



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

熊野, 雅仁 ; 有木, 康雄 ; 上原, 邦昭 ; 下條, 真司 ; 春藤, 憲司 ; 塚田, 清志

(Citation)

電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, J85-D-1(7):592-602

(Issue Date)

2002-07-01

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

Copyright (c) 2002 IEICE

(URL)

<https://hdl.handle.net/20.500.14094/90000834>



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

熊野 雅仁[†] 有木 康雄[†] 上原 邦昭^{††} 下條 真司^{†††}
春藤 憲司^{††††} 塚田 清志^{††††}

Automatic Shot Size Discrimination for a Video Editing Support System

Masahito KUMANO[†], Yasuo ARIKI[†], Kuniaki UEHARA^{††}, Shinji SHIMOJO^{†††},
Kenji SHUNTO^{††††}, and Kiyoshi TSUKADA^{††††}

あらまし デジタル時代の到来により、映像コンテンツの不足が問題となっている。この問題を解決するためには、映像コンテンツの制作において最も時間を必要とする編集作業の効率化が必要である。一般に、放送用の映像は、映像内容を一意に伝えるための普遍的な規則である映像文法に従って表現されている。したがって、編集作業を効率良く行うためには、この映像文法を反映した映像編集支援システムを開発する必要がある。この映像編集支援システムを実現するためには、映像文法が適用できるように、カメラマンが撮影した素材映像に対して、あらかじめ索引情報を付与しておく必要がある。映像文法の中で特に重要な規則は、ショットの接続に関する規則である。この点から論文では、映像文法に基づく映像編集支援システムを提案するとともに、ショットサイズを自動付与する手法を提案する。

キーワード アクティブ探索, 映像文法, ショットサイズ, 映像編集

1. ま え が き

デジタル時代の到来により、放送のデジタル化をはじめ、インターネットや携帯端末への映像コンテンツ配信など、映像関連の新たな情報環境が創出されつつある。このような環境においては、これまで以上に大量の番組作成が必要となる。また、アナログ時代、デジタル時代を問わず映像の制作において最も時間を要する作業は編集である。しかし、デジタルコンテンツの作成に対しては、これまでの編集作業に新たなサービスを提供するための索引付与作業など、多大な作業時間を要する作業が加わるため、映像コンテンツの制作効率が悪化する。このような背景のもと、コンテンツ不足と人材不足が問題となっている。映像コ

ンテンツ提供者がこうした問題に対応していくためには、従来より効率的な番組制作技術、または新しいコンテンツ制作技術の開発と人材の育成が必要となる。したがって、従来より効率的な番組作成を行うためには、この映像編集作業を効率良く行う必要がある。

ところで、放送用映像を制作するまでの編集作業には、熟練した技術と労力を必要とする。編集作業における熟練した技術とは、長年の経験によって培われ、大衆に受け入れられる映像表現の基本技法であり、映像文法と呼ばれる。映像は極端に言えば、10年の話を10分の世界に押し込むことも可能であるが、我々はこの10分の映像を自然に受け入れたい。このためには、番組制作者の伝えたい意図が誤解を招かないように表現され、意図を自然にかつ一意に大衆へ伝える映像文法に従って番組を制作しておく必要がある。

図1は、映像文法の一例を示している。この例では、Shot 1で男女が向き合っている。次のShot 2では、女性がアップショットとなっている。これは女性が男性に話かけている状況を想定させ、自然な接続となっている。Shot 3では二つの接続例が示されている。上のショットは特に問題はないが、下のショットでは違和感を感じる。この理由は、Shot 1で男性が左を向いていたのにもかかわらず、Shot 3では右を向い

[†] 龍谷大学理工学部電子情報学科, 大津市
Faculty of Science and Technology, Ryukoku University,
Otsu-shi, 520-2194 Japan

^{††} 神戸大学都市安全研究センター, 神戸市
Research Center for Urban Safety and Security, Kobe Uni-
versity, Nada, Kobe-shi, 657-8504 Japan

^{†††} 大阪大学サイバーメディアセンター, 茨木市
Cyber Media Center, Osaka University, 5-1 Mihogaoka,
Ibaraki-shi, 567-0047 Japan

^{††††} (株) 毎日放送メディア開発局, 大阪市
Mainichi Broadcasting System, Inc., 17-1 Chayamachi,
Kita-ku, Osaka-shi, 530-0013 Japan

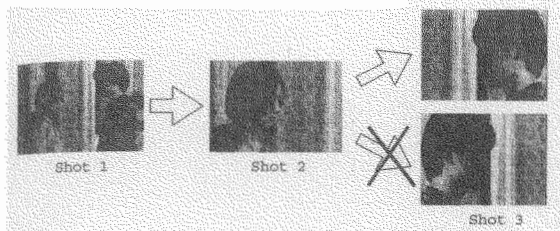


図1 映像文法の例
Fig. 1 Example of video grammars.

ているからである。こうした接続を行うと、空間的な関係が崩れ、複数の解釈や何か特別な意図があるのではないかと視聴者を混乱させるため、よくない接続とされる。映像文法とは、こうした規則の集大成である。

このような接続規則の中で最も基本的な索引情報とされるのはショットサイズである。ショットサイズには、対象を遠くから撮影したルーズショット (LS)、近くで撮影したミドルショット (MS)、より対象に接近したタイトショット (TS) があり、これらの接続によって時空間の抽象化が行われるのである。

本論文では、この映像文法をコンピュータ上で表現し、映像文法に基づいて映像編集を支援するシステムを提案する。この支援システムは、編集者の人材育成を目的とした映像文法の学習支援システムにも拡張可能である。また、このシステムにより、映像作成現場で培われた表現技法を、一般家庭における映像編集に用いることができるため、現状の市販ソフトよりレベルの高い編集が可能となる。

映像編集支援システムにおいて、映像文法を使用するためには、図1に示されるようなショットのサイズや、人間の顔の方向といった映像から計算されるメタデータを抽出しておく必要がある。この理由から、本論文では映像文法を使用する上で必要となるショットサイズを自動付与する方法について提案する。

まず、2.では、素材映像と編集作業の特徴について述べ、3.では映像編集支援システムの概観を述べる。また、4.では、ショットサイズとショットサイズに関する映像文法について、5.ではショットサイズを付与する方法について述べ、6.で実験の結果をまとめる。

2. ショットと素材映像の特徴

2.1 ショットの定義

編集された映像は、概念的に図2のような階層構造を形成している。

編集された映像は、シークエンスの接続により構成

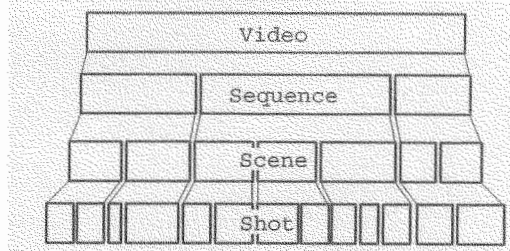


図2 編集された映像の構造
Fig. 2 Structure of edited video.

され、シークエンスはシーン、シーンはショットの接続により構成されるという構造である。シークエンスは1映像内の1エピソードとして定義され、ある程度の話がわかる単位である。シークエンスの一部としてのシーンは通常、場面と定義されるが、シーン単独では話がわからないような単位である。シーンの明確な定義ははっきりしていないが、色調の変化がシーンを区別するという見方がある。映像に関する言葉の定義においては、文献により微妙なニュアンスの違いが存在するため、最も正しい定義が何であるかを定めることはできない。以下には本論文で用いている定義について述べる。

ショットは構図や撮影対象に依存した意味をもっており、時間方向にとぎれのない一連の連続したフレームの集合である。このショットは、映像の最小単位と定義されている。また、1ショット内にはカメラワークの存在が許されている。ここでいうカメラワークとは、カメラの軸位置を固定した上でカメラの撮影方向を変化させるパン (PAN: チルトはパン動作の中で上下の動作に特化した呼び方であり、ここでは使用しない)、またカメラのレンズ操作により、撮影対象を拡大縮小するズーム (ZOOM) に大別される。カメラワークが存在していない部分はフィックス (FIX) と定義する。

1ショット内にカメラワークが存在すると、構図や撮影対象が変化するため、カメラワーク後にはショットが変化すると見ることもできるが、本論文ではカメラワークに依存したとらえ方として図3に示すような二つのタイプをショットと呼んでいる。

Type 1はFIXのみのショットであり、Type 2はカメラワークの前後にFIXが存在するものである。映像文法ではカメラワークの前後に必ずFIXを入れることを規定している。Type 2はこの映像文法を具体化したショットである。ショットの接続において、ショットの切り換わり後、突然カメラワークがある場合、視

聴者は急激な変化に対応できないか、何らかの違和感を覚える。この規定はこうした状況避けるために設けられている。

2.2 素材映像の特徴

カメラマンは映像文法を意識しながら、ディレクターの意図を伝えるためにショットを撮影する。そのカメラマンが撮影した素材映像にはいくつかの特徴がある。図4は素材映像の構造的な特徴を表している。

編集された映像では、カット点はショットを区切る変化点である。しかし、カメラマンの撮影した素材映像では、カット点の意味が編集された映像と異なり、カメラのONとOFFに対応している。本論文では、素材映像中のカット点で挟まれた区間をパート（Part）と呼んでいる。また、カメラマンはあるショットに対して、テイク（Take: 撮り直し）を複数回行う場合がある。このテイクの仕方には、一つのテイクを撮影するごとにカメラをON、OFFする場合もあれば、カメラをON、OFFする間に複数回テイクを行うこともある。この場合、テイクとテイクの間は、カメラ調節やカメラの引き戻しに使用されるため、放送映像用には使用できない使用不能区間となる（図4のCamera adjustment）。使用不能区間としては、このほかにも実際には何も撮影されていないブランクの部分（図4のBlank）や、撮影時に起こったカメラマンの手ぶれ（図4のCamera shake）、撮影に同行しているスタッ

フの手などが入った部分などがある。

別の特徴として、撮影は基本的にシーンごとに行われるため、素材映像中、あるシーンに含まれるショットは隣接しているという特徴がある。また、別のシーンを撮影する場合、必ずカメラを移動するため、カメラのスイッチがOFFとなる。つまり、素材映像上のカット点中、いくつかはシーンの変化点であると考えられる。したがって、素材映像をシーンの変化点で区切れば、シーンごとの区間を抽出できることになり、この区間ごとにショットサイズの自動付与を行えばよい。

2.3 編集作業とショット

編集作業は単純化すると、次の二つの処理で構成されている。(1) 番組作成者の意図に従ってカメラマンが撮影した素材映像から、放送用映像で使用される映像の最小単位であるショットを切り出す処理。(2) 番組作成者の意図に沿うようショット集合の中からショットを選択し、これらを接続する処理。

図5は、図3に示した二つのショットタイプに対するショットの接続形態を示している。図1に示した例は、図5中、Shot 1とShot 2の関係に相当する。しかし、ショットのタイプにはカメラワークを含んだType 2があり、その場合、単純にショット同士を比較するのではなく、例えばShot 4のカメラワーク後のFIX部と、Shot 5のFIXを接続判定することになる。

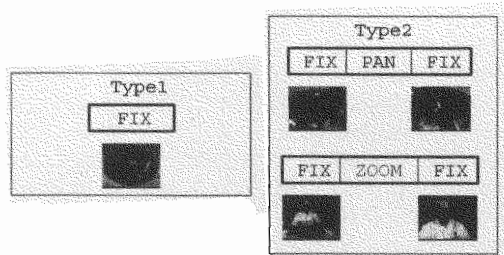


図3 ショットのタイプ
Fig. 3 Shot type.

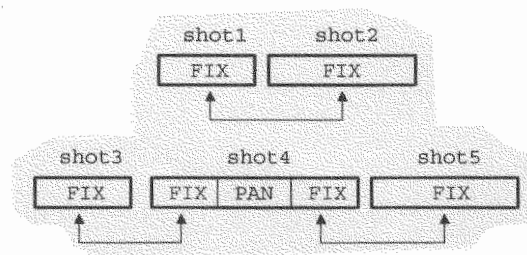


図5 ショットの接続形態
Fig. 5 Types of shot connection.

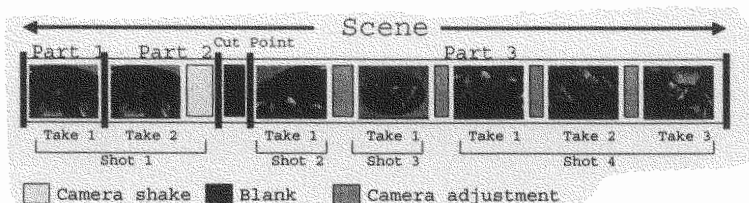
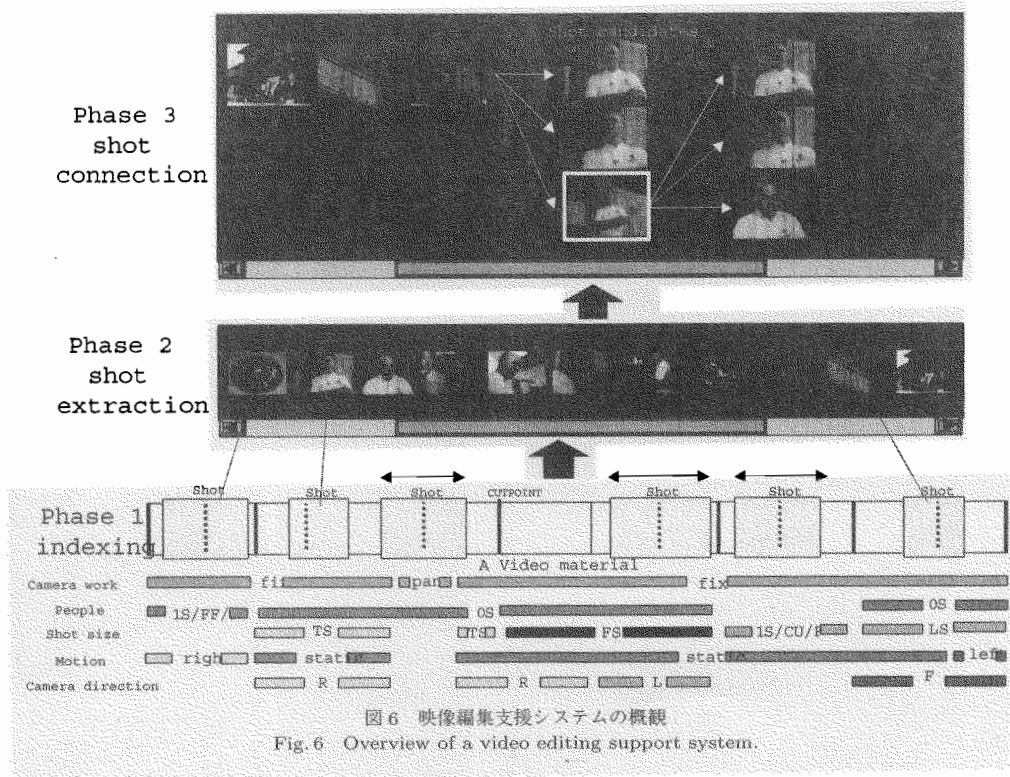


図4 素材映像の特徴
Fig. 4 Feature of video material.



3. 映像編集支援システム

3.1 概 要

映像編集支援システムの概観を図6に示す。2.では、編集作業の概観を述べたが、提案する映像編集支援システムは、これらの編集作業を総合的にサポートすることを目指しており、大きく以下の3点に集約できる。

- Phase 1 素材映像に索引付けを行う
- Phase 2 素材映像からショットを切り出す
- Phase 3 映像文法によりショット接続の支援を行う

図6は、この三つのフェーズの概観を示している。編集支援の立場からは、これらの作業工程において、編集者が介在する作業をできる限り自動化することが最終目的となる。

編集支援システムの目的は、編集者の仕事をショットの接続に集中させることである。ショットの接続によって生成される単位はシーンであり、Phase3で行う作業は、基本的にシーンを作成する作業となる。映像文法に「シーンの冒頭はマスターショットで始まる」という規定がある。マスターショットとは、シーンの

全体像を表現し、比較的長めのショットである。例外がない限り、Phase3のボード上、一番左にあるショットはマスターショットである。そのためPhase3で編集支援システムが最初に編集者へ提示するものは、マスターショットの候補群である。その後、編集者がいずれかのマスターショットを選択することで、シーンの第1ショットが決定される。システムが次に行うことは、選択した第1ショットに接続できるショット候補だけを一覧表示することである。

このような映像編集支援システムを実現するためには、Phase1, 2を自動化しておく必要がある。Phase1を自動化するためには、あらかじめ素材映像に対して編集のために必要な種々の索引を付与しておく必要がある（索引内容の詳細は[1]に記載している）。また、Phase2を自動化するためには、使用不能区間を推定し、使用可能な区間についてのみ、適切なショットの候補を切り出す必要がある。

このようなシステムを実現することで、編集者はショットの接続に関してのみに、作業を集中することができるようになる。本論文では、Phase1の中からショットサイズの自動付与について提案する。

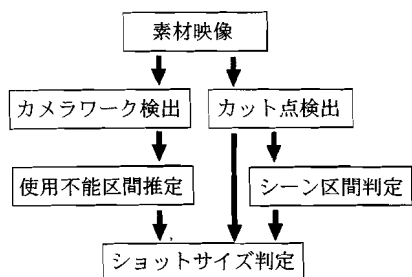


図7 ショットサイズ索引付けシステム
Fig. 7 Shot size indexing system.

3.2 Phase 1 の処理過程

ショットサイズの自動付与を行うためには、カメラマンの撮影した素材映像をあらかじめ前処理しておく必要がある。ただし、処理時間も重要な要素である。素材映像に索引付けを行う処理時間があまりにもかかるようであれば、放送に間に合わない。そこで、Phase 1 には処理の高速性が求められる。本研究では、完全自動化を目指して多大な処理時間を必要としたり、逆にシステムの実現化を見送るのではなく、編集者の介入を許すことによって現実的な編集支援システムを構築する立場をとる。前処理を含めたショットサイズ自動付与システムにおける処理の流れを図7に示す。

ショットサイズを自動的に付与するために、まず、カメラワークを抽出して、手ぶれやブランクなどの使用不能区間を取り除く。また、カメラが連続して撮影しているパートを切り出すためにカット検出を行い、隣接するパートを接続してまとまりのあるシーン区間を判定する処理を行う。これらの情報を統合し、シーン内の各ショットに対してショットサイズを判定して付与する。

カット点検出については、従来より多くの研究が行われている。精度については100%に近い結果が出ている[2]が、処理時間がかかり、高速な処理を必要とするPhase 1には向いていない。MPEG-1ファイルに対しては高速なカット点検出法がいくつか提案されているが[3], [4], MPEG-1は、画質より転送効率を重視しているため、画像サイズがハーフサイズ以下となる。このため、画質が問われる放送用映像としては適切でない。同一話者区間を検出する方法として用いられているGLR法[5]を簡略化したバッファリング法[6]では、フルサイズ・フルフレームのMotionJPEG映像において、実時間の3倍程度で処理が可能であり、正解率94%、適合率92%を達成している。ここで、正

解率と適合率は次のように定義される尺度であり、正解率はすべての正解を漏れなく検出（完全性）できたか、また適合率は必要なものだけを検出（正確性）できたかを表している。

$$\text{正解率} = \frac{M}{M+D} \cdot 100, \quad \text{適合率} = \frac{M}{M+I} \cdot 100$$

M : 正解検出数, D : 未見出数, I : 湧き出し数

シーン区間の判定処理では、パートの平均的な色調変化からシーン判定を行う[6]。

カメラワーク検出については、従来よりカメラワークの区間検出が行われている[7]が、素材映像では、カメラマンの手ぶれなどを検出する必要がある。精度の高いカメラワークの変化量を求める必要がある。この変化量を取り出す方法としては、仮動きベクトルを利用してグローバルなカメラワークの動き量を推定する方法[8]、投影法を利用した方法などがある[9]。しかし、素材映像では、被写体が画面一杯に大きく映りながら動いているものや、その被写体の動きとともにカメラが被写体を追いかける部分があり、この点でいずれの手法も正しいカメラワーク移動量を抽出することは困難である。使用不能区間の推定については、カメラワークの変化量をもとに、手ぶれ等を検出している[10]。

本論文では、ショットサイズの自動付与に関し、カット点とシーン変化点、カメラワークの情報が理想的に得られたことを想定し、これらの情報を用い、ショットサイズを自動的に決定する手法について研究を行った。また、この手法によってもショットサイズが決定できない、二つの解釈が可能な場合については、最終的に映像文法を用いてショットサイズの付与を行う。

4. ショットサイズと映像文法

映像による時間・空間の表現においては、実際には長い時間を短い映像で表現したり、空間的にかけ離れた場所をショットの接続によって違和感なく表現できる。これは、映像制作者の意図を視聴者に伝えるために、冗長な部分を省いて時空間の抽象化を行っているためである。このような時空間の抽象化を行っても視聴者に違和感を抱かせないショットの接続規則が存在し、これが映像文法の1要素となっている。

映像文法における最小の単位はショットであるが、ショットには図8に示すような対象を遠くから撮影したルーズショット（LS）、近くで撮影したミドルショッ

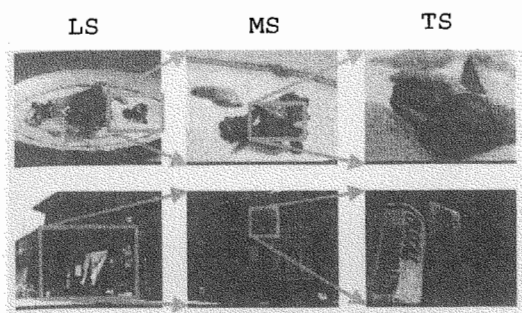


図8 ショットサイズ
Fig.8 Shot size.

表1 ショットサイズに関するルール
Table 1 Rules for shot size.

rule-S1:	シーンの冒頭は空間の全体像を表現するマスターショットから始まり、MS、TSへ接続することで注目する対象を明らかにしていく。
rule-S2:	ショットサイズが急激に変化するLSからTSなどへの接続は視聴者に不要な違和感を与えたり、視点がどこからどこへ移ったのか追いかけれなくなるなどの問題が生じるため、通常、接続できない。
rule-S3:	MSからMSへの接続は、不要な冗長性を生むため接続できない。
rule-S4:	TSとTSの接続は認められる。

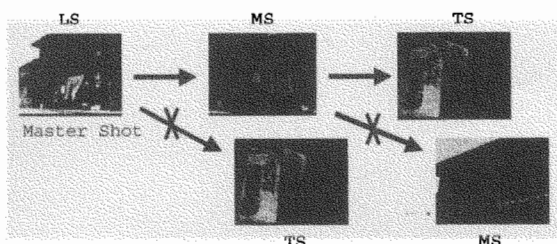


図9 ショットサイズの接続に関する映像文法
Fig.9 Video grammars for shot size connection.

ト (MS), より対象に接近したタイトショット (TS) といった3種類の相対的関係を示すショットサイズが存在する。編集作業において、ショットサイズに関する映像文法中の接続規則には、表1に示すような規則がある。また、表1の関係を図9に示す。

この相対的関係をもつショット同士を接続することで、撮影対象の全体から部分、客観から具体、外見から内面、概要から核心など様々な意図を視聴者へ自然に伝達することが可能になる。もし映像全体でショットサイズが変化しなければ、映像の内容はどこが焦点であるかがはっきりわからないだけでなく、非常に単調ですぐに飽きてしまうような映像となってしまう。

このショットサイズの変化は、映像の制作を行う上で、ディレクターの意図を表現するために用いる映像文法である。

5. ショットサイズの索引付け

5.1 ショットサイズ付与における問題点

人に関しては、フルフィギュア、ニーショット、ウェストショット、バストショット、タイトバストショット、アップ、クローズアップなどのショットサイズが存在する[11]。この場合、対象が人であることを事前に知っているの、人を検出し、その大きさを測ることによって、ショットサイズを判定することができる。

しかし、顔以外の一般的な物に関して、それを検出し、認識して大きさを測ることは極めて困難である。例えば、図8の上にあるものは、皿に乗った料理である。この場合、サイズの異なる三つのショットを比較することにより、皿の上にある料理の一部が対象であることを人間は推測することができる。しかし、この対象は、いつも形が同じであるとか、色が同じであるなど、どの素材映像においても共通する特徴を備えていない。また、図8の下にあるものは、料理店の建物である。図より、店の入り口にあるちょうちんに書かれた店の名前が撮影対象であるが、LSをみただけではどこに焦点をあてるか判断することはできない。また、この料理店の建物を含めた風景ショットが存在し、このショットを接続する場合、料理店の建物ショットはMSであるかもしれない。この意味で、ショットで撮影されている対象を認識することなく、ショットの集合に対してショットサイズを付与することは極めて困難である。カメラマンは、あるシーンを撮影する場合、このショットサイズを意識しており、放送映像に必ずしも使わない場合であっても、一般的にLS、MS、TSそれぞれのショットを撮影することが慣習となっている。したがって、シーン内でFIX部の相対的関係を結ぶことが素材映像に対し、ショットサイズを付与する有効な方法論となる。

5.2 パート内でのショットの包含関係

撮影されている対象を認識することなく、ショットサイズを判定するためには、ショットの包含関係を判定する方法が有効である。図10は、一つのパート内でショットの包含関係を判定している例である。

この図で、PAN前後では、ショットサイズが変化しないので、FIX1とFIX2、FIX3とFIX4は同じサイズとなる。またZOOM前後では、ショットサイズが

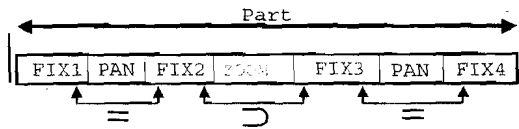


図 10 パート内でのショットの包含関係
Fig. 10 Including relation of shots within parts.

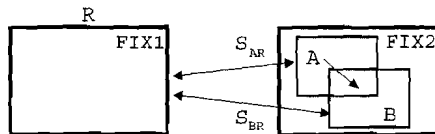


図 11 アクティブ探索
Fig. 11 Active search.

変化するので、FIX2 と FIX3 は包含関係にあることがわかる。

5.3 パート間でのショットの包含関係

一つのパート内では、カメラワークの情報を用いることで、ショットの包含関係を判定することができた。しかし、パート間ではカメラワークが途切れているため、カメラワークの情報を使ってショットの包含関係を判定することができない。そこで、パート間では、カメラワークの情報を使わずに、ショットの FIX 部同士で包含関係を判定する必要がある。

一般に、ある画像が、別の画像の一部としてどこに存在しているかを判定する問題は、探索問題と呼ばれている。村瀬らは、この画像における探索問題を高速実行できるアルゴリズムとして、アクティブ探索法 [12] を提案している。本研究では、このアクティブ探索法を適用して、パート間におけるショットの包含関係を判定した。図 11 は、アクティブ探索法の概念図である。

この図は、FIX2 の一部をズームアップして FIX1 が生成された場合である。この場合、アクティブ探索法は、FIX1 を参照画像 R とし、これを縮小して FIX2 上で最も類似度の高い部分を探索している。アクティブ探索法では、参照画像 R の縮小画像と窓 A との類似度 S_{AR} が求まると、次に類似度を計算すべき窓 B の位置にスキップすることができる。

探索画像中の窓 A の大きさは、もとの画像サイズを 1 としたとき、(1, 0.59, 0.35, 0.21, 0.125) の 5 種類となる。この 5 種類の窓の大きさ比率は、正解データの中から包含関係にあるショットサイズの組合せを取り出し、ヒューリスティックに包含関係を決定した。アクティブ探索では、比較する二つのフレームを F_1 ,

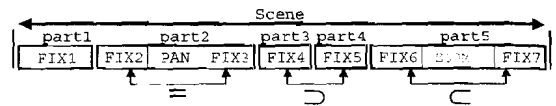


図 12 包含関係だけでショットサイズが決定できない
Fig. 12 Example of impossible case in discriminative shot size using including relation.

F_2 としたとき、 F_1 が F_2 を含む、また F_2 が F_1 を含むという二つの仮説が成り立つため、両方の探索を行う。結果として、二つのフレームについて 10 種類のアクティブ探索を行うが、アクティブ探索の類似度がしきい値より高いものの中で最も高い類似度を示したものを包含関係があったとし、その比を包含の比率として類似度とともに記録する。

5.4 包含関係に基づくショットサイズの自動付与

ショットの包含関係がパート内とパート間で判定できても、ショットサイズが一意に決定できるとは限らない。

例えば、図 12 で、Part1 は孤立する FIX 部となるため、隣接している FIX 部に包含関係が判定されない場合は、US (Unknown Shot) となってしまう。また Part2 では、パート内でショットサイズが同じであることはわかるが、LS, MS, TS いずれであるかを決定できない。Part3 と Part4、また Part5 も、包含関係はわかるが、LS-MS か、MS-TS かのいずれであるか決定できない。このように、素材映像中、隣り合うショット同士が必ずしも包含関係にあるとは限らないため、カット点を越えた場合、シーン内の全ショットに関する包含関係を結ぶことができないという問題がある。そこでシーン内に存在するすべての FIX 部について包含関係を総当りで判定し、あるパート内の FIX 部が別パート内の FIX 部とつながるよう、繰り返しグループ化を行う必要がある。これらの処理を含めた処理過程を図 13 に示す。

図中、A のブロックでは、シーン内のすべての FIX 部同士の包含関係を総当りで判定し、表として作成しておく。B のブロックでは、パート内の包含関係を求めて表としておく。C のブロックは、これらの表を参照して繰返し総当り法により、パート間で包含関係を判定する。表の参照を行うだけであるため、処理時間は高速である。

図 14 は、パート間で包含関係を判定する方法を示している。ここで、 P_i は、 i 番目のパートを表している。Pattern B の場合、シーン内のすべての FIX が包

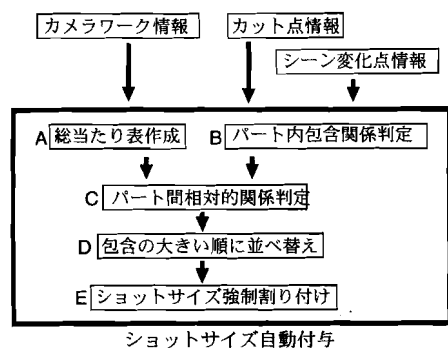


図 13 ショットサイズ自動付与部の処理過程

Fig. 13 The process of automatic shot size indexing.

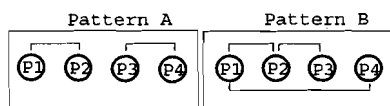


図 14 パート間の包含関係を判定する方法

Fig. 14 Discrimination of including relation between parts.

含関係で結ばれ、ショットサイズを付与できる。しかし、Pattern A の例では包含関係が途切れている。この場合には、P1 と P2 をグループ化するとともに、P3 と P4 をグループ化し、このグループ間で FIX 同士の総当たりによる類似度計算を行う。その中で、最も高い類似度がしきい値を超えていれば、その FIX 部同士に新たな包含関係があると判定する。こうして、パート間で包含関係を決定することができる。

D のブロックでは、C のブロックで得られた包含関係をもとに、包含の比率の大きい順に並べ換える。E のブロックでは、包含の比率を三つに分割し、上位を LS、中位を MS、下位を TS とする。分割のしきい値は、窓の大きさ比率より 0 以上 0.126 未満までに含まれる FIX を TS、0.126 以上 0.59 未満までを MS、0.59 以上から 1 までを LS とした。これによりショットサイズが自動的に付与されたことになる。

もし、2 段階の包含関係しか得られないような場合、これを強制的に 3 段階のショットサイズに割り振るとショットサイズの判定誤りを生じる。この場合、表 1 の rule-S1 を適用する。rule-S1 は、シーンに必ずマスターショット、つまり LS が一つは存在することを規定している。図 12 は一つのシーン内でのパートの関係を表しており、相対的に一番大きいものが必ず LS である。(LS, MS), (MS, TS) 2 通りの解釈がある場合は、シーン内の包含関係の中で最も包含の比率が大き

表 2 ショットサイズ自動判定の結果

Table 2 Results of automatic shot size discrimination.

Scene 番号	Scene 内 FIX 部数	Shot Size 正解判定数	自動判定 正解率 (%)	Shot Size 未付与数
1	2	0	0.0	0
2	14	11	78.6	0
3	7	2	28.6	0
4	11	3	27.2	0
5	4	4	100	0
6	3	3	100	0
7	16	16	100	0
8	5	3	60.0	0
9	2	2	100	0
10	22	16	72.7	0
11	21	16	76.2	0
12	12	8	66.7	1
合計	119	84	70.6	1

い FIX 部を LS とする。

6. ショットサイズ自動付与実験

実験は以下の環境で行った。

- ・ 計算機： SGI Octain, CPU: R12000
- ・ 映像： 毎日放送、あまからアベニュー素材映像
料理店名バスカルベニョ
640 × 480, 30 frame/s
Motion-JPEG, Quality: 75%
25 分 27 秒, 45763 frame

表 2 に前章の実験を行った結果を示す。表では、素材映像をシーンごとに分割し、各シーンごとに含まれる FIX 部の数を「Scene 内 FIX 部数」に示している。また、ショットサイズ自動判定の結果、正解した FIX 部の数を「Shot Size 正解判定数」に示す。ショットサイズ自動判定処理の正解率は、全体の FIX 部数に対する正解判定数の比として求め、その結果 70.6% となった。表中、シーン 1 は正解率 0 であるが、これはアクティブ探索による包含関係判定の段階で誤ったためである。また、表中、Shot Size の未付与数は、5.4 で述べたグループ化によっても包含関係を結ぶことができなかった孤立ショットの数である。表より、孤立ショットは全体の中でシーン 12 中 1 個のみであった。パート内で包含関係を判定する処理では、多くの FIX 部にショットサイズを付与することができなかったが、本手法により、119 個の FIX 部中、孤立ショットを 1 個だけに減少させることができた。これにより、全 FIX 部に対するショットサイズの付与率は、99.2(118/119)

表3 窓比ごとの包含関係判定率
Table 3 Decision rate related to the inclusion relation as a function of window ratio.

窓比	包含関係判定率 (%)
1	77.8
0.59	85.7
0.35	83.3
0.21	75.0
0.125	55.0

%となった。システムの性能としては、ショットサイズを正確かつ数多く付与することが求められるため、自動判定正解率と付与率がともに高い値をもつことが理想となる。

次に、窓比ごとに包含関係判定率を算出した結果を表3に示す。

包含関係の判定を行う上で、設定したしきい値以下の類似度は、理論上関係ないFIX部同士であることを示す。包含関係判定率は、包含関係の正解数とともに、関係のないものを関係なしとして判定できた正解数を考慮したものである。その定義を以下に示す。

$$\text{包含関係判定率} = \frac{\text{包含関係の正解検出数} + \text{関係なし正解検出数}}{\text{包含関係・関係なしの正解総数}}$$

表3中、窓比1は参照画像と窓サイズが同じとなるため、ヒストグラムインターセクションによる判定結果となる。また、窓比0.59以下は、参照画像より窓サイズが小さくなるため、アクティブ探索を行った判定結果である。これにより、アクティブ探索は窓サイズが小さくなるほど精度が落ちることがわかる。ショットサイズの判定では、この探索窓が小さくなるほど誤判定に強く影響を与えている。

また、表4に図13の処理時間の大半を占めるAブロックについて、シーンごとの処理時間を示す。

この処理は、シーン内に含まれるFIX部の包含関係を総当りで求めるものである。単純に処理時間を積算すると809分(約13.5時間)となるが、これらは二つのフレーム間に対する処理を組み合わせ数分で行っているだけであり、単純に並列処理を適用できる。近年、PCの価格低下により、PCクラスタが注目されているが、PCの台数を増やすことで要求される実用レベルの処理時間に収めることが可能である。

ここで、本手法によるショットサイズ判定は、ショットを切り出した上で行っているわけではない。判定されたショットサイズは、素材映像上のFIX部に付与されるため、システムがどの部分をショットとして抽出し

表4 シーンごとの処理時間
Table 4 Processing time for each scene.

Scene 番号	Fix 部数	処理時間 (分)
1	2	5
2	14	156
3	7	30
4	11	51
5	4	13
6	3	4
7	16	34
8	5	16
9	2	3
10	22	300
11	21	144
12	12	53
合計	119	809

ても、そのショットに含まれるFIX部に関してショットサイズを調べるだけでよい。これにより、編集者に対しショット長の微調整を許すことが可能であるなど、柔軟な対応が行える編集支援システムを構築できる。

7. む す び

本研究では、映像編集支援システムの一部として、ショットサイズを素材映像に対して自動付与する方法を提案した。ショットサイズには、人に関するショットサイズがある。これについては、人を検出することでショットを付与することができる。しかし、顔以外の一般的なものに関しては、ショットサイズを判定することが困難であった。そこで、本研究では、画像の探索問題を適用することでこの問題を解決した。

今回、探索アルゴリズムとしてアクティブ探索を適用したが、表3に示すとおり、アクティブ探索は、探索窓が小さくなるごとに推定精度が低下することが知られている。また、暗い画像で精度が悪くなる傾向も見られた。ショットサイズの判定では、パート間の包含関係を求める際、関係を結んだFIX部のショットサイズを一つ誤ると、同パート内に含まれる相対的關係で結ばれたFIX部のショットサイズを連鎖的に誤判定する。そこで、アクティブ探索の精度を改善することにより、ショットサイズの自動判定精度が大きく向上する可能性がある。

放送局の現場からは、素材映像に情報が何もついていないと幾度となく映像を見直す必要性が生じ、多大な編集時間を要するが、映像が構造化されるとともに、50%の精度でも内容について情報が付与されていれば十分役立つという声がある。我々は、この精度を最低限とし、精度を100%に近づけていくことで編集の支

援度を上げていくという立場をとっている。よって、今後はショットサイズの索引付けに関して精度を向上させる必要がある。

また、ショットサイズ以外の索引についても自動付与を行い、映像編集支援システムの Phase 2 では、最適なショットの切出し法、また、Phase 3 では、編集者の作業をより効率良く行える技術を開発し、システムの完成度を高めていく予定である。

謝辞 本研究を行うにあたり、実験に協力頂いた倉橋政徳君、坂江伸悟君、林義文君に感謝します。

文 献

- [1] 徳平征恒, 山口悟士, 永田博靖, 熊野雅仁, 有木康雄, 春藤健司, 塚田清志, “映像内容の解析に基づく映像編集に向けて,” 画像電子学会第 182 回研究会, 00-05-04, pp.21-28, 2000.
- [2] 鈴木賢一郎, 中嶋正臣, 坂野 鋭, “識別的な手法に基づく映像データからのカット検出法,” 画像工学研究会, IE2001-27, pp.63-70, July 2001.
- [3] 中島康之, 氏原清乃, 米山暁夫, “部分複合を用いた MPEG データからのカット点検出,” 信学論 (D-II), vol.J81-D-II, no.7, pp.1564-1575, July 1998.
- [4] 金子敏充, 堀 修, “ゆ一度比検定を用いた MPEG ビットストリームからの動画像カット検出手法,” 信学論 (D-II), vol.J82-D-II, no.3, pp.361-370, March 1999.
- [5] D. Liu and F. Kubala, “Fast speaker change detection for broadcast news transcription and indexing,” Erurospeech99, pp.1031-1034, 1999.
- [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, Nov. 2001.
- [7] M.A. Smith and T. Kanade, “Video skimming and characterization through the combination of image and language understanding techniques,” CVPR1997, pp.775-781, 1997.
- [8] 興沼正克, 村岡洋一, “グローバルなアフィン動きパラメータの実時間推定手法,” 信学論 (D-II), vol.J82-D-II, no.7, pp.1161-1170, July 1999.
- [9] 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,” 信学論 (D-II), vol.J82-D-II, no.10, pp.1572-1580, Oct. 1999.
- [10] 倉橋政徳, 熊野雅仁, 有木康雄, 春藤憲司, 塚田清志, “素材映像中のカメラワークによる使用不能区間の推定,” 電気関係学会関西支部連合大会, G18-7, Nov. 2001.
- [11] 永田博靖, 徳平征恒, 山口悟士, 山本哲也, 熊野雅仁, 有木康雄, 春藤憲司, 塚田清志, “映像編集支援システムのための人物に関するインデキシング,” 2001 信学総大, D-12-138, 2001.
- [12] 村瀬 洋, V.V. Vinod, “局所色情報を用いた高速物体探索—アクティブ探索法,” 信学論 (D-II), vol.J81-D-II, no.9, pp.2035-2042, Sept. 1998.
- [13] 熊野雅仁, 林 義文, 有木康雄, 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

イズ判定,” 2001 信学ソ大 (情報・システム), D-12-20, Sept. 2001.

- [14] 林 義文, 坂江伸悟, 熊野雅仁, 有木康雄, 春藤憲司, 塚田清志, “素材映像中のショットサイズ判定と関連ショットの検出,” 電気関係学会関西支部連合大会, G18-6, Nov. 2001.

(平成 13 年 11 月 5 日受付, 14 年 1 月 15 日再受付)



熊野 雅仁 (正員)

平 2 立命大・理工学情報系コース卒。平 2 龍谷大・理工・実験助手, 現在に至る。映像編集に関する研究に従事。情報処理学会, 日本音響学会会員。



有木 康雄 (正員)

昭 49 京大・工・情報卒。昭 51 同大学院修士課程了。昭 54 同大学院博士課程了。昭 55 京大・工・情報助手。平 2 龍谷大学・理工・電子情報助教授, 平 4 教授, 現在に至る。工博。昭 62~平 2 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。情報処理学会, 日本音響学会, 人工知能学会, 画像電子学会, IEEE 各会員。



上原 邦昭 (正員)

昭 53 阪大・基礎工・情報卒。昭 58 同大学院博士課程単位取得退学。同大産業科学研究所助手, 講師, 神戸大・工・情報知能工学科助教授を経て, 同大都市安全研究センター教授, 情報知能工学科を兼任。現在に至る。工博。平 1~2 オレゴン州立大学客員研究員。平 6~8 神戸大総合情報センター副センター長。人工知能, 特に機械学習, マルチメディアデータベース, 自然言語によるヒューマンインタフェースの研究に従事。人工知能学会, 計量言語学会, 日本ソフトウェア科学会, AAAI 各会員。



下條 真司 (正員)

昭 56 阪大・基礎工・情報卒。昭 61 同大学院基礎工学科博士課程了。工博。同年大阪大学基礎工学科情報工学科助手。平 1 同大学大型計算機センター講師。平 3 同助教授。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 43 関西学院大・社会学部卒。昭 44 (株) 毎日放送入社。テレビ番組制作に従事。平 10 同メディア開発局。平 13 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 54 阪大・工・通信卒。同年 (株) 毎日放送入社。放送技術に従事。平 13 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。

クト委員。