映像文法に基づく映像編集支援システム

天野美紀[†1] 上原邦昭[†2] 熊野雅仁[†3]
有木康雄[†3] 下條真司[†4]
春藤憲司[†5] 塚田清志[†5]

映像の編集とは，素材映像の中から編集に用いることができるショットを選択し，それらを接続する作業である．これらのショットの接続の仕方は無限に存在する．しかし，作者側の意図することを視聴者に正確に伝えることを目的として編集した場合，ある普遍的な規則が存在する．これを「映像文法」と呼ぶ．本稿では，編集作業を支援することを目的として，映像文法に基づいた自動編集システムを提案する．本システムでは，まず，素材映像からショットの切り出しと，切り出した個々のショットに対して属性値の付与が行われる．次に，映像文法をルール化したプロダクションシステムを用い，推論を重ねることによって，属性値を付与された素材映像集の中から適切なショットを選択し編集を行うようになっている．

Video Editing Support System Based on Video Grammar and Content Analysis

Miki Amano,[†1] Kuniaki Uehara,[†2] Masahito Kumano,[†3]
Yasuo Ariki,[†3] Shinji Shimojo,[†4] Kenji Shunto[†5]
and Kiyoshi Tsukada[†5]

The video editing is a work to produce the final video with certain duration by finding and selecting appropriate shots from material videos and connecting them. In other to produce the excellent video, this process is generally conducted according to the set of special rules called “video grammar”. In order to make video grammar applicable, the metadata such as shot size or camera work included in shots have to be extracted and indexed. The purpose of this study is to develop an intelligent support system for video editing system where these metadata are extracted automatically and then the video grammars are applied to them.

1. はじめに

テレビ番組やコマーシャルの素材となる映像は，ある現実の断片を記録したものにすぎず，写っている事実以外，何の意味も持たない．編集者はその事実の断片をつなぎ合わせて，ある「意味を持った」まとまりのある映像を作りあげていく．我々がコミュニケーションに用いる言語の場合，最小単位は「a」や「あ」などの文字であるが，単純に考えればその組合せは無限に存在する．しかし，意味を伝える場合，単語レベル，構文レベルでルールが存在し，制限がある中で文により意味が伝えられる．これらの単語のスペルの間違いや構文上の誤りが存在すれば，文章の意味を相手へ正確に伝えることができない．これと同じように，映像においても断片の接合の仕方は無限に存在するが，制作者側の意図することを視聴者に正確に伝えることを目的として編集した場合，ある普遍的な規則が存在する．これを映像文法[1)]と呼ぶ．逆にいうと，もしこの映像文法からはずれてショットをいいかげんに接合すると，見ている側は混乱するか，あるいは十分な情報を受け取ることができなくなる．

プロの編集者は，まず映像文法に従い，視聴者に伝える情報を確実に表現する適切な長さを持ち，冗長性

†1 神戸大学大学院自然科学研究科情報知能工学専攻
Department of Computer and Systems Engineering, Graduate School of Science and Technology, Kobe University
†2 神戸大学大学院自然科学研究科情報メディア科学専攻
Department of Information Media Science, Graduate School of Science and Technology, Kobe University
†3 龍谷大学理工学部電子情報学科
Faculty of Science and Technology, Ryukoku University
†4 大阪大学サイバーメディアセンター
Cyber Media Center, Osaka University
†5 株式会社毎日放送メディア開発局
Mainichi Broadcasting System, Inc.

のないショットを抽出して，文法上間違いを起こさないショットの接続を行う．これらの条件を満たしたうえで，映像に意味や表現，ストーリーを与えることが求められる．我々は，映像制作にかかわるプロの人々との議論を通じて，これまで開発してきた映像解析の技術を駆使すれば，映像文法を表現する要素を抽出するための属性が自動的に取得できそうなことが分かってきた☆．そこで，本稿では，編集作業上，最も高度な技術を要する意味レベルの接続に編集者が集中できるよう，それまでに必要となる，構文レベルまでの処理の自動化を行う手法に着目した．

本システムでは，まず映像文法を背景にして撮影された素材映像からカットの開始点や，カメラワークがあるかどうかといった索引の付与をほぼ自動で行う．次に編集に適したショットを自動抽出し，映像文法に従って，文法上間違いのない映像を自動編集する．これにより，編集者は，編集作業の中で多大な作業量を必要とし，非効率的な作業から開放され，知的な作業にだけ集中することができる．

本稿では，まず 2 章で関連研究について述べ，3 章で「映像文法」について説明する．次に，4 章でショットの自動抽出と索引付けの手法を説明し，5 章で提案する映像編集支援システムの実装方法について述べる．6 章で提案手法の実験，評価を行う．最後に 7 章でまとめとする．

2. 関 連 研 究

映像の編集支援システムは，これまで多くの研究がなされている．Chiueh ら[2]は，「edit history abstraction」というデータ構造を用いて，対話型ビデオオーサリングシステムを構築している．「edit history abstraction」とは，編集過程の履歴を木構造で表現するものである．「edit history abstraction」を用いることによって，編集のやり直しや，これによって生成されたビデオストリームからシーン，ショット検出を容易に行うことができる．しかしながら，このシステムでは編集に必要なショットの選択や接続に関してユーザに何の助言も与えることはできない．

Girgensohn ら[3]は，素材映像から編集に使用可能なショットを選択し，適当な開始点と終了点を与える半自動的な編集システムを提案している．このシステムでは，あるショットに対して適切な持続時間を与える判断基準として編集ルールが用いられている．しかし，ショットの接続に関しては何の編集ルールも与えられていない．

Sundaram ら[4]は，「visual complexity」および「film syntax」という 2 つの観点からビデオの要約を行っている．「visual complexity」では，ショットの複雑さを計算し，ショットごとの再生時間に上限と下限を設けて要約を生成している．「film syntax」では，「1 つの会話は，最低 3 つのショットから構成される」，「x 人による会話を表すためには，最低 $3x$ 個のショットが必要である」などのルールに基づいて，1 つの場面から不要なショットを削除して，要約を生成している．しかしながら，これらの文法的なルールはシーンを要約するためであり，編集に用いられるものではない．

本稿で提案する映像編集支援システムは，「映像文法」に基づいてショットの接続が自動で行われる．具体的には，素材映像からショットの切り出しと，切り出した個々のショットに対して属性値の付与がほぼ自動で行われる．次に，映像文法をルール化したプロダクションシステムを用い，推論を重ねることによって，属性値を付与された素材映像集の中から適切なショットを選択し編集を行うようになっている．

3. 映 像 文 法

映像はシーン，カット，ショットの 3 つを用いて論理的に構造化される．具体的には，カメラが回ってから止まるまでの連続した映像の一区切りをカットと呼ぶ．そして，同一の空間または被写体を撮影したカットの集合をシーンと呼ぶ．さらに，カメラワークの開始点や終了点を境界として，カットをさらに細かく分類したものをショットと呼ぶ．これは，ズームやパンなどのカメラワークによって，被写体の大きさや被写体自体が変わり，同一のカット内でも開始点と終了点ではまったく違う映像となることがあるためである．したがって，1 つのカット内に複数の論理的なショットが含まれることがある．これらの関係を**図 1** に示す．この図では，4 つのカットで 1 つのシーン A が構成されており，カット 2 内に，カメラが静止している状態のフィックスショット，カメラが動いているパンショット，カメラが静止したフィックスショットの 3 つのショットが含まれている．

ショットサイズは，被写体とカメラの距離，すなわち撮影されている人物，または物体の大きさに比例して決定される．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）に分類される．これらは相対的な分類であり，あるショット

☆ もちろん，映画などの高度な編集の場合，あえてこの規則をはずそうとする．しかし，TV での映像，ドキュメンタリー，バラエティなどの場合，用いられる映像文法の規則数は限られたものとなる．

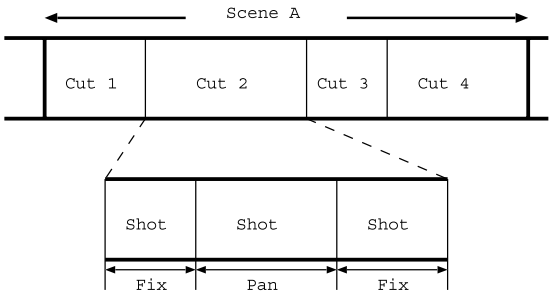

図 1 カット内に含まれる論理的なショットの例
Fig. 1 Example of a cut including three shots.

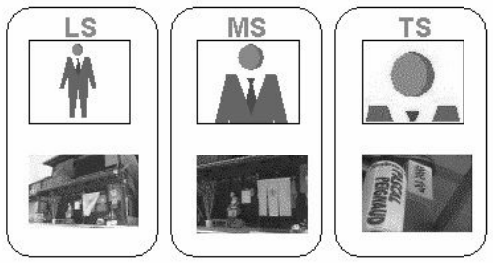

図 2 ショットサイズの例
Fig. 2 Example of shot size.

より対象に近寄ったショットをタイトショット，引いて撮ったショットをルーズショットと呼ぶ．両者の中間となるショットをミドルショットと呼ぶ．また，シーンを構成するショットの中で，シーン全体の様子が分かるようなショットのことをマスターショットと呼ぶ．たとえば 3 人が会話しているシーンの場合，3 人の位置関係が一度に分かるようなショットが，マスターショットとなる．例を**図 2** に示す．

たとえば，人物を撮影した場合，全身が写っているショットを LS，上半身が写っているショットを MS，顔しか映っていないショットを TS と分類する．建物や物の場合，全体が写っていれば LS，半分ほど写っていれば MS，ある一部分だけしか写っていなければ TS と分類する．また，これらの例の場合，両方とも LS のショットがマスターショットとなる．

視覚的に見やすい映像を作り上げる場合，映像のリズムは非常に重要な要素となる．映像のリズムとは，ショットサイズの変化と，個々のショットの継続時間によって，映像に視覚的，時間的アクセントを加えるものである．映像のリズムを考慮せず，似通ったショットサイズが多数連続したり，どのショットも長短のない秒数で編集したりした場合，視聴者はどのショットが重要なのかを判断しにくく，編集者側の伝えたい意図を映像から得ることは難しい．また，動きや変化のないショットが不必要に続く場合，視聴者は映像を見ることに疲れ，飽きてしまう．逆に，情報量が多いショットをあまりにも短く編集すると，編集者側の意図したことが視聴者に十分に伝わらないことがある．したがって，ショットサイズの変化や，個々のショットにどれだけの秒数が必要かを考えて，リズム感のある映像として編集する必要がある．

カメラワークには，カメラが静止している状態のフィックス，カメラを固定してフレームを変化させるパン，ズームなどがある．フィックスは曖昧さを残さず，見やすく分かりやすい映像である．逆に，パンやズームなどはきわめて曖昧さを多く残す映像になるので，カメラが動く前の開始点と動いた後の終了点は，フィックスを用いて映像を静止させなければならない．このように，編集作業には映像の意図を正確に伝えるために，従わなければならない規則が存在する．以下に本稿で使用する編集規則を示す．

規則 (1) ショットサイズが急激に変化するものはつなぐことができない．

規則 (2) シーンの冒頭はマスターショットで始まる．

規則 (3) LS は 6 秒，MS は 4 秒，TS は 2.5 秒程度の長さとする．

規則 (4) パン，ズームなどは開始点と終了点をフィックスさせて編集する．

たとえば LS の次に TS を接続すると，両者の関係がつかみにくく，見にくい映像となってしまう．このため規則 (1) のような規則がある．また，LS の方が TS よりも多くのものを写しこんでいるため，情報量も多いものとなる．したがって，規則 (3) のようなリズムで編集する必要がある．

4. ショットの自動抽出と索引付け

カメラマンは，映像文法を背景としながら，映像制作者の伝えたい意図を反映する映像をシーンごとに撮影する．本稿では，1 つの映像を作成するために撮影された未編集の映像の素材を「素材映像」と呼んでいる．また，素材映像上のシーンとは，完成された映像上の意味的なシーンではなく，その意味的なシーンを撮影するための候補となる，同じ空間または同じ被写体を撮影したカットの集合である．この素材映像上のシーンは，シーンごとに撮影が行われる関係上，例外がないかぎり，素材映像中，連続している．本稿で提案する映像編集支援システムは，事前に行われる処理として，この素材映像を入力とし，カット点，使用可能または使用不能なカメラワーク区間，フィックス区間の検出とショットサイズ付与などの素材映像に対する索引付けの大部分が自動的に行われる．

素材映像上のカット点はカメラのON，OFFに対応するが，カット点の自動検出については，我々が提案するバッファリング法[5]を用い，現在リアルタイム処理で，再現率94.3%，適合率90.9%を達成している．ここで，C を検出対象の正解数，D を正解を検出できなかった数として「未検出数」，E を正解でないものを過剰に検出した数として「過剰検出数」とするとき，再現率 $= C/(C+D)$，適合率 $= C/(C+E)$ と定義される．再現率は，検出対象を漏れなく検出できたかという完全性を表現し，適合率は，検出結果の中にどれだけ必要な対象が存在するかという正確性を表現する指標である．システムの性能を評価する場合，漏れがなく，必要な対象だけを抽出することが目的となるため，この再現率と適合率ともに高い値を示すことが求められる．

素材映像上では，場面ごとにカットが撮影されるため，素材映像上のカット点のうち，いくつかは，素材映像上でのシーンの切れ目を表している．この特徴を用い，我々が提案する，色調と平均画像を用いたシーンの切れ目判定法[6]を適用した結果，再現率92%，適合率48%を達成している．ただし，後段の処理である，ショットサイズの自動付与やショットの自動接続にとって，素材映像上のシーンの切れ目は非常に重要である．そこで，シーンの切れ目判定を行った後，手動で訂正を行う．素材映像は，長くて25分程度であり，シーン数は多くても10前後である．適合率が50%を切っているが，シーン数が少なく，過剰検出されたシーンの切れ目はたかだか10前後となるため，作業量は少ないものと考えられる．

また，映像文法を背景としてショットを切り出す場合，フィックス区間を切り出すだけでなく，カメラワークの前後をフィックスするショットも必要となる．しかし，素材映像上には，手ぶれやカメラワークの失敗，また無意味である無造作なカメラワークなど，放送用に使用できない使用不能区間が存在する．ショットの自動接続を行うためには，これらの使用不能区間を推定し，放送用の映像には採用されないように判定する機能が必要となる．使用不能区間は，主に不安定なカメラワークに起因するため，使用不能なカメラワークを推定するためには，カメラワークの変化量について高い精度を必要とする．そこで，現在，高い精度を持ち，リアルタイム処理も可能である長坂らによる投影法[7]をカメラワーク検出法として用いている．この検出結果を用いて使用可能・不能区間を推定する手法[5]により，現在，手ぶれ区間については，再現率81%，適合率55%，また，使用できないカメラワークについては，再現率88%，適合率100%を達成しており，フィックス区間の抽出では，再現率91%，適合率94%を達成している．

本稿で提案する映像編集支援システムでは，索引情報として特にショットサイズが重要な役割を果たすことになるが，このショットサイズの自動付与法については，文献8)で報告しており，ショットサイズ付与正解率70.6%，ショットサイズ付与率99.2%を達成している．本稿では，これらの手法を用いて索引付けされた素材映像から自動的にショットを抽出して処理が行われる．

5. 映像編集支援システム

5.1 属 性 値

前章までに述べた映像の文法的要素と自動索引付け，ならびに規則を考慮して，素材映像集に付与すべき属性値を検討した．**表1** に属性値を示す．属性CutID，ShotID，CameraWork，StartFrame，EndFrame，ShotSizeはそれぞれ4章で述べた手法を用いてすべて自動インデクシングされる．SceneIDは4章で述べたとおり，色調と平均画像を用いたシーン判定を行い，過剰に検出されたシーンの切れ目を手動により訂正したものを使用した．また，マスターショットかどうかの判断は，シーンを構成するショットの集合

表1 属性値の一覧
Table 1 Table of attributes.

属性名	説明	型	例
SceneID	シーンの識別番号	Integer	8, 8
CutID	個々のカットの識別番号	Integer	3, 4
ShotID	個々のショットの識別番号	Integer	2, 1
StartFrame	ショットが開始するフレーム番号	Integer	1250, 1500
EndFrame	ショットが終了するフレーム番号	Integer	1350, 2000
CameraWork	カメラワーク	Text	Zoom, Fix
ShotSize	ショットサイズ	Text	-, LS
Master	マスターショットとして使えるかどうか	boolean	0, 1
Used	すでにその映像を編集したかどうか	boolean	0, 0

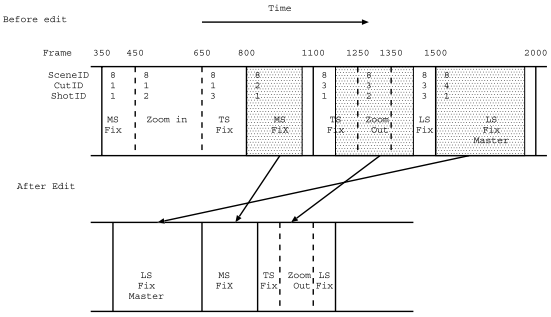

図 3 編集過程の概要
Fig. 3 Editing process.

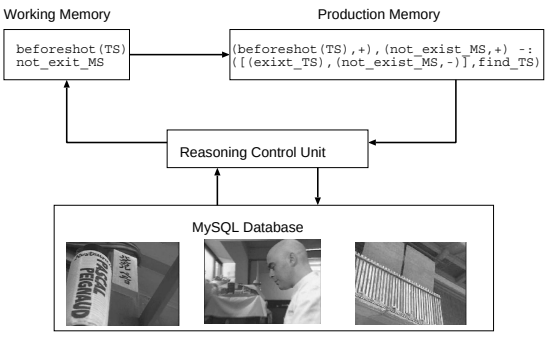

図 4 編集支援システムの概要
Fig. 4 Video editing support system.

に何が映っているかに依存するため，自動で厳密に判断するのは困難である．しかしながら，3 章の図 2 で示したように，マスターショットは LS として撮影されることがほとんどである．そこで本研究では，すべての LS をマスターショットとして使用可能だと判断し，属性 Master を自動インデクシングした．

表 1 の例は**図 3** に対応している．素材映像中の各ショットは，属性値 SceneID，CutID，ShotID の組によって一意に識別される．たとえば，図 3 中の右端のショットは SceneID 8，CutID 4，ShotID 1 である．このショットは，マスターショットとして使用可能なため，属性 Master の値は 1 になる．さらに，ルーズショットなので ShotSize は LS，カメラが固定されているので CameraWork は Fix である．

5.2 編集過程

図 3 をもとに編集過程の概要について述べる．図中で影になっている部分が編集に用いる映像を表している．まず初めに，SceneID 8，CutID 4，ShotID 1 のショットが規則 (2) より選択される．このショットはカメラが固定しており，ルーズショットなので，ルール (3) より素材映像から 6 秒間抜き出して編集に用いる．LS に接続可能なショットサイズは MS だけなので（規則 (1)），次のショットは CutID 1，ShotID 1 または，CutID 2，ShotID 1 である．規則 (1) を満たすショットが複数存在した場合，1 つ前に編集されたショットと時間的に一番近いショットが編集に用いられる．よって 2 つ目のショットは CutID 2，ShotID 1 のショットとなる．このショットは MS なので，素材映像から 4 秒抜き出して編集に用いられる（規則 (3)）．MS に接続可能なショットサイズは LS と TS である（規則 (1)）．3 番目のショットとして CutID 3，ShotID 1 のショットが選ばれる．このショットには，連続して ShotID 2 のズームショットと ShotID 3 のフィックスショットがあるので，規則 (4) より ShotID 2 の前後 1 秒が抜き出されて編集に利用される．この場合，ショット終了時のショットサイズは LS になるので，次に接続可能なショットサイズは MS である．このように，規則に合致する編集可能なショットがなくなるまで編集を繰り返し，すべてのシーンを編集する．

5.3 編集支援システムの構成

属性値を付与した素材映像集を用いて，映像文法に基づいて次に接続すべき映像を素材映像集から選択し，自動的に編集する手法について述べる．編集支援システムのエンジンには前向きプロダクションシステムを用いている．個々の素材映像集は，表 1 に示した属性値とともに，ショット単位で MySQL データベースに格納されている．編集支援システムの概要を**図 4** に示す．Production memory は編集規則の集合，Working memory は編集過程の各時点で求まる事実の集合である．Reasoning Control Unit では，編集規則の競合を解消した後，適切な編集規則を実行する．

本システムでは，“if **condition** then **action**” というプロダクションルールを，3 章で述べた編集規則に対応付けている．たとえば，規則 (1) は「前のショットが LS ならば，次のショットは MS である」，「前のショットが TS ならば，次のショットは TS である」などいくつかの細かいルールに分けられ，これらのルールは以下のように表現される．

- beforeshot(LS) −: ((exist_MS,+),find_MS)
- beforeshot(TS) −: ((exist_TS,+),find_TS)

このように，4 つの編集規則をすべてプロダクションルールで表すと，総数 27 個となった．

action 部は，アクションのリストと，MySQL データベースに接続するための述語から構成される．個々の述語は真か偽かを返し，述語が真を返したときにのみ，そのルールが実行される．たとえば，述語 find_MS は，MySQL データベースに接続し，まだ編集されていないミドルショットがあるかどうかを調べる．もし，ミドルショットがあった場合，find_MS は真を返し，新

たな事実 exist_MS が Working memory に加えられる．この操作を繰り返すことによって推論を進め，素材映像集から次に接続するショットを選択し，映像の自動編集を行うようになっている．

また，本システムでは編集規則に優先度をつけている．1つは，ショットサイズの接続に関する優先度である．たとえば，TS の次には TS と MS が接続可能だが，TS から TS の接続より，TS から MS への接続の方が優先度が高い．なぜならば，TS から TS の接続の優先度を高くしてしまうと，TS が多数あった場合，TS が連続してしまいショットサイズの変化がなく単調な映像になってしまうからである．もう1つはカメラワークに関する優先度である．図3 の SceneID 8，CutID 3，ShotID 1 のショットのように，フィックスショットの次にパンショットなどのカメラワークがあるショットが続く場合，そのショットをフィックスショットとして規則 (3) のように 2.5 秒抜き出して編集するよりも，次に続くカメラワークを含んで編集する方が優先される．これは，素材映像を撮影したカメラマンは，カメラワークがある部分をメインに撮影したかったと考えられるからである．競合が生じた場合は，この優先度が高い編集規則が優先的に実行される．

なお，プロダクションシステムは Prolog で実装し，Prolog Cafe というトランスレータを用いて Prolog から Java へ変換している．プロダクションシステムと MySQL データベースのインタフェース部は Java で記述している．

6. 映像文法による編集支援システムの実験

6.1 実験対象

本研究では，毎日放送（株）より提供された放送用の素材映像を実験対象として用いている．映像のフォーマットは H.263 形式で，1 秒間は 30 フレームで構成される．映像文法を用いた自動編集の実験には 4 つの素材映像を用意し，material 1〜4 の番号を割り振った．素材映像の詳細を**表 2** に示す．これらの素材映像は，飲食店の紹介を行う番組用に撮影された映像であり，それぞれ店の外観，内装，料理，調理方法などのシーンが含まれている．

6.2 素材映像の使用率による編集評価

素材映像は，1つ1つのショットが不必要に続いていたり，リテイクのショットが多数あったり，冗長的でかなり映像時間が長い．そこで素材映像からどれだけ冗長性をなくし，編集結果を凝縮できたかを示すために，素材映像ごとにショット使用率，フレーム使用率を算出したものを**表 3** に示す．編集に使用したショットの割合を Shot，編集に使用したフレームの割合を Frame と定義する．素材映像の総ショット数を AS，総フレーム数を AF，編集したショット数を ES，編集したフレーム数を EF としたとき，Shot Rate，Frame Rate は式 (1)，(2) で算出される．

表 2 実験対象
Table 2 The experiment movies.

	mat1	mat2	mat3	mat4
シーン数	5	4	5	19
カット数	31	28	34	41
ショット数	180	54	88	141
フレーム数	45,440	20,352	20,385	34,177
内容	焼肉店	カフェ	ラーメン店	ラーメン店

表 3 カットとフレームにおける利用率
Table 3 The rate of use in cut and frame.

material number	Shot Rate (%)	Frame Rate (%)
material 1	44.1	12.8
material 2	72.5	11.9
material 3	53.6	16.9
material 4	41.0	15.3

$$Shot\ Rate(\%) = \frac{ES}{AS} \times 100 \quad (1)$$

$$Frame\ Rate(\%) = \frac{EF}{AF} \times 100 \quad (2)$$

表 3 の Frame Rate より，もともとの素材映像の10〜20%の長さで編集できていることが分かる．Shot Rate については 40〜70%の間でばらつきがある．これは，素材映像の各シーン内のショットサイズのバラつきが原因となっている．たとえば，規則 (1) より，ショットサイズが急激に変化するショットを接続することはできないため，LS と TS は接続不可能である．よって，素材映像内に MS が極端に少なかった場合，編集に用いられなかった LS と TS が多数残り Shot Rate は低いものとなる．また，規則 (3) より，各ショットにはショットサイズに見合った持続時間を与えなければならないが，素材映像内の各ショットがそれより短かい場合，そのショットを編集に用いることはできない．このようなショットが素材映像に多数存在した場合，Shot Rate が低くなる．

表 3 より，素材映像をかなり短縮できたことが分かる．しかしながら，単純に短く編集できたというだけで，本システムの有用性を示すことはできない．たとえば，素材映像のシーン内に TS が大量にあった場合でも，TS と TS との接続は可能なので，TS が連続した見にくい映像に編集される可能性がある．

また，ショットサイズと持続時間に関して規則 (3) があるが，これは各ショットの情報量を考慮したもの

である．一般的に，LS は TS よりも多くの対象物を 1 ショット中に写しこんでいるため，含まれる情報量も多いと考えられる．よって，LS には 6 秒と長い継続時間を与えられている．しかしながら，ショットサイズは相対的な尺度なため，情報量の少ないショットでも他のショットとの関係によって，LS となりうる．その結果，ショット中に含まれる情報量によって，ある LS では 6 秒という長さが冗長であると受け取られたり，短すぎると受け取られたりする場合がある．このように，もし各ショットの情報量が持続時間に見合っていなければ，見にくい映像となってしまう．そこで，これらの特性を考慮して，編集された映像の「品質」と「情報量」を定義する．

6.3 品質による評価

先ほども述べたように，ショットサイズの接続に関する規則には優先度が定められている．優先度の高い規則が用いられる割合が高い方が，映像的に見やすく品質が高いと考えられる．この品質を Quality Rate として定義する．すなわち，編集された映像の全ショット間の接続数のうち，優先度の高い規則が用いられている数を High，優先度の低い規則が用いられている数を Low とすると，Quality Rate は式 (3) で表される．

$$Quality\ Rate(\%) = \frac{High}{High + Low} \times 100 \quad (3)$$

素材映像ごとの Quality Rate を算出した結果を**表 4** に示す．Quality Rate が極度に低くなるようであるならば，優先度の低い規則を使用する限度数が必要であったか，あるいは，優先度のつけ方に問題があったなど，システム異常が考えられる．しかし，表 4 から，3 分の 2 以上のショット接続において優先度の高い接続が行われており，システム上の問題はなかったといえる．また，ショットサイズに変化があり，見やすい映像ができているといえる．

6.4 情報量による評価

次に，情報量の観点から編集された映像を評価するために，映像の保持する情報量として，画像の視覚的な複雑さを取り上げる．まず，情報量が多く視覚的に複雑なショットほど，長い持続時間が与えられていると考えられる．そこで，編集された映像の各ショットに対して，視覚的な複雑さに応じた持続時間が与えられているかどうかを検証する．いい換えると，リズムに関する規則では，フィックスの LS，MS，TS に対してそれぞれ，6 秒，4.5 秒，3 秒の持続時間が与えられているが，そのことと視覚的複雑さとの対応について検証する．

映像の要約に関する研究[4]では，画像の視覚的複雑性と視聴者の理解時間の相関について述べられている．この研究では，ショットの視覚的な複雑さとショットの非圧縮性が比例することが示されている．また，個々のショットの非圧縮性は，普遍的なデータ符号化手法である Lempel-Ziv 圧縮（LZ 圧縮）アルゴリズムを用いて算出することができる．したがって，ショットの LZ 圧縮を計算すれば，ショットの視覚的複雑さを定義できることになる．いい換えると，LZ 圧縮しにくいショットほど，情報量が多く視覚的に複雑なショットであると判断できる．

そこで，編集された映像に含まれる全ショットのうち，カメラワークを含まないものについて先頭フレーム画像を取り出し，LZ 圧縮による圧縮率を算出した．圧縮前の画像の容量を Normal，圧縮後の容量を Compressed とすると，圧縮率 Compressibility は式 (4) で表される．

$$Compressibility(\%) = \frac{Compressed}{Normal} \times 100 \quad (4)$$

Compressibility の値が高いほど圧縮されにくく，情報量が多いショットだといえる．LS，MS，TS ごとに平均をとったものを**表 5** に示す．カメラワークを含むものについては，カメラワークのすべてを含むように編集せよという規則があり，ショットサイズに関係なく持続時間が決定されるために，評価対象から除外している．

表 5 より，material 2，3 では LS の値が一番高く，圧縮されにくいことが分かる．よって，LS の情報量は多いと考えられる．しかし，material 4 では TS に最も高い数値が出ている．この原因として，自動索引によるショットサイズの判定に誤りが含まれているこ

表 4 品質の評価値
Table 4 The evaluation value of quality.

material number	Quality Rate
material 1	68.2
material 2	66.7
material 3	78.6
material 4	72.7

表 5 ショットサイズごとの圧縮率の平均
Table 5 The average of the compressibility for every shot size.

material number	LS (%)	MS (%)	TS (%)
material 1	79.6	73.8	75.4
material 2	93.4	80.4	74.6
material 3	85.5	78.8	76.9
material 4	88.7	88.0	91.6
total	87.2	83.4	82.6

表 6 被験者から得た回答結果
Table 6 Test scores from 14 users.

	mat1	mat2	mat3	mat4
ショットの選択	2.4(3.2)	3.3(3.5)	2.4(3.3)	3.1(3.4)
ショットの継続時間	3.4(3.8)	3.9(3.9)	3.9(3.6)	3.6(3.6)
内容の一貫性	1.9(3.5)	3.0(3.6)	2.0(3.3)	2.7(3.6)
映像の見やすさ	2.9(3.3)	4.0(3.9)	2.9(3.4)	3.3(3.6)
全体の長さ	1.6(2.6)	3.3(3.9)	1.9(3.6)	2.0(3.3)

と，また，ラーメンの細かい具などが映しこまれていて，TS にもかかわらず情報量が多いものが多数あったためである．全体としては，LS の値が一番高く，LS が一番圧縮されにくい視覚的に複雑なショットとなっている．したがって，LS に長い持続時間を与える必要がある．このように，規則 (3) を用いて，LS を 6 秒，MS を 4 秒，TS を 2.5 秒として編集した結果，各ショットサイズが保持している情報量と見合った持続時間が与えられ，リズムのある見やすい映像に編集されたといえる．

6.5 主観評価

自動編集された映像を 5 人の編集の専門家と 14 人の一般被験者に視聴してもらい，以下の 5 つの指標から本手法を評価した．第 1 の指標は，編集に適したショットを選択したかどうかを示す．2 番目に，1 つ 1 つのショットの継続時間の妥当性，3 番目に内容の一貫性，4 番目に映像の見やすさ，5 番目に編集された映像全体の長さの妥当性を評価してもらった．これらの 5 項目について，被験者から得た回答を**表 6** に示す．括弧内の数値は一般被験者の評価値を表している．ただし，評価値はそれぞれの平均値であり，5 が最も良い評価値である．

表 6 より，mat1 では「ショットの選択」,「内容の一貫性」,「全体の長さ」が低いことが分かる．これは，素材映像中に肉を切ったり，焼いたりしているなど，動作が含まれているショットが多く，本システムでは，被写体の動作を示す属性値を付与していなかったため，動作の途中でカットが切れるショットが多数あった．そのため，視聴者に不自然な印象を与え，ショットの選択が低くなったと考えられる．また，mat1 は焼肉店の映像であり，本来なら肉の種類によってシーンを区別しなくてはならないが，どの肉も赤色のため，色調を用いたシーン判定では肉の種類の区別は難しく，すべて 1 つのシーンとしてインデクシングされてしまった．そのため，複数の種類の肉が交互に編集されてしまい，内容の一貫性が低くなった．さらに，表 5 に示すように，すべてのショットサイズでは，他の素材映像よりも情報量が少ないにもかかわらず，同じリズムで編集している．その結果，1 つ 1 つのショットの継続時間では評価に影響はなかったが，全体としてみると，冗長的で長すぎる印象を与えてしまった．

mat2 では，「ショットの継続時間」,「映像の見やすさ」,「全体の長さ」の 3 つの項目において高い評価を得た．これは，表 5 に示すように，各ショットサイズにおいて情報量の差が顕著に表れている．その結果，情報量にみあった編集がなされ，リズムのある見やすい映像に編集できたからだと考えられる．mat3 では，自動編集された映像中に，カメラのパンニングをやめて引き戻しをしているカットや，パンニングの途中で映像が切れてしまっているカットなどがあった．これは，使用不能区間の推定でカメラワークの失敗を検出できなかったショットや，カメラワークを正しく検出できなかったショットがあったためである．そのため，「ショットの選択」が低い評価となった．mat4 では他の項目が良好なのにもかかわらず，「全体の長さ」において低い評価を得てしまった．これは，素材映像にリテイクのショットが多数あり，本システムではリテイクを示す属性がなかったため，同一シーン内できわめて類似したショットが何回も連続して編集されてしまった．その結果，視聴者に冗長な印象を与えてしまったと考えられる．

また，全体的に「ショットの継続時間」と「映像の見やすさ」において，高い評価を得ている．これは，規則 (1) や規則 (3)，または優先度を用いて確実に定義付けができていたからであると考えられる．一方，「シーンの一貫性」や「ショットの選択」はショット内に重要な動作や音声があった場合，それを示す属性と規則がなかったため，低い評価となってしまった．今後これらを示す属性や規則を増やし，より高い評価を得るのが課題である．

以上の結果から，ショットのリズムや接続順序を考慮した映像文法を用いることによって，見やすい映像が編集できたといえ，本システムの有用性を示すことができた．

7. おわりに

本稿では，「映像文法」に基づく自動編集支援システムを構築した．また，映像文法を用いるためにショットの自動抽出，自動索引付けを行う手法を提案した．自動編集支援システムは，映像文法をルール化し，前向きプロダクションシステムを用いて実現した．自動索引付けにより文法要素を抽出するのに十分な索引付けが可能なこと，また，映像文法を用いて自動編集を行い，見やすい映像が編集されることを示した．今後の

課題としては，音声や動作にもカメラワークと同じように開始点と終了点を示す属性を付与し，必ず音声や動作がある部分を取り出して編集するような規則を導入することや，リテイクのような類似しているショットが多数あった場合，それを判別して編集に用いない規則などを導入することがあげられる．また，制作者が編集映像の希望収録時間を設定した場合，その時間間隔に可能な限り近い範囲で編集できるようなシステムを考えている．

参 考 文 献

1) ダニエル アリホン，岩本，出口（訳）：映画の文法，紀伊国屋書店 (1980).
2) Chiueh, T.-C. and Mitra, T.: Zodiac: A History-Based Interactive Video Authoring System, *Proc. ACM Multimedia '98*, pp.435–443, ACM Press (1998).
3) Girgensohn, A. and Borecxk, J.: A Semi-automatic Approach to Home Video Editing, *Proc. UIST '00*, pp.81–89, ACM Press (2000).
4) Sundaram, H. and Chang, S.-F.: Condensing Computable Scenes Using Visual Complexity and Film Syntax Analysis, *Proc. ICME 2001*, pp.389–392 (2001).
5) Kumano, M. and Ariki, Y.: Automatic Useful Shot Extraction for a Video Editing Support System, *MVA2002*, pp.310–313 (2002-12).
6) 坂江伸悟，林義　文，熊野雅仁，有木康雄，春藤憲司，塚田清志：素材映像中のカット点検出と色調によるシーン判定，電気関係学会関西支部連合大会，G18-5 (2001-11).
7) 長坂晃朗，宮武孝文：時間変化領域の画像相関に着目した実時間ビデオモザイク，電子情報通信学会論文誌 D-II，Vol.J82, No.10, pp.1572–1580 (1999).
8) 熊野雅仁，有木康雄，上原邦昭，下條真司，春藤憲司，塚田清志：映像編集支援システムのためのショットサイズ自動付与，電子情報通信学会論文誌 D-I，Vol.J85, No.7, pp.592–602 (2002).

(平成 14 年 6 月 6 日受付)
(平成 15 年 1 月 7 日採録)

天野　美紀

平成 13 年神戸大学工学部情報知能工学科卒業．同年同大学大学院自然科学研究科情報知能工学専攻博士前期課程入学．映像編集に関する研究に従事．

上原　邦昭（正会員）

昭和 53 年大阪大学基礎工学部情報工学科卒業．昭和 58 年同大学大学院博士後期課程単位取得退学．大阪大学産業科学研究所助手，講師，神戸大学工学部情報知能工学科助教授，同都市安全研究センター教授を経て，同大学院自然科学研究科教授．平成元年より 2 年まで Oregon State University，Visiting Assistant Professor．工博博士．マルチメディア処理の研究に従事．1990 年度人工知能学会研究奨励賞受賞．人工知能学会，電子情報通信学会，計量国語学会，日本ソフトウェア科学会，AAAI 各会員．

熊野　雅仁（正会員）

平成 2 年立命館大学理工学部情報系コース卒業．平成 2 年龍谷大学理工学部実験助手，現在に至る．映像編集に関する研究に従事．電子情報通信学会，日本音響学会各会員．

有木　康雄（正会員）

昭和 49 年京都大学工学部情報工学科卒業．昭和 51 年同大学大学院修士課程修了．昭和 54 年同大学院博士課程修了．昭和 55 年京都大学工学部情報工学科助手．平成 2 年龍谷大学理工学部電子情報学科助教授，平成 4 年教授，現在に至る．工学博士．昭和 62 年〜平成 2 年エディンバラ大学客員研究員．画像処理，音声情報処理に従事．日本音響学会，人工知能学会，画像電子学会，IEEE 各会員．

下條　真司（正会員）

昭和 56 年大阪大学基礎工学部情報工学科卒業．昭和 61 年同大学大学院基礎工学科博士課程修了．工学博士．同年大阪大学基礎工学部情報工学科助手．平成元年同大学大型計算機センター講師．平成 3 年同助教授．平成 10 年同教授．平成 12 年同大学サイバーメディアセンター教授．現在に至る．LAN のアクセス方式の性能評価，分散処理システムの性能評価，オペレーションシステムの研究に従事．

春藤　憲司

昭和 43 年関西学院大学社会学部卒業．昭和 44 年（株）毎日放送入社．テレビ番組制作に従事．平成 10 年同メディア開発局，平成 13 年同ソフト企画局兼メディア開発局チーフ・プロデューサー．映像制作の知的構造化，権利ビジネス開発に従事．

塚田　清志

昭和 54 年大阪大学工学部通信工学科卒業．同年（株）毎日放送入社．放送技術に従事．平成 13 年デジタル計画部長．データ放送・インターネット等デジタルメディアに関する研究，放送のデジタル化全般の開発業務に従事．映像情報メディア学会，サイバー関西プロジェクト委員．