ選択候補のカットを読み込み、優先度を算出する必要がある。

その際、優先度を算出したカットを選択するかどうかの判断基準として、閾値を用いることが考えられる。しかし、閾値を用いる方法では、優先度が閾値以上となるカットが採用されるため、現在選択されているカットに対して次に採用するカットがどこで選ばれるかが予想しにくく、すなわちダイジェスト全体のカット数が予想しにくい。結果として、ダイジェストの時間があらかじめ目標としている時間と大きく異なる恐れがある。閾値の取り方によっては、極端に長かったり短いダイジェストが生成されることも考えられる。それに

対して、候補となるカット数のある範囲で限定し、それらの中から一番優先度の高いものを採用する手法をとれば、選択されるカットの数が目標よりも大きく増減することはない、ユーザの望んだ時間に近いダイジェストを提供することができる。

そこで次カットの候補に対して、先読みを行い優先度を算出する手法を導入する。本稿では、次に選択するカットの候補の数のことを先読み数と記述する。先読み数を固定にした場合、目標とするダイジェスト時間を平均のカット時間で割ることにより目標カット数を求めることができ、さらに目標カット数で総カット数を割ることにより先読み数が決定する。ここで、全カットを先読み数ごとにグループに分割し、各グループから最も優先度の高いものを選択する手法(図 4(a))を用いたとすると、ダイジェストのカット数は目標と同じ数のためダイジェスト時間は保障されるが、優先度の高いカットが近くに密集していた場合、それらを見逃してしまう恐れがある。また、決定されたカットの次カットから先読み数だけを先読みする手法(図 4(b))を用いたとすると、先頭に近いカットでダイジェストが構成される可能性が高いが、シナリオ通りにダイジェストを生成するには、必要なカットをバランスよく配置する必要がある。

そこで、先読みするカットの数を動的に変化させる手法を提案する。具体的には、上記の手法 1,2 で先読み可能なカットを組み合わせることを考える。これを用いることにより、ダイジェスト時間を保証し、優先度の高いカットを見逃すこともなく、重要度の高いカットを多く含むダイジェストを生成することができる。

しかし、カットの繋ぎが一行ならば先読みする

カットの数を算出することは容易だが、実際の映像では映像内の時間順序やダイジェストシナリオから、構成が単純になることは少ない。そこで、複雑な構成に対してカットの数を算出する手法を以下で説明する。

先読み数が3の場合



(a) 先読み手法 1

(b) 先読み手法 2

図 4: 先読み数固定の先読み手法

#### 4.4.1 先読み候補カット集合

ダイジェストにおける各カットの順序は、映像内の時間順序とシナリオの順序に従う。これにより各カットには選択すべき順序が存在する。従って、各カットは次に先読みできるカットの集合を持つ。それを、先読み候補カット集合と呼ぶ。カット先読みの際には、先読み数を超えない限り、先読み候補カット集合からカットの先読みを行う。

本システムでは、各カットの選択すべき順序を、各シーンが次に接続できるシーンの集合を持つことにより表される。ここでのシーンはカットの集合で表すことができ、シーンに対して

- 同一シーン内においては、あるカットの次に選択できるカットは常に一つであり、その順序も一意に決定する
- 一度あるシーンの中のカットが読み込まれ、そこから他のシーンの中のカットを読み込んだとき、元のシーン内の全てのカットはその後先読みの対象外となる

といった制約を課す。すなわち、この条件を満たさないカットが存在した場合、それらは別シーンとなる。具体的には、あるカット A について考えると、カット A の属するシーン内のカット A 以降のカット全てと、カット A の属するシーンから選ぶことのできる次のシーンに属するカット全てがカット A の先読み候補カット集合となる。このような手法を用いることにより、先読み候補カット集合は先読み候補シーンとみなすことができ、記述を簡略化することができる。

#### 4.4.2 先読み数の決定

先読み数を動的に決定するために、新たに最大先読みカット数  $L_{max}$  という値を定義する。これは、各カットを起点として先読み候補カット集合を辿った場合に最大どれだけの数のカットを読み込めるかを表したものである。計算方法は、起点となるカットが属するシーン中でそのシーンより後ろに来るカットの数と、そのシーンから選択できる各シーンの先頭カットの最大先読みカット数の最大値を足し合わせることにより計算できる。各シーンの先頭カットの最大先読みカット数は同様に再帰的に求めることができる。

ユーザによりあらかじめ決められたダイジェスト時間を  $T$ 、1 カット当たりの平均カット時間を  $T_c$  とすると、目標とするダイジェストのカット数  $C$  は、

$$C = T/T_c$$

で与えられる。また、ソース映像の全カット数  $C_{all}$  から、仮の先読み値  $R_t$  は

$$R_t = C_{all}/C$$

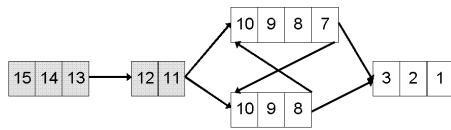
で与えられる。さらに、 $C$  はすでに出力カットとして決定しているカット数 ( $C_{after}$ ) と現在検索中のカット ( $C_{now} = 1$ ) と未検索カット数 ( $C_{before}$ ) に分割できる。ここで、先程求めた  $R_t$  に対して、

$$L_{max} > R_t * C_{before}$$

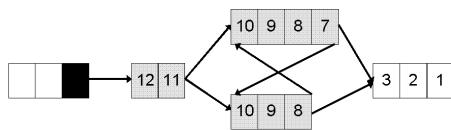
を満たすカットを先読み可能カットとする。このように先読み数を決定することにより、初めに決定されたダイジェストのカット数が変わることなく、検索漏れの無いようにすべてのカットに対して先読みが行われることになる (図 5)。つまり、未決定カット数から後で必要になる候補カットを除き、残りを選択候補とする。

ダイジェストに用いられるカットが 1 つ決定するたびに全カットの最長カット数を算出し直す必要があるため、算出のための計算量を考慮しなければならない。一般に、計算量は並列なシーン、すなわちダイジェストに用いる時にそれらのシーンのうちどのシーンが先に使われても残りのシーンをその後用いることができる、といったシーンの並列数が多いほど、また並列なシーンの登場回数が多いほど計算量は増えてしまう。しかし、今回用いたような「お店の紹介映像」では、映像内における時間の制約とダイジェストのシナリオの制約から、あまりに複雑な形の並列なシーンは登場しないと考えられる。

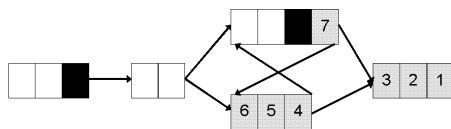
$C = 15$      $R_i = 5$   
 : 採用カット     : 先読みカット



(a)  $C_{before} = 2$



(b)  $C_{before} = 1$



(c)  $C_{before} = 0$

図 5: 最大先読みカット数と先読みカットの推移

#### 4.5 映像文法を用いた優先度算出

次のカットを求める際には、映像文法を適用する。ただし、映像文法に反するものを使わないことにすると、映像文法に反するものはすべて使えなくなるので個人の嗜好に応じたダイジェストの提供が難しくなる。また、ソースとなる映像に偏りがある場合にもダイジェストを生成することが困難になる。そこで、映像文法への適合度を計算し、優先度に反映させることにより、なるべく映像間の整合性をとりつつも、個人の嗜好を満たすダイジェストを生成する(表 3)。

映像文法は映像を繋げるための規則の集合から成り立っている。各々の規則に対して、現在のカットのインデックス情報と次候補のカットのインデックス情報からその規則に対する適合度を計算する。そのカットの繋がりが規則に適合するようであれば大きな値が、適合しないときには小さな、時には負の値が与えられる。また、映像文法同士においても優先される規則とそうでないものが存在すると考えられるので、各規則の重要度に比例して適合度の値も大きくなる。

カット間の優先度は、あらかじめインデックス情報として与えられた重要度に、映像文法の各規

則に対応する適合度をすべて足し合わせた値となる。すべての次候補カットに対してこの優先度の算出を行い、最終的に一番優先度の高かったものを次カットとして採用する。

## 5 実験

### 5.1 実験 1

生成したダイジェストの一例から、ショットサイズとカメラワークを抜き出して表 4, 5 に示し、映像文法を使わない場合および放送された映像の場合と比較して、映像の文法を再現できているかを評価する。

LS	フォロー
LS	フィックス
TS	パン
TS	パン
TS	パン
TS	パン

LS	フォロー
LS	フィックス
TS	パン
TS	パン
MS	フィックス
TS	フィックス

表 4: 映像文法なし

表 5: 映像文法あり

映像文法を用いていないものはアップでかつカメラの動くカットが続いており、視聴者が見にくく映像内容を理解しにくいダイジェストになっている。映像文法を用いることにより、ショットサイズやカメラワークの異なるカットがバランスよく選択されていることが確認できる。

### 5.2 実験 2

次に、ユーザによる主観的な評価を行った。評価方法は、ユーザに対して表 6 の 3 種類の映像を提示し、

Q1: どの映像が一番自然に見えたか?

Q2: どの映像が一番テレビに近いと感じたか?

Q3: 映像間の繋がりに一番違和感を感じたのはどれか?

という 3 種類の基準に対して、映像 A ~ C に 1 位から 3 位まで順位をつけてもらった。各映像のダイジェスト時間はほぼ同じである。このアンケートは放送局の社員 22 名、そうではない一般の人 19 名の計 41 名に対して行った。結果を表 7 に示す。

アンケート結果を見ると、映像の専門家がである放送局社員と一般の人との間で結果が大きく異

映像文法の例	適合度
LSのカットからMSのカットに繋いでいる	+5
同じショットサイズのカットを繋いでいる	-3
同じカメラワークのカットを繋いでいる	-5
途中切断不可のカットで、一定の長さをオーバーしている	オーバーしている長さ(秒) × -2

表 3: 映像文法の適合度の例

	Q1			Q2			Q3		
	A	B	C	A	B	C	A	B	C
1位選択数	8	1	13	2	2	18	8	14	0
2位選択数	7	6	9	13	5	4	5	7	10
3位選択数	7	15	0	7	15	0	9	1	12
平均順位	1.95	2.64	1.41	2.23	2.59	1.18	2.05	1.41	2.55
一位選択率 (%)	36.36	4.55	59.09	9.09	9.09	81.82	36.36	63.64	0.00

(a) 放送局社員のみアンケート結果

	Q1			Q2			Q3		
	A	B	C	A	B	C	A	B	C
1位選択数	4	7	8	3	8	8	6	6	7
2位選択数	8	5	6	9	4	6	7	6	6
3位選択数	7	7	5	7	7	5	6	7	6
平均順位	2.16	2.00	1.84	2.21	1.95	1.84	2.00	2.05	1.95
一位選択率 (%)	21.05	36.84	42.11	15.79	42.11	42.11	31.58	31.58	36.84

(b) 一般の人のみアンケート結果

表 7: アンケート結果

映像 A	本システムで、映像文法を用いて生成したダイジェスト
映像 B	本システムで、映像文法を用いずあらかじめ定義されていた映像の優先度のみに基づいて生成したダイジェスト
映像 C	実際の放送で用いられたダイジェスト

表 6: 実験に用いた映像

なっていた。これは、日頃映像についてどのように考えているかという意識の違いの差が表れたものと考えられる。特に、放送局社員のみアンケートに注目すると映像の自然さという点ではテレビ放送のものにも近く、使わないものよりもより自然である、という評価であった。しかし、放送局員以外の一般の人アンケートに注目してみると、映像文法を用いたものは用いなかったものよりも総じて悪い。このことから、今回使った映

像文法が放送局のものであるため、放送局の人にとっては違和感が無いが、文法の定義者ではない一般の人にとっては見るべき視点が違う可能性があることが分かる。

また、今回使用した映像文法は基本的なものであり、実際の映像ではこの文法をわざと外した映像を含めることでインパクトを出したり、強調したりする。特にテレビ的なものは、そういった意図がよく表れているものと考えられるが、今回の実験の映像ではそういった意図を反映することができなかったものと思われる。

## 6 まとめ

既存の自動ダイジェスト生成システムではカット間の整合性を考慮していないため、視聴者にとって見づらく、理解しにくい映像になる可能性があった。そこで本稿では、映像文法の導入とそのための



先読み機構を備えた自動ダイジェスト生成システムの提案し、実装を行った。

実験を行うことによって、映像文法を使えば放送局で作られるダイジェストにより近いものが自動生成できることが証明できた。しかし、これは放送局における映像文法を用いたからであり、一般の人からの評価はあまりよくはなかった。これは、映像文法自体が基本的なものであるため、一般の人が普段目にする映像とは異なっていた、文法通りでありすぎるためにインパクトや強調などの「意図」が失われていた、ということが考えられる。

今回実験に伴って放送局側の意見を聞いてみたところ、映像文法を用いて自動生成したダイジェストに対して「映像の流れが見えない」や「編集意図が感じられない」といった意見が聞かれた。そこで、自動生成のダイジェストに意図を入れる手法を考案も今後の課題とする。

## 謝辞

本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(Project No.JSPS-RFTF97P00501) によるものです。

また、本研究を行うにあたり、実験に用いた映像と映像文法の資料の提供と実証実験へのご協力をしていただいた毎日放送株式会社に深く感謝致します。

## 参考文献

- [1] 岡本道也, 植田和憲, 鎌原淳三, 下條真司, 宮原秀夫, “教師付学習を利用したストーリー性を持つシナリオテンプレートによるダイジェスト自動生成機構,” in 第 11 回データ工学ワークショップ (DEWS2000), March 2000.
- [2] 橋本隆子, 白田由香利, 飯沢篤志, “時空間情報を利用したサッカー番組のダイジェスト作成方式,” in 第 12 回データ工学ワークショップ (DEWS2001), March 2001.
- [3] 熊野雅仁, 林義文, 有木康雄, 上原邦昭, 下條真司, 春藤憲司, 塚田清志, “アクティブ探索を用いた映像編集支援のためのショットサイズ自動判定,” 電子情報通信学会技術研究報告,(オフィスシステム研究会 OFS2001-24), 2001.

- [4] 森山 剛, 坂内 正夫, “ドラマ映像の心理的内容に基づいた要約映像の生成,” in *Proceedings of ACM Multimedia 2000 Workshops*, pp. 191–194, November 2000.