映像間の関連性を考慮したダイジェスト自動生成手法の提案

西澤　尚宏† 　鎌原　淳三†† 　下條　真司††† 　宮原　秀夫†††

† 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
†† 神戸商船大学 情報処理センター　〒 658–0022 兵庫県神戸市東灘区深江南町 5–1–1
††† 大阪大学大学院 情報科学研究科　〒 565–8531 大阪府豊中市待兼山町 1–3
E-mail: †nisizawa@ics.es.osaka-u.ac.jp, ††kamahara@cc.kshosen.ac.jp,
†††{simojo,miyahara}@ist.osaka-u.ac.jp

あらまし　本稿では，ダイジェスト生成の際にシーン間の意味的な関係を考慮するために，映像内容の関連性を定義し，それを用いることで自然な流れのダイジェストの生成を可能とする手法の提案を行い，有効性の評価を行った．
キーワード　ダイジェスト，自動生成，映像インデックス情報，映像文法，映像内容の関連性，個人化

Automatic Generation of Video Digests Considering Relationship between Scenes

Takahiro NISHIZAWA†, Junzo KAMAHARA††, Shinji SHIMOJO†††, and Hideo MIYAHARA†††

† Graduate School of Engineering Science, Osaka University　Machikaneyamatyou 1–3, Toyonaka-si, Osaka, 560–8531 Japan
†† Information Processing Center, Kobe University of Mercantile Marine　Fukaeminamityou 5–1–1, Higashinada-ku, Hyogo, 658–0022 Japan
††† Graduate School of Information Science and Technology, Osaka University　Machikaneyamatyou 1–3, Toyonaka-si, Osaka, 560–8531 Japan
E-mail: †nisizawa@ics.es.osaka-u.ac.jp, ††kamahara@cc.kshosen.ac.jp,
†††{simojo,miyahara}@ist.osaka-u.ac.jp

Abstract　In this paper,we defined semantic relationship of scenes and proposed automatic video digest generation method with this relationship for digest movie of a natural flow. And we evaluated validity of this method.
Key words　Digest,Automatic Generation,Video Index,Video Syntax,relationship between scenes,Personalization

1. はじめに

近年の計算機技術の発展により，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．映像で情報を得ることにより，一度に多くの情報を得ることが可能となる．一般的に，映像情報はある長さを持った映像として与えられ，その映像の中には視聴者が求めている情報以外の場面も多く存在する．そこで，有用な場面だけを抜き出し繋ぎ合わせたダイジェスト映像は，情報を効率的に得るために非常に有用な手段である．

現在ダイジェスト映像は，映像提供者である放送局で制作し，テレビで放映されているものがほとんどである．テレビ局で制作されたものは一般的に重要であると思われる場面を採用し，それらの場面の組み合わせで構成されているため，視聴者のそれぞれが欲しいと思う情報がダイジェスト映像内に含まれているとは限らない．そこで，計算機を用いて視聴者各々の好みに合ったダイジェストを自動で生成する手法が検討されている [1].

ダイジェスト映像を自動で生成するためには，いくつかの手順を行う必要がある．まず素材となる映像に，各時間においてどのような内容が映っているかという情報を付与する必要がある．次に，付与された情報を元に，映像を意味のある単位映像に分割する．さらに，単位映像毎に映像の重要度を求め，最後に重要度の高い映像を繋ぎ合わせる．以上の作業を行うことにより，ダイジェスト映像は生成される．

映像内容の情報の付与については，映像がどのように撮影された映像であるか，どのような内容であるかといったダイジェストの生成に必要と思われる細かい情報を，素材映像に与える．客観的な内容情報，例えばショットサイズやカメラワーク，カット割り等は画素値などを解析することにより，情報を自動で与える研究が行われている [2-4]．客観的には判断できない情報，すなわち具体的に映像に何が映っているか，例えば「車」「花」などといった意味情報も．ダイジェストを自動生成するためには不可欠である [5]．しかし，このような情報は映っている対

象の候補が無限に存在するため，機械で自動的に判断することは難しく，情報の付与者の主観による部分がある．この付与作業は手動で行われるため，非常にコストがかかる．ただし，意味的な映像内容の情報の付与は，映像と同期の取れた映像以外の素材，例えば音声やタイムテーブルのようなものが存在すれば，それらを用いることによって自動で与えることが可能となる [6].

これらの情報を映像に対する索引 (インデックス) として利用するには，ある映像範囲を指定して，そこに情報を付与する形をとる．この映像範囲を決定するには，それ以外に分割できない単位映像が基準となる．単位映像とは連続する画像の中で意味や性質が変わらない部分であり，従って同じ内容情報を付与することができる．この単位映像は，シーンチェンジなどから自動的に検出することができる [2]．また，内容情報を元に単位映像の範囲を類推する研究も行われている [7].

これは，ダイジェストに用いる映像を選択する際に，場面毎の重要度に加えて，各視聴者の嗜好に合っているかどうか，また出力されるのテーマに場面が合っているか，ということを考慮し，映像の優先度を算出する手法である．しかし，従来の手法では個々の映像単位ごとに意味や重要度が決定されるといった観点に立っていたため，複数の場面を並べた際の整合性や，並べられた場面の関連性が考慮されていない．そのため，並び方によっては映像の意図がうまく視聴者に伝わらず，違和感のあるダイジェスト映像が生成される可能性があった．

我々の研究グループでは，インデックス情報を用いたダイジェスト自動生成システムの提案を行ってきており，特にユーザプロファイルを用いることによるダイジェスト映像の個人化や，シナリオテンプレートを用いることでダイジェスト映像に流れをつくる，といった研究を行ってきた [8, 9]．これは，ダイジェストに用いる映像を選択する際に，場面毎の重要度に加えて，各視聴者の嗜好に合っているかどうか，また出力されるのテーマに場面が合っているか，ということを考慮し，映像の優先度を算出する手法である．しかし，従来の手法では個々の映像単位ごとに意味や重要度が決定されるといった観点に立っていたため，複数の場面を並べた際の整合性や，並べられた場面の関連性が考慮されていない．そのため，並び方によっては映像の意図がうまく視聴者に伝わらず，違和感のあるダイジェスト映像が生成される可能性があった．

そこで，映像文法と呼ばれる映像と映像を繋ぐ際の規則を用いることにより，映像の断片が違和感無く接続されたダイジェストを自動生成する研究を行った [10]．映像文法を導入することで，ダイジェスト映像の違和感を薄れさせ，自然なダイジェストの生成を手助けすることができることが示された．しかし，映像文法は映像内容については触れていないため，ダイジェスト映像の内容に関しての違和感は考慮されていない．このため，「シーン A を説明するシーンがあるのに，シーン A が存在しない」といった整合性が取れておらず，視聴者が理解しにくいダイジェスト映像が生成されてしまうということがあった．また，生成されたダイジェスト映像に対して「映像の意図が見えない」といった意見も聞かれ，これが内容が理解しにくい原因の 1 つであることが考えられた．

本研究では映像内容の関連性を考慮したダイジェスト自動生成システムの提案を行う．素材となる映像にはシーン間にそれぞれ関係が定義されており，それを考慮したダイジェストの自動生成を行う．マルチメディアデータに関連性を用いたインデックス情報を定義する研究はさまざまなものが行われている．キーワードの関連性を用いたものや [11, 12]，ストリーム内の構造の関連性を用いたもの [13] があるが，映像のみの素材に対して，テキスト情報を元にした手法を用いることはできない．したがって，本研究では映像内容に対する関連性を定義するデータ構造を提案する．また，映像制作者の意図を取り入れることにより，生成されるダイジェスト映像に意図を与え，視聴者が理解しやすいダイジェストを自動生成することを考える．さらに，制作者の意図を具体的に定義し，定式化することでシステムに取り入れる手法を提案する．これらの手法を用いることにより，映像制作者の意図するシーンや映像の流れが表現でき，かつ映像内容に関して違和感無く整合性のとれたダイジェストを生成することが可能となる．この出力されるダイジェストは，使用したいシーンを指定することにより，そのシーンを中心にした流れのある映像を表現できるため，手動編集の荒編集としても利用できる．この提案手法をシステムに適用し，実際にダイジェスト映像を生成することにより，提案手法の有効性を評価した．

2. 映像内容の関連性

映像内容に対する違和感を取り除くためには，素材となる映像にどのようなものが映っているかといった映像内容に関する詳細な情報が必要となる．本研究では，映像のみを素材として扱うため，音声等の他の情報源から映像内容に関する情報を得ることはできない．また，各単位映像ごとに手動で絶対的な映像内容情報を付加することは，情報付加のコストが大きくなってしまうことが考えられる．そこで，映像と映像の相対的な関係を用いることを考える．相対的な関係の記述方法であったならば，記述する情報をパターン化することができ，さらに映像と映像との関係が分かるために，関係の近い映像を用いることで映像内容に関する違和感をとりのぞくことが可能と考えられる．しかし，既存の方法では映像の構造は表していても，その中の映像間の意味的な関係は明らかではなかった．

そこで本研究では，映像内容に関する記述として，映像内容の関連性を用いる．これは，関連性を定義してそれをダイジェスト生成に生かすことで，より関係の強い映像同士を接続し，映像の流れをわかりやすくすることができると考えられるからである．また，各映像毎にどのような内容が映っているかを記述する手法であると，情報を付加する人によって付けられる言葉が異なり，システムで扱いにくいということが挙げられる．そこで，抽象化した映像間の関連性を定義し，システム内で関連性を利用することで整合性の取れた理解しやすい映像を生成することを考える．これにより，類似した映像には同じ関連性を与える，といったことが可能になるとも考えられる．

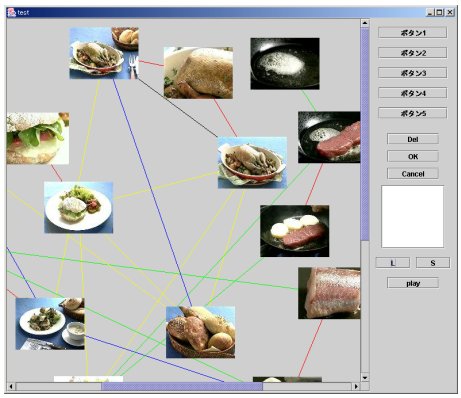


図 1　関連性の入力例

2.1　関連性の定義

映像内容の関連性は，シーン単位を基準にして行われる．ここでのシーンとは，意味的に一致する映像の集合を意味する．映像内容に関する関連性は，次の 5 つの抽象的な概念で定義する．

（1）　強関連性

映っている対象の間に関連性がある．2 つの映像の間に他の映像を挟むことはできない．

（2）　弱関連性

映っている対象の間に関連性がある．2 つの映像の間に他の映像を挟むことができる．

（3）　連鎖関連

一方の映像が選択されたときには，もう一方の映像もなるべく選択した方が望ましい関係．

（4）　並列関連

映っている対象の間には関連性が無いが，シチュエーションが類似している関係．ここでのシチュエーションとは，映像の構図や映像構成上のその映像の位置づけのことを指す．

（5）　排他関連

一方の映像が選択されたときには，もう一方の映像は選択しないことが望ましい関係．

（6）　無関連

2 つの映像間には関連性が無い

この中で，関連性 1〜3 には方向性が存在し，4 と 5 には方向性は存在しない．また，1 と 2，3 と 5，4 と 5 は同じシーン間にそれぞれ同時に成立することはできない．関連性を付与する手間の軽減のため，1 と 2，3 の関係にはそれぞれ推移律が成り立つものとする．さらに，1〜5 の関連性を辿ったときに 2 つの映像間が到達不能であったならば，その 2 つの映像は無関連であると定義する．

2.2　関連性の付与

関連性は．データベース上の映像に対して，映像提供者側が手動で与える (図 1)．一度与えた関連性はデータベース内に映像と共に格納され，ダイジェスト生成の際に用いられる．また必要に応じて書き換えることもできる．

3.　映像提供者の意図

視聴者がどのような映像を見たいか，といった嗜好情報と同様に，どのような映像を見せたいか，といった映像提供者による意図も存在する．既存のダイジェスト自動生成システムでは，視聴者の嗜好情報を用いたダイジェストの個人化は行われていたが，映像提供者の意図は考慮されていなかった．そこで，映像提供者の意図を定式化し，システムで利用することを考える．

映像提供者の意図とは，自身の考える映像の意味を正確に視聴者に理解してもらうことである．一般的な映像内容の関連性だけでは，映像提供者の考える意図が正しく視聴者に伝わるとは限らない．そこで，映像の意味を正しく視聴者に伝えるための映像提供者の意図を次のように考える．

- 映像提供者の意図するシーンがダイジェスト内に必ず含まれる
- 映像提供者の意図した映像の使われ方，映像の流れが実現できる

これらを次のような命令で定義する．

（1）　select(A)

（2）　exist(A, B, locate, direct)

ここで，A と B はシーンを表す．命令 1 は，ダイジェスト内にシーン A を含めるための命令である．命令 2 は，もし A がダイジェスト内で使われていたら B もダイジェストに用いる命令であり，locate で A と B の前後関係を，direct で A と B を直接繋ぐか，間接的に繋ぐかを指定する．この 2 種類の命令を映像提供者はシステムに与えることによって，提供者の意図するシーンや流れが反映されたダイジェストの生成を実現する．

4.　映像内容の関連性を用いた映像の構成手法

本章では，映像文法 [10] に加えて映像内容の関連性を考慮したダイジェストの構成手法についての提案を行う．

まず，想定するシステムでは，素材映像がデータベース内に格納されており，その素材映像には映像内容を表すインデックス情報が映像提供者によって付加されているものとする．視聴者は，自分が見たいと思うダイジェストの希望をダイジェスト自動生成システムに与える．システムは，視聴者からの要求，例えばある内容を多く含むであるといったものや，出力して欲しいダイジェストの再生時間，といったものを受け取ると，その要求を満たす映像をデータベースから検索し，映像を抽出する．データベースから抽出された映像をシステムで再構成し，ダイジェストとして視聴者に出力する．

システムでの流れは，

（1）　シナリオや視聴者個人の嗜好情報から，目標とするダイジェストの再生時間や使用する映像の種類，数を決定する

（2）　1 の情報を元に，各単位映像に優先度を与える

（3）　候補となる映像の中から，優先度の高い映像を出力映像として選択する

（4）　2,3 を繰り返し，決定された映像を繋ぐことでダイジェストを生成する

といった流れとなる．

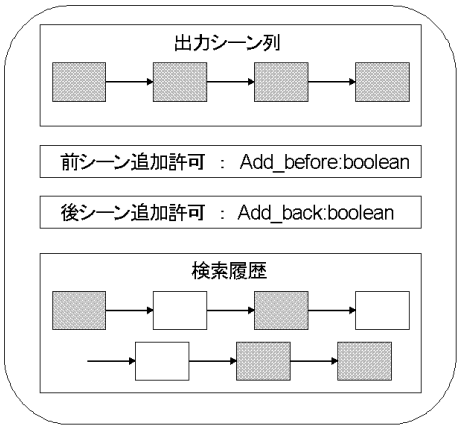


図 2　出力シーングループ

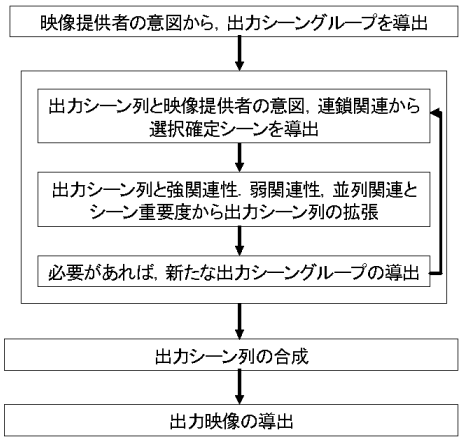


図 3　出力映像決定の流れ

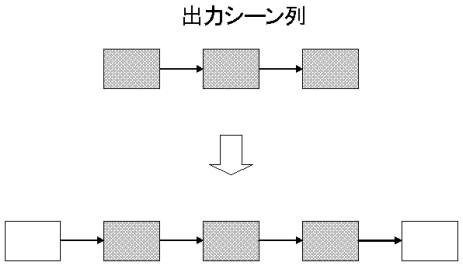


図 4　出力シーン列の拡張例

映像を構成する際に用いるデータ構造として，出力シーングループという構造体を用いる．出力シーングループは，出力シーン列，前シーン追加許可パラメータ，後シーン追加許可パラメータ，検索履歴から成る (図 2)．出力シーン列とは，実際に出力されるシーンの種類と順序を表す．前シーン追加許可パラメータ，後シーン追加許可パラメータは現在保持している出力シーン列の前後に，新たにシーンを追加できるかどうかをチェックする論理形パラメータである．検索履歴とは，出力シーン列を求めるために検索が行われた候補シーンの集合が，シーン列の形で格納されている．

主な流れとして，まず視聴者の嗜好情報をからダイジェスト映像の再生時間を決定し，その映像に用いる目標カット数を決定しておく．次に，映像提供者の意図命令 select() から，選択シーンが存在する場合には，そのシーンを元に出力シーングループを生成する．選択シーンが存在しない場合には，映像内容の関連性を考慮し，出力シーン列の先頭シーンを探し，そのシーンを元に出力シーングループを生成する．各出力シーングループの出力シーン列の前後から順に出力シーンを決定していく．目標シーン数に到達する前に出力シーン列に追加できなくなった場合には，新たに出力シーングループを生成し，繰り返し出力シーンの検出を行う．最終的に全ての出力シーンが決定したところで，出力シーン列の合成を行う．そうして出来たシーン列からダイジェストを生成し，視聴者へと提供する (図 3)．詳細を以下で説明する．

4.1　出力シーングループの生成

映像提供者の意図命令 select() から選択シーンが存在する場合，そのシーンを出力シーン列の中心シーンとする出力シーングループを生成する．また，映像提供者の意図命令 select() が存在しない場合，出力シーン列の先頭候補シーンの検索を行う．これは，全ての未選択シーンに対して，

- 強関連性，弱関連性をもった上位シーンが存在しないまたは選択済みである
- 並列関連シーンが存在しない，または並列関連シーンすべてにおいて，強関連性，弱関連性をもった上位シーンが存在しないまたは選択済みである

をチェックし，この項目を満たすシーンを先頭候補シーンとする．全ての先頭候補シーンにおいて一定数のシーン先読みからシーン優先度を計算し，それらの中で最も，優先度の高かったシーンを出力シーン列の中心として出力シーングループを生成する．この際，前シーン追加許可パラメータは”不可”となる．

4.2　出力シーン列の拡張

全ての出力シーン列について，前後のシーン追加許可パラメータが”不可”でなければ，出力シーン列の前後に加えることのできるシーンを選択する (図 4)．ここでは，後シーン追加許可パラメータが”可”である場合を例に挙げて説明する．

まず，出力シーン列の最後のシーンを取り出し，これを S_{last} とする．次に，S_{last} と強関連性，弱関連性を持ち，かつ S_{last} の方が上位シーンであるシーン群 S が存在するとする．また，S_{last} と並列関連を持つシーンも S に加える．S の中の要素 S_i に対して，シーン優先度 S^i_{pri} を求める．S^i_{pri} は，

$$S^i_{pri} = S^i_{significance} \cdot R(S_{last}, S_i)$$

で求められる．ここで，$S^i_{significance}$ は S_i 自身の持つ重要度であり，$R(S_{last}, S_i)$ は S_{last} と S_i の関係の種類から与えられる重みである．重みは，強関連性が最も大きな値が与えられ，以下順に弱関連性，並列関連，無関連の順に値は小さくなる．S_{last} と S_i が直接の関係を持たない場合においても，推移律から関係を導き出すことができるので，値が与えられる．その関係の導出アルゴリズム CheckRelation() を以下に示す．

シーン間の関連性の導出アルゴリズム

```
function CheckRelation(S1,S2:シーン) : 関連性;
var r,max:関連性;
begin
    if(S1 と S2 が直接関連性 R で
                    繋がっているならば) begin
        CheckRelation = R;
    end;
    else begin
        max = null;
        for (S1 から直接関連性を持つシーン S
                            全てに対して) begin
            r = Connect(CheckRelaton(S1,S),
                        CheckRelation(S,S2));
            if(max より r の方が強い関係ならば)
                                    max = r;
        end;
    end;
end:

function Connect(R1,R2:関連性) : 関連性;
begin
    Connect = R1 と R2 のうち弱い方の関連性;
end:
```

求められた優先度の中で最も高い値を持つシーンが候補シーンとなるが，先読み数に満たない場合には，さらに下位のシーンの検索を行う．この際の先読み数 L_{ave} は，出力される目標シーン数を Sum_{output}，現在選択されているシーン数を Sum_{select} と，全ての出力シーングループの検索履歴に含まれていない (まだ選択できる) シーン数を S_{rest} とすると，

$$L_{ave} = S_{rest}/(Sum_{output} - Sum_{select} - 1)$$

で表される．この数を満たすまでは，再帰的に下位シーンを検索し，シーン優先度を求める．全ての検索されたシーン優先度の中で最も高い値を持つシーンが，出力シーン列の最終シーンとして加えられる．ただし，先読み数に満たないのに検索できるシーンが存在しなくなった場合には，その出力シーングループの後シーン検索可能性を不可とし，優先度を求められたシーンの中から選択する．

また，新たに決定された出力シーンに対して，提供者の意図命令 exist()，連鎖関連，排他関連のチェックを行う．決定された出力シーンが提供者の意図命令 exist()，連鎖関連の上位シーンであった場合，下位シーンにあたるシーンの検索を行い，そのシーンが未検索シーンであればそのシーンの重要度を上げる．また既検索シーンで，選択されていない場合にはそのシーンを出力シーンに加える．場所は，出力シーングループの検索履歴を元に決定される．決定されたシーンが排他関連を持つシーンが存在する場合には，そのシーンの重要度を下げる．

4.3 出力シーン列の合成

目標とするシーン数の出力シーンが決定されたならば，それぞれ別々の出力シーン列を合成して 1 つのシーン列にする必要がある．ここで考慮すべき条件は，映像提供者の意図命令 2 と，映像のシナリオである．映像のシナリオはおおまかな映像の流れを指定するものであり，これに従うことによって意味の通った映像となり，視聴者に内容を理解してもらう助けとなる．

映像提供者の意図命令 exist() に関係している出力シーンが存在したならば，その順序に従うような合成を行う．但し，強関連以外の部分で出力シーン列は分割することが可能であるので，分割することによってよりシナリオに適合する組み合わせ方が存在するなら，その組み合わせを採用する．映像提供者の意図命令 exist() に関係ない出力シーン列同士の結合に関しては，強関連以外の部分が分割できることを利用し，シナリオとの適合度が高い組み合わせを行う．最終的に 1 つのシーン列となるまで，この作業を繰り返す．

実際の合成においては，シーンの属性を用いたシナリオに基づいて組み合わせが決定するため，合成方法の詳細については 4.4.4 節で説明する．

4.4 出力映像の決定

シーン列の各シーンにおいて，使用するカットを選択することで出力される映像を決定する．1 つのシーン列で与えられているために，順序は固定となるので，先頭から順に選択していく．選択には，映像文法への適合度を計算し，最も適合度が高いものを出力カットとして選ぶ．カット i の優先度を C^i_{pri}，カット i におけるインデックス情報を I_i，規則 k でのカット i とカット j の繋がりに対する映像文法の適合度を $R(k, I_i, I_j)$，規則 k の重要度を W_k とすると，カットの優先度は

$$C^i_{pri} = \sum_{k=1}^{N} W_k \cdot R(k, I_{base}, I_i)$$

と表される．ただし，N は映像規則の総数，I_{base} は出力された直前カットのインデックス情報を表す．

5. ダイジェスト自動生成システムの実装

4 章で提案した手法を用いたダイジェスト自動生成システムの実装を行う．本システムでは素材映像として,「飲食店の紹介番組」用に撮影された編集前の映像を用いた．

5.1 システムの概要

撮影者によって撮影された映像に，映像提供者がインデックス情報と映像内容の関連性，更には映像提供者の意図を与え，データベースへ格納されている．視聴者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．システムは視聴者の嗜好情報と素材映像，さらにそのインデックス情報と映像内容の関連性を入力として出力カットを決定し，それらを繋ぎあわせて視聴者にダイジェストとして提供する．

5.2 映像インデックス情報

本システムで，インデックス情報としてシーン情報 (表 1)

シーン番号
シーン重要度
シーンの種類
シーン内カット数
シーン開始時間
シーン終了時間

表 1　シーン情報

カット番号
カット開始時間
カット終了時間
カメラワーク
開始時ショットサイズ
終了時ショットサイズ

表 2　カット情報

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 3　状 況 属 性

図 5　ユーザ入力画面

とカット情報 (表 2) の 2 種類の情報を与える．

本システム内ではシーンとカットを扱う場所はそれぞれ異なるので，別々のインデックス情報を用意するほうが効率的である．

5.3　システムの個人化

本システムでは，撮影対象の状況を表 3 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変化する．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 5)．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となる．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．

5.4　出力シーン列の合成

出力導出段階で求められた出力シーン列をどのように合成するかについて具体的に説明する．本システムでは素材となる映像を外観，内装，材料，調理，料理という 5 種類に分類している．これを用いて，本システムで出力されるダイジェストのシナリオを作成する．そのシナリオは，拡張 BNF 記述で

シナリオ::=外観∗　内装∗　(材料∗　調理∗　料理∗)∗

と表される．ここでの合成の方針として，いかに出力シーン列を崩さずに，シナリオと映像提供者の意図命令 exist() に従う映像列を生成できるかを考える．

以下，出力シーン列の合成の説明を行う．合成を行う 2 つの出力シーン列を L_1，L_2 とする (図 6(a))．

まず，L_1 と L_2 を分類に応じてシーングループに分割する．具体的には，外観部分のシーン列 ($L^1_{outside}$, $L^2_{outside}$)，内装部分のシーン列 (L^1_{inside}, L^2_{inside})，それ以外のシーン列 (L^1_{else} , L^2_{else}) に分割する (図 6(b))．このとき，別のカテゴリのシーンが切断不可な関係 (強関連性など) で接続されていた場合，切断不可な関係で接続されているシーンは全て前のシーン，例えば外観-内装間であるならば $L_{outside}$ に，内装-調理間であるならば L_{inside} のシーンに含める．

ここで，映像提供者の意図命令 exist() の検査を行う．この命令が適用できる，すなわちこの命令で指定された 2 つの映像が共に出力シーンとして含まれている場合，分割されたシーン群が意図命令に従った順序になるよう，順序の入れ替えを行う．

次に，(L^1_{else} と L^2_{else}) の結合を行う．ここで，シーンの属性値として，便宜的に材料=1，調理=2，料理=3 という値を与える．まず (L^1_{else} と L^2_{else}) の最終シーンの属性値を求め，その属性値に差があれば，属性値の小さな方のシーングループを前にしてグループ同士の結合を行う．最終シーンに差が無い場合には，先頭シーンの属性値を比較する．ここでも，属性値に差があれば，属性値の小さな方のシーングループを前にして，グループ同士の結合を行う．最終シーンの属性値も先頭シーンの属性値も等しい場合には，(L^1_{else} を前シーングループとして結合する．

続いて，($L^1_{outside}$ と $L^2_{outside}$)，(L^1_{inside} と L^2_{inside}) 同士の結合をそれぞれ行う．この際の接続順は，(L^1_{else} と L^2_{else}) の接続順と同じ順序とする (図 6(c))．

各カテゴリ毎の合成が行われたなら，最後に各カテゴリをシナリオに応じて結合し，一つのシーン列とする (図 6(d))．すべてのシーン列に対してこの作業を繰り返すことにより，1 つの出力シーン列を生成してゆく．

6.　システムの実験と評価

視聴者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 4　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: どの映像が一番自分の好みに合うか？

という 3 種類の基準に対して, 映像 A〜D に 1 位から 4 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 5 名, そうではない一般の人 12 名の計 17 名に対して行った. 結果を表 5 に示す.

アンケートの結果を見ると，動画 A は総じて違和感なく自然

	Q1				Q2				Q3			
	A	B	C	D	A	B	C	D	A	B	C	D
1 位選択数	3	0	0	2	3	0	1	1	2	0	1	2
2 位選択数	0	0	4	1	0	0	1	4	1	0	1	3
3 位選択数	2	1	1	1	2	0	3	0	2	0	3	0
4 位選択数	0	4	0	1	0	5	0	0	0	5	0	0
平均順位	1.8	3.8	2.2	2.2	1.8	4.0	2.4	1.8	2.0	4.0	2.4	1.6
一位選択率 (%)	60	0	0	40	60	0	20	20	40	0	20	20

(a) 放送局社員のみのアンケート結果

	Q1				Q2				Q3			
	A	B	C	D	A	B	C	D	A	B	C	D
1 位選択数	4	5	2	1	3	2	3	4	4	3	1	4
2 位選択数	4	1	2	5	3	4	2	3	3	2	4	3
3 位選択数	2	1	5	4	4	1	5	2	3	2	4	3
4 位選択数	2	5	3	2	2	5	2	3	2	5	3	2
平均順位	2.6	3.0	3.3	3.1	2.9	3.3	3.0	2.8	2.7	3.3	3.3	2.7
一位選択率 (%)	33.3	41.6	16.7	8.3	25.0	16.7	25.0	33.3	33.3	25.0	8.3	33.3

(b) 一般の人のみのアンケート結果

表 5　アンケート結果

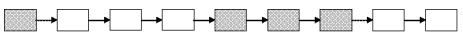
映像 A	本システムで，映像提供者の意図を用いずに生成したダイジェスト
映像 B	本システムで，映像提供者の意図を用いて生成したダイジェスト
映像 C	文献 [10] のシステムで映像文法のみを用いて生成したダイジェスト
映像 D	実際の放送で用いられたダイジェスト

表 4　実験に用いた映像

(a) 合成するシーン列の準備

(b) 各シーン列をカテゴリに応じて分割

(c) 各カテゴリごとに結合



(d) 各カテゴリを結合

図 6　シーン列の合成の推移

な動画であるという評価であった．これは関連性を用いて映像を構成した結果，自然な映像を作ることができたことを示していると思われる．また，同じ関連性を用いて生成した動画 B に関しては，一般の人の映像の自然さの評価は大きく 2 つに分かれた．これは，映像提供者の意図は映像の関連性やダイジェストのシナリオよりも強く出力映像に働くため，その提供者の意図を視聴者が受け取れるかどうかによって，映像の視聴者に与える印象が変化し，意図をうまく理解できなかった人にとっては違和感だけが残ってしまったのではないかと考えられる．また，放送局社員の評価を見ると，テレビで放映された映像よりもシステムで生成した映像の方が時に良い評価を受けている．これは，テレビで放映されている映像は一般的に音声やテロップが使われているものであり，それを前提として編集されている映像の音声とテロップを取り除いたものでは，映像内容の説明が不足し，視聴者が内容が理解しにくく，違和感のある映像になってしまったのではないか，といった評価がアンケートの際のコメントから得られた．

結果として，映像のみの素材を再編集してダイジェストとする際には，映像内容の関連性をあらかじめ定義しておき，それ

を用いることによって視聴者にとって違和感のなく，自然なダイジェストが生成できることが示された．また，映像提供者の意図を取り入れた映像に関しては提供者の意図の与え方によっては，視聴者に意図が伝わらず違和感のある映像になってしまうために，意図の与え方を注意する必要がある．また，違和感なく自然な映像が生成できるということは，ダイジェストのみならず手動編集の際に，前段階の荒編集としても用いることができると考えられる．今後は，映像のみの素材だけではなく，音声やテロップを組み合わせることを考え，それらとの関連性からより違和感の無い映像を生成できる方法を考えていくことが必要である．

7. おわりに

映像を計算機上で効率よく扱うためには，映像内容が容易に理解できるようなメタデータが必要となり，そのようなメタデータを用いた処理の一つとして自動でのダイジェスト生成がある．重要な情報が含まれている映像を効率的に取得できるため，ダイジェスト映像は非常に有用である．従来のダイジェスト自動生成手法では，再構成される際の映像の繋がり，すなわち複数の映像を並べた際の整合性や並べられた映像の関連性が考慮されていなかったため，並び方によっては映像の意図がうまく視聴者に伝わらず，違和感のある映像が生成される可能性があった．

そこで，本研究では映像内容に関する関連性を考慮したダイジェスト自動生成手法の提案を行った．素材となる映像のシーン間にそれぞれ関係を定義するデータ構造の提案を行い，さらにそれを用いて映像提供者の意図を記述する手法を提案した．映像文法に加えてこれらの手法を用いたダイジェスト自動生成システムの実装を行い，アンケートによる評価を行った．評価の結果，関連性を考慮することで違和感なく自然な流れのダイジェストが生成できることを示した．しかしながら，意図に関しては期待した結果が得られず，意図の指定の方法などを再検討する必要があると思われる．

今後の課題として，映像提供者の意図の指定方法の再検討を行う必要がある．どのような指定方法ならば提供者の意図を正確に記述でき，それを視聴者に正しく伝えることができるのかを考えなければならない．また，映像文法を個々の視聴者ごとにカスタマイズすることを考えられる．本研究で用いた映像文法は放送局で定義されたものであり，それが一般視聴者全員に当てはまるものであるとは言えない．そこで，提供される映像の満足度を上げるために，視聴者それぞれが好みの映像文法を用意し，それに応じたダイジェストの生成が可能となるシステムが必要であると考えられる．また，本研究では素材となるものは映像だけであったために出力されるダイジェストも映像のみであった．しかし，これに別の要素，例えば音声や背景音楽，テロップなどを組み合わせ，総合的な作品として出力できれば，より視聴者にとって理解しやすく，また映像提供者の意図を反映しやすいものとなる．また，映像間の関連性を表すデータ構造に関して，本研究における構造ではデータを付与する際に，出力される映像の流れが想像しにくいという問題があった．そのため，関連性のデータ付与者が多少付けにくい，ということが考えられた．そこで，今後はより分かりやすいデータ構造やインターフェースを検討していく必要があると考えられる．

謝　辞

本研究を行うにあたり，素材映像の提供と実証実験への協力をしていただいた毎日放送株式会社に深く感謝いたします．

文　献

[1] 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，” 第 12 回データ工学ワークショップ (DEWS2001), Mar. 2001.

[2] M. Kumano, Y. Ariki, K. Shunto, and K. Tsukada, “Video Editing Support System Based on Video Content Analysis,” in *Fifth Asian Conference on Computer Vision (ACCV2002)*, vol. II, pp. 628–633, Jan. 2002.

[3] Sofia Tsekeridou and Ioannis Pitas, “Content-based Video Parsing and Indexing based on Audio-Visual Interaction,” in *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, 2001.

[4] S.-F. Chang and H. Sundaram, “Structual and Semantic Analysis of Video,” in *Proceedings of IEEE International Conference of Multimedia & EXPO (ICME 2000)*, July 2000.

[5] N. Dimitrova, “The Holy Grail of Content-Based Media Analysis,” *IEEE Multimedia*, vol. 9, no. 2, pp. 6–10, 2002.

[6] R. Leonardi and P. Migliorati, “Semantic Indexing of Multimedia Documents,” *IEEE Multimedia*, vol. 9, no. 2, pp. 44–51, 2002.

[7] 吹野直紀, 角谷和俊, 田中克己, “キーワード毎のショット長分布を用いたビデオ映像シーン検索，” 情報処理学会研究報告, vol. 2002, pp. 49–56, May 1998.

[8] M. Okamoto, K. Ueda, J. kamahara, S. Shimojo, and H. Miyahara, “An Architecture of Personalized Sports Digest System with Scenario Templates,” in *Proceedings of 7th International Conference on Database System for Advanced Applications (DASFAA2001)*, pp. 170–171, Apr. 2001.

[9] M. Okamoto, J. kamahara, S. Shimojo, and H. Miyahara, “Automatic Production of Personalized Contents with Dynamic Scenario,” in *Proceedings of 2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM'01)*, vol. 1, pp. 91–94, Aug. 2001.

[10] 西澤尚宏, 鎌原淳三, 春藤憲司，塚田清志，有木康雄，上原邦昭，下條真司，宮原秀雄, “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，” 第 13 回データ工学ワークショップ (Dews2002), Mar. 2002.

[11] 是津耕司, 上原邦昭, 田中克己, “時刻印付オーサリンググラフによるビデオ映像のシーン検索，” 情報処理学会論文誌, vol. 39, pp. 923–932, Apr. 1998.

[12] H. Maeda, K. Koujitani, and T. Nishida, “Utility of weakly structured memory organization for integrating heterogeneous information,” in *Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing (ASC'97)*, pp. 284–287, Nov. 1997.

[13] A. Duda and R. Weiss, “Structured Video: A Data Type with Content-Based Access,” Tech. Rep. 580, MIT LCS Technical Report, Sept. 1993.