映像文法に基づいた撮影・編集支援技術と映像コンテンツの自動撮影・自動生成技術に関する研究

熊野, 雅仁

(Degree)
博士（工学）

(Date of Degree)
2008-03-05

(Date of Publication)
2009-06-26

(Resource Type)
doctoral thesis

(Report Number)
乙2980

(URL)
https://hdl.handle.net/20.500.14094/D2002980

神 戸 大 学 博 士 論 文

映像文法に基づいた
撮影・編集支援技術と映像コンテンツの
自動撮影・自動生成技術に関する研究

平成20年 1月

熊 野　雅 仁

内容梗概

情報通信技術の発達により，映像配信方式が多様化し，新たな技術が次々に導入される中，映像コンテンツの制作現場では，いくつかの問題が起きている．第一に，多チャンネル化によるコンテンツ不足，第二に，制作コストの制限により実現が困難な潜在的映像コンテンツを有効活用できる技術の不足，第三に作業コストや作業時間の増大による，コンテンツ制作を行う人員の不足，第四に，現場では，新人の指導時間の確保が困難であることから，新人の育成環境・教育プログラムの不足がその問題である．サービスを増やす一方で，コンテンツ不足や人材不足の問題を同時に解決するためには，新たなサービスの提供に関し，新たな人材を必要とない自動化技術，作業を効率化する支援技術が必要となる．また，第四の問題においては，民間・一般家庭の素人が映像作品を制作する環境が発展しており，その解決策が素人にも潜在的に必要とされつつある．例えば，一般家庭・個人への映像撮影機器の普及，You Tube などの素人の映像作品共有投稿サイトの出現などがあり，個人で楽しむ映像の制作環境，一般公開の基盤が整いつつある．しかし，映像リテラシーは，まだ十分な体系化がおこなわれておらず，特に，素人にとって，映像の撮影・編集作業や概念に関して書籍を通じた概念習得は敷居が高い．また，映像の撮影・編集において，具体的な問題を解決する手段としては，直感に頼るしかなく，潜在的に支援が必要とされている．この支援の問題を解決する手段として，本論文では，映像文法を用いる．映像文法とは，編集が行われるようになった 20 世紀初頭からの数多くの実験的映像表現の中で，大衆に受け入れられ，効率的な内容伝達を担い，映像の品質維持に貢献する規範的映像組織法である．本論文では，映像文法を背景とする撮影・編集支援技術と，映像コンテンツ自動生成・生成支援技術を用いて，この四つの問題を解決するための研究を行う方法を提案している．

本論文は，十章で構成され，第一章は緒論である．第二章では，文法概念を整理し，映像の組織法に文法概念を用いる理由と，本研究で抜粋して用いる映像文法について述べる．

第三章と第四章では映像撮影支援技術について論述する．映像撮影支援技術に関し

て，初心者が抱える撮影上の主な問題は，特に初心者の場合，不適切なカメラワークの使用が映像の品質を落とす主要な原因となっている．これらの問題は，撮影中に指摘することが効果的である点に注目し，撮影中に実時間オンライン処理で映像文法に従った撮影法を誘導する，映像撮影訓練法を提案している．この手法により，新人カメラマンの初期の技能育成に必ずしも人が介在する必要がなくなり，初心者の撮影技術の向上や，個々人の問題点に合わせた問題解決の支援を行うことも可能となり，第四の問題「人材育成」に貢献する．

第三章では，映像撮影支援技術のうち，撮影中にオンライン処理でカメラワークの問題を指摘するための，カメラワーク高速解析手法を提案している．カメラワーク高速解析法については，ビデオカメラから時間分解能の高い画像が得られる DV モードにて，時空間画像を構成し，輝度投影相関法と二分化テンソルヒストグラムを併用した提案手法を用いることで，典型的な 30fps のフレームレートの 3 倍に相当する，平均で 94fps の高速なカメラワーク解析法を実現した．また，この高速なカメラワーク解析法を用いた FIX，PAN の上下左右，ZOOM のイン・アウトの判定精度として，91.9%の精度を達成し，従来法に比べ 10.1%の精度向上を達成した．

また，第四章では，第三章の高速カメラワーク解析法を用いて，手ぶれ，速度超過，蛇行といった不適切なカメラワークを引き起こすユーザに撮影訓練システムを使用させることにより，システムを使用しなかった者よりも，大きく問題が解決される結果を得ており，オンライン処理で問題を指摘することの有効性を示した．

次に，第五章，第六章，第七章では，映像編集支援技術について論述する．映像の編集作業は，編集を行う前のショットの切り出し作業が多大な時間を浪費し，必ずしも人が介在する必要のない作業となる．この観点に着目し，第五章では，撮影の失敗や取り直しが含まれる素材映像を対象に，映像文法に従って使用不能区間を推定し，使用可能なショット区間の自動切り出しを可能とする，第三の問題「作業コスト・人材不足」の解決に貢献する手法を提案している．これは，文法概念における連続的表現体を分節し，規範的表現単位を選定することに相当する．この使用可能・不能区間推定法としては，映像文法に従い，カメラワークの速度量の変化を評価している．実験の結果，再現率の平均値で 91.8%，適合率で 95.0%を達成した．

また，映像文法では，相対的な関係を持つショットサイズに基づいてショットの接続を行う文法的規則があり，これに基づいて編集を行うことから，ショットサイズの自動

付与を行えば，編集者は知的な編集作業のみに集中することができ，第三の問題「作業コスト・人員不足」に貢献する．これは，文法概念における分節された単位のカテゴリー化の問題である．この観点から，第六章では，相対的ショットサイズの自動付与手法について提案している．ショットサイズの自動付与法としては，プロのカメラマンが撮影した素材映像を対象とするが，カット点検出を行って得られる映像の区間は，編集上のショットに必ずしも対応しておらず，一つの断片に複数のショットに相当する部分が含まれる場合が多い．そこで，カメラワーク解析情報を用いて，FIX 区間の推定を行い，FIX 区間の代表フレームどうしの包含関係から，相対的ショットサイズを決定する手法を用いた．この実験の結果，平均値で，70.6%の自動判定正解率を達成し，有効性を示した．

第七章では，第五章と第六章で得られたメタ情報を用いて，あるショットを定めたとき，次に接続可能なショットは映像文法によって限定されるという規則を用いて，自動的に接続可能なショット候補を選択・接続し，映像文法に従って複数のショットの構成から一つのシーンを生成する，前向きプロダクションシステムによる自動編集手法を提案している．本研究で用いる映像文法では，ショットの継ぎ目となる編集は意識されず，内容に没入可能な見やすい自然な接続が望ましく，各ショットの時間長によるリズムや接続順序を考慮した映像の見易さが評価の指標となる．主観評価実験によって，シーンの見易さが高い評価を得ており，有効性を示した．

一方，映像コンテンツの不足を補うコンセプトの一つとして，アマチュアのスポーツ映像コンテンツが注目されている．映像コンテンツ産業では，例えば 100 万人規模の大衆に向けた映像の制作を行うことで，制作コストに見合った撮影・編集が行える．しかし，潜在的には，10 万人，1 万人，更に小規模な同好者の小集団や個人に向けた映像コンテンツのコンセプトが存在するとされている．アマチュアのスポーツ映像は，こうしたいずれかの小集団規模の映像コンテンツの一つである．アマチュアのスポーツは，大衆ではなく，同一の興味を所有する小集団向けの映像コンテンツとなるため，人件費を伴うプロの映像制作者が介在すれば，制作コストに見合わず，具現化が難しい領域の一つである．この点に注目し，第八章では，サッカー映像を対象として，デジタルシューティングの観点から映像自動撮影技術に関する手法を提案している．これは，第二の問題「コスト・技術不足で実現できないコンテンツ」の解決に貢献する．大衆向けの映像コンテンツではないが，同好者小集団から見て映像コンテンツのチャネ

ル数を増やす点では，第一の問題「コンテンツ不足」の解決にも貢献する．また，必ずしも人が介在する必要がないため，第三の問題「作業コスト・人員不足」の解決にも貢献する映像コンテンツ自動生成支援技術となる．デジタルカメラワークでは，高解像度の映像の一部を切り出して映像を自動生成する際，何らかの対象の動きに基づいて仮想のフレーム枠の動きを操作する必要がある．本研究では，小刻みに動くボールを対象としながらも，映像文法に従う安定した仮想カメラワークを実現している．被験者による主観評価を行ったところ，十分許容できる映像ではないが，許容できるとする評価が得られ，その有効性を示した．

また，価値が高い映像コンテンツとして，プロのスポーツ中継映像がある．ただし，外出中のファンにとって，常に全ての中継映像を見ることのできない環境に置かれることも想定される．この点に注目し，第九章では，速報として，スポーツ中継映像のハイライトシーンを自動的に映像再生の可能な携帯端末へ配信するための，実時間ハイライトシーン自動抽出法について提案している．ただし，本研究では，この実時間ハイライトシーン自動抽出法の画像処理部に焦点を当てる．これについても，新たなサービスを提供する意味で，コンテンツ不足に応える点から第一の問題「コンテンツ不足」の解決に貢献し，必ずしも人が介在する必要がなく，第三の問題「作業コスト・人員不足」解決にも貢献する映像コンテンツ自動生成支援技術となる．ハイライトシーンの決定を行う手法としては，野球中継映像の構造に着目し，ピッチャーとキャッチャーが同時に映る PC ショットを高速に安定して検出する手法として，ブロックの輝度分散値が安定した領域をマイニングして判別に優れたブロックを学習により選択する手法により，最大で 97.2%の検出精度が得られ有効性を確認した．

最後に，第十章では，本研究をまとめ，将来への応用・展望を述べる．

以上のように、本研究によって，映像コンテンツの不足，通常ではコストの見合わない潜在的映像コンテンツの具現化，作業・作業時間の増大，人員不足，また人材育成に果たす貢献は，これまで人の介在を必要とするが必ずしも人が介在する必要のない課題について，自動支援技術を導入することで，問題の解消に貢献することにある．また，映像コンテンツ業界に果たす貢献は，映像コンテンツを生成する枠組みを広げ，コンテンツ不足や新しい観点の映像コンテンツ生成に向け支援技術を提出した点にある．これらの技術により，映像制作，映像自動生成，また人材育成を支援することで，映像コンテンツ産業，また映像コンテンツの教育産業の進展に貢献することができる．

目次

図目次

表目次

第1章

緒論

1.1 背景

映像は，多人数が同時に同じ映像を視聴できる投影式映像提示装置の発明(1895)以来，物語や情報を語る新しい表現メディアとして映画とともに発達し，1940年代から本格的に放送が開始されたテレビを通じ，第二次大戦後1950年代から一気に一般大衆化した．そして近年，1980年代の衛星放送の開始やケーブルテレビの普及による多チャンネル化，さらにインターネットへの映像配信技術の発達，無線技術の発達による無線携帯端末向けのワンセグ放送など，この100年余りの間に，大衆への同時視聴から個人の視聴へ，また室内固定機器から移動体・外出先まで，映像の視聴形態は，多様化し続けるとともに，映像を提供するチャネル数は多角的に増え続けている．また，デジタル放送技術は，映像コンテンツにメタ情報を埋め込むことを可能とし，新たなサービスの提供や，視聴者との双方向通信も可能にしたが，制作者側に映像コンテンツ制作作業の増大を招いている．そして，技術的には同一嗜好の小集団や，個人の嗜好へも対応した放送形態の技術的可能性も研究されており，今後，新規に導入されるサービス技術も多様化し，増え続けるものと思われる．

このような情報通信技術の発達により，映像コンテンツの制作現場では，いくつかの課題を抱えている．第一に，多チャンネル化によるコンテンツ不足への対処，第二に，制作コストの制限により，実現が困難な映像コンテンツを実現可能にする打開策への対処，第三に，作業コストや作業時間の増大への対処，またコンテンツ制作を行う人員不足への対処，第四に，現場では，新人指導の時間確保が困難であることから，

新人の育成法・教育プログラムの不足問題への対処である．

一方，近年，ポータブル・ビデオカメラが普及し，一般家庭のユーザが映像を撮影する機会が増えており，一般ユーザの撮影によるスクープ映像や映像作品が放送用のニュース番組やバラエティ番組に用いられる事例も増えている．また，ディジタルカメラでは映像撮影機能が一般化し，携帯電話でも映像が撮影できるようになっており，携帯電話の映像撮影機能を用いたディジタルシネマが制作される試みも行われている [1]．特に，日本のコンテンツが海外で評価され，国策として映像コンテンツの制作や制作者の育成を支援する動きがあり [2]，映像制作に関してディジタル時代に即した人材育成法も模索されている [3, 4]．また，You Tube など，映像共有ウェブサイトが出現し，素人作品の投稿も行われており，You Tube の映像を視聴する専用の家電も販売されつつある．このような流れを受けて，放送によって提供される受動的な映像コンテンツだけではなく，個人で楽しむ映像の制作環境や，自ら映像コンテンツを制作・公開する環境が整備されつつある．

しかし，素人は，撮影・編集概念の無い状況で，直感に頼らざるを得ず，編集を行う時点で，無計画に撮影した映像が編集に適さない状況に陥ることも多く，素材となる映像は，撮り溜めのまま放置される傾向がある．これは，映像リテラシーの不足からくる問題とも言えるが，撮影・編集概念を本格的に学ぼうとすれば，書籍等での学習や，専門学校等での講義の受講により習得する方法が考えられるが，社会人を含めれば，時間や労力の点で敷居が高い．また，プロの制作者や素人を問わず，編集という知的な作業に至る前に，使用可能な区間を選定・抽出するという，時間と労力を要し，必ずしも知的とはいえない作業に直面しなければならない．この観点からすれば，コンテンツ産業・制作者側の四つの課題のうち，第三や第四の課題は，素人による映像の撮影・編集作業においても言える面があり，これらの作業を支援する技術や映像リテラシー教育プログラムが潜在的に必要とされている．また，その支援技術や教育プログラムは，集団的扱いよりも，個人ごとの特性や要求に合わせた支援や指南が望ましく，必ずしも人が介在しない自動化技術を実現すれば人員不足も解消すると考えられる．

本論文では，この四つの課題を対象とし，必ずしも人が介在する必要のない，映像撮影支援技術，映像編集支援技術を提案し，またコンテンツ不足の解消や個人化へつながる映像の自動撮影・自動生成技術に関する提案を行う．これらは，映像文法 [5] に基づいているため，技術内容を述べる前に，映像文法について 2 章で詳しく述べる．

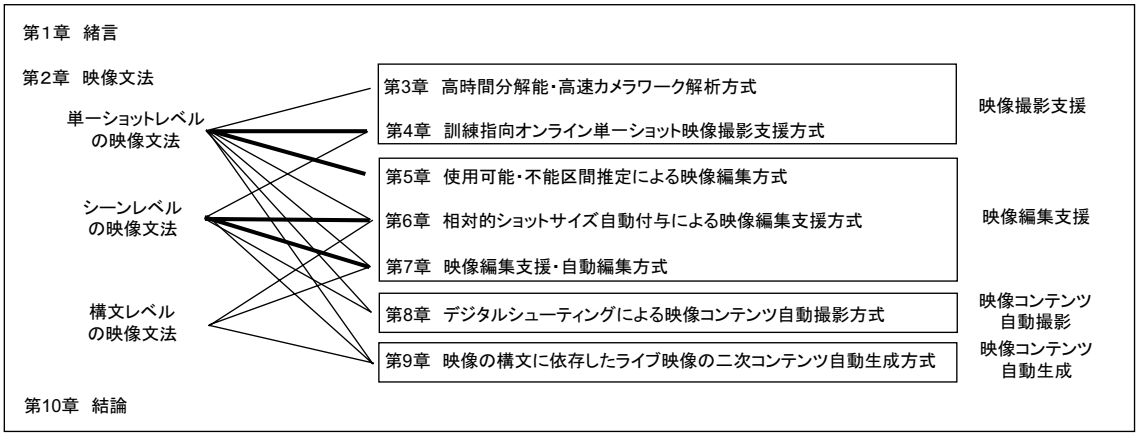


図 1.1 本論文の構成

1.2 本論文の構成

図 1.1 に，本論文の構成と映像文法との関係を示し，また，各章と映像撮影支援，映像編集支援，映像コンテンツ自動撮影，映像コンテンツ自動生成技術との関係を示す．本研究は，10 章で構成され，第 1 章は，緒論であり，これは本章のことである．

本論文は，映像文法に基づき，映像文法の教示法を研究するのではなく，必ずしも人の介在を必要としない，映像の撮影・編集支援法，また映像コンテンツ自動撮影・自動生成に関する研究の提案を行う．提案する技術と映像文法との関係を説明するため，本研究の映像文法の定義を第 2 章にて行う．ただし，映像文法には，第 2 章で述べるように，広義のモンタージュ，広義のデクパージュで説明される背景概念に加え，狭義のモンタージュ，狭義のデクパージュとなる概念や，映画文法も含まれる．また，この文法という言葉の根源となる西洋文法における規範文法の概念や，文学的視点が含まれる．本論文で用いる映像文法は，映像文法の中で，特に狭義のデクパージュとなる古典的デクパージュに背景概念が偏っている．第 2 章では，これらの関係について詳細を述べ，本論文で用いる映像文法の一部を抜粋する．また，この抜粋した映像文法を，三つのクラス，1. 単一ショットレベルの映像文法，2. シーンレベルの映像文法，3. 構文レベルの映像文法，に分割する．図 1.1 は，本論文で用いる映像文法について，三つのクラスと各章の関係を示している．太線は，特に関係の強いものを現している．

次に，3 章と 4 章で，映像撮影支援技術に関する提案を行う．映像撮影支援技術に関し，初心者が抱える撮影上の主な問題は，映像文法による観点から単一ショットの撮影

において，不適切なカメラワークの使用が，のちの編集に問題となり，映像の品質を落とす主要な原因となっていることである．この問題は，特に初心者の場合，撮影中に問題を指摘することが効果的であるという結論に至り，撮影中に実時間オンライン処理で映像文法に従った単一ショットの撮影法を繰り返し練習する訓練指向型の映像撮影ナビゲーションシステムを提案する．この映像撮影ナビゲーションシステムでは，撮影者のカメラワークの動きを評価する必要があるため，カメラワークの動作量変化を時間分解能の高い状態で得る手法が必要となる．そこで，3 章では，その時間分解能が高く，動作量変化が得られるカメラワーク解析法として，輝度投影相関法と二分化テンソルヒストグラムを用いたカメラワーク解析部の手法を提案する．また，4 章では，3 章で提案したカメラワーク解析手法を基盤とし，映像文法に従うショットについて，撮影訓練を繰り返すことで，映像文法の概念を教えることなく，映像文法に従うショットの撮影技術を習得できる，撮影ナビゲーションシステムに関する提案を行う．このシステムでは，ビデオカメラ自体が訓練を実施することになるため，これにより，撮影法を指導する人員が必要ではなくなる．つまり，人材育成に貢献しながら，人の介在を必要としない，第三と第四の問題解決に貢献することになる．

次に，5 章から 7 章において，映像の編集支援技術に関する提案を行う．映像の編集作業は，編集を行う前のショットの切り出し作業に多大な時間を浪費することから，必ずしも人が介在する必要のない作業であると言える．この点に注目し，5 章では，プロのカメラマンが撮影した，撮影の失敗や取り直しが含まれる素材映像を対象に，映像文法に従って使用不能区間を推定し，使用可能なショット区間の自動切り出しを可能とする手法の提案を行う．この支援システムにより，作業コストが軽減し，この作業を行う人員の削減にもつながることから，第三の問題解決に貢献する．また，映像文法では，相対的な関係を持つショットサイズに基づいてショットの接続を行う文法的規則があり，これに基づいて編集を行うことから，ショットサイズの自動付与を行えば，編集者は知的な編集作業のみに集中することができ，第三の問題解決に貢献する．これは，文法概念における分節された単位のカテゴリー化問題であり，この観点から，6 章では，プロのカメラマンが撮影した素材映像を対象に，ショットサイズの自動付与手法について提案する．

また，ショットが切り出され，ショットにショットサイズが付与されていれば，編集作業を行う際，あるショットを定めたときに映像文法によって接続できるショットが限

定される．この規則に着目し，さらに，接続可能なショットを自動接続し，シーンを構成するショットの接続候補を生成すれば，編集者は，出来上がった複数のシーンを選択するだけでよく，作業効率はより高くなる．この観点に着目し，7 章では，映像文法に従って複数のショットの構成から一つのシーンを自動生成する，自動編集手法を提案する．この支援システムにおいても，作業コストが軽減し，この作業を行う人員の削減にもつながることから，第三の問題解決に貢献する．

次に，8 章と 9 章において，第一，第二の問題解決に貢献する映像コンテンツの自動撮影・自動生成に関する基盤技術を提案する．映像コンテンツの不足を補うコンセプトの一つとして，アマチュアのスポーツ映像コンテンツが注目されている．映像コンテンツ産業では，例えば 100 万人規模の大衆に向けた映像の制作を行うことで，制作コストに見合った撮影・編集が行える．しかし，潜在的には，10 万人，1 万人，更に小規模な同好者の小集団や個人に向けた映像コンテンツのコンセプトが存在すると言われている．アマチュアのスポーツ映像は，こうしたいずれかの小集団規模の映像コンテンツの一つである．アマチュアのスポーツは，大衆ではなく，同一興味を所有する小集団向けの映像コンテンツとなるため，人件費を伴うプロの映像制作者が介在すれば，制作コストに見合わず，具現化が難しい映像コンテンツ領域の一つである．この点に注目し，8 章では，サッカー映像を対象として，ディジタルカメラワークによる映像の自動撮影技術に関する手法を提案する．これは，第二の問題解決に貢献し，大衆向けの映像コンテンツではないが，同好者小集団に対して映像コンテンツのチャネルを増大する点では，第一の問題解決にも貢献する．また，必ずしも人が介在する必要がないため，第三の問題解決にも貢献する映像コンテンツ自動生成支援技術となる．

また，価値が高い映像コンテンツとして，プロのスポーツ中継映像がある．ただし，プロのスポーツ映像は，価値が高いことから，プロの映像コンテンツ制作者が介在するため，自動化技術の導入は，大きな価値を持つ可能性が低い．しかし，外出中の野球ファンにとって，常に全ての中継映像を見ることのできない環境に置かれることも想定される．この点に注目し，9 章では，情報通信技術が可能とした外出先の個人向け映像配信基盤技術を想定した新しいサービスの形態として，スポーツ中継映像のハイライトシーンを映像再生の可能な携帯端末へ自動配信するための，速報的実時間ハイライトシーン自動抽出法について提案する．ハイライトシーンを自動的に切り出し配信する作業を行うため人員の増員を必要としない点において，提案手法は，第三の問

題解決に貢献している．新しい映像配信チャンネルに対する映像コンテンツの穴埋めとなるため，第一の問題解決にも貢献する．これは，新しいサービスを提供するにも関わらず，人材不足の問題を新たに引き起こさない効率的な方法である．ただし，本研究では，この実時間ハイライトシーン自動抽出法の画像処理に焦点を当てる．ハイライトシーンは，到着が遅れた場合，価値が急激に低下する．このため，構造化を高速に精度良く処理する画像処理手法が望まれた．この点で，本研究では，PC ショットを高速に判定する手法を提案する．本研究の手法により，ピッチャーとバッターが同時に撮影される PC ショットに基づいた映像の構造化を行い，アナウンサーの音声認識によるキーワードの検出と合わせることで，ハイライトシーンを決定する．

以上のように、本研究は，(1) 映像コンテンツの不足，(2) 通常ではコストの見合わない潜在的映像コンテンツをコンテンツとして提供するための技術の不足，(3) 作業・作業時間の増大とそれに伴う人員不足，(4) 人材育成といった問題に対して，自動支援技術を導入することで，問題解決に貢献することにある．これらの技術により，映像の作業過程を高能率化し，映像コンテンツの制作を支援することで，映像制作のハードルを低くすることが可能となる．最後に，10 章では，映像文法の可能性や展望について述べ，まとめを行う．

第2章

映像文法

2.1 緒言

映像文法という言葉は，1940年代に広汎する映画文法という概念が分析・体系化の対象とした，20世紀初頭の映画の技法を祖型とし，1950年代頃からテレビ放送業界に継承され，一気に世界に広汎したアメリカ式の映像撮影・編集技法や，ヨーロッパの撮影・編集概念の総体である．ただし，映画文法の文法という言葉は，今日の文法概念の祖型といえる古代ギリシア・ローマ時代より用いられた，言葉に関する規範という観点が源流となり，現在も語学学習に生き残る伝統的な規範文法の概念が参考にされたという経緯がある．映像文法は，その絶え間なく積み重ねられた映像制作経験により，修正・淘汰され，大衆に浸透・慣習化したため，目立った存在感を現してはいないが，時間と空間の次元を持ち，映像の断片を組織する映像の撮影・編集に関する基本概念として，影を潜めつつも広汎に根付き，映像制作の根底に浸透している．

今日の映像文法は，広義のモンタージュ，広義のデクパージュで説明される背景概念に加え，狭義のモンタージュとなるソヴィエト・モンタージュ理論，狭義のデクパージュとなる古典的デクパージュの概念や，映画文法の概念も含まれる．本章では，映像文法の中で，特にテレビの放送局で用いられる規則から抜粋を行い，規範文法の概念や，背景として狭義のデクパージュとなる古典的デクパージュに偏った，文学的美学の観点から特徴を述べる．また本章では，映像文法の全体像に関連する概念の歴史的位置づけを簡潔に述べ，抜粋された本論文の映像文法について位置づけを明らかにする．図2.1は，映像文法の全体像につながる編集概念の歴史的な変遷を示している．

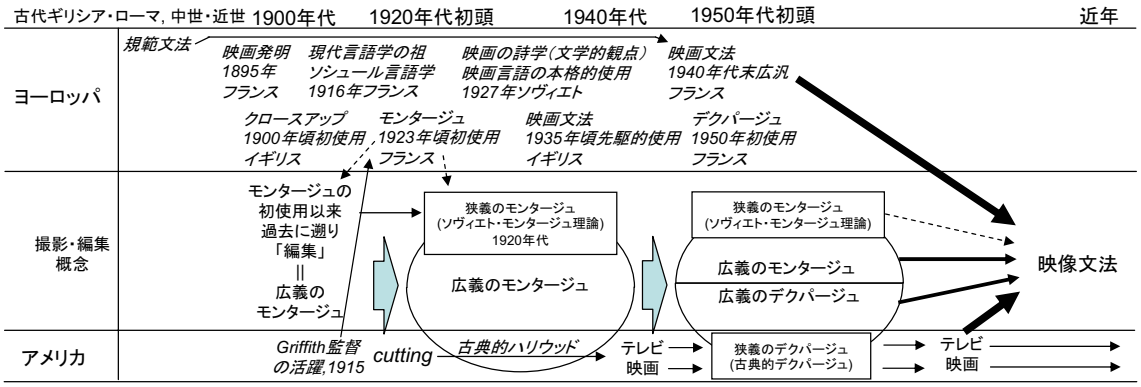


図 2.1　編集概念の変遷

2.2　モンタージュとデクパージュ

2.2.1　広義のモンタージュと狭義のモンタージュ

映像編集の史的変遷は，映像の萌芽と発展が映画により進行するため，映画の理論が中心となる．その映画理論は，萌芽時代からヨーロッパの諸研究者によって論じられ，その中心がフランスにあったことから，フランス語の概念が今日にも広汎している．特に，フランス語の「組み立てる」を語源的意味とする構成的概念のモンタージュ($montage$)[6]や,「切り分ける」を語源的意味とする分節的概念のデクパージュ($decoupage$)[7] は，映画の理論を語る上で，欠かせない概念である．ただし，これらの語は，初めて編集行為が行われた頃に存在した訳ではなく，同時に生まれたものでもない．

図 2.1 上，モンタージュを初めて用いたのは，1923 年頃，フランスの映画理論家 Moussinac とされ，映像断片の時間長を操作することで，編集がリズムを生み出す点に着目し，その組み立てる行為が生み出す映画の特性をモンタージュと呼んだ [6]．そして，このモンタージュという概念が注目され，過去に遡って事例が調査された．それ以後，その全ての技法はモンタージュとして語られるようになった．その中で，最も古いものは，1900 年にイギリスで用いられたクロースアップ (大写し) の初使用とされている [6]．この点で，デクパージュの観点が現れる 1950 年代を迎えるまで，映画に関する操作は，過去に遡り，すべてモンタージュで語られ，この映像断片の組み立てによって操作できる映画の特性が「広義のモンタージュ」の意味となる．

一方，同時代となる 1920 年代初頭，ソヴィエトの理論家達は，短い映像断片が数多

く接続されるアメリカ映画の特徴を分析し，特にアメリカ映画の父とされるGriffithが制作した長編大作「国民の創生」(1915)や「イントレランス」(1916)に影響され，独自のモンタージュ理論を発展させた．その独自のモンタージュ理論が「狭義のモンタージュ」であり,「ソヴィエト・モンタージュ理論」と呼んで，本章では広義のモンタージュと区別する．また，映画を言語とする本格的な取り扱いもソヴィエトの理論家が始めており，文学的視点を導入している．ただし，いずれにせよ，この時代の「編集」を意味するフランス語は，アメリカの技法も含め，すべてモンタージュとされた．

2.2.2　広義のデクパージュ

デクパージュという言葉を初めて映画の編集概念に適用したのはフランスの先駆的映画評論家Bazin(1950)である．その概念は，図2.1の狭義のデクパージュとなる古典的デクパージュの概念として語られるが，本節では，今日，映画に関して使われる広義のデクパージュの意味を述べる．デクパージュの語源的意味は「切り分ける」であったが，現在，映画用語としてのデクパージュは，表2.1に示す分類が可能である[8].

表 2.1　広義のデクパージュ

1. 撮影に必要な技術的指示を含む撮影台本（絵コンテ）
2. シナリオをショットやシークェンス分割する操作
3. できあがった作品の内在的構造
4. 作品の記号学的分析における単位の分割

表2.1の1と2は映画制作上の用語であり，3と4は映画研究の分野に属する[8]．これらは，いずれも「切り分ける」（分節）から派生できる概念である．例えば，2や4は，切り分けるがそのまま適用できる．3は，構想上の物語(未分節の連続体）が選択された語り口に従い，断片に切り分けられ，その接続結果を作品と見なすとき，その内在的構造がどのように切り分けられたかに着目した構造を示すと見ることで解釈できる．1.も同様に推測できる．1.は，構想上の物語連続体を，その物語がうまく伝達されるよう「分節」し，台本に仕上げることから，その台本がデクパージュと呼ばれるようになったこともさほど不思議ではない．また，1940年代までの古典的ハリウッドの編集概念は，分節された断片が，もともと同じ空間であったことを認知上維持で

きるよう，連続性を維持した分節を計画することに特徴がある．その計画により仕上がる構成物が台本であり，絵コンテとして仕上げられる場合もある．この絵コンテの「コンテ」とは continuity からきた言葉であり，絵コンテとは，断続されてしまった断片を視覚的に自然に連続しているように見せる「連続性」を計画することである．以上のように,「切り分ける」から派生する概念が「広義のデクパージュ」である．

2.2.3　表裏の関係にある広義のモンタージュと広義のデクパージュ

このデクパージュの概念は，それまですべての編集行為がモンタージュで特徴づけられた時代に対し，編集概念を「切り分ける」「組み立てる」という二つの観点を持つ概念に分化させた．つまり，図 2.1 の 1950 年代以降は，編集操作のうち，切り分けることをデクパージュ，組み立てることをモンタージュとして，技術的に対置しうる概念として捉えることが可能となった．このような対立する関係を見出し，深い洞察に至り，より高みとなる真理を追求する過程は，弁証法的過程である．

ただし，広義のモンタージュと広義のデクパージュの関係は単純ではない．例えば，台本を作成するデクパージュの思考過程において，連続体としての物語空間の分節を構想する際，時空間を様々に分節し，頭の中で，そのあらゆる断片の接続に思考をめぐらす行為には，一種の構成的過程としてのモンタージュが含まれる．つまり，デクパージュの思考過程の中で，組み立てる操作としてのモンタージュが含まれることになる．一方，映像断片の接続 (モンタージュ）時に，前後の接続を良くするためフィルムを切る行為は，一種のデクパージュである．つまり，広義のモンタージュと広義のデクパージュは，撮影・編集の計画段階から現実の作業までに含まれる，分節と構成といった，それぞれの側面を強調的に表現した術語であり，表裏一体の関係にある．

2.2.4　狭義のデクパージュとしての古典的デクパージュと映画の美学

一方，Bazin が映画にデクパージュという言葉を持ち込んだ理由は，アメリカの Griffith 以来，その技法を受け継いだ 1930,40 年代の古典的ハリウッドにおける編集の古典的スタイルに特別な名前を与えるためであり，そのスタイルの背後に見出した映画の美学を論じるためであった．その映像の特性は「古典的デクパージュ」と呼ばれた．この古典的デクパージュという概念が，図 2.1 の狭義のデクパージュのことである．

Griffith がそのほとんどの基礎を築いたと言われる古典的ハリウッド・スタイルの特徴は，映画の萌芽時代である 1900 年代に，新しい表現を大衆に普及させるため,「わかりやすさ」が求められ,「なめらかさ」「流動性」「ぜい肉のなさ」[9] が重視された．Bazin は，それらの特徴を文学の美学における「技法の不可視性」や「リアリズム」に照らし合わせ，論じた．技法の不可視性とは，優れた作家ほど，巧みな技法が用いられる事実を見せつけず，巧みな技法が用いられているにもかかわらず，その技法の存在は無意識の中に送られる．また，技法の不可視性によって実現され得ることは，物語の世界への自然な没入を誘導し，架空の物語世界を現実の世界のように錯覚させうることであり，それがリアリズムと呼ばれる．Bazin は，映画が語る世界を日常の現実のように錯覚させることに映画の本質を見たのであり，古典アメリカ映画の技法に，そのリアリズムを阻害する編集行為による断絶を隠蔽しようとするスタイルを見たのである．技法の不可視性とリアリズムを追求する文学の美学とは，巧みな技法を用いているのに，技法の存在が感知されず，物語世界が現実であるように読者を錯覚に陥らせ得ることであるとした．

古典的アメリカ映画の制作手法には，膨大な映像素材を編集して必要な区間を残し，不必要な区間は切り捨てる，冗長性の排除に力を注ぐ意味での Cutting，複数のショットの時間・空間を最大限，連続，かつ自然であるかのように見せるための分節を計画する Continuity だけでなく，その別側面として，映画がそもそも編集されているという事実を隠蔽し，継ぎ目がないように見せる Invisible が含まれている．その包括的特徴は，物語（内容の伝達）を優先すること，および，その映像を現実（すなわち，巧妙に計算された戦略によって自らの制作方法を隠蔽する高度に構築された現実）のように見せることである．ここに，文学と同じ「技法の不可視性」があり，Bazin はそれを「透明性」と呼んだ．そして，透明性によって実現される錯覚的現実が「リアリズム」であり，透明性 (技法の隠蔽とリアリズム) を「映画の美学」としたのである．

Bazin は，古典的アメリカ映画の手法が，もともと連続体としての物語空間の連続性を崩さずに分節する点に強く配慮される点において，もともと独立して存在する構成素を構成的に組み上げるモンタージュというよりは，うまく切り分けるという観点に力点が置かれる点でデクパージュに着目したものと思われる．Bazin は，デクパージュ (切り分ける) という言葉だけでは語りきれない，その透明性とリアリズムで説明される古典的アメリカ式のモンタージュを，古典的デクパージュとして呼び直したのである．

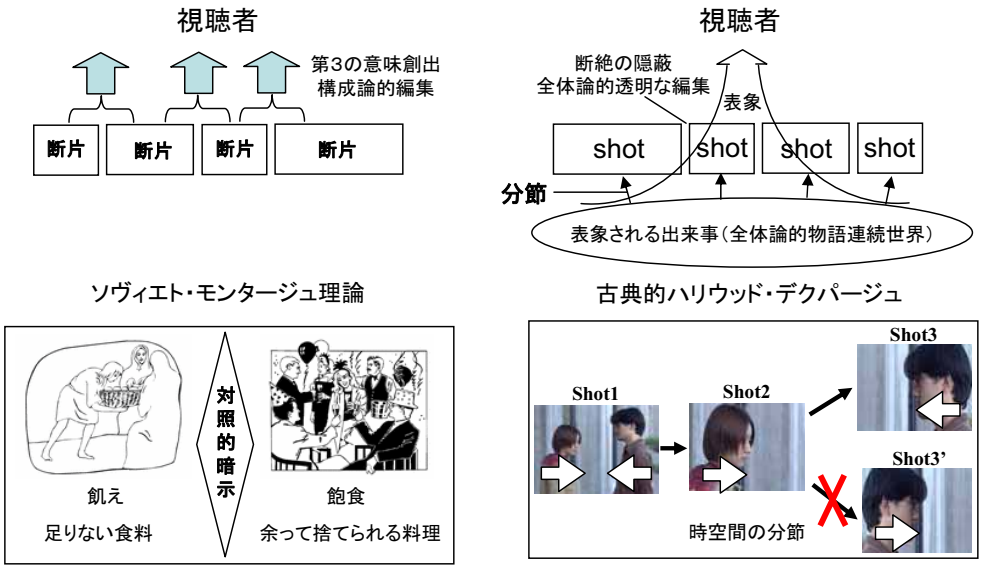


図 2.2　空間の分節と Shot size (Blocking size)

この Bazin の古典的デクパージュは，モンタージュの構成的側面をより強調したソヴィエト・モンタージュ理論と概念上，対立する．この対立は，広義のモンタージュと広義のデクパージュとの対立以上に編集に関する立場として明確な違いが現れる．

2.2.5　ソヴィエト・モンタージュ理論と古典的デクパージュの違い

図 2.2 は，ソヴィエト・モンタージュ理論と，古典的デクパージュの違いを対比させるための図である．まず，ソヴィエト・モンタージュ理論にとって，映像の断片は，物語空間から分節されることで現れた断片ではなく，全く関係のない映像から抜粋された断片でも良い．つまり，言葉の単語が，もとの物語世界を考慮して取り出された単位ではないように，第一に自律する断片があるとする．そして，言葉が自律した単語の構成的組み合わせにより，高次の意味を生み出すように，その映像断片の衝突が高次の意味を生み出すという，モンタージュの構成的側面を強調した編集概念である．

図 2.2 左の例では,「飢え」を示す自律的な断片と,「飽食」を示す自律的な断片があり，それを接続し，断片の意味を衝突させることで，同じ人間であるにもかかわらず，その違いを見せつける高次の意味を創出した例となっており，このような形式は，対照モンタージュと呼ばれる．つまり，編集行為を，ここに存在するぞとばかりに見せつけ，モンタージュにより現れる意味を見る者に意識的に読ませる技法と見なす [7].

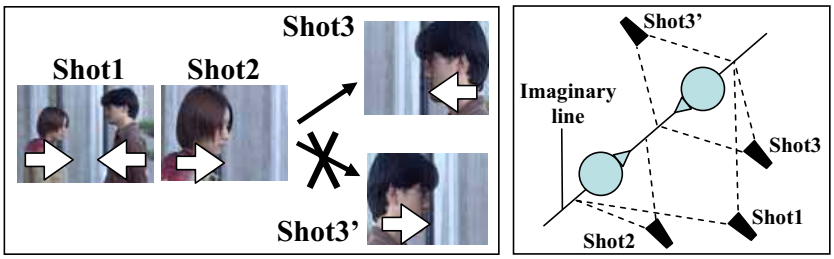


図 2.3　古典的デクパージュの典型例

一方，図 2.2 右に示す古典的デクパージュは，編集行為はむしろ読まれない (隠蔽される) ことが望ましく，編集行為 (断絶) が透明になることによって，背後にある物語世界が技法を透過し，物語世界への没入を誘引するという技法である．断片としてのショットは，物語世界の所有物であり，ショットどうしは切り出される前から，概念上，依存関係にある．この Shot1〜Shot3 の接続は，編集行為を読ませるのではなく，気づかせない (無意識下に送る) ことを目論む技法となる．この例は，映像文法として今日にも映画やテレビの領域で受け継がれる，古い時代から存在する技法である．

2.2.6　古典的デクパージュの特徴

図 2.3 は，図 2.2 右例と同じであり，古典的デクパージュのうち，最も典型的な例である．これは，Shot1 で示される男女が向き合う空間的関係を崩さないよう，被写体との距離を変更しても，Shot2，Shot3 のように向いている方向を一致させる規則である．これを守らない Shot3' では，空間的関係が知覚上曖昧になり，何らかの解釈が必要となるため編集行為が意識にのぼり，自然な無意識的内容理解が阻害される．この規則を守るには，実際の撮影上，図 2.3 右に示す，二人の撮影対象に想像上の線（imaginary line）を引き，その片側を出ないよう撮影を行えばよい．この規則には撮影の効率を上げる特徴もある．それは，規則の意味を知らずとも，この手法に従うことで，想像上の線のどちらか一方でカメラの設置を計画すればよく，同様の撮影パターンでの撮影方法のパターン化が可能となり，撮影作業の分業化にも貢献するものと思われる．

このような「一致 (Match)」に関する概念は，古典的デクパージュの中で特に特徴的な位置を占めており，他にもいくつか存在する．ここでは，それらのうち，いくつかを抜粋して説明を行う．図 2.4 はその抜粋である．

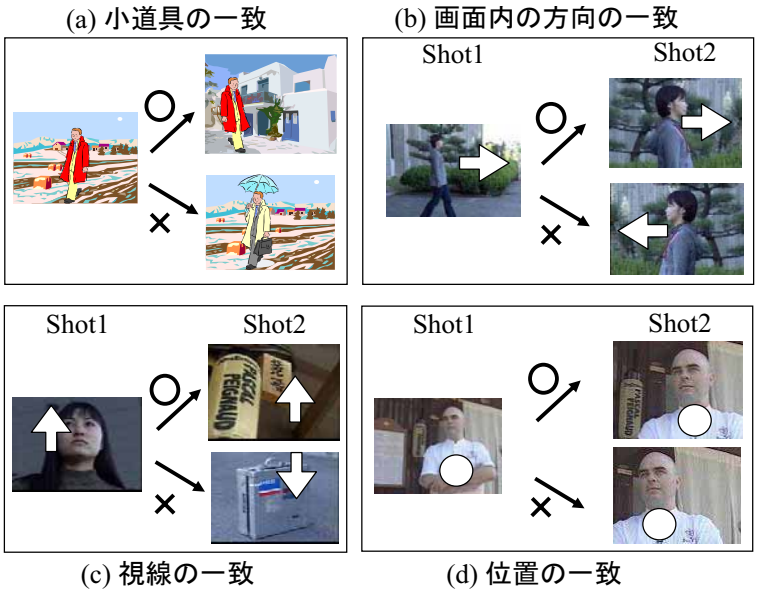
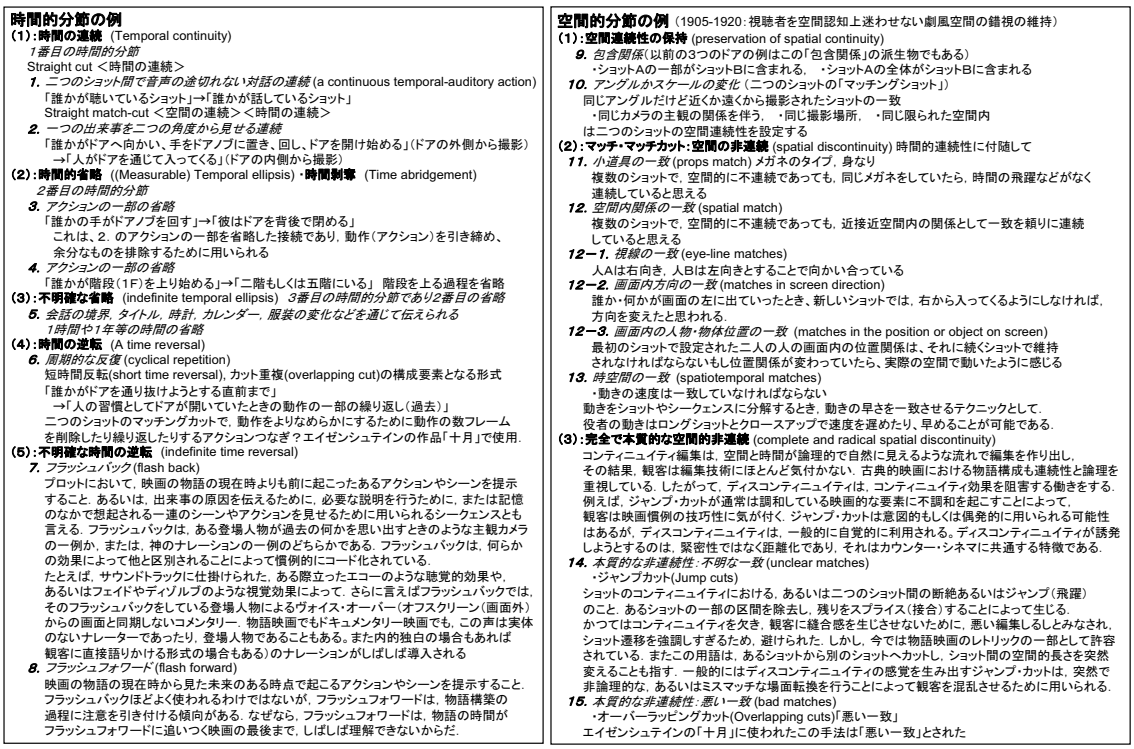


図 2.4　デクパージュの Match Rule（一致則）抜粋

通常，背景が異なれば時空間が異なる非連続の印象を与えるが，(a) は「小道具の一致」と言い，登場人物の服装や持ち物が二つの Shot で一致していれば，同じ空間と認知されることを示す．逆に言えば，同じ背景でも，登場人物の何かが変化すると，時間・空間的断絶という印象を与える．(b) は「画面内方向の一致」（この例では動作方向の一致）といい，Shot Size を変更した場合，動作の方向が一致していないと，うろうろしているなど，もとからその解釈を意図していないかぎり，別の解釈を与えてしまい，内容が曖昧になるため，それを避けるための規則である．(c) は「視線の一致」の一種で，登場人物が上を見ている場合，次の Shot は，カメラの方向を上向きで撮影し，登場人物の視線の先にある対象を映すべきとする規則である．これを破り，カメラの向きを下向きにすれば，この接続は意味が不明か曖昧になる．(d) は「位置の一致」の一種で，登場人物の構図として，右寄りの Shot の次に，左寄りの Shot が接続されると，移動したような余計な印象を与えるため，それを避けるための規則である．

後年，Burch は，古典的デクパージュの分類を行っており，1981 年に出版した文献 [10] の中で，五つの時間的分節と三つの空間的分節の大分類に含まれる 15 の小分類を行っている (図 2.5)．図 2.4 の四つの例は，連続・隠蔽を実現するための Match Rule（図 2.5 空間的分節大分類 (3)）から抜粋したものである．

この古典的デクパージュの技法は，1930 年代のハリウッドで至高の統治状態にあっ



時間的分節の例
(1):時間の連続 (Temporal continuity)
1番目の時間的分節
Straight cut ＜時間の連続＞
1. *二つのショット間で音声の途切れない対話の連続* (a continuous temporal-auditory action)
「誰かが聴いているショット」→「誰かが話しているショット」
Straight match-cut ＜空間の連続＞＜時間の連続＞
2. *一つの出来事を二つの角度から見せる連続*
「誰かがドアへ向かい、手をドアノブに置き、回し、ドアを開け始める」(ドアの外側から撮影)
→「人がドアを通じて入ってくる」(ドアの内側から撮影)
(2):時間的省略 ((Measurable) Temporal ellipsis) **・時間斜塞** (Time abridgement)
2番目の時間的分節
3. *アクションの一部の省略*
「誰かの手がドアノブを回す」→「彼はドアを背後で閉める」
これは、2. のアクションの一部を省略した接続であり，動作（アクション）を引き締め、余分なものを排除するために用いられる
4. *アクションの一部の省略*
「誰かが階段（1F）を上り始める」→「二階もしくは五階にいる」 階段を上る過程を省略
(3):不明確な省略 (indefinite temporal ellipsis) *3番目の時間的分節であり2番目の省略*
5. *会話の境界，タイトル，時計，カレンダー，服装の変化などを通じて伝えられる1時間や1年等の時間の省略*
(4):時間の逆転 (A time reversal)
6. *周期的な反復* (cyclical repetition)
短時間反転(short time reversal), カット重複(overlapping cut)の構成要素となる形式
「誰かがドアを通り抜けようとする直前まで」
→「人の習慣としてドアが開いていたときの動作の一部の繰り返し（過去）」
二つのショットのマッチングカットで，動作をよりなめらかにするために動作の数フレームを削除したり繰り返したりするアクションつなぎ？エイゼンシュテインの作品「十月」で使用．
(5):不明確な時間の逆転 (indefinite time reversal)
7. *フラッシュバック* (flash back)
プロットにおいて，映画の物語の現在時よりも前に起こったあるアクションやシーンを提示すること．あるいは，出来事の原因を伝えるために，必要な説明を行うために，または記憶のなかで想起される一連のシーンやアクションを見せるために用いられるシークェンスとも言える．フラッシュバックは，ある登場人物が過去の何かを思い出すときのような主観カメラの一例か，または，神のナレーションの一例のどちらかである．フラッシュバックは，何らかの効果によって他と区別されることによって慣例的にコード化されている．
たとえば，サウンドトラックに仕掛けられた，ある際立ったエコーのような聴覚的効果や，あるいはフェイドやディゾルブのような視覚効果によって．さらに言えばフラッシュバックでは，そのフラッシュバックをしている登場人物によるヴォイス・オーバー（オフスクリーン（画面外）からの画面と同期しないコメンタリー．物語映画でもドキュメンタリー映画でも，この声は実体のないナレーターであったり，登場人物であることもある。また内的独白の場合もあれば観客に直接語りかける形式の場合もある）のナレーションがしばしば導入される
8. *フラッシュフォワード* (flash forward)
映画の物語の現在時から見た未来のある時点で起こるアクションやシーンを提示すること．フラッシュバックほどよく使われるわけではないが，フラッシュフォワードは，物語構築の過程に注意を引き付ける傾向がある．なぜなら，フラッシュフォワードは，物語の時間がフラッシュフォワードに追いつく映画の最後まで，しばしば理解できないからだ．

空間的分節の例（1905-1920：視聴者を空間認知上迷わせない劇風空間の錯視の維持）
(1):空間連続性の保持 (preservation of spatial continuity)
9. *包含関係*（以前の3つのドアの例はこの「包含関係」の派生物でもある）
・ショットAの一部がショットBに含まれる，・ショットAの全体がショットBに含まれる
10. *アングルかスケールの変化*（二つのショットの「マッチングショット」）
同じアングルだけど近くか遠くから撮影されたショットの一致
・同じカメラの主観の関係を伴う，・同じ撮影場所，・同じ限られた空間内
は二つのショットの空間連続性を設定する
(2):マッチ・マッチカット:空間の非連続 (spatial discontinuity) 時間的連続性に付随して
11. *小道具の一致* (props match) メガネのタイプ，身なり
複数のショットで，空間的に不連続であっても，同じメガネをしていたら，時間の飛躍などがなく連続していると思える
12. *空間内関係の一致* (spatial match)
複数のショットで，空間的に不連続であっても，近接近空間内の関係として一致を頼りに連続していると思える
12−1. *視線の一致* (eye-line matches)
人Aは右向き，人Bは左向きとすることで向かい合っている
12−2. *画面内方向の一致* (matches in screen direction)
誰か・何かが画面の左に出ていったとき、新しいショットでは，右から入ってくるようにしなければ，方向を変えたと思われる．
12−3. *画面内の人物・物体位置の一致* (matches in the position or object on screen)
最初のショットで設定された二人の人の画面内の位置関係は、それに続くショットで維持されなければならないもし位置関係が変わっていたら、実際の空間で動いたように感じる
13. *時空間の一致* (spatiotemporal matches)
・動きの速度は一致していなければならない
動きをショットやシークェンスに分解するとき，動きの早さを一致させるテクニックとして．
役者の動きはロングショットとクロースアップで速度を遅めたり、早めることが可能である．
(3):完全で本質的な空間的非連続 (complete and radical spatial discontinuity)
コンティニュイティ編集は，空間と時間が論理的で自然に見えるような流れで編集を作り出し，その結果，観客は編集技術にほとんど気付かない．古典的映画における物語構成も連続性と論理を重視している．したがって，ディスコンティニュイティは，コンティニュイティ効果を阻害する働きをする．例えば，ジャンプ・カットが通常は調和している映画的な要素に不調和を起こすことによって，観客は映画慣例の技巧性に気が付く．ジャンプ・カットは意図的もしくは偶発的に用いられる可能性はあるが，ディスコンティニュイティは，一般的に自覚的に利用される。ディスコンティニュイティが誘発しようとするのは，緊密性ではなく距離化であり，それはカウンター・シネマに共通する特徴である．
14. *本質的な非連続性：不明な一致* (unclear matches)
・ジャンプカット(Jump cuts)
ショットのコンティニュイティにおける，あるいは二つのショット間の断絶あるいはジャンプ（飛躍）のこと．あるショットの一部の区間を除去し，残りをスプライス（接合）することによって生じる．
かつてはコンティニュイティを欠き，観客に縫合感を生じさせないために，悪い編集しるしとみなされ，ショット遷移を強調しすぎるため，避けられた．しかし，今では物語映画のレトリックの一部として許容されている．またこの用語は，あるショットから別のショットへカットし，ショット間の空間的長さを突然変えることも指す．一般的にはディスコンティニュイティの感覚を生み出すジャンプ・カットは，突然で非論理的な，あるいはミスマッチな場面転換を行うことによって観客を混乱させるために用いられる．
15. *本質的な非連続性：悪い一致* (bad matches)
・オーバーラッピングカット(Overlapping cuts)「悪い一致」
エイゼンシュテインの「十月」に使われたこの手法は「悪い一致」とされた

図 2.5　ノエル・バーチの古典的デクパージュ諸分類「Theory of Film Practice」1981 年

た．至高の統治状態とは「守らなければならない規則」として，作品制作上，編集規則が強制された状態を示す．しかし，1940 年代には，これらの強制的扱いは次第に緩められ，第二次世界大戦の前に実験が行われていたテレビ放送が，戦後でも国力を維持していたアメリカで再開された．このとき，古典的デクパージュの技法は，1940 年代末からハリウッドで制作されたテレビ放送番組を通じてテレビ放送業界に流入し，その技法を見習った諸外国にも広汎し始めたものと思われる．このため，古典的デクパージュは，今日の映像技法の根底に浸透しており，映像制作の基本技法となっている．

2.3　映画文法と規範文法

2.3.1　先駆的映画文法

Michael(1970)[11] によれば，言葉について，自国語，外国語のものであれ，また書かれたものであれ話されたものであれ，言葉の意味の「わからなさ (unintelligibility)」

が文法学を生む動因であると述べている [12]．物語を語り始めた映画での，初期の編集された映像が，当時の大衆に正しく理解（正読）されたかと言えば，必ずしもそうではなかった．そのためか，1900 年代後半から 1910 年代の古典的ハリウッドの映画スタイルは，わかりやすさが重視された．その技法は，のちに映画文法と呼ばれる概念に含められるが，映画文法という言葉は，第二次世界大戦が終結した 1940 年代後半に，映画が芸術として広汎に認知され，映画が特有の言語活動を備えた完全な芸術と見なされはじめた頃，学校の教科書を思わせる多数の映画に関する書物が出版されることを通じて広汎した．その出現理由は，映画という未知の言語活動をより良く知り，より良く扱うための主要な形態を探求することが要請されたためである [13]．

ただし，この映画文法という言葉を先駆的に使用したのは，イギリス人 Spottiswoode であるとされている．著書「A Grammar of the Film」[14](1935) [1] は，1920 年代のソヴィエト・モンタージュ理論を牽引した Eisenstein の理論や，同じソヴィエトの Pudovkin が示したモンタージュ分類，それをさらに発展させたドイツの Arnheim によるモンタージュの研究などをいち早く取り入れ，教育的見地から映画文法を体系化した試みであったとされる [7]．このソヴィエト・モンタージュ理論は，1917 年のロシア革命以後，1922 年に設立されたソヴィエト連邦におけるレーニンのもと，映画を娯楽ではなく，用いる言語が異なる民，字の読めない農民，暇のない労働者に，労せず社会主義の思想を魅力的な方法で伝達する新しい言葉として，むしろ映画を映像言語に仕立てるもくろみにより，実験・研究・実践が試みられた理論である．Spottiswoode の映画文法は，図 2.1 の史的順序からもわかるように，デクパージュの概念が現れる前の時代であり，広義や狭義のモンタージュ概念に影響された内容となっている [7]．

2.3.2　ロジェ・オダンによる映画文法史と規範文法

Odin は，フランスの映画理論家 Berthomieu の「映画文法試論」(1946) や Bataille の「映画文法」(1947) が，学校教育用の規範文法を手本にして作られていることを明確に示しているという (1978)[13]．Odin の説明によれば，映画言語はラングとではなく,「文学」と比較検討されており，それが目論むことは，映画の言語を「模範的作家」の作法と合致させることであるとする [13]．映画文法の目的は，映画作品の組み立てを

[1] この邦訳版「映畫の文法：映畫技巧の分析」[15] が 1936 年に日本で出版されている．

支配する「基本原則」や「普遍の規則」を学ばせることによって,「正しい映画スタイル」や「調和の取れたスタイル」を習得させることにある．そのため，規範文法として排除すべき「間違い」や「重大な誤り」が列挙され，それらは監督が特殊な「スタイル上の効果」を生み出そうとする場合にのみ許容される [13]．その編著を行う上で，伝統文法の規範的な様態を踏襲することになるが，それらが掲げる「文学における美学」となる，技法を見せつけない「技法の不可視性」と，物語世界を実在と錯覚させる「リアリズム」も語られている．これは，1950 年以降 Bazin がデクパージュと共に語る概念となるが，Berthomieu や Bataille の映画文法の概念に，既に現れている．

Odin は映画文法について次のように語っている「規範的文法は，言葉による言語に関する多くの学校文法と比べて，より優れているわけでも，より劣っているわけでもない．これは，そこでの立脚点が，真に言語学的であるよりも,「文体論的」であったということである．そして，概念の誤った比喩化を行いながらも，それらの文法は時に，映画的言語活動を記述するための手がかりを提供し，その成果を踏まえて，後に数多くの分析が展開されることになったのである」[13].

では，規範文法とはいかなる概念なのか．ここで，規範の源流と，文法を語る上で無視できない，品詞の概念とともに，文法学の興りを概観する．

2.4　文法

2.4.1　プラトン・アリストテレスと文法・規範・品詞

Platon と Aristoteles は，今日の言語・文法にも影響を与え続ける紀元前 5〜紀元前 4 世紀の古代ギリシア哲学者である．Platon は，言語を (普遍的な) 自然の理という観点で考察した．しかし，結果的に言語の理を明らかにできなかった Platon を見て Aristoteles は，理による言語観を据え置き，言語を「学」と「術」に分離した上で，言語は，話者と聞き手の取り決めによって成立する相互の伝達という経験が生み出す「術」であるとした [16]．Aristoteles も，Platon と同様に言語と (理に通じる) 認識の関係を考察しているが，Aristoteles の言語考察の対象は，言語の普遍性や認識の表現である前に，まず，思想を伝達するための形式に向けられた [16]．Aristoteles は，形式は多様であるために，言語表現を明瞭にする「標準」に目を向け，弁論や詩作をより効果的に伝達す

るためには，最も伝達されやすい多数派の共通性が反映され規範となる「標準語」に習うべきだと考えた [16]．このように，Aristoteles は，言語を「伝達の術」として，その形態を分類し，標準を知るという言語研究の課題と方法を確立した．この課題と方法は，その後のギリシア語文法，ローマ帝国の政治的支配と結びついてラテン語の規範文法に発展し，西洋の伝統文法に引き継がれ，東洋にも影響を与えていく．この伝統文法の基礎を築いた点において，Aristoteles は，古典的ヨーロッパ文法の最大の創始者と言われる [16]．つまり，Aristoteles の形態分類は，弁論や詩作をより効果的に伝達する標準を知るためのものである．形態の分類とは，今日の品詞分類の基礎となる．

品詞分類は，文法という概念と容易には切り離せない概念であるが，その源は，古代ギリシアの言語の考察において，言語をまず最初に「構成要素」に分析することから始まった [17]．Platon は「文」を「オノマ」（「名詞」または「主部」と言えるもの）と，「レーマ」（「動詞」または「述部」と言えるもの）に大きく二分し，Aristoteles は，「シュンデスモイ」（前置詞，接続詞などそれ自身では意味を持たないもの）を付け加えたという．そして，これらの構成要素は，Platon 以来「$\mu\varepsilon\rho\eta\ \lambda o\gamma o\upsilon$」と呼ばれた [17]．

これは，ギリシア文字を代用するローマ字表記で表現すると「*merê logû*」となる．*merê* は「部分」，*logû* は「実際しゃべった言葉」という意味でよく使用される．これを，ラテン語では partes orationis（発話の部分）と翻訳し，英語ではラテン語の翻訳として parts of speech があてられたが，この日本語への翻訳語が「品詞」である．

2.4.2　最初のギリシア語文法の規範・品詞

現存する最初のギリシア語文法は，（紀元前 2 世紀末〜紀元前 1 世紀初め）頃の Thrax によるものとされ，この時点で八品詞が立てられている．これが今日の概念と親和性の高い品詞概念の始まりであるが，この時代，品詞は最も重要な部門ではない．

Thrax の文法は，以下の六つの項目がある．1．韻律法にしたがった正しい読み（音読），2．テクストに出てくる文学的技巧の説明，3．不明瞭な語句の注解と，言及されている神話や歴史的出来事の注解，4．語源の解明，5．規則的範型（analogia: 語形変化などの規則性）の詳述，6．詩の批判的評価．このうち，6 の文学的観点が当時の文法学で最も重要な部門であるとされるが，現代の文法学に主に引き継がれているのは 5 のみであるという [12]．

宮脇によれば，Thrax にとっての文法は，自分の言葉遣いを良くするといった「アウトプット型」の文法観 (2.4.3 で述べる) ではなく，先行する詩人や散文作家が残した作品を観察することによって，あるがままの言葉遣いを引き出し，そこに何らかの「規則性」を見出そうとする営みであり，いったん規則性が確立された後では，それを「規範」として用いるという，著作家の慣用の検討であるとしている [12].

2.4.3 規範文法

規範は，まず第一に，言語表現を明瞭にする多数派の共通性を反映した「標準」を知るため，先人の慣用を検討した結果であり，多数派が用いるがゆえに何らかの特性を持ち，多数派が受け入れた慣用であり，多数派が共有することで自然な対話ができ，多数派の一員となるためのお手本である．その反映である文法は,「正しく読み理解する」の部門をインプット,「正しく書き話す」の部門をアウトプットとして特性が論じられることがある．ただし，概念は，その焦点に偏りを見せることがある．インプット型は「正読，知るための文法，学」に偏る文法観，また，アウトプット型は「正話，実践・表現のための文法，術」に偏る文法観である [12].

この観点において，19 世紀に興った科学的言語研究の方法論は，現代言語学を含めて言語表現のなぜに焦点を当て探究するインプット型に偏る．一方，近世 (15,16 世紀) から 18 世紀末までの規範文法には，受動的なインプット型の聞く・読む能力よりも能動的な書き話す能力がより高級な知的作業と見なされ，文法を art（技）とし [4, 18],「文法とは，正しく書き，話す技術である」とするアウトプット型の観点が重視される．宮脇によれば，このアウトプット重視型文法観の源流は，古代ローマの Quintilianus の「雄弁家教育論」(紀元前 90-95) にあるという [19]．この時代の文法学にも，現代の理論的・語学的要素だけでなく，詩学・文学的要素が含まれ，むしろ，その詩学・文学的要素が重視された．Quintilianus の文法は，雄弁家を育てるための文法であり，雄弁家として「正しく話す」ためには，話す内容を豊かで効果的なものにするための「詩学・文学」の素養が必要であると述べている [7]．この偏りを見せたアウトプット型は，正しく書き話す (技) を基礎として，良い表現 (技) を探究するが，なぜその表現が良いのかという科学的視点については気に止めず，説明もできない．一方，偏りを見せたインプット型は，慣用表現の理 (学) を探究するが，良い表現とは何かを気に止めない．

また，規範には，気をつけなければならない偏りがある．規範文法は，古代ローマ，ラテン語の文法形式に基づくが，ラテン語の知識は，イギリスにおけるある特定の社会階級のグループの慣例と結びつき，それが見本とされるようになった事例がある．その場合,「文法の正確さ」の概念は，言語自体に内在する特徴から生じるものではなく，ある限られた階級グループの社会的慣例から形成されることになる [20]．さらに，規範は，こう「するとよい」「すべき」から「しなければならない」に到ることもある．

Odin によって，1940 年代後半に現れた映画文法は，規範文法を参考にしたことが明らかとされた．この点で，映画文法の概念は，20 世紀前半，アメリカやヨーロッパで制作された，先人の作品を観察し，規則性の体系化を目指す慣用の検討が第一にあり，文学の概念が取り入れられていることも，古来の文法学に沿う姿勢と言える．ただし，古典的デクパージュにおける規範は，1930 年代，お手本というよりも，厳格な規則として強制された時代もあった．しかし，1940 年代以降はその傾向は緩み，映画では，新しい技法が次々に創造される時代に進み，慣用が何であるかは明瞭でなくなった．

一方，テレビ放送に浸透した規範は，一般大衆という多数派が受け入れる性質を維持しなければならず，初期の規範が色濃く受け継がれることになる．本論文で抜粋する映像文法は，テレビ放送に受け継がれ，多数派に許容される技法という制約を受けながら進展を遂げた技法であり，古典的デクパージュの概念が色濃く残る．また，映画では撮影者と編集者は，一体となって映像が制作されるが，テレビでは，特に取材などにおいて，撮影者と編集者が分離し，独立に独自の発想で撮影や編集が行われることもある．このような場合，綿密な計画に基づいた撮影ではないために，最低限の守るべき約束として，映像文法に従う撮影・編集は映像の品質を維持する効力を発揮する．では，その規範をどのような方法で共有するのか．文法の記述は，インプット型の観点が参考になる．そのインプット型の文法観で重視される品詞は，何故に文法にとって必要なのか．その大局的本質を次に述べ，本論文の映像文法記述に利用する．

2.4.4　文法・品詞分類の大局的本質の一面

古来の文法が品詞を最重要部門にしなかったとしても，インプット型文法観の話題に品詞の概念は欠かせない．ここでは，品詞と映像文法の関係を述べる．名詞や動詞などの性質を詞性と呼ぶならば，日本語の「品詞」には，初めから詞性が含まれてい

る印象がある．しかし，parts of speech のもとになるオノマとレーマは，分節された構成要素を名詞や動詞という観点だけでなく，主部と述部という単位で性質を分類する観点も内包している．さらに指摘するならば，parts of speech は，発話の部分であり，詞性の印象はなく，発話に構造がある観点が第一に主張されている．その次に，各部分のカテゴリー分類があり，詞性の観点が色濃く表れたカテゴリー分類が品詞である．

ただし，こうしたカテゴリー化では，一つの規準で多数のカテゴリーを抽出することが望ましい．しかし，近代の様々な個別言語の研究において，形態的特徴だけでは，名詞，動詞，その他の三つ程度にしか分類できず，一般的な 10 前後の品詞分類では，三つの異なる規準が主観的に組み合わされる点で，非科学的とされる問題指摘があり，その組み合わせ方や比重に世界の言語が共通する普遍性は存在しないことなどが指摘されている．また，中国語などでは，語形変化がないため，品詞分類を無理矢理持ち込む必要は必ずしもないとの指摘もあり，すべてのヒトの言語が品詞に基づく文法記述を前提とする必要はないとの可能性も示唆されている．

しかし，西洋文法に依存する文法から品詞を排除すると困った問題が発生する．それは，代替的概念がないかぎり，言語や文法の説明を行う際，その言語記述に適切なレベルの抽象度を持つカテゴリーを用いた説明ができないため，無数に存在する言語表現の数だけ，異なる説明を行わなければならないという，途方もない事態を招くためである．文法学では，できるかぎり言語表現を網羅的に簡潔な少ない記述で説明する指針を持つために，適切な抽象レベルのカテゴリーが必要なのである．つまり，古代ギリシアの文法観に従うかぎり，言語表現を適切なレベルの抽象カテゴリーに分類し，カテゴリーに基づいた簡潔な説明を行うことに，その大局的本質があると思われる．

映像文法は，ショットサイズに関連して記述されることが多いため，ショットサイズは，品詞の代替として，映像文法を記述するための抽象カテゴリーの有力な候補である．

2.5　文法記述の抽象カテゴリーとショットサイズ

2.5.1　絶対的ショットサイズと空間的分節

映像を撮影するカメラは，世界を「フレーム」（四角い枠）によって空間的に分節する．フレームが動くと構図が変化することになるが，映像文法では面積の 5% の変化

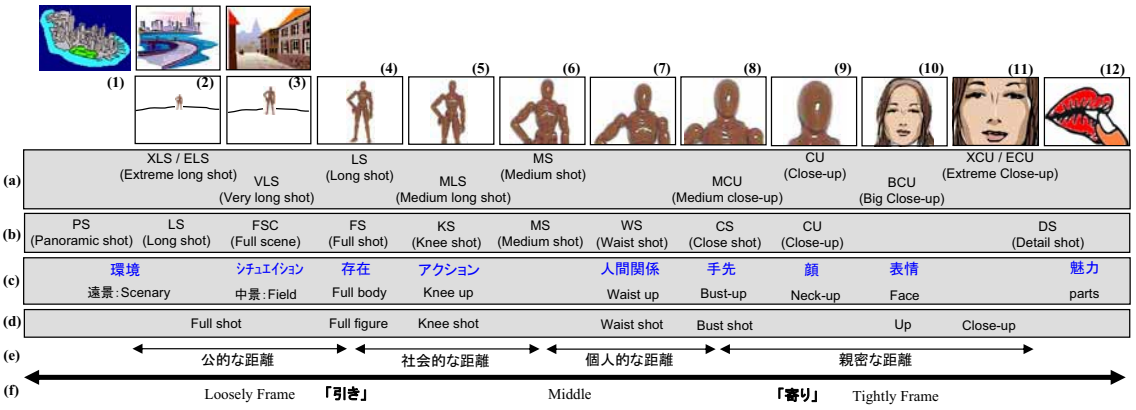


図 2.6　空間の分節と絶対的 Shot size (Blocking size)　(a):[21]，(b):[22]，(c):[23]，(d):[5]

が起これば異なるフレームであるとされている．このショットの空間に関する観点は，時間的区間としてのショットではなく，フレームを通して見える撮影対象に着目した観点である．また，空間的にフレームが固定された状態をフィックス（FIX）と呼ぶが，フィックス以外に，映像撮影用のカメラの動きによって，空間的にフレームから見える映像全体の視点が変化し続けることに対応する，カメラワークが含まれる場合がある．

映像を用いた内容伝達法では，フレームで切り取られる世界のあらゆるパターンの中で，変動項を含みつつも，ある共通する特性に基づいた記号的・カテゴリーを用いて映像を意識的に空間分節する．それが図 2.6 に示す「ショットサイズ」(Shot Size) である．別名「ブロッキング・サイズ」(Brocking size) ともいい，この空間からフレーム枠を切り取る行為を「フレーミング」(Framing) と呼ぶ場合がある．また，カメラと被写体との距離と考えることもできる．ただし，カメラは，レンズ操作によるズームという機能を持った時代から，このカメラと被写体の距離という定義は必ずしもふさわしい定義ではない．しかし，見た目に依存した感覚としては，カメラと被写体との距離で説明してもさほど問題とはならない．

図 2.6 の (a)〜(d) までは，ショットサイズの分類に対応する「絶対名」であるが，映像の現場により若干の方言的違いによる，呼び方にずれを生じている場合もある．映像の撮影対象としては，人もしくは擬人化された対象が多いためか，ショットサイズは，人の体の特徴に基づいて分類される傾向がある．次に，図 2.6(e) はそれぞれのショットサイズに対応する個人・社会上の距離である [23]．(f) は，「相対的な関係」を示し，被写体の周囲が入るものほどルーズな Shot，被写体に近づくほどタイトな Shot であり，そ

の中間がミドルなShotとなる．ショットサイズの利用法として，人の場合は，心理的な距離の描写に使われる場合がある．撮影対象が物体の場合は，対象の全体像を伝えたり，焦点をどこにあてるかといった表現に用いられる．また，ショット内でカメラワークを含まず，カメラが固定され，フレームの動きがない一定の時間的区間を「フィックス・ショット」と呼ぶが，ショットサイズは，基本的にフィックス・ショットに対して規定される．ショットサイズは，空間をフレーム枠で切り取ることで得られる分節要素と考えることができるが，映像文法では，フレームが，5%の枠取りの違いにより異なるフレームとするために，カメラワークが存在する時間的区間は，常にショットサイズが変化しているため，ショットサイズが規定できないと考えることができる．

このショットサイズを映像での「品詞分類」に相同させる者の一人が，Berthomieuである．Berthomieuは，映画文法を記述するその着想において，映画言語の分析に，自然言語の文法をかなり忠実に反映しており，実際（「単語」に相当するものとして）ショットから出発し，（「品詞分類」のように）ショットサイズの分類目録が作成され，ショットが（映画的な文として）「シークェンス」に構造化される際の規則が説明されているという[13]．例えば，「Long ShotからClose Upに飛躍することは，その意外さと視覚的なショックによって観客の注意を引く，意図的な誤りとなる場合もある」と述べているという[13]．映画史家Sadoulによれば，映像におけるClose Upを初めて用いたのは，1900年ごろのイギリス人Smithによるとされる[6]．この時代には，全景（Long Shot）を固定で撮影する方法が一般的であったため，初期の映像には，Long Shotしかなく，そこにClose Upが加わったと見ることができる．そして，ショットサイズは，新たな技法の発明とともに増やされたと思われるが，初期の技法では，この少ない絶対的ショットサイズに基づいて接続規則等が発明されていったと思われるため，その接続技法を文法の一部とするならば，文法記述は，絶対的ショットサイズに基づいて記述されても問題はなかったものと思われる．しかし，図2.6が示すように，現代では，ショットサイズが細分化されているため，絶対的ショットサイズに基づいて文法的記述を行うことは文法記述を煩雑化する．毎日放送の映像文法では，絶対的ショットサイズを映像文法の記述基盤とせず，相対的ショットサイズを用いて文法規則が記述される．

現代言語学では，品詞分類について，古代ギリシアからのカテゴリー化に関する問題点の指摘から，新しいカテゴリー概念による品詞分類の見直しが，特に認知言語学において始まっている．例えば，語を特定したとき，その語の品詞は，語の特徴だけ

では品詞を特定できない場合も数多く存在する．特に中国語などの孤立語（古典的言語類型論の術語）では語が語形変化しないため，語だけでは品詞が特定できない．孤立語では，語形よりも，語順が品詞分類に貢献していると説明されるが，その語順を使って分類した語も，表記上は，複数の品詞にまたがることになり，文脈によって複数の品詞に使い分けられるため，語を規定する境界は，人工言語でないかぎり，古代ギリシア以来の二項分類的に明晰に分類できないことが指摘されている．認知言語学の礎となった新しいカテゴリー概念に，要素の重複を許し，カテゴリーに典型例と非典型例に渡るレベルを導入した Rosch のプロトタイプ理論がある [24, 25]．このプロトタイプ的カテゴリーの考え方は言語における品詞の概念自体にも適用され，品詞分類が再検討されている．例えば，名詞や動詞のような品詞も，境界は明確でなく，それぞれ典型的な事例と，非典型的・周辺的な事例をもつと考える．このプロトタイプ理論を参考とした相対的ショットサイズの分類体系を次に示す．

2.5.2　相対的ショットサイズ

ショットサイズは，絶対的・相対的を問わず，フレーム内に占める被写体の占める程度，見方を変えれば被写体とカメラ (または，映像を見る側) との感覚的距離を表している．絶対的なショットサイズに対し，映像文法では，相対的なショットサイズがある．図 2.6(f) が示すように，ショットサイズにおける，被写体が余裕をもってフレーム枠に含まれるショットをルーズショット (LS:Loose Shot) と呼び，フレーム枠が被写体の一部を撮影し，被写体全体が写らないショットをタイトショット (TS:Tite Shot) と呼ぶ．このルーズショットとタイトショットの中間に位置するのがミドルショット (MS:Middle Shot)，である．古典的デクパージュでは，もともと同じ空間内の対象を複数のショットサイズに空間的に分節する観点，またそれらのショットが連続している印象を与えるための一致則（Match Rule）に従う．一致則は，空間的に断絶し，フレーム枠を通じて見た世界でも，何かを頼りに連続した空間であると感じさせるプロット[2]の関係を利用する技法であり，基本的に，三つの相対的ショットサイズは，何らかのプロット（ここでは撮影対象）による包含関係を持っている．例えば，ルーズショットで撮影したプ

[2]プロット (plot) とは，上映中のスクリーン上で提示されるすべての対象．物語世界的な出来事に加え，キャプションや日付，タイトルなど非物語的な素材も包含し，現在では音楽も含まれており，そのすべての扱いが映画の物語上の戦略となる．つまり，物語全体が進む方向に視聴者を導くもの．

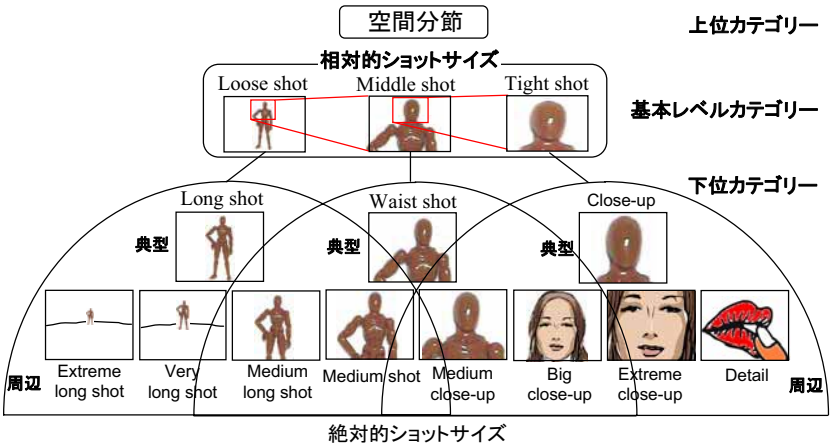


図 2.7　プロトタイプ理論に基づく相対的ショットサイズの分類体系

ロットから，次のミドルショットでは，同じプロットの一部分，そしてそのミドルショットの一部がタイトショットとして撮影されるなどの関係となる．

Rosch は，古来の二項的カテゴリー分類法では無理が生じる多数のカテゴリー化の事例から，典型事例と周辺事例を同一カテゴリー内に含め，上位・基本レベル・下位などの階層を用いた分類を行う新しいカテゴリー論として，プロトタイプ理論を提唱した．これが認知言語学の契機の一つとなっている [24]．

本論文では，この手法を参考とし，下位カテゴリーを絶対的ショットサイズ，基本カテゴリーに相当するレベルを相対的ショットサイズとし，上位カテゴリーを空間分節（Spatial articulation）として，絶対的ショットサイズと相対的ショットサイズの関係を述べる．図 2.7 は，プロトタイプ理論に沿ってショットサイズを分類した分類体系である．図 2.6 から包含関係にある三つの相対的なショットサイズを無作為に取り出すパターンは数多く存在する．ただし，ルーズショット，ミドルショット，タイトショットの境界は曖昧である．しかし，各相対的ショットサイズは，それぞれ絶対的ショットサイズの典型事例を持ち，ルーズショットでは Long shot，ミドルショットでは Waist shot，タイトショットでは Close up が典型事例として該当する．つまり，図 2.6(f) が示すように，各カテゴリーの成員は，図 2.6 の左側に近いほど，よりルーズであり，右側に近いほどタイトであるため，境界は曖昧あるが，より典型的なものと，より非典型 (周辺的) なものという，カテゴリーの成員らしさに差が生じる．本章では，文法記述用抽象カテゴリーに，この相対的ショットサイズを用いて映像文法の規則を記述する．

表 2.2　カメラワークに依存した単一ショットレベルの映像文法

Rule(1-1)	カメラワークを用いるときは前後に 1 秒以上の FIX ショットが必要
Rule(1-2)	カメラワーク（PAN）は滑らかに動かすこと
Rule(1-3)	カメラワーク（ZOOM）は一定速度であること
Rule(1-4)	カメラワーク（FIX）はしっかり止めること
Rule(1-5)	FIX ショットで LS,MS,TS の時間長はそれぞれ基本的に 6s, 4s, 2.5s
Rule(1-6)	被写体の動きがない FIX ショットは最大でも 15 秒まで

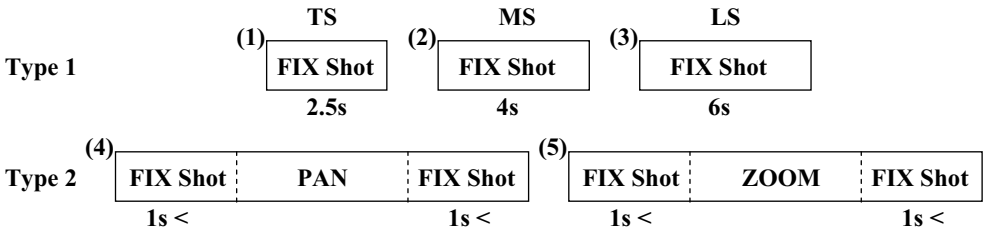


図 2.8　映像文法から導かれるショットのタイプ

2.6　本論文で用いる映像文法

2.6.1　映像文法における三つのクラス

本章では，抜粋した映像文法を三つのクラスに分ける．一つ目は，単一のショットに関する規則であり，編集への考慮と，余計な解釈を誘発しないカメラワークの品質に基づいた規則である．二つ目は，シーン内での相対的ショットサイズに基づいた規則であり，相対的ショットサイズの遷移モデルが記述できる規則である．三つ目は，シーン内の遷移モデルに制限・付随事項を与え，構文的パターンの形成に寄与する規則である．

2.6.2　単一ショットの映像文法

Rule(1-1)

表 2.2 は，単一ショット（ワンショット）に関する映像文法である．特に，表 2.2 の Rule(1-1) は，編集を考慮した規則である．この Rule(1-1) から，典型的な単一ショットの形式が少なくとも二つの時間的分節のタイプとして定義できる．その一つが Type

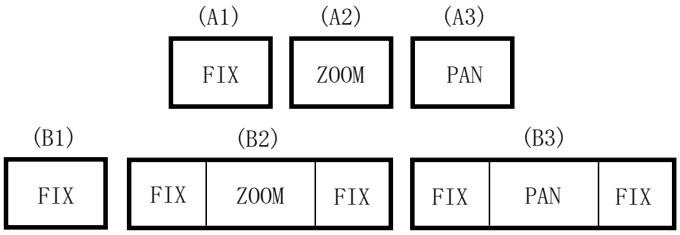


図 2.9　ショット接続の禁則

1(図 2.8 上段) であり，もう一つは Type 2(図 2.8 下段) である．この Rule(1-1) は，古典的デクパージュにおける「継ぎ目の隠蔽」で説明できる．

図 2.9 の A1,A2,A3 の接続を考察すると，A1 から A2 への接続では，カット点を挟み，FIX から突然 ZOOM が起こることになり，A2 から A3 への接続では，ZOOM から突然 PAN が起こることになる．これは，視覚上，目に付く映像断片の転換となり，継ぎ目が強調されるため，古典的デクパージュにおける編集行為の「隠蔽」の観点からは避けるべき接続 (編集) である．つまり，カメラワークの前後にのりしろとして少なくとも 1 秒の FIX を撮影しておけば，編集の際，継ぎ目の存在を和らげ，隠蔽することに貢献する．編集の際には，最終的にのりしろの部分を排除する場合もあるが，基本的に撮影の時点では Rule(1-1) を守ることで，編集上の選択肢が得られる．

また，視聴者にとって，カメラワークの動きが示す意味はきわめて曖昧である．特に，何らかの意図を持ってカメラワークを用いないかぎり，プロのカメラマンは，カメラワークの使用を必要最小限に留める．仮にカメラワークを使用する場合でも，Rule(1-1) に基づいた撮影は，カメラマンにとって，カメラワークの動きの始点・終点を意識し，余計な動きが入り込まない撮影を心がけることに効果があり，安定なカメラワークの動きを生み出す撮影技能の習得にも貢献する．

Rule(1-2)〜Rule(1-4)

表 2.2 の Rule(1-2)〜Rule(1-4) は，カメラワークに依存する当然の規則と言えるが，カメラワークは極めて曖昧であるため，不要な動作は不要な意図を発生させてしまう．例えば，FIX では，手ぶれなどが起こっていると，地震などの意味を与えてしまったり，腕の悪いカメラマンの存在が感知されてしまうことから，映像の世界に没入することを阻害することにもつながる．

この映像世界への没入を阻害する観点では，PAN や ZOOM の動作が乱れていると，同様の状況となる．特に，ZOOM の動作では慎重性が必要である．このレンズによる視覚的変化は，人間の視覚機能にはないものであるため，ZOOM は，自然性を阻害することへの影響力が高い．そのため，基本的には，一定の速度で撮影することが望ましいとされる．一方，PAN は，人間の首を動かすときに目が見る世界と同様の動きとなるため，ZOOM よりはその動きが許容されると言われる．また，プロの制作者の間では，ZOOM と PAN をできるかぎり使わないことが規範的指針とされている．このことから，カメラワークを乱用しがちな初心者には特に必要な規則である．

Rule(1-5)

表 2.2 の Rule(1-5) は，相対的ショットサイズに依存して，ショットの空間的側面としての画面情報量に応じて，視聴者が内容理解に必要とする時間長を考慮した規則である．冗長性は cutting によって排除されるが，短くなりすぎると展開が早すぎて連続編集の観点から自然に理解できる映像とならない．ハリウッドの編集概念には，素材映像を cutting してショットを生成する場合，画面情報量の少ない寄り (TS) の時間長は短く，画面情報量の多い引き (LS) は長くカットすることで内容と時間のバランスを取るカット情報一定則 [23] がある．この具体的な時間長としては，古典映画の時代から七五三理論 [26] が知られている．この理論では，映画の 35mm フィルム 90 フィートを 1 分とした場合，経験から，引き (LS)，中間 (MS)，寄り (TS) はそれぞれ 7，5，3 フィートが良いとされた．ただし，同じ「引き」でも Plot の無い漠然とした Shot では「引き」でも時間は短くする [23]．これはそれぞれ 5 秒，3.5 秒，2 秒に相当する．本研究では，放送局の映像文法 [5] に従い，LS は 6 秒，MS は 4 秒，TS は 2.5 秒が良いとされる，この時間長を目安とする．

Rule(1-6)

Rule(1-6) は，冗長性を排除する切削編集と時間の自然な連続性を問う連続編集を背景とした規則である．冗長な部分が存在すると，映像内容の展開が起こらないことに視聴者が待ちきれず，映像への没入や集中力が途切れてしまうだけでなく，全体の時

表 2.3　シーンにおける映像文法抜粋

Rule(2-1):	ショットサイズの急激な変化 (LS-TS) は避ける（自然性の保持）
Rule(2-2):	LS と LS は接続できる
Rule(2-3):	TS と TS は接続できる
Rule(2-4):	MS と MS は接続できない（冗長性の排除）
Rule(2-5):	ショットサイズの接続は単調にせず変化を優先すること（冗長性の排除）

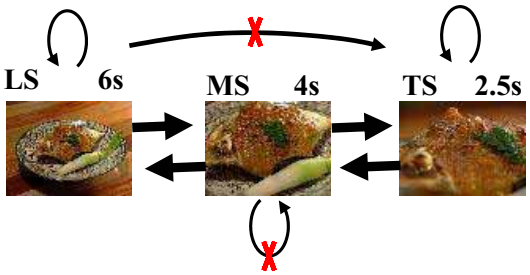


図 2.10　シーンレベルの映像文法による相対的ショットサイズ遷移図

空間の簡潔な流れを阻害する．放送局の映像文法では，特別な理由がないかぎり，被写体の動かない FIX ショットは，最大の時間長を 15 秒としている [5].

2.6.3　シーン内の相対的ショットサイズに関する映像文法

表 2.3 は，シーン内の相対的ショットサイズに基づいたショットの規範的接続規則である．表 2.3 は，特に相対的ショットサイズの遷移規則について，最も基本的な規則を抜粋している．このため，この規則から，図 2.10 のような相対的ショットサイズに基づく基本的な状態 (相対的ショットサイズ) 遷移図が記述できる．

Rule(2-1)

Rule(2-1) は，古くからある古典的デクパージュの一種で，Berthomieu の映画文法にも表れる規則である．急激な変化は視聴者を驚かせることになるため，それを視覚的効果とする意図がないかぎり，禁止される．また，接続のなめらかさや流動性の観点から，そして，ショットサイズの大きな変化が，連続性の意味において視覚的迷子を引き起こす可能性もあるため，いずれの観点からも禁止の対象となる．

表 2.4　映像文法の構文規則抜粋

Rule(3-1):	シーンの冒頭はマスターショット (シーンの全体像:通常 LS) で始める
Rule(3-2):	シーン中最初と最後のショットは 2 秒増やす（時間遷移的なめらかさ）

Rule(2-2)〜Rule(2-4)

Rule(2-2)〜Rule(2-4) は，Rule(2-1) 以外の接続可能性から導かれる．Rule(2-1) や Rule(2-2) において，LS どうし，TS どうしは接続が可能となっている．しかし，MS を 2 回以上続けてはいけない規則 Rule(2-3) がある．MS は，LS と TS の接続という大きなショットサイズの変化を避け，なめらかな接続となるよう，その緩衝的役割を担うために導入された中間の相対的ショットサイズである．つまり，本来の役割からすれば，LS と TS をつなぐためのものであり，MS が連続すると，本来到達するショットサイズへ移動せず，冗長なシーンとなるからである．その意味では，Rule(2-2) や Rule(2-3) も LS や TS の状態に長く停留すれば冗長となるが，緩やかに認められている．

Rule(2-5)

Rule(2-2) や Rule(2-3) によって，緩やかに LS や TS が連続することは認められているが，LS もしくは TS が続きすぎると単調となる．Rule(2-5) は，もし，接続候補として，同じショットサイズと異なるショットサイズが存在するとき，変化を生む異なるショットサイズが選択されるよう，接続の優先度を与えることに貢献する．これによって単調性が排除される．

2.6.4　シーン内の構文に関する映像文法

構文が，二つ以上の要素からなる構成体が持つ，パターンであるとするなら，表 2.3 も同じシーン内の構造を定める一種の構文とする観点も考えられる．しかし，表 2.3 は，前後のショット接続に関して制限を与える規則にすぎない．表 2.4 の規則は，表 2.3 の規則に対して，シーンの先頭や末尾について，明確にシーン構造を規定する役目を担う．この観点から，これらの文法クラスを構文に関する映像文法として位置づけた．

Rule(3-1)

古典的デクパージュの概念を継承する技法では，シーンの冒頭はマスター（シーンの全体が映る）ショット（相対的に普通はLoose Shot）で始めるスタイルを規範とした．これは，シーン内のプロットの空間的関係を全体像として提示しておき，複数のプロットがある場合にはそれぞれのプロットに焦点を当てる，つまり，全体から焦点へ向かう表現スタイルである．このRule(3-1)によって，シーンは，常にLSで始まるという構造を持つことになる．映像の視聴は，読書と違い，視聴者側に読み直しに相当する巻き戻し等の権限がないかぎり基本的に見直しはできない．映像のシーンを組織する際，構文的パターンは，見慣れれば理解の仕方はパターン化し，素早く無意識的に理解できるようになり，自然な技法として受け入れられるようにもなる．

古典的デクパージュでは，全体から細部へという親しみやすく誰もが受け入れやすい語り口を啓蒙しパターンを定着させたのである．つまり，古典的デクパージュの概念を引き継いで新しい映像対象を表現する際は，いかにわかりやすい構文的パターンを生み出すかが課題となる．テレビ放送の領域では，映像は物語だけではなく，様々な情報番組，スポーツ中継などの映像がある．このため，断片の遷移について解釈を容易にする技法は，それぞれの映像ジャンルごとに存在する．9章の野球中継映像に適用されたPC(ピッチャー・キャッチャー)ショットで始める構文的構造もその一例である．

Rule(3-2)

Rule(3-2)は，シーンの冒頭と最後に位置するショットは，時間長を2秒増やす規則であるが，これは，シーン全体の時間遷移のなめらかさを生むための規則である．これも経験則としての規範であるが，シーンが始まるという印象や，終わるという印象を感じやすくなる効果がある．

2.7 結言

文法が出現する必要条件は，第二言語として語学学習をする者にとって，学習対象の言語の意味の仕方，さらには，その言語を扱うための組織法がわからないときである．そして，語学学習者にとっては，学習対象となる言語の文法の良さが語学学習の

効率を左右する．同様に，映像の撮影・編集法がわからない者にとって，映像メディアの組織法を学ぶ良い文法的記述が求められている．その文法にも，ただ先人の編み出した優れた技法を見習う段階から，より良い表現力を発達させることを目指したアウトプット型重視の文法観と，その技法の背景となる理論を追求し，説明を可能とするインプット型重視の文法観が存在した．いずれの文法観においても，もし，適度に抽象化されたカテゴリーに基づく文法記述がなければ，無数の表現形式について，一つ一つの技法を説明しなければならない．そのような混沌とした無限に思える行為を要求するようなものに取り組もうとする者は，教える側にも学ぶ側にもいない．もし，それが少ない規則とそれに基づく説明で学習できるなら，それにこしたことはない．

また，映像文法の文法性と言語の文法との比較を行うことが重要なのではない．まず映像表現が言語や文法であるかを論じる前に，映像文法は，先人の見本となる技法をまねるための，規範的文法である．それらが，伝達内容の自然な理解を促す技法であるならば，人にとっての自然な情報伝達機構を反映する概念として，普遍的な資質を持つ可能性がある．その点で多数派に受け入れられる技法は，普遍的であるとも考えられる．それは，空間的解釈の迷子にさせず，自然な理解を促す技法である．

一方，言語学の文法とう概念そのものも，20 世紀後半から新展開がある．映画文法の議論が盛んであった 1940 時代後半には，言語学史上，統語論の発展の契機となった Chomsky の生成文法論や，意味論を進展させ，脳科学や認知科学の知見を取り入れた認知言語学は存在しておらず，1950 年代後半からの動きである．また，生成文法や認知言語学は，言語表現の正しい読みや理解の背後となる理論の追求を行っている点で，インプット型に偏っているが，選択体系機能文法など，アウトプット型に傾倒した文法概念も 20 世紀末から発展しており，今後の言語学は，インプット型とアウトプット型の研究が進展する可能性もある．映像文法は，このような言語学の研究成果から，インプット型，アウトプット型の新しい体系化を行える可能性があり，工学との融合によって，コンテンツ業界の制作支援・人材育成の進展に貢献する可能性がある．

映像文法は，概念的に，古典的デクパージュ，ソヴィエト・モンタージュ理論，テレビの世界等で扱われる技法を含め，映像の組織に関わる表現体系を取り入れた概念であるが，本論文では，歴史的にも最も古い古典的デクパージュの概念に偏る映像文法の一部を工学的に応用し，その第一歩として，撮影・編集の支援，映像コンテンツの自動撮影，自動生成支援技術に関する研究に応用する．

第3章

高時間分解能・高速カメラワーク解析方式

3.1 緒言

映像文法 [5] を工学的に応用するシステムとしては，映像の撮影・編集を支援する方法が考えられる．3 章と 4 章では，映像の撮影を支援する方法として，人材育成の観点から，映像文法に従った撮影法を学習する映像撮影学習システムの一つとして，訓練指向のオンライン単一ショット撮影ナビゲーションシステムを提案する．本研究は，4 章で提案する訓練指向のオンライン単一ショット撮影ナビゲーションシステムの部分システムとして必要となる，実時間オンライン処理向きの，高い時間分解能を実現する高速なカメラワーク解析法に焦点を当て，輝度投影相関法と二分化テンソルヒストグラム法を用いたカメラワーク解析手法の提案を行う．

一般的に，初心者の撮影では，不適切なカメラワークが問題となりやすく，撮影者の乱れたカメラワークを解析し，問題点を減らすよう指導する手法が必要となる．本研究では，映像撮影学習システムの中で，映像文法を背景とし，撮影中にオンラインで規範となる撮影スタイルに誘導する訓練指向型オンライン映像撮影ナビゲーションシステム [27] の実現を目指しており，本章では特に，その部分システムとして，オンライン処理向けカメラワーク解析法の精度を向上させる手法に焦点を当てる．カメラワークとは，カメラの首振りに対応する PAN，カメラのレンズ操作による焦点距離の変更に対応する ZOOM であり，本論文ではカメラの固定状態に対応する FIX も動きのないカメラワークとして，カメラワークの概念に含めるものとする．

このような撮影中のカメラワークを解析し，ユーザへ即時に問題点を指摘するシス

A: 大分類	FIX	PAN	ZOOM
B: 中分類	FIX	左方向PAN　右方向PAN 上方向PAN　下方向PAN	寄り(IN) 引き(OUT)
C: 小分類	カメラの 動き無し	360度のPAN方向 ＋速度量	寄り・引き ＋速度量

図 3.1　カメラワークの大・中・小分類

テムでは，実際の撮影動作と提示された映像，カメラワークの解析結果に時間的ずれを感じさせないよう，オンライン処理上での即時応答性が求められる．特に，オンライン映像撮影ナビゲーションシステムでは，システム全体の処理能力として実時間性が要求されるため，その部分システムとしてのカメラワーク解析部には，処理の高速性が求められ，同時に精度の高い手法が望まれる．

ここで，図 3.1 にカメラワークの大分類・中分類・小分類を示す．図 3.1 の A は，カメラワークを FIX，PAN，ZOOM という三つに分類する方法であり，PAN の方向や ZOOM の寄り (IN) もしくは引き (OUT) については問わない分類である．また，図 3.1 の B である中分類は，ZOOM について寄りか引きかを区別し，PAN の方向をいくつかの方向に区分する分類である．図 3.1 の中分類では，PAN を上下左右，四つの方向に区分している．また，図 3.1 の C は，PAN の方向を 360 度の角度として表現し，各方向での動作量（速度量）を付随させ，ZOOM についても，中分類の寄りか引きかを区別するだけでなく，動作量（速度量）を付随させる分類である．

本研究でのカメラワーク [27] は図 3.1 の A に示したような大分類ではなく，4 章で提案される訓練指向型オンライン単一ショット撮影ナビゲーションシステムでは，中分類と小分類を用いて処理が行われる．中分類は，小分類が得られれば判定処理を用いて分類が行えるため，本研究では，小分類を可能とする手法に着目する．訓練指向型オンライン単一ショット撮影ナビゲーションシステムでは，初心者の不適切なカメラワークの動きを検出するため速度の連続量が必要であり，低速から高速，小さな動きから大きな動きを扱うことができ，PAN，ZOOM ともに時間分解能の高い速度量に関する情報を得る必要がある．このため，訓練指向型オンライン単一ショット映像撮影ナビゲーションシステム [27] では，(1) オンライン処理向きの高速性，(2) 時間分解能の高

い動作の速度量，(3) 中分類のカメラワーク判定法，をすべて獲得可能な小分類の解析手法が求められる．

3.2　カメラワーク解析法

近年，主にオフライン環境を前提として，映像に付与するアノテーション情報の解析・付与法を競う TRECVID[28] では，2005 年までカメラワークの解析法が競技の項目に含まれており，数多くの研究 [29] が行われているが，カメラワークの解析法については，従来より，動きベクトルを基盤にする手法と，時空間画像を基盤とする手法に大別される．

3.2.1　動きベクトルを用いる手法

画像処理によって動きベクトルを算出する方法は，Horn[30] が画像処理にオプティカルフローを導入して以来，数多くの研究 [31] が報告されている．しかし，動きベクトルを推定する段階で高い計算コストが必要となるため [32]，現時点ではオフライン処理向きである．カメラのグローバルな動きを得るため，この動きベクトルの推定を高速化する試み [32] も行われているが，繰り返し推定による値の収束が必要であり，(1) の観点で高速化には限界がある点が問題となる．

一方，MPEG ファイルの取得を前提にすれば，内部に格納された動きベクトルを利用することで，高速なカメラワーク推定法の実現が期待される．このアプローチとしては，映像の検索や要約を可能にするためのメタ情報としてカメラワークを推定する手法 [33] をはじめ，数多くの方法が提案されている [34, 35, 36]．しかし，MPEG はオフライン環境での再生が本来の目的であり，1 フレームごとの動きベクトルを取得するには，GOP(Group of Picture) ごとに異なるフレーム間に分散した情報からの復号が必要となるため，GOP ごとの取得を待つ必要があり，MPEG を対象としたカメラワークの解析法は，オンライン処理向きとは言えない．このため，即時応答性が問われる理由により，動きベクトルの直接送受信機構がないかぎり，先読みができないことから高速化に限界があり，これも (1) の観点で不利となる．

これに対し，一般的な DV 端子を持つポータブルビデオカメラには，ビデオカメラ

側が送信できる最大限の可変 Frame rate で画像 (DV 形式で圧縮) を送るモードがある．このため，フレーム画像の段階から高速にカメラワークを推定する手法があれば，高 Frame rate でのカメラワーク解析法が実現可能となる．本研究では，高 Frame rate(非同期) の環境を利用し，高速性が期待できる時空間画像に基づく手法に着目する．

3.2.2　時空間画像を用いる手法

時空間画像とは，画像の xy 平面と時間 t の時空間で，例えば x もしくは y を固定して得られる時空間上の断面画像であるが，その断面をそのまま利用する時空間スライス [37] と，投影を行う時空間投影画像 [38] がある．

時空間投影画像に基づいた方法では，ハフ変換を用いてカメラワークの判定を行う手法がある [38]．この手法は (2) の時間分解能の高い動作量が得られるものの，(3) の観点では図 3.1 の大分類が行われているだけであり，ハフ変換は高速化の観点で限界があるため，(1) の観点でも問題が残される．

この時空間投影画像を用いた方法で (1) と (2) の観点を満たす方法に，輝度投影相関法 [39] がある．ただし，輝度投影相関法では，PAN と ZOOM を独立に計算するため，双方の過剰検出の扱い方と改善法が課題であり，(3) のカメラワーク中分類の判定法が別途必要となる．特に，輝度投影相関法では，ZOOM 検出器が低速と高速の ZOOM で検出精度が低く，フレーム内の動きに過敏に反応するため，ZOOM の過剰検出の多さが問題となっていた．天野らは，この (1) と (2) の観点を満たす輝度投影相関法を基盤として (3) のカメラワーク中分類判定法を導入したオフライン映像撮影ナビゲーションシステム [40] の一例を構築した．天野らの提案したカメラワーク中分類判定法では，FIX，PAN の上下左右，ZOOM イン・アウトを一つずつ順番に判定を行うことで，ZOOM の過剰検出が抑えられ精度が向上した．しかし，この判定法は，先に行われた過剰検出による誤判定が後にまわる正解判定を覆い隠す問題がある．PAN の検出では判定順序に依存しない方法，ZOOM は過剰検出を抑え，検出性能の高い手法が必要となっていた．また，(1) の観点を満たし，時空間スライスを用いるカメラワーク解析法としては，時空間画像の局所領域に対応する構造テンソルの主方向ヒストグラムを用いた手法が知られている [41]．これは，カメラワークの大分類の手法として提案されているが，(2) の観点での時間分解能は比較的高く，動作量を推定する可能性を有している．

3.2.3　提案手法

本研究では，観点 (1) オンライン処理向きの高速性と (2) 時間分解能の高い動作量が期待できる輝度投影相関法を PAN の解析手法として用いる．輝度投影相関法は，PAN と ZOOM を独立して解析することから，ZOOM の解析で精度が低い輝度投影相関法の代替として別手法を用いることも可能である．この代替手法として，観点 (1) と観点 (2) を満たす構造テンソル・ヒストグラム法に着目する．構造テンソルは，処理の対象となる時空間画像に時空間投影画像を用いることも可能であり，輝度投影相関法と親和性が高く，時空間投影画像を共有できるため，高速性が維持できる．構造テンソルを用いた従来の構造テンソル・ヒストグラム法では，大分類の手法として提案されていたため，ZOOM IN・OUT の判別法や ZOOM の動作量（速度量）推定法が不足していた．そこで本研究では，構造テンソル・ヒストグラムを時空間画像の中央で左右に二分化することにより，ZOOM IN・OUT の判別が可能となり，ZOOM 動作量も推定できる手法として，二分化テンソルヒストグラム法を提案し，輝度投影相関法で問題となっていた ZOOM 検出法の代替として用いる．

次に，(3) のカメラワーク中分類を判定する手法として，輝度投影相関法から得られる x，y 方向の動作推定量から直接分類を行い判定順序の影響がある天野らの従来法ではなく，PAN の動作量をまずベクトル化し，PAN ベクトルの方向 (角度) で PAN 動作の上下左右を判定するカメラワーク中分類判定手法を提案する．PAN の動作方向判定では，この手法により，方向が平等に扱われるため，判定順序に依存して精度が低下していた問題の改善を行う．これら手法により，ZOOM 検出器の過剰検出が減少し，早い ZOOM，また非常にゆっくりとした ZOOM の検出精度が高まったため，PAN と ZOOM の判定順序の検討も行う．

このように，本研究では，天野らの従来法に対して，輝度投影相関法から得られる PAN の動作量をベクトル化し，ZOOM の解析には二分化テンソルヒストグラム法を用いて ZOOM の小分類を行い，このオンライン処理向きの高速性が期待できる二手法から得られる小分類のカメラワーク解析情報を利用して，観点 (3) のカメラワーク中分類について判定順序の影響が少ない手法を提案する．

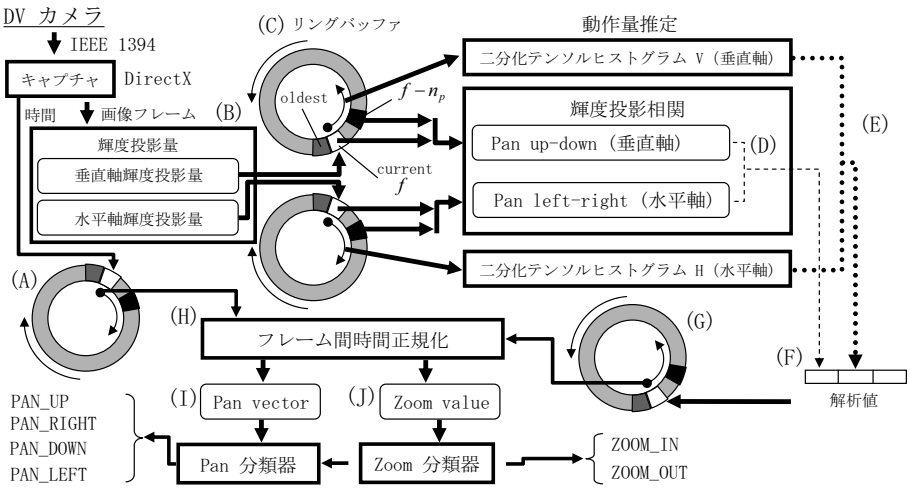


図 3.2　処理過程

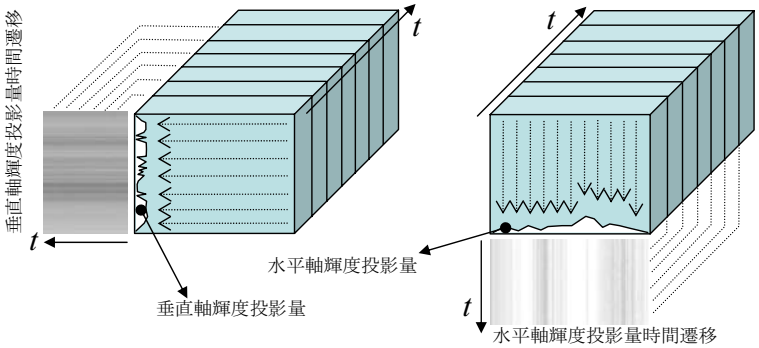


図 3.3　垂直軸輝度投影量と水平軸輝度投影量

3.3　オンライン処理用のカメラワーク解析法

3.3.1　輝度投影量と時間分解能のゆらぎ

図 3.2 は，提案手法の処理過程を示している．まず，DV カメラから非同期で画像が１フレームごとにキャプチャされ濃淡画像に変換される．この濃淡画像１フレーム分の画素について，水平方向に平均化した輝度値を垂直方向の輝度平均値列として見たものが，図 3.3 の垂直（縦軸）投影量 (Vertical projection) であり，これを時間方向に並べたものが垂直方向時空間投影画像となる．また，画素の輝度値を垂直方向に平均化した値を水平方向の輝度平均値列として見たものが水平投影量 (Horizontal projection) であり，これを時間方向に並べたものが水平方向の時空間投影画像となる．１フレー

ムごとの画像を取得した時間が図3.2(A)のリングバッファに蓄積される．フレーム画像は輝度投影量を算出する図3.2(B)に送られ，式(3.1)で求められるフレーム画像の横軸の輝度投影量 $P_H(f,i)$ と式(3.2)で求められる縦軸の輝度投影量 $P_V(f,j)$ がそれぞれ独立した図3.2(C)の上下にあるリングバッファに蓄積される．ただし，式(3.1),(3.2)の $Gray(f,i,j)$ は，縦 h，横 w のフレーム f に対応する画像の位置 i,j での輝度値(濃淡値)である．

$$P_H(f,i) = \frac{1}{h}\sum_{j=1}^{h} Gray(f,i,j) \tag{3.1}$$

$$P_V(f,j) = \frac{1}{w}\sum_{i=1}^{w} Gray(f,i,j) \tag{3.2}$$

この輝度投影量が蓄積されたリングバッファを展開すると，それが時空間投影画像となる．ただし，フレームの取得方式は非同期であることから，各輝度投影量に対応する時間の間隔は変動する．**図 3.4** は，PentiumM1.4GHz を搭載したノートパソコンを用い，図3.2(A)以外の処理を無効として，キャプチャのみを行った場合，各フレーム画像の取得時間の差をヒストグラムにした実験例である．この例では，最小で8ms，最大で28msであり，平均では10.3msとなった．このように，Frame rateにはゆらぎがあることがわかる．動きが自然に見える映像では30fps程度のFrame rateであり，フレーム間の時間間隔は約33msとなるが，10.3msは，97fps程度の高いFrame rateに相当する．しかし，これが逆に10fps程度に落ちても，相対的にカメラワークの動きが速すぎなければ体感的に遅れを感じない場合もある．これは，時間分解能の低下した実時間処理であると考えられるが，本研究では，時間分解能の低下をいかに抑えるかが課題となる．一方，このFrame rateの変動によるカメラワーク動作量の過大もしくは過小な推定は，不適切なカメラワークとしての誤判定につながる．そこで，以降の節において，カメラワークの解析手法を述べ，推定動作量を時間正規化により補正する手法を示す．

3.3.2　判定順序の影響を受けないPAN動作推定量

左右，上下方向のPAN動作量は，現在(current)の投影量と過去の投影量を用い，式(3.3)，(3.4)を距離関数として式(3.5)と式(3.6)によって求められる．ここで，図3.2(D)

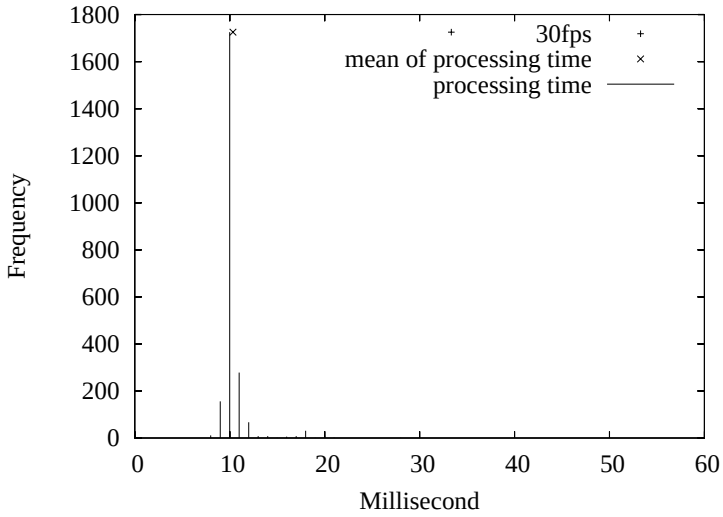


図 3.4　処理時間のヒストグラム (Capture 処理のみ)

に対応する縦と横方向の動作推定量は，従来法と同じ $\hat{\delta}_{P_{lr}}$ が水平方向，$\hat{\delta}_{P_{ud}}$ が垂直方向の PAN 動作推定量となる．なお，$n_p = 3$，$-\delta^{max}_{P_{lr}} < \delta_{P_{lr}} < \delta^{max}_{P_{lr}}$，$-\delta^{max}_{P_{ud}} < \delta_{P_{ud}} < \delta^{max}_{P_{ud}}$，また $\delta^{max}_{P_{lr}} = w/9, \delta^{max}_{P_{ud}} = h/9$ を用いた．

$$
\begin{aligned}
&D_{P_H}(f, i, \delta_{P_{lr}}) \\
&\quad = \quad |P_H(f, i) - P_H(f - n_p, i - \delta_{P_{lr}})| \qquad (3.3)\\
&D_{P_V}(f, j, \delta_{P_{ud}}) \\
&\quad = \quad |P_V(f, j) - P_V(f - n_p, j - \delta_{P_{ud}})| \qquad (3.4)\\
&\hat{\delta}_{P_{lr}}(f) \\
&\quad = \quad \arg\min_{\delta_{P_{lr}}} \sum_{\substack{i=1+\delta^{max}_{P_{lr}} \ (\delta_{P_{lr}} \geq 0) \\ i=1 \ (\delta_{P_{lr}} < 0)}}^{\substack{w \ (\delta_{P_{lr}} \geq 0) \\ w-\delta^{max}_{P_{lr}} \ (\delta_{P_{lr}} < 0)}} D_{P_H}(f, i, \delta_{P_{lr}}) \qquad (3.5)\\
&\hat{\delta}_{P_{ud}}(f) \\
&\quad = \quad \arg\min_{\delta_{P_{ud}}} \sum_{\substack{j=1+\delta^{max}_{P_{ud}} \ (\delta_{P_{ud}} \geq 0) \\ j=1 \ (\delta_{P_{ud}} < 0)}}^{\substack{h \ (\delta_{P_{ud}} \geq 0) \\ h-\delta^{max}_{P_{ud}} \ (\delta_{P_{ud}} < 0)}} D_{P_V}(f, j, \delta_{P_{ud}}) \qquad (3.6)
\end{aligned}
$$

図 3.2(D) の推定された縦軸・横軸のカメラワーク推定量は，解析値を格納する図 3.2(F) の配列に格納された後，図 3.2(G) のリングバッファに蓄積される．ただし，この推定

値 $\hat{\delta}_{P_{lr}}(f)$ や $\hat{\delta}_{P_{ud}}(f)$ は，時間変動を考慮していないため，図 3.2(A) のリングバッファに格納された時間情報に基づき，図 3.2(H) の過程で，$Pan_{lr}(f) = \hat{\delta}_{P_{lr}}(f)/(t_f - t_{f-n_p})$，$Pan_{ud}(f) = \hat{\delta}_{P_{ud}}(f)/(t_f - t_{f-n_p})$ のようにミリ秒単位での時間正規化を行い，式 (3.7) により速度関数 $P_S(f)$ として計算する．また，PAN の方向 $P_\theta(f)$ は PAN 動作推定量 $Pan_{lr}(f)$ を x，$Pan_{ud}(f)$ を y としたとき，式 (3.8) により計算され，PAN 動作量は最終的に式 (3.9) の極座標ベクトルとして表現される (図 3.2(I))．この Pan_f を用いることで，提案手法となる判定順序に依存しない 360 度の方向と速度量が得られる．

$$P_S(f) = \sqrt{x^2 + y^2} \tag{3.7}$$

$$P_\theta(f) = \begin{cases} deselection & (x = 0 \wedge y = 0) \\ arctan\frac{y}{x} & (x > 0 \wedge y \geq 0) \\ \frac{\pi}{2} & (x = 0 \wedge y > 0) \\ arctan\frac{y}{x} + \pi & (x < 0 \wedge y \geq 0) \\ arctan\frac{y}{x} + \pi & (x < 0 \wedge y < 0) \\ \frac{3\pi}{2} & (x = 0 \wedge y < 0) \\ arctan\frac{y}{x} + 2\pi & (x > 0 \wedge y < 0) \end{cases} \tag{3.8}$$

$$Pan_f = (P_S(f), P_\theta(f))^T \tag{3.9}$$

3.3.3　構造テンソル

次に，本研究で提案する二分化テンソルヒストグラムについて述べる前に，その基盤となる構造テンソルについて述べる．ここでは，構造テンソルによる窓領域の主方向推定法について説明する．まず，時空間投影画像を I として，画像 I 上のある点における空間軸 s 方向の偏微分 I_s と時間軸 f 方向の偏微分 I_f を要素に持つベクトル $z = (I_s, I_f)^T$ を考え，ある窓領域 w 中の z を対象とした分散・共分散行列として得られる式 (3.10) の Γ を構造テンソルとする．本章では窓サイズ w は 5x5 とした．

$$\begin{aligned}\Gamma &= \sum_w zz^T = \begin{bmatrix} \sum_w I_s^2 & \sum_w I_s I_f \\ \sum_w I_s I_f & \sum_w I_f^2 \end{bmatrix} \\ &= \begin{bmatrix} J_{ss} & J_{sf} \\ J_{fs} & J_{ff} \end{bmatrix} \end{aligned} \tag{3.10}$$

そして，この Γ は，対称行列であるため，式 (3.11) のような主軸を (l_1, l_2) とする回転行列 R によって式 (3.12) の Λ ように対角化が可能である．

$$R = [l_1, l_2] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.11}$$

$$\Lambda = \begin{bmatrix} \lambda_s & 0 \\ 0 & \lambda_f \end{bmatrix} = R^T \ \Gamma \ R \tag{3.12}$$

θ は Γ の窓 w における主方向を表しており，式 (3.12) から導出される式 (3.13) により，主方向の回転角が計算できる．また，θ では，水平方向が 0 度になるため，垂直方向を 0 度となるよう，式 (3.14) により主方向を ϕ(図 3.2(E)) に変換し，図 3.2(F) を通じてリングバッファ (G) に蓄積される．

$$\theta = \frac{1}{2} arctan \frac{2J_{sf}}{J_{ss} - J_{ff}} \tag{3.13}$$

$$\phi = \begin{cases} \theta - \frac{\pi}{2} & \theta > 0 \\ \theta + \frac{\pi}{2} & otherwise \end{cases} \qquad \phi = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{3.14}$$

時空間投影画像上の部分領域 w(図 3.6) を時空間投影画像の s 軸方向へ 1 画素ずつシフトして得られる主方向 ϕ の頻度分布が式 (3.15) のテンソルヒストグラムである．

$$M(\phi, f) = \sum_{\Omega(\phi, f)} c(\Omega) \tag{3.15}$$

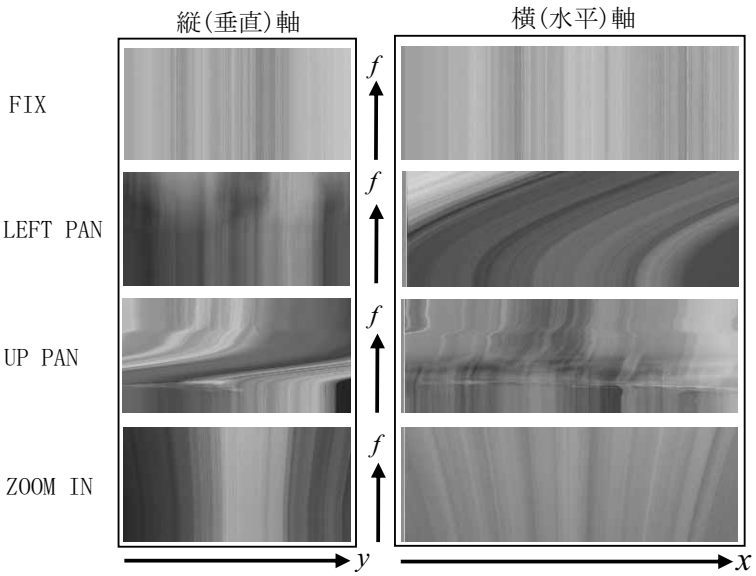


図 3.5　時空間投影画像とカメラワークの関係 (f:フレーム番号)

式 (3.15) に用いられる c は，式 (3.16) により導かれる重みであるが，$\lambda_s = 0$ または $\lambda_f = 0$ のように空間軸，時間軸に対して主方向が一方向に定まっているときは $c = 1$ となり，$\lambda_s = \lambda_f$ のように，方向が空間軸にも時間軸にも同程度存在するときは $c = 0$ となる確信度として利用でき，$c = [0, 1]$ となる．ただし，式 (3.15) では，実質的に c を頻度値として計算することになり，各 ϕ に対応する c の総和がフレーム f でのテンソルヒストグラム上の頻度となる．

$$c = \frac{(J_{ss} - J_{ff})^2 + 4J_{sf}^2}{(J_{ss} + J_{ff})^2} = \left(\frac{\lambda_s - \lambda_f}{\lambda_s + \lambda_f}\right)^2 \tag{3.16}$$

3.3.4　テンソルヒストグラムとカメラワーク

図 3.5 は，図 3.3 に示した縦軸 (Vertical line) と横軸 (Horizontal line) の投影量を時間方向であるフレーム番号の方向へ並べて作成した時空間投影画像 (Projection-temporal image) とカメラワークとの対応関係を示している．図 3.5 において，FIX の場合，この時空間投影画像は，縦軸・横軸ともに垂直方向の縞模様となる．また，横方向の PAN

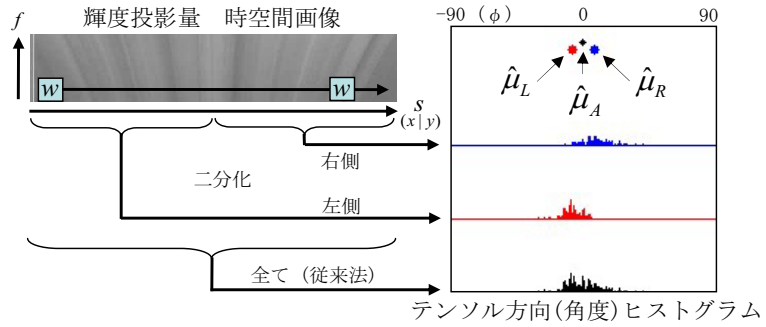


図 3.6　二分化テンソル・ヒストグラム

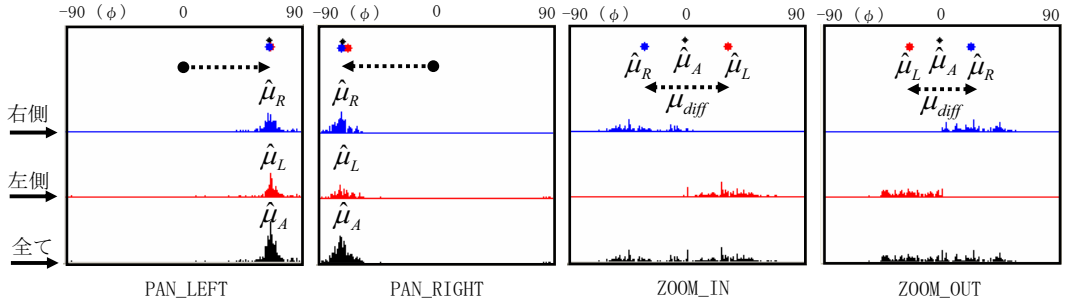


図 3.7　従来のテンソル・ヒストグラムと二分化テンソル・ヒストグラム

では，基本的に横軸の時空間投影画像のみ，その縞模様が傾斜するような画像となるため，縞模様の傾斜角度がほぼ同じになる．同様に，縦方向の PAN では，縦軸の時空間投影画像のみ，その縞模様が傾斜し，縞模様の傾斜角度がほぼ同じになる．画面の中心を軸とする ZOOM では，縦軸，横軸いずれの時空間投影画像も角度分布が広がりを見せる．この特徴を生かしてカメラワークを分類する手法はいくつかあるが，計算コストが少なく，精度の良い手法として，構造テンソルヒストグラム [41] に着目する．構造テンソルとは，図 3.6 の時空間投影画像中に示した微小領域 w で縞模様の主方向 ϕ を算出する手法であり，w の位置をシフトして得られた ϕ の頻度分布がテンソルヒストグラム (式 3.15) である．

図 3.7,「全て」の矢印が示すテンソルヒストグラムが従来手法である．これは，図 3.6 の場合，時空間投影画像の f を固定し，$s(x$ または $y)$ 軸に沿って得られた領域 w の ϕ_Aを 全て用いて得られたテンソルヒストグラム $M(\phi_A, f)$ である．この ϕ_A の平均値を $\hat{\mu}_A(\phi_A) = mean(\phi_A)$ とするとき，FIX，PAN の場合は ϕ_A がほぼ一致するため，ヒストグラムの尖度の高さを前提に，$\hat{\mu}_A$ が 0 のときは FIX，正負の値を持つときは PAN と

判定できる．これに対し，ZOOM では，角度の分布が広がり，分布の尖度が小さくなるため，従来のカメラワーク判定法では，これらの特徴から FIX，PAN，ZOOM の大分類が行われていた [41]．ただし，この $\hat{\mu}_A$ を用いれば，PAN の動作量は，角度の大きさと速度量が比例することから，$\hat{\mu}_A$ の値そのものを用いることができ，横軸の $\hat{\mu}_A$ を $\hat{\mu}_{A_H}$，縦軸の $\hat{\mu}_A$ を $\hat{\mu}_{A_V}$ として，$Pan_{lr}(f) = \hat{\mu}_{A_H}$,$Pan_{ud}(f) = \hat{\mu}_{A_V}$ とすることもできる．しかし，図 3.7 の「全て」で ZOOM IN・OUT の分布が示すように，いずれも同様の分布となり，このままでは ZOOM の動作量，ZOOM IN・OUT の判別が定義できなかった．

3.3.5　二分化テンソルヒストグラムによる ZOOM の解析法

そこで，本研究では，図 3.6 のように，時空間投影画像を左右の領域に二分化してテンソルヒストグラムを求めることで，ZOOM 動作量の定義，ZOOM IN・OUT の判別を可能にする手法を提案する．左右の領域の各分布の平均値を $\hat{\mu}_L$，$\hat{\mu}_R$ とすると，図 3.7 のように，PAN では $\hat{\mu}_A$ と一致し，ZOOM IN・OUT では，$\hat{\mu}_L$ と $\hat{\mu}_R$ の符号が入れ替わることから，判別が可能となる．これは，フレームの中心を軸とする ZOOM では，ZOOM IN 時に縞模様が左右で外向きとなり，ZOOM OUT 時に内向きになるためである．

そこで，縦 (V) 横 (H) の時空間投影画像で，二分化した各領域 w の主方向をそれぞれ ϕ_L^V,ϕ_R^V,ϕ_L^H,ϕ_R^H とするとき，各平均値を式 (3.17) から (3.20) に示す．ただし，ロバスト性を考慮して，各 ϕ は，一度平均値 μ を計算し，標準偏差 σ の 3σ 内に入る ϕ により，それぞれの平均値 $\hat{\mu}$ を求めた．

$$\hat{\mu}_L^V(\phi_L^V) = mean(\phi_L^V) \quad (\mu_L^V - 3\sigma_L^V < \phi_L^V < \mu_L^V + 3\sigma_L^V) \tag{3.17}$$

$$\hat{\mu}_R^V(\phi_R^V) = mean(\phi_R^V) \quad (\mu_R^V - 3\sigma_R^V < \phi_R^V < \mu_R^V + 3\sigma_R^V) \tag{3.18}$$

$$\hat{\mu}_L^H(\phi_L^H) = mean(\phi_L^H) \quad (\mu_L^H - 3\sigma_L^H < \phi_L^H < \mu_L^H + 3\sigma_L^H) \tag{3.19}$$

$$\hat{\mu}_R^H(\phi_R^H) = mean(\phi_R^H) \quad (\mu_R^H - 3\sigma_R^H < \phi_R^H < \mu_R^H + 3\sigma_R^H) \tag{3.20}$$

次に，遅い ZOOM では，時空間投影画像上の縞模様の角度が垂直に近づき，速い ZOOM では，角度が大きくなることから，$\hat{\mu}_L$ と $\hat{\mu}_R$ の距離を $\hat{\mu}_{diff}$ とすれば，$\hat{\mu}_{diff}$ は，

遅い ZOOM では小さく，速い ZOOM では大きくなる．この性質を利用して，$\mu^V_{diff} = |\hat{\mu}^V_L(\phi^V_L) - \hat{\mu}^V_R(\phi^V_R)|$, $\mu^H_{diff} = |\hat{\mu}^H_L(\phi^H_L) - \hat{\mu}^H_R(\phi^H_R)|$ としたとき，縦と横の ZOOM 動作量 Z^V_f,Z^H_f は，ZOOM IN･OUT の判別法を導入し，$\hat{\mu}^V_L$ と $\hat{\mu}^V_R$ の距離として，式 (3.21),(3.22) のように定義される．

$$Z^V_f = \begin{cases} -\mu^V_{diff} & (\hat{\mu}^V_L{>}0 \wedge \hat{\mu}^V_R{<}0) \\ \mu^V_{diff} & (\hat{\mu}^V_L{<}0 \wedge \hat{\mu}^V_R{>}0) \\ 0 & otherwise \end{cases} \tag{3.21}$$

$$Z^H_f = \begin{cases} -\mu^H_{diff} & (\hat{\mu}^H_L{>}0 \wedge \hat{\mu}^H_R{<}0) \\ \mu^H_{diff} & (\hat{\mu}^H_L{<}0 \wedge \hat{\mu}^H_R{>}0) \\ 0 & otherwise \end{cases} \tag{3.22}$$

また，式 (3.21),(3.22) によって得られた縦横の ZOOM 動作推定量は，フレームの中心を軸とするズームの場合，基本的に同程度の値となるはずである．そこで，縦横の ZOOM 動作推定量の片方が 0 である場合，また符合が一致しない場合は矛盾するものとして，検出された ZOOM 動作量を無効とする．したがって，最終的な ZOOM 動作推定量 $Zio(f)$(図 3.2(J)) は，式 (3.21),(3.22) と時間正規化の適用を含め，式 (3.23) によって速度量として計算される．これにより，$Zio(f)$ は，ZOOM IN・OUT の判別と，連続的に変化する速度量が得られるため，速度量の変化を検証することも可能となる．

$$Zio(f) = \begin{cases} 0 & (Z^H_f{=}0 \vee Z^V_f{=}0) \\ 0 & (Z^H_f{>}0 \wedge Z^V_f{<}0) \\ 0 & (Z^H_f{<}0 \wedge Z^V_f{>}0) \\ \frac{Z^H_f + Z^V_f}{2(t_f - t_{f-n_z})} & otherwise \end{cases} \tag{3.23}$$

3.4　カメラワーク中分類の判定法

3.4.1　従来法のカメラワーク判定順序に依存する問題

天野らはオフライン方式の映像撮影ナビゲーションで，輝度投影相関法のみを基盤とし，以下に示す判定条件を上から順番に適用し，条件が一致した際，左側に対応す

るカメラワークを判定結果とする中分類法 [40] を採用している．

FIX	$\hat{\delta}_{P_{ud}}(f) = 0$ && $\hat{\delta}_{P_{lr}}(f) = 0$ && $\hat{\delta}_{Z_h}(f) = 0$
PAN_LEFT	$0 < \hat{\delta}_{P_{lr}}(f) < \theta_{lr}$
PAN_RIGHT	$-\theta_{lr} < \hat{\delta}_{P_{lr}}(f) < 0$
PAN_UP	$0 < \hat{\delta}_{P_{ud}}(f) < \theta_{ud}$
PAN_DOWN	$-\theta_{ud} < \hat{\delta}_{P_{ud}}(f) < 0$
ZOOM_IN	$-\theta_z < \hat{\delta}_{Z_h}(f) < 0$
ZOOM_OUT	$0 < \hat{\delta}_{Z_h}(f) < \theta_z$
OTHER	$otherwise$

輝度投影相関法では，PAN と ZOOM の動作量が独立に計算され，相互に過剰検出が発生するが，そのような状況下でこの判定方式を採用した場合，先に判定が行われた条件式の影響力が強くなる．特に，輝度投影相関法では，実際には PAN の動作に対して，ZOOM 検出器が敏感に反応する傾向があり，こうした場合に ZOOM の判定順序を後方にまわすことで，判定上，過剰検出を抑えることが可能となる．しかし，実際には ZOOM の動作に対して PAN 検出器が反応した場合，この判定順序では，逆に過剰検出した PAN の動きが判定上採用され，ZOOM 検出器が正しい動きを検出しても，判定順序による弊害で埋もれてしまうため，カメラワーク解析手法が持つ検出能力を最大限に引き出せない問題がある．これは，式 (3.5)$\hat{\delta}_{P_{lr}}$ と式 (3.6)$\hat{\delta}_{P_{ud}}$ の間でも生じることから，カメラワーク解析法の検出能力を最大限に引き出すためには，相互の検出器の過剰検出を抑えつつ，判定順序に影響されにくい判定法が必要となる．

3.4.2　本研究でのカメラワークの判定法

本研究では，PAN の中分類判定で，従来法のように，$\hat{\delta}_{P_{lr}}$ と $\hat{\delta}_{P_{ud}}$ を直接用いて判定を行わず，式 (3.9) のように一度ベクトル化し，角度によって上下左右の方向を判定する方式を提案する．この手法によって，PAN の方向は，式 (3.8) により一意に定まるため，従来手法で判定順序により過小評価される事例がある場合は精度の改善が期待できる．式 (3.24) に PAN の角度による判定式を示す．式 (3.24) では，PAN ベクトルの角度を用いて四つの方向へ分類するが，特に上下の PAN として判定される範囲を狭めている．これは，上下の PAN を意識してカメラを動作させる場合は比較的垂直にカメラを動かす傾向があるのに対し，左右の動作については大きく斜め方向に動作させて

いても左右に動作させている感覚があるため，左右の判定の角度範囲を広げたためであり，本研究では $\alpha = \frac{\pi}{18}$ を用いている．

$$J_P = \begin{cases} PAN\_UP & (\frac{\pi}{4}+\alpha<P_\theta(f)<\frac{3\pi}{4}-\alpha) \\ PAN\_LEFT & (\frac{3\pi}{4}-\alpha\leq P_\theta(f)\leq\frac{5\pi}{4}+\alpha) \\ PAN\_DOWN & (\frac{5\pi}{4}+\alpha<P_\theta(f)<\frac{7\pi}{4}-\alpha) \\ PAN\_RIGHT & otherwise \end{cases} \tag{3.24}$$

次に，ZOOM の判定式を式 (3.25) に示す．本研究では，従来採用していた輝度投影相関法による ZOOM 動作量推定法の代替として，式 (3.21)，式 (3.22) による縦・横軸の動作量に制約を与える手法で過剰検出の抑制にも配慮した，二分化テンソルヒストグラムによる ZOOM 動作量推定法を提案した．その式 (3.23) の $Zio(f)$ は，符号により一意に ZOOM IN・OUT が判定されるため，式 (3.8) と同様，判定順序によらない．

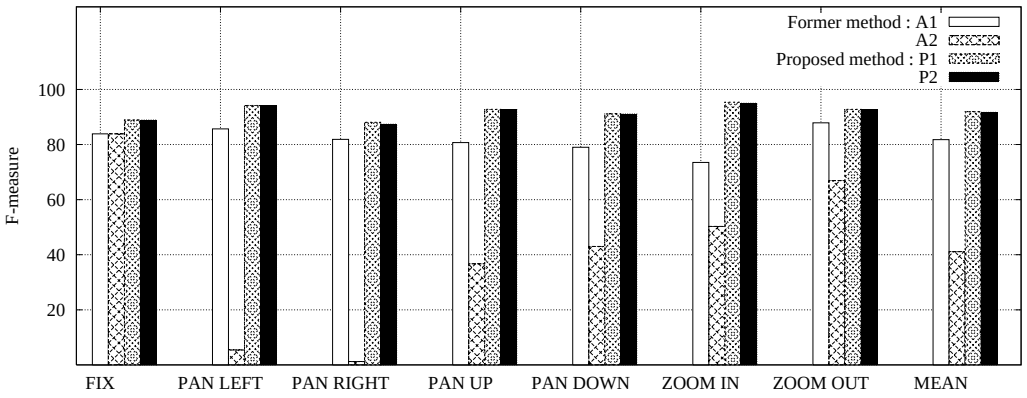
$$J_Z = \begin{cases} ZOOM\_IN & (Zio(f)>0) \\ ZOOM\_OUT & (Zio(f)<0) \end{cases} \tag{3.25}$$

また，式 (3.26) は，最終的なカメラワーク中分類の判定法であり，3.5 の実験結果から FIX，PAN，ZOOM の順序を採用した．

$$J_{cw} = \begin{cases} FIX & (P_S(f)=0 \ \wedge \ Zio(f)=0) \\ J_Z & (|Zio(f)|>0) \\ J_P & (|P_S(f)|>0) \\ OTHER & otherwise \end{cases} \tag{3.26}$$

3.5　評価実験

我々は，システムの野外等への携帯性を考慮し，Pentium M (1.4GHz) のノートパソコン上にオンライン映像撮影ナビゲーションシステムを実装した．評価用映像としては，毎日放送のプロのカメラマンが撮影した番組用の素材映像 8 本 (各 30 分) から，



	A1	A2	P1	P2
カメラワーク解析法 PAN	LPCM	LPCM	VLPCM	VLPCM
カメラワーク解析法 ZOOM	LPCM	LPCM	STHM	STHM
分類判定順序	FIX PAN_{LEFT} PAN_{RIGHT} PAN_{UP} PAN_{DOWN} $ZOOM_{IN}$ $ZOOM_{OUT}$	FIX $ZOOM_{IN}$ $ZOOM_{OUT}$ PAN_{LEFT} PAN_{RIGHT} PAN_{UP} PAN_{DOWN}	FIX J_z J_P	FIX J_P J_z

LPCM: Luminance projection correlation method
VLPCM: Vectorized luminance projection correlation method
STHM: Split tensor histogram method

図 3.8　従来法・提案法の F-measure の比較と判定順序の影響

放送用に使用可能で被写体にできるかぎり動きのない FIX:25 ショット，PAN_LEFT:14 ショット，PAN_RIGHT:19 ショット，PAN_UP:14 ショット，PAN_DOWN:17 ショット，ZOOM_IN:12 ショット，ZOOM_OUT:23 ショット，総数 27,035 フレームを抽出し，ショットごとに AVI ファイル化して評価実験を行った．

3.5.1　実験条件

図 3.8 の下側は四種類 (A1,A2,P1,P2) の実験の条件である．まず，図 3.8 の下側上部にカメラワーク解析法を PAN と ZOOM に分けて示す．比較実験では，PAN の解析法として，従来法の LPCM(輝度投影相関法) と，提案法となる VLPCM(輝度投影相関の

ベクトル化手法) を比較し，ZOOM の解析法では，従来の LPCM と，提案手法となる STHM(二分化テンソルヒストグラム法) を比較する．また，図 3.8 下側中央部に判定順序 (Judgment order) の条件を示す．FIX から判定を始め，下方向に次の判定対象が示されている．従来法は 4.1 に示した A1 の順序であり，A2 は，PAN と ZOOM の判定順序を入れ替えたものである．一方，提案法 P1 は，判定式 (3.26) の J_{cw} に従い，FIX の次に式 (3.25) の J_Z，その次に式 (3.24) の J_P を判定するものであり，P2 は J_Z と J_P の判定順序を入れ替えたものである．

3.5.2　従来法と提案法の判定順序による影響

図 3.8 の上図は，カメラワーク解析法と判定順序を変更した四種類の実験結果から得られた F-measure による判定精度 (縦軸) を，七つのカメラワーク中分類ごとに横に並べ，またその平均精度 (MEAN) を右端に示したものである．

Fmeasure は，本研究では再現率を R，適合率を P として，Fmeasure=2·R·P/(R+P) と定義され，この R と P の指標を統合する評価法であり，二つの指標が 100%に達したとき，100%となる．本研究では，未検出がなく，過剰検出のないシステムが理想的であるため，Fmeasure の精度が 100%に近いことが望ましい．また，再現率と適合率については，C を検出対象の正解フレーム数，N を検出できなかった正解フレーム数として未検出数，E を過剰検出したフレーム数として過剰検出数とするとき，再現率 (Recall)=C/(C+N)，適合率 (Precision)=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．

図 3.8 から，従来法 A1 に対し，判定順序のみを変更した A2 は，精度が大きく悪化することがわかり，従来法は判定順序に影響されやすい手法であることがわかる．これは，A2 において，LPCM による ZOOM 判定部が実際には PAN の動きを ZOOM と判定する，過剰検出の多さが原因である．一方，提案法 P1 と，判定順序のみを入れ替えた P2 では，ほとんど精度が変化しないため，PAN の解析に VLPCM，ZOOM の解析に STHM を用いる提案法では，過剰検出が少なく PAN 解析と ZOOM 解析の独立性が高まったことを示している．ただし，P1 の平均精度 91.9%に対し，P2 は 91.7%であり，精度が 0.2%落ちる．これは，P2 の VLPCM による PAN 判定部において，実際に

表 3.1　従来法と提案法のカメラワーク判定精度一覧

	FIX	PAN_l	PAN_r	PAN_u	PAN_d	$ZOOM_i$	$ZOOM_o$	MEAN
A1	83.9	85.7	81.9	80.7	79.0	73.5	87.9	81.8
P1	88.9	94.2	88.1	92.8	91.1	95.4	92.8	91.9

はZOOMの動きについてPANと判定する過剰検出がわずかではあるが存在することを示す．本研究ではP1の判定順序を提案法とする．

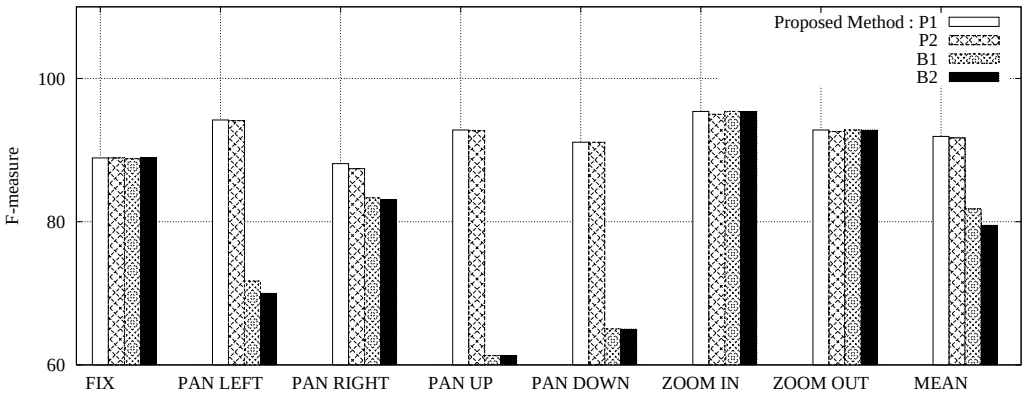
3.5.3　従来法と提案法のPAN・ZOOM解析法の比較

表3.1にA1とP1の判定精度を列挙する．表3.1より，従来法A1と提案法P1のカメラワーク判定精度において，P1は，七つのいずれのカメラワーク中分類についてもA1より精度を改善させている．PANのみの精度を平均した場合，P1はA1に対し9.8%精度改善を達成し，ZOOMのみの精度を平均した場合，P1はA1に対し，13.4%の精度改善を達成している．精度の全体平均による改善度としては，A1の81.8%に対し，P1は91.9%となるため，平均精度で10.1%の改善を達成している．以上より，PAN解析法ではベクトル化による効果，ZOOM解析ではSTHMを用いる提案法の効果が示された．

3.5.4　PAN解析にテンソルヒストグラムを用いない理由

本研究では，PAN解析にVLPCM，ZOOM解析にSTHMという二手法の併用法を提案しているが，3.3.4節で述べたように，PAN解析法にTHM(テンソルヒストグラム法)を用いる方法も考えられる．なお，PAN解析にSTHMを用いないのは，二分化する必要がないからである．

ここで，もしPAN解析にTHMを用いれば，THMという同一基盤の手法でシステムを構築でき，システムを簡潔化できるが，本研究ではVLPCMを用いている．これはTHMを用いたPAN解析の精度が良くないためである．それを確認するため，B1とB2の評価実験を導入する．ただし，P1のPAN解析精度はLPCMのベクトル化によって精度が向上しているため，THMについても，ベクトル化したVTHMを用いて実験



	P1	P2	B1	B2
カメラワーク解析法 PAN	VLPCM	VLPCM	VTHM	VTHM
ZOOM	STHM	STHM	STHM	STHM
分類判定順序	FIX J_z J_P	FIX J_P J_z	FIX J_z J_P	FIX J_P J_z

VTHM: Vectorized tensor histogram method

図 3.9　テンソル・ヒストグラムの PAN 検出器への適用に関する考察

を行う．B1 は，P1 の PAN 解析部を VTHM に置き換えた手法である．また，B2 は，B1 の判定順序を変更したもので，P2 の PAN 解析部を VTHM に変更したものに対応する．その実験条件を図 3.9 下側に示す.

図 3.9 上側より，P1 と比較して，B1 と B2 ともに，ZOOM の判定精度はほぼ同じであるが，PAN 解析の精度が大きく悪化している．一方，従来法の図 3.8A1 は，平均精度で 81.8%であるが，B1 は 81.8%，B2 は 79.5%であったため，従来法と同程度かそれ以下に精度が悪化する．ただし，B1 と B2 では，精度が判定順序にほとんど影響しないことから，提案法と同様に，PAN 解析と ZOOM 解析の独立性が高められる点では A1 より優位である．以上の結果から，従来手法に対して，PAN の解析に VLPCM，ZOOM の解析に STHM を併用する本研究の提案法 P1 の優位性が示された.

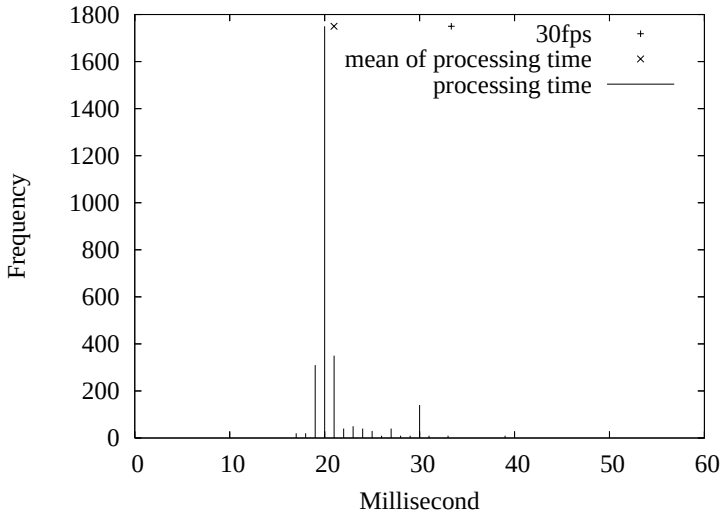


図 3.10　処理時間のヒストグラム

3.5.5　処理速度

図 3.10 は，図 3.4 で示した capture のみの処理時間と，フレームごとのカメラワーク解析処理時間を含めて算出した，フレームごとの処理時間をヒストグラム化した結果である．映像は，フレーム内に動きのない室内について FIX，PAN 上下左右，ZOOMIN・OUT の動きで撮影した 2850 フレームについて評価した．capture 処理とカメラワーク解析処理を合わせた各フレームの処理時間の平均値は 21.0ms であり，最頻値は 20ms であった．最小値は 17ms であるが最大値は 39ms となり，33.3ms を越える結果となった．ただし，21.0ms 以内に 86%のフレーム間処理が含まれる結果であり，実時間処理として一般的な 33.3ms(30fps) を大きく上回る高速性を実現している．

また，capture のみの処理時間は，平均値として 10.3ms であったが，カメラワーク解析処理のみの処理時間平均値も 10.7ms(94fps) となることから，部分システムとしてのカメラワーク解析処理は，高速性を実現していることになる．本章でのカメラワーク解析法を基盤として実現化される訓練指向型オンライン映像撮影ナビゲーションシステム [27] では，すべての処理を含めたフレームごとの処理時間平均が 29.8ms(34fps)，98.0%の処理時間が 30fps 内に含まれる結果を得ており，ほぼ 30fps の時間分解能を達成している．

3.6　結言

映像撮影支援技術に関し，初心者が抱える撮影上の主な問題は，映像文法による観点から単一ショットの撮影において，不適切なカメラワークの使用が，のちの編集に問題となり，映像の品質を落とす主要な原因となっている．本研究では，DV カメラを用いた撮影中に撮影を誘導する，オンライン映像撮影ナビゲーションシステムのカメラワーク解析部として，PAN 解析に輝度投影相関法に基づくベクトル化された動作量と，ZOOM 解析に二分化テンソルヒストグラムによる手法を用いることで，3.1 の観点 (1) オンライン処理向きの高速性，(2) 時間分解能の高い動作の速度量を得ることのできるカメラワーク小分類法を提案し，(3) 中分類用のカメラワーク判定法について精度を改善する手法を提案した．時間分解能の観点では，86%の処理が平均で 21.0ms で行われる結果に達しており，映像で一般的な 30fps の 33.3ms を大きく上回る結果が得られた．

従来手法によるカメラワークの判定精度に対しては，FIX で 5%，PAN のみの平均精度で 9.8%，ZOOM のみの平均精度で 13.4%，全体の平均精度では，10.1%の改善を行うことができた．また，提案手法では，PAN と ZOOM の解析で過剰検出を抑えることができたため，解析器の独立性も向上し，判定順序の影響が少ない手法となった．

4 章で提案する訓練指向型オンライン映像撮影ナビゲーションシステムでは，解析された PAN の速度量と方向，また ZOOM の速度量とズームイン・アウトを反映する矢印を画面上に実時間で同時に表示する．PAN と ZOOM の過剰検出が多ければ，この矢印も過剰に表示されるため，訓練中の支障となる．しかし，提案手法により，この問題についても改善されたことになる．訓練指向型オンライン映像撮影ナビゲーションシステムの実時間オンライン処理は，本研究において精度が改善され，時間分解能が高く，高速なカメラワーク解析法に強く依存して成立するため，本研究は，訓練指向型オンライン映像撮影ナビゲーションシステムの実現に大きく貢献する手法となっている．

第4章

訓練指向オンライン単一ショット映像撮影支援方式

4.1 緒言

3章では，映像撮影学習システムの部分システムとなる高時間分解能，高速カメラワーク解析法について提案を行った．一般的に，初心者の撮影では，不適切なカメラワークが問題となりやすく，撮影者の乱れたカメラワークを解析し，問題点を減らすよう指導する手法が必要となる．本研究では，映像撮影学習システムの中で，映像文法を背景とし，撮影中にオンラインで規範となる撮影スタイルを学習できる技量を身につけられるよう誘導する，訓練指向型オンライン単一ショット映像撮影ナビゲーションシステムを提案する．

Barryは，一般家庭のビデオ撮影を支援する上で，社会的な常識を利用し，例えば，マラソンなどの撮影を例として，想定できる撮影上のポイントを提示し，何をどのように撮影するかという撮影概念の構築を助けるシステムの提案を行っている[42]．また，Adamsにより，特定のパターンが想定できる撮影の場合，映画で一般的になっている撮影パターンを利用して，ガイドとなる構図をＣＧ等で例示し，撮影の仕方を提案するモバイルシステム[43, 44]など，映像品質を高め，撮影者を支援するシステムの研究が行われている．

しかし，以上の研究では，一般ユーザのカメラの操作技術が一定のレベルを越えていることが前提となり，初心者の映像で問題となりがちな，映像断片の接続時に生じる基本的な問題の考慮や，映像の動きの乱れなどの問題を改善させる課題が抜け落ち

表 4.1　カメラワークに依存した単一ショットの映像文法

Rule(1-1)	カメラワーク（FIX）はしっかり止めること
Rule(1-2)	カメラワーク（PAN）は滑らかに動かすこと
Rule(1-3)	カメラワーク（ZOOM）は一定速度であること
Rule(1-4)	カメラワークを用いるときは前後に１秒以上の FIX ショットが必要
Rule(1-5)	FIX ショットで LS,MS,TS の時間長はそれぞれ基本的に 6s, 4s, 2.5s
Rule(1-6)	被写体の動きがない FIX ショットは最大でも 15 秒まで

ている．また，映像の撮影・編集は，多くの課題を同時に解決しなければならないため，初心者にはハードルが高く，効率的な学習法が必要となっている．

我々は，放送局で培われた映像文法 [5] に基づき，高次の撮影・編集支援だけでなく，初めてビデオカメラを操作するユーザを含めた映像撮影学習システムの実現を目指している．本稿では，映像文法を知らず，ビデオカメラの撮影経験がない初心者でも，映像文法で求められるカメラワークの基礎品質を満たし，同時に映像文法に従った単一ショットの撮影スタイルを自然に習得させる手法に焦点をあてる．

ここでの単一ショットとは，映像編集で扱われる時空間が連続した映像の最小単位であり，カメラを固定して撮影した FIX と，カメラの首振りに対応する PAN(水平方向を PAN，垂直方向を TILT とする場合もあるが本稿ではいずれも PAN とする)，カメラのレンズ操作による焦点距離の変更に対応する ZOOM により構成される．本論文では，FIX，PAN，ZOOM を総称してカメラワークと呼ぶ．

特に，本稿では，まず撮影中にオンライン処理で撮影の問題箇所の指摘を行う手法を採用するため，輝度投影相関 [39] と，テンソルヒストグラム [41] を応用した二分化テンソルヒストグラムによる実時間処理向きのカメラワーク解析法 [45]（3 章）を用いる．この結果，不適切なカメラワークとして，1. 手ぶれ，2. 速度超過，3. 蛇行など，初歩的な三つの問題点の判定を行う．また，本稿で提案する訓練指向型オンライン映像撮影ナビゲーションシステムでは，映像文法に従った型通りのショットの撮影を誘導するため，問題箇所の提示を行って，問題箇所を減らすよう撮影を繰り返すことにより撮影スタイルを教え込む手法を採用する．

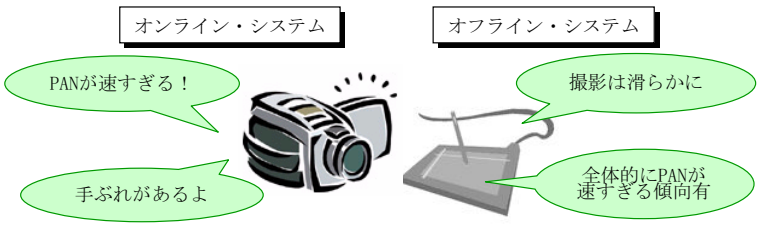


図 4.1　映像撮影ナビゲーションカメラ (左) と助言端末 (右)

4.2　映像文法を基盤とする映像撮影学習システム

4.2.1　初心者の問題点

放送用の映像では，ある定められた時間長を基本とするFIXを映像の基本単位として多用することが望ましいとされ，カメラワークは乱用せず，用いる場合でも，その動きが制限される．しかし，初心者の場合，FIXは時間長が短く，しっかりとカメラが固定されていないことも多く，カメラワークが乱用され，その動きは速く乱れており，それが問題であると気づきにくい．また，映像文法では，編集で単一ショットを接続する場合の問題を考慮した撮影スタイルが導入されている．しかし，初心者の撮影では，編集までを意識して撮影が行われないため，編集した映像が見にくく，対処策もわからないため，無編集での放置が多くなり，映像制作が普及しにくい問題の一因となっている．

4.2.2　オンライン学習とオフライン学習

このような問題を解決する手法として，撮影者個人に密着した撮影学習システムを実現する方法が考えられ，次のような特徴の異なる２つのシステムが想定できる．図4.1は，このコンセプトを図示したものである．一つ目は，図4.1(左)のように，実時間で撮影を評価し，撮影中，正しい撮影技法に誘導するオンライン型のビデオカメラシステム，二つ目は，図4.1(右)のように，いくつかのショットの撮影終了後，撮影の指導を行うオフライン型の端末などのシステムである．オンライン型のシステムは，撮影中，動的に状況に対応できるという利点がある反面，短時間内での過多な情報提示は撮影者に負担となるだけでなく，実時間処理を実現することが困難となるため，機

能は限定される．これに対し，オフライン型のシステムは，動的な対応はできないが，必ずしも実時間性に執着する必要がなく，撮影後に，多角的な分析を行って撮影上の癖や問題点を指摘するなど，きめ細かい指導を実現することも可能である．我々は，規範となる撮影スタイルに誘導する観点から，これらのシステムを総称して映像撮影ナビゲーションシステムと呼ぶ．

4.2.3　オフライン学習による従来法の問題点

天野らは，数分間，ユーザが連続して撮影した映像を，撮影終了直後に解析し，カメラワークの速度超過区間の一覧提示と，文章による問題箇所の指摘を行い，また，映像文法に従った使用可能な区間を見せる手法により，オフライン映像撮影ナビゲーションシステムの一例を構築した [40]．この方法により，意識の高いユーザは映像文法に従う映像が増加する傾向にあるが，その映像量は少なく，ユーザの多くは，撮影に自由度があるため，ユーザに訓練して欲しい撮影スタイルを強制できず，映像文法に従う区間はほとんど撮影できなかった．そのようなユーザでは，映像に FIX が少なすぎるばかりか，必要な時間，カメラを止めていられず，映像の大部分はカメラワークで構成される結果となった．

また，カメラワークの品質では主に速度超過を指摘するだけで他の問題を指摘できていないことも問題となった．

4.2.4　提案する映像撮影ナビゲーションシステム

従来法では，表 4.1:Rule(1-1)〜Rule(1-3) のようなカメラワークの品質だけでなく，FIX や表 4.1:Rule(1-4) に従う撮影が，映像文法を熟知していても意外に難しいことが判明した．このため，映像文法に従った単一ショットを確実に撮影させ，繰り返し修練させる手法が問題解決に必要であると思われた．また，映像の制作では，映像作品の時間長の制約も考慮する必要があり，定められた時間枠内での撮影を前提としてカメラワークの動きを計画し，撮影するセンスを持つことも望ましい．これは必ずしも時間通りに撮影できずとも，時間枠を意識して撮影を修練させることに意義がある．このように，映像文法に従った撮影技法を定着させるためには，時間長の制約がある環

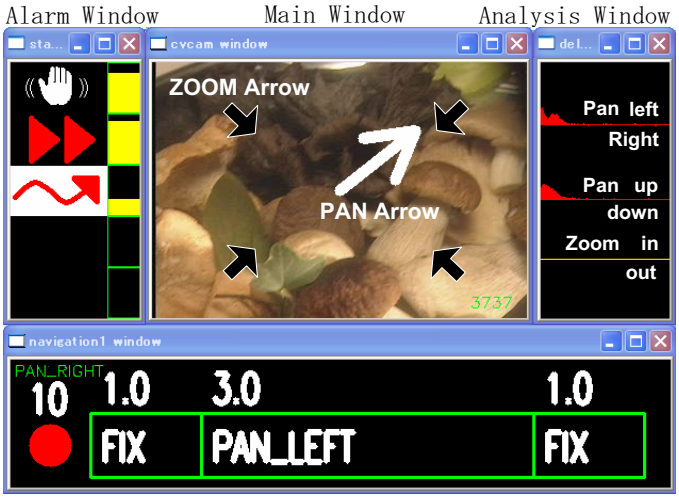


図 4.2　システムの GUI

境下で，カメラワークの品質，撮影スタイルを同時に修練させる効率の良い学習手法が必要となる．

特に，初心者では，撮影の問題点を事後的に指摘されるより，問題が発生した時点で指摘を行わなければ，その問題点を意識させにくい．また，これらの撮影法は，映像文法を背景とするが，概念を教えずに済むなら，初心者の負担が軽減する．そこで，本研究では，初心者が事前に映像文法を知らずとも，単一ショットの撮影スタイル，カメラワーク，時間の感覚を確実に修練させる方法として，訓練指向型オンライン映像撮影ナビゲーションシステムに着目する．本研究で提案する訓練指向とは，ある型に沿って繰り返し修練することで，映像文法という概念を説明せず，意味を考える前に，映像文法に従った単一ショットの撮影技法を自然に体得させてしまおうとする指導法の指針を示す．

4.3　訓練指向型映像撮影ナビゲーションシステム

4.3.1　ナビゲーションシステムの GUI

図 4.2 は，学習者がビデオカメラのファインダーの代わりに見る訓練指向型撮影ナビゲーションシステムの GUI，図 4.3 は全体の処理過程を示している．図 4.3 に示すよう

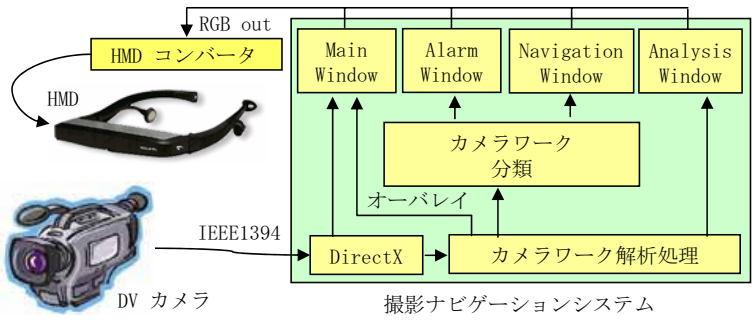


図 4.3　システムの処理過程

に，ビデオカメラの映像は，DirectX 経由で撮影ナビゲーションシステムにて処理が行われ，学習者は図 4.2 の画面を HMD(Head Mount Display) を通じて見ることになる．GUI には，四つのウィンドウがあり，上部中央の Main Window はビデオカメラのファインダーの替わりとして見る映像である．

上部右側 Analysis Window は，カメラワーク検出部の結果 [45] から，時間正規化された水平方向の動作量関数 $Pan_{lr}(f)$，垂直方向の動作量関数 $Pan_{ud}(f)$ と，ZOOM の動作量関数 $Zio(f)$ による速度関数が時間同期で表示される．また，$Pan_{lr}(f)$ と $Pan_{ud}(f)$ から計算された，PAN ベクトル Pan_f，また $Zio(f)$ の動作量は，オーバレイによって図 4.2:Main Window に示した矢印のように提示される．矢印の長さは速度量に対応する．

次に，上部左側がカメラワークの不適切な動きを警告する Alarm Window であり，現在，上から 1. 手ぶれ，2. 速度超過，3. 蛇行に対応する三つのアイコンが並んでいる．それぞれの警告は，独立に計算され，ある警告が発生した場合，アイコンの背景が黒から白に反転し，それに合せて各アイコンの右側にある黄色で満たされた四角形の高さが低くなる．この四角形の高さの度合いによって，どの警告が数多く発生しているかを把握することができる．

最後に，下部が映像文法に従うショットの撮影を誘導し，問題区間を明示する Navigation Window である．

4.3.2　Navigation Window

図 4.4 は，Navigation Window の四つの状態 (状態 A〜状態 D) の変遷を示している．図 4.4(a) には，訓練対象となる映像文法によって導き出された FIX ショット (Type1)，もし

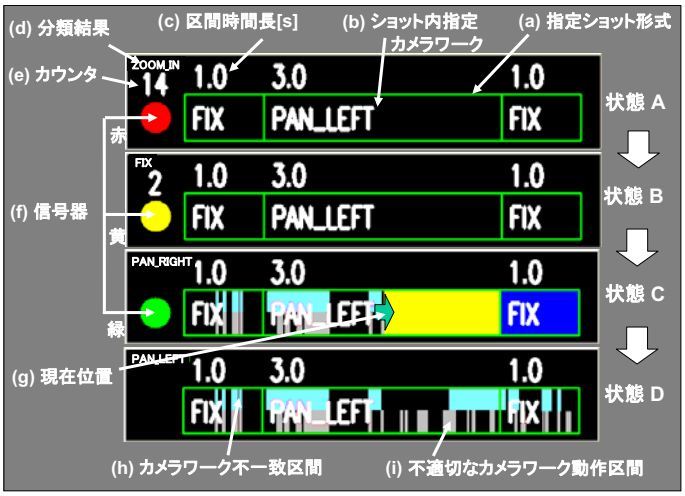


図 4.4　Navigation window の表示過程

くはカメラワークの前後にFIXが含まれるショット (Type2) の型が提示される．図4.4(b) は，Type2 ショットのカメラワーク部について，PAN_LEFT，PAN_RIGHT，PAN_UP，PAN_DOWN，ZOOM_IN，ZOOM_OUT の 7 種類が示される．また，図 4.4(c) のように，各カメラワーク区間の時間が秒数で提示される．図 4.4(d) には，実時間処理で毎フレームごとに判定されたカメラワークの分類結果が随時提示される．

次に，図 4.4(e) の数値は，秒数に同期したタイムカウンタを示しており，図 4.4(f) の円形の信号は，状態 A では赤，カウンタが 10 秒を切ると黄色 (状態 B)，0 秒をきると緑 (状態 C) となって撮影ナビゲーションが開始される．

状態 C で図 4.4(a) 内の各区間は，FIX 区間の背景が青，カメラワーク区間の背景は黄色で満たされる．図 4.4(g) は現在位置を示しており，塗りつぶされた背景が経過時間とともに同期して右方向へ減少していく．学習者は，この現在位置に対応するカメラワークを実現するのである．この例では，最初に FIX 区間を 1 秒，左方向の PAN 区間を 3 秒，最後に FIX 区間を 1 秒現在位置に合わせて実現すればよい．もし，学習者が実現しているカメラワークの判定結果 (図 4.4(d)) と，現在位置の指定カメラワークが異なる場合，図 4.4(h) のように，上半分に赤紫の区間 (カメラワーク不一致区間) が提示される．また，同様にカメラワークの何らかの警告があった区間は，図 4.4(i) のように，下半分に赤の区間 (不適切なカメラワーク区間) として提示される．

このように，学習者は，提示されたショットの右端にインジケータが到達した状態 D で，図 4.4(h)，(i) のように赤紫と赤の区間が残らないよう，ゲーム感覚で撮影を行

う．この撮影を繰り返し行う訓練指向の方法により，学習者は，映像文法の概念を知らなくても，自然に FIX 区間の時間感覚や撮影スタイルを意識し始める．特に，初心者にとって，カメラワークの前後でビデオカメラをしっかり止めることは難しく，図 4.4(h)，(i) の区間を残さないようにする試行錯誤が撮影技術の向上に貢献することになる．

4.4　不適切なカメラワークの判定法

4.4.1　不適切なカメラワークの判定項目選定

人では，同時に追従できる視覚的要素は四つ程度とされる [46]．また，初心者は，一度に多くの課題を与えると大きな負担となる．提案システムでは，第一に Navigation window の指示に注意を払う必要があるため，以上の観点に従えば，不適切なカメラワークの判定項目は上限で三つ程度となる．特に初心者が抱えるカメラワークの問題は，手ぶれ，速度超過である．手ぶれは FIX の撮影で問題となるが，手ぶれと PAN の速度超過が加わると，蛇行となる．つまり，PAN 動作中の手ぶれを指摘するために蛇行の判定が必要となる．この他にも，速度のむら，急激な速度変化，急激な方向転換などもあるが，これらはプロが撮影した素材映像にも現れる失敗で使用不能区間 [47] と呼ばれ，むしろカメラワークの品質をより向上させるための課題であると位置づける．本研究では，初心者向けに，三項目に的を絞る観点から，手ぶれ，速度超過，蛇行を判定対象とする．

なお，判定に用いるカメラワークの動作推定量は，輝度投影相関と二分化テンソルヒストグラムを用いた実時間処理向きのカメラワーク解析法 [45] (3 章参照) による結果を用いた．PAN の動作推定量は，輝度投影相関法により水平方向 $Pan_{lr}(f)$ と垂直方向 $Pan_{ud}(f)$ の動作量が計算され，それに基づき速度量を示すベクトルの大きさを $P_S(f)$，角度を $P_\theta(f)$ として PAN ベクトルが計算される [45]．また，PAN ベクトルの角度差分は $\Delta P_f^\theta = P_\theta(f) - P_\theta(f-1)$ とする．一方，ZOOM の動作推定量は，二分化テンソルヒストグラムにより計算される [45] が，ここでは $Zio(f)$ と表記する．

4.4.2 手ぶれ (Hand Shake)

手ぶれは，PAN の小さな動きが小刻みに孤立して現れる状況を示しているため，式 (4.1),(4.2) のように，縦横それぞれ，あるフレーム f の直前の PAN 動作量が 0 かつ，フレーム f の PAN 動作量が閾値 $(T_{hs}^{lr}, T_{hs}^{ud})$ 内の場合を $hs_f^{lr} = 1$，$hs_f^{ud} = 1$ としたとき，時間窓 N 内 $\bar{hs}_f^{lr} = \sum_{n=0}^{N-1} \frac{hs_{f-n}^{lr}}{N}$，$\bar{hs}_f^{ud} = \sum_{n=0}^{N-1} \frac{hs_{f-n}^{ud}}{N}$ の大きい方の出現率 hs_f(式 (4.3)) が閾値 T_{hs}^{min} と T_{hs}^{max} 間に含まれる場合として，式 (4.4) により計算される．ただし $N = 15$ とした．

$$hs_f^{lr} = \begin{cases} 1 & \left(Pan_{lr}(f-1)=0 \ \wedge \ 0<Pan_{lr}(f)<T_{hs}^{lr}\right) \\ 0 & otherwise \end{cases} \tag{4.1}$$

$$hs_f^{ud} = \begin{cases} 1 & \left(Pan_{ud}(f-1)=0 \ \wedge \ 0<Pan_{ud}(f)<T_{hs}^{ud}\right) \\ 0 & otherwise \end{cases} \tag{4.2}$$

$$hs_f = \begin{cases} \bar{hs}_f^{lr} & (\bar{hs}_f^{lr} \geq \bar{hs}_f^{ud}) \\ \bar{hs}_f^{ud} & (\bar{hs}_f^{lr} < \bar{hs}_f^{ud}) \end{cases} \tag{4.3}$$

$$HS_f = \begin{cases} true & (T_{hs}^{min} < hs_f < T_{hs}^{max}) \\ false & otherwise \end{cases} \tag{4.4}$$

4.4.3 速度超過 (Too Fast Motion)

速度超過は PAN，ZOOM 動作量が閾値 T_{tfm}^{min} と T_{tfm}^{max} 内にある場合とし，式 (4.5) によって判定される．

$$TFM_f = \begin{cases} true & (T_{p,tfm}^{min} < P_S(f) < T_{p,tfm}^{max}) \\ true & (T_{z,tfm}^{min} < |Zio(f)| < T_{z,tfm}^{max}) \\ false & otherwise \end{cases} \tag{4.5}$$

4.4.4　蛇行 (Serpentine Motion)

蛇行は PAN の動作方向にゆらぎがある場合であるが，ある方向に向かっている際，その方向から閾値以上の正 (positive) の角度差が得られた場合を $k_f^p = 1 \ (\Delta\hat{P}_f^\theta > T_{sm}^k)$，閾値以上の負 (negative) の角度差が得られた場合を $k_f^n = 1 \ (\Delta\hat{P}_f^\theta < -T_{sm}^k)$ とし，それぞれの時間窓 $F(F = 30)$ 内での総和をそれぞれ $sm_f^p = \sum_{n=0}^{F-1} \frac{k_{f-n}^p}{F}$，$sm_f^n = \sum_{n=0}^{F-1} \frac{k_{f-n}^n}{F}$ としたとき，正と負の角度差総和が同時に閾値を越えた場合として，式 (4.6) にて判定される．

$$SM_f = \begin{cases} true & (T_{sm}^p < sm_f^p \wedge T_{sm}^n < sm_f^n) \\ false & otherwise \end{cases} \tag{4.6}$$

4.5　提案システムの性能

4.5.1　実験装置と撮影環境

我々は，Pentium M (1.4GHz) のノートパソコン上で訓練指向型オンライン映像撮影ナビゲーションシステムを実装した．映像のキャプチャについては，DirectX を直接用い，GUI については OpenCV を利用した．また，DV カメラとしては，Victor 社のハイビジョンカメラ GR-HD1 の DV モードを用いた．撮影実験では，パソコンの画面を見ながらビデオを操作することが難しいため，ビデオカメラのファインダの代わりともなる図 4.2 の GUI を HMD の画面にフル表示し，撮影者は HMD の画面をファインダの代わりとして外界を見る方式の撮影訓練を行った．HMD には，両眼方式で，2m 先に 42 インチ相当で解像度 640x480，フルカラー画面が得られる Icuiti 社の V920 を用いた．撮影環境は，室内で，視野内に動物体のない環境を用いた．

4.5.2　不適切なカメラワークの判定実験

不適切なカメラワークの検出精度を確認するために，三つ (手ぶれ，速度超過，蛇行) の不適切なカメラワークそれぞれに該当すると思われる動きを 6 秒程度 3 回ずつ撮影

表 4.2　不適切なカメラワークの判定結果

	C	D	E	Recall C/(C+D)	Precision C/(C+E)
手ぶれ	10	0	5	100%	67%
速度超過	9	1	2	90%	82%
蛇行	10	0	4	100%	71%

し，合計 18 ショットについて評価を行った．ただし，三つの警告内容は，それぞれ瞬時的に発生するものと，ある時間窓を通じて判定できるものが混在している．そこで，警告内容の判定においては，フレーム単位ではなく，長すぎず短すぎない区間長を実験的に 20 フレームと定め，区間単位で判定精度を計算した．**表 4.2** は，その 3 種類の不適切なカメラワークを判定した実験結果である．

表 4.2 中，C を検出対象の正解数，D を正解が検出されなかった数として「未検出数」，E を正解でないものが過剰に検出された数として「過剰検出数」とするとき，再現率 ($Recall$) は C/(C+D)，適合率 ($Precision$) は C/(C+E) で定義される．ただし，表 4.2 中，C，D，E に提示された数値は，フレーム区間数である．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．

表 2 より，過剰検出よりも未検出の少なさを優先した精度となっている．これは，撮影者が発生させた，警告が必要なカメラワークの確実な検出を優先したためである．

4.5.3　処理速度

DV カメラから IEEE1394 経由で DirectX の機能を用いてフレーム画像を取得する場合，Frame rate にはゆらぎがあり，フレームをキャプチャするだけの処理で，各フレーム間の時間間隔を計測した例では，平均 10.3ms，最小が 8ms で最大は 28ms であった [45]．また，カメラワーク解析部 [45] と本研究での不適切なカメラワークの判定，GUI の処理を含めた結果は，平均値で 29.8ms となった．映像のフレームレートとして一般

的な 30fps の場合，フレーム間の時間差は 33.3ms となり，この時間差が短いほど高い時間分解能になることを示す．本研究のシステムは，この 33.3ms の時間差よりも短い 29.8ms が平均値であり，最も時間差が短い例として 22ms に至るものもあった．また，最も長いでは 120ms を要する場合もあるが，最頻値が 30ms であり，約 30ms に集中していることがわかる．

本研究のシステムは，IEEE1394 経由のキャプチャによる画像取得は非同期であるため，隣り合う画像フレーム間の時間差は，変動する．このような非同期のフレーム取得が前提となる場合，Frame rate が低下しても，得られるフレームの取得時の時間は，その時点での現在時間であり，遅れが生じるのは，常にその現在時点から結果を提示するまでの経過時間である．したがって，その遅れが後続に積算されて，現在のカメラワークと提示したカメラワークのずれが増大していくことはない．この観点では，Frame rate として一般的な 30fps や 15fps であっても，それは時間分解能の異なる実時間処理と見なせるが，激しい動きを適切に判定するためには高い時間分解能であることが望ましい．

また，本研究のシステムでは，非同期によるフレーム間時間差が変動するだけでなく，カメラワークの動作量推定，カメラワークの判定処理，GUI 処理等で必要となる処理時間や計算機内の状態によって，この画像フレーム間の時間差は変動し，長くなる．

しかし，実時間オンラインシステムとしては，画像フレーム間の時間差は可能なかぎり短く，変動動があるとしても，基本的にはすべての画像フレーム間が少なくとも 33.3ms 以下であることが望まれる．この性能が実際にどの程度実現されているかを検証するため，図 4.5 を用いる．

図 4.5 の横軸はフレームレート (Frame rate)，縦軸は式 (4.7) で示されるある処理時間間隔の累積含有率 (RC) である．

$$RC(fps) = \frac{1}{N} \sum_{i=0}^{fps_M - fps} x_{fps_M - i} \tag{4.7}$$

式 (4.7) の，N は全フレーム区間数，fps_M は想定される最大のフレームレート，x_{fps} はフレームレート fps となったフレーム間処理の数である．また，提案システムを稼働させた際に隣り合う画像フレーム間の時間差を計測し，その時間差をフレームレートに換算した値を fps とする．

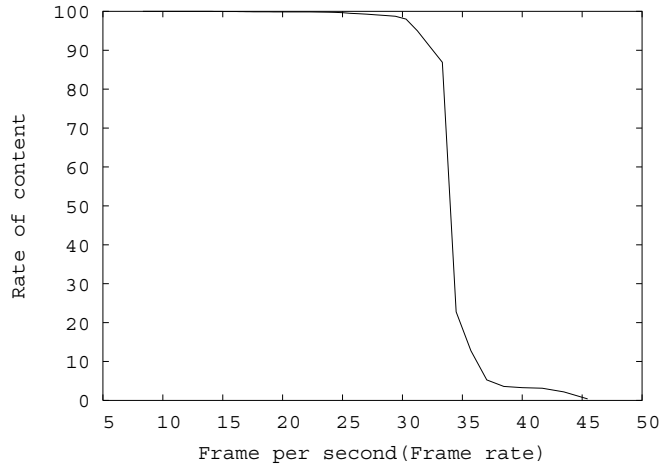


図 4.5 Frame rate と処理時間含有率との関係

式 (4.7) は，累積として高いフレームレート fps_M から fps までに含まれるフレーム間が全フレーム間に対し，どれほど含有するかを比率を示しており，ある fps のとき，$RC(fps)$ により，fps より高いフレームレートとなるフレーム区間がどれだけ含まれるかを知ることができる．図 4.5 より，少なくとも 98.0%フレーム間の処理時間が一般的な 30fps 以上のフレームレートに含有される結果を得ており，ほぼ 30fps での実時間とみなせる．

4.6 被験者によるシステムの評価実験

4.6.1 提案システムの評価法

ここで，提案システムを評価するための着眼点を述べる．表 4.3 は，実験に用いるショットの種類である．表 4.3 の Type1 ショットの撮影は指定された時間長の間，確実に止めることが難しく，特に，Type2 ショットの前後を止めることは，映像文法を概念的に熟知していても意外に難しい．そこで，ユーザの目標はカメラワークの前後をしっかり止め，同時に手ぶれを起こさない撮影スタイルの技術修得となる．また，初心者には，カメラワークの速度が大きく，手ぶれを抑える意識の弱い者も少なくないため，ビデオカメラ先端の揺れに PAN が加わると，PAN は蛇行のような動きとなる．ただし，映像作品を制作する場合，一般的には 10 年の話を 1 時間に収めるなど，基本的に

表 4.3　被験者実験用の指定ショット

No.	ショットの型	1 セットに含まれるショットの種類
1	Type1	FIX(6.0[s])
2	Type1	FIX(4.0[s])
3	Type1	FIX(2.5[s])
4	Type2	FIX(1.0[s]), PAN_LEFT(4.0[s]), FIX(1.0[s])
5	Type2	FIX(1.0[s]), PAN_RIHGT(4.0[s]), FIX(1.0[s])
6	Type2	FIX(1.0[s]), PAN_UP(4.0[s]), FIX(1.0[s])
7	Type2	FIX(1.0[s]), PAN_DOWN(4.0[s]), FIX(1.0[s])
8	Type2	FIX(1.0[s]), ZOOM_IN(4.0[s]), FIX(1.0[s])
9	Type2	FIX(1.0[s]), ZOOM_OUT(4.0[s]), FIX(1.0[s])

はある定められた時間内に映像全体の時間長を収める必要があるため，カメラワークは時間制約の中で適切な動作となるよう，計画を練って撮影する感覚も必要である．

このように初心者は，不慣れなカメラ操作とともに，時間制約の中，撮影スタイルを考慮しつつ，カメラワーク動作を安定させる課題を同時に克服する必要がある．しかし，特に初心者の場合，ある課題に集中すれば，他の課題を忘れがちになるため，提案システムの評価としては，総合的に課題を向上させられるかが焦点となる．以上の観点から，評価においては，表 4.3 のショットのうち，Type1 の FIX，Type2 のカメラワーク前の FIX(PRE FIX)，後の FIX(POSTFIX) のカメラワーク一致率と，手ぶれ，速度超過，蛇行の改善率の総合的な傾向を評価する．

ただし，従来手法となるオフライン映像撮影ナビゲーションシステム [40] の一例では，撮影後に「カメラワークの前後に FIX が必要」などの文章が指示される誘導方式であるため，初心者に表 4.3 のショットを確実に撮影させる強制力がないことから，Type1 や Type2 のようなショットがほとんど撮影されなかった．そこで，従来手法との比較ではなく，提案システムを使用した場合と使用しなかった場合の比較によって評価を行う．

表 4.4 実験者グループ

	Try						
Group	1	2	3	4	5	6	7
A(システム使用)	S	S	S	S	S	N	–
B(システム使用)	N	S	S	S	S	S	N
C(システム未使用)	N	N	N	N	N	–	–
D(映像文法の概念のみを知る)	N	S	S	S	S	S	N

4.6.2 撮影実験の条件と環境

本研究では，三脚を使うなど，設備の整った撮影環境だけではなく，手持ちで撮影する際にも対応することが目標となる．そこで，ビデオカメラは手持ちで撮影をしてもらった．また，3 章で提案したカメラワーク解析法を含め，画像処理を用いてカメラワークを解析する多くの手法では，フレーム内に動物体が含まれるとカメラワークの解析精度が低下する．本研究では，撮影技法を習得させることが主な目的であるため，被験者の問題のある撮影技法として誤検出される可能性が高い動物体を含んだ映像の撮影を避けるため，撮影環境としては，動物体のない室内で実験を行った．また，撮影対象は特に指定しなかった．

１回の実験では，表 4.3 に示した映像文法に従う 9 種類の指定ショットを 1 セットとして，１セットごとに休憩をとる方法で行う．また，本稿で提案するシステムの効果を調べるために，16 名の被験者を 4 名ごとに四つの Group A〜D に分けた．被験者は，いずれも映像の撮影経験がない，もしくは乏しく，Group A〜C は映像文法に関する知識がない 20 代の大学生であり，Group D は映像文法を概念としてのみ熟知した同様の大学生もしくは大人である．

表 4.4 に，各 Group と，撮影実験回数 (Try) を示す．表 4.4 の S とは，システムを用いたことを示し，N は，システムを使用せず１セットの撮影を行うことを示す．N では，１ショットごとにビデオカメラの録画を開始し，指定されたショットを撮影して，各被験者の時間感覚で指定の時間が来たと判断すれば録画を終了する方式で実験を行った．評価は，録画した映像を AVI ファイル化し，提案するシステムと同じ手法で評価した．ただし，時間感覚の長短を考慮して，Type2 のショットは，表 4.3 の時間配分

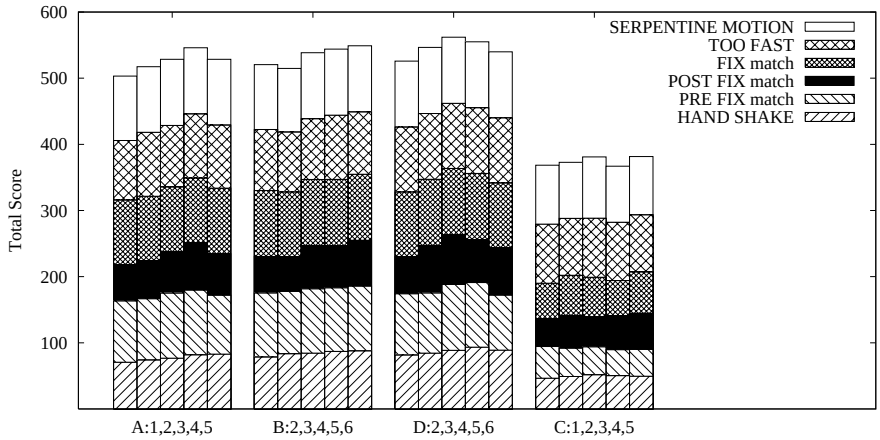


図 4.6　システムの使用・未使用による実験結果

の割合から 1 秒に相当する区間を算出し，AVI ファイルの前後を PRE・POST FIX の区間とした．表 4.4 より，Group A は，初回から 5 回の撮影実験でシステムを使用し，Group B,D は，システムを使用せずに表 4.3 の九つのショットを一回撮影した後，システムを用いて 5 回の撮影実験を行うことを示す．Group C は，5 回の撮影実験でシステムを使用しないが，一つのショットを撮影するたびに，表 4.3 を見せ，撮影対象を指定した．また，Group B,D は，7 回目にシステムを用いない N の撮影実験を 1 回行った．

4.6.3　5 回の撮影実験によるシステムの効果

図 4.6 は，5 回分の撮影実験を Group ごとにまとめて横に並べ，縦軸は，総合点 (Total Score) を各構成要素ごとに色分けして示した実験結果の経緯である．Group A,C が 1〜5 回目，Group B,D が 2〜6 回目の実験経緯であり，これは，システム S を使用した場合と，Group C の全く使用しない場合の 5 回分の実験結果を比較したことになる．

縦軸の総合点は，不適切なカメラワークとして，手ぶれ (HAND SHAKE)，速度超過 (TOO FAST)，蛇行 (SERPENTINE MOTION) の改善率 $Score^{I}_{n,i}$ と，表 4.3，各ショット内のカメラワーク区間のうち，FIX，PRE FIX，POST FIX のカメラワーク一致率 $Score^{M}_{n,s}$ を縦に積算した．

ここで，実験回数を n，被験者を e，各ショット s のフレーム総数を $A_{n,e,s}$，不適切なカメラワーク i と判定されたフレーム数を $E_{n,e,i,s}$ として，不適切なカメラワーク i ごとの適切性は，$I_{n,e,i,s} = 100 \cdot (1 - E_{n,e,i,s}/A_{n,e,s})$ とするとき，$Score^{I}_{n,i} = mean_e(mean_s(I_{n,e,i,s}))$

として求められる．カメラワークの問題が改善する傾向は，この適切性が向上することで判定することができる．次に，実験回数を n，被験者を e，各ショットのカメラワーク種類別区間 s のフレーム総数を $A_{n,e,s}$，指定されたカメラワークと被験者の実現するカメラワークが一致したフレーム数を $E_{n,e,s}$ として，$M_{n,e,s} = 100 \cdot E_{n,e,s}/A_{n,e,s}$ とするとき，カメラワーク区間の種類 s ごとの一致率は $Score^{M}_{n,s} = mean_e(M_{n,e,s})$ として求められる．

図 4.6 より，システムを用いた Group A,B,D では，4, 5 回目あたりで集中力が切れ，総合点で下がる部分もあるが，すべての問題について取組み，Group C よりは高い得点で少しずつ全体的な改善の傾向を得ることができている．Group A,B の比較では，Gruop B がシステム未使用実験を 1 回目に行ってカメラ操作に若干慣れたせいか，Group A よりは少し高い総合点から始まっているが，Group A,B ともに，映像文法を熟知した Group D に及ばずながら大差はなく，映像文法の概念を知らずとも，映像文法による撮影スタイルを着実に修得する様子が伺える．

4.6.4　システム使用前後の変化

次に，提案システムを用いる前後の撮影能力の変化を検証するため，Group B,D のシステム使用前 (1 回目) と使用後 (7 回目) の，システムを使用しなかった実験の総合点と，システムを使用しなかった Group C の 1 回目と 5 回目の総合点を並置した．その実験結果を図 4.7 に示す．図 4.7 より，映像文法を知る Group D は FIX, PRE・POST FIX の改善度が高く，総合点で最も高い値を示した．また，Group C では，ショットの撮影ごとに表 4.3 を見せているにもかかわらず，問題の指摘が行われないためか， 総合的な改善が少なく，PRE FIX の Score が悪くなっているのに対して，Group B では各課題の Score を維持もしくは改善を同時に導く総合的な傾向があり，システムの効果が得られている．

4.6.5　手ぶれ・速度超過・蛇行の改善

図 4.8 は，図 4.7 と同様に，案システムを使用する前と後において，システムを用いないで撮影した 1 回目の実験と 7 回目の実験実験結果を抜粋したものであり，システ

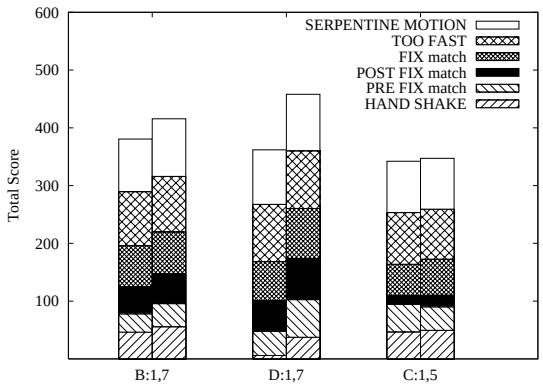


図 4.7　システムの使用前と使用後の実験結果

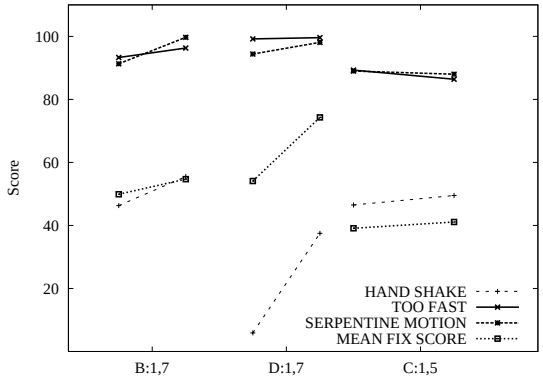


図 4.8　システム使用前後の各スコアの変遷

ムを用いない状態で，手ぶれ，速度超過，蛇行の Score の変遷を判定するためのものである．また，図 4.8 の MEAN FIX SCORE は，図 4.7 の FIX，PRE FIX，POST FIX のカメラワーク一致率に関する Score を平均して，システム使用前後の変遷を示している．本研究の目的である，映像文法に従った撮影スタイルを修得させつつ，手ぶれ，速度超過，蛇行という不適切なカメラワークの問題を改善させるという観点で図 4.8 を見た場合，システムを用いた Group B,D では，いずれも改善傾向が示されているのに対し，システムを用いなかった Group C では，手ぶれが改善傾向を示すものの，速度の超過や蛇行は改善傾向を示さず，同程度か，むしろ悪化する傾向が示された．手ぶれは，一般的に普及しているカメラやビデオカメラでも問題点として知られているため，特に指摘を行わなくても自己解決する対象として意識にのぼるものと考えられる．しかし，Group C で，速度超過や蛇行が改善傾向を示さない原因は，それが問題だと

いう意識がなく，指摘もされないため，改善対象にならないためであると考えられる．

ただし，映像文法を概念的に熟知したGroup Dは，図4.7で示したように，システム使用後の撮影で，高い改善度とTotal Scoreで最も高い値を獲得しているが，手ぶれのScoreがもともと低く，改善度が高くても，Group B,CのScoreに到達していない．本研究の手ぶれは，FIX区間での小刻みな動きを判定しているが，実は，Group B,Cが撮影したFIX区間では，撮影の方向が非常に緩やかに上下左右の動きを伴う超スローPANが多数ある．本章では，主に手ぶれと速度超過に依存する問題に焦点を当てるため，この問題点は今回の指摘の対象とはしておらず，これが手ぶれの判定を逃れる結果に一部関係している．

一方，Group Dでは，図4.8のMEAN FIX SCOREが示すように，FIX区間を指示通りに撮影するカメラワーク一致率のScoreが，他のGroupより大きく改善している．これは，映像文法の形式を満たす撮影技術の能力が大きく向上することを示しているが，この形式を満たすことに集中する傾向がある．また，映像文法を熟知するがゆえに，画面を止める意識も高い．このため，映像文法の形式に集中しつつ，画面を止めようとする意識が，逆に肩に力が入り，小刻みな手ぶれを生む結果につながるように思われた．このように，映像文法を概念として熟知していても，同時に問題解決を行う撮影の難しさが反映されているものと思われる．ただし，Group Dでは，映像文法を満たすFIX区間の撮影を大きく改善しつつ，手ぶれも同時に改善傾向を示しているため，概念だけではわからない問題に直面させ，本システムにより，その問題の改善を導く結果が得られていると思われる．また，映像文法を知らないGroup Bにおいても，手ぶれ，速度超過，蛇行を改善させつつ，映像文法に従う撮影技能を改善させる結果が得られている．

4.6.6　ショットの時間長に関する感覚の変化

最後に，システムの使用前後でシステムを用いないで撮影を行ったGroup B,Dの撮影実験1と7，また比較のため，Group Cの撮影実験1と5で，指定されたショットの時間長と，撮影者の時間感覚を矯正する効果が得られるかを検証するため，その誤差の平均M，標準偏差σ，最大Max・最小値Minの変化を図4.9に示す．Group Aは，システム使用前にシステムを使わない撮影を行っていないため，参考までにシステム

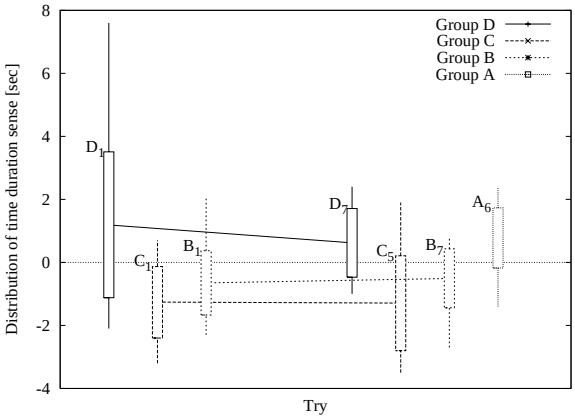


図 4.9　Group ごとのショット時間長の分布

を使用した後の撮影実験 6 のみを提示した．

図 4.9 の縦軸は 0 を基準として，与えられたショットの時間長と，ビデオカメラの録画ボタンの ON,OFF 間の時間長との誤差を示しており，負の値は時間が短かった場合，正の値は長かった場合を示す．縦軸は，例えば Group D の場合，D_1 が Group E の 1 回目，D_7 が 7 回目の撮影実験を示しており，D_1 と D_7 を結ぶ横の線の端点が誤差の平均値を示している．また，E_1 の平均値 M の周りを囲む四角形の上端が $M+\sigma$，下端は $M-\sigma$，四角形を貫く縦の線の上端が Max，下端が Min を表している．図 4.9 より，提案システムを用いた Group B は若干ながら分散が減り誤差も 0 に近づき，Group D では，時間長の大きなズレが改善している．しかし，システムを使わない Group C は平均値は変わらず，標準偏差は増える傾向を示した．提案システムでは，時間長に関するズレの矯正訓練は行っておらず，訓練対象ではないが，撮影者が時間感覚の補正を意識しつつ，図 4.7 の改善結果が得られていることになる．

4.7　結言

本研究では，映像文法に従う撮影技法を学習する映像撮影ナビゲーションシステムの中で，単一ショットの訓練指向型オンライン映像撮影ナビゲーションシステムを提案した．提案システムは，3 章の高い時間分解能を有するカメラワーク解析法を用いて，手ぶれ，速度超過，蛇行といった不適切なカメラワークを引き起こすユーザに撮影訓練システムを使用させることにより，システムを使用しなかった者よりも，大きく問

題が解決される結果を得ており，オンライン処理で問題を指摘することの有効性を示した．

本提案システムにより，初心者による不適切なカメラワークの問題点として，手ぶれ，速度超過，蛇行を克服する課題に直面させ，また同時に映像文法に従う時間長に従い，FIX やカメラワーク前後の FIX を確実に撮影させて，難しい撮影課題に直面させつつ，継続的にショットの品質を改善させる傾向も得られた．

また，初心者の癖として，画面の方向が上下左右緩やかに動いてしまう問題をどう扱うかが検討課題の一つとして残されるが，これらはテレビ映像としてある程度許容されているようである．ただし，撮影訓練では，こうした問題を解消するよう指摘し，この課題に取り組ませることにも意義があると思われる．

一方，放送に耐える映像の品質を問題にする場合，他にも速度のむらや急激な速度変化，急激な方向転換などの問題もある．これらは，手ぶれ，速度超過，蛇行などの問題が解決した後に，次の訓練課題として取り上げることができる．つまり，プロの撮影技法として訓練すべき課題は数多く存在するが，本研究の意義は，専門の指導員という人員を必要とせず，撮影技法を学習・訓練する枠組みを提供したことであり，新たな人員を必要としない意味において，第三の問題「人員不足」に影響を及ぼすことなく，個別指導を可能にし，第四の問題「人材育成」に貢献できることにある．

第5章

使用可能・不能区間推定による映像編集支援方式

5.1 緒言

映像の編集は，映像を制作する上で，根幹を成す作業である．映画などの制作では，綿密な計画のもとに撮影が行われる．しかし，特にテレビの世界では，必ずしもコンテを用意することなく，カメラマンとレポーターが現地に赴き，撮影を行い，編集者は，別に存在する場合もある．テレビの世界では，映像文法の概念がカメラマンや編集者によって共有されているために，映像の品質が保たれる関係があり，編集者は，映像文法に従ったショットが素材映像に含まれることを前提に，映像文法に従う完パケの編集作業に従事できる．しかし，その作業には，単純作業と知的な作業がある．本研究では，作業に多大な時間を要する編集作業を支援するために，この作業の一部を代行する，編集支援システムの一部に焦点を当てる．本研究で素材映像に索引付けされたメタ情報は，7章で提案する編集支援システムにおいて使用される．

映像編集作業を非効率にしている主要な問題の一つは，カメラマンが撮影した素材映像中に，放送用には使用できない使用不能区間が存在することである．使用不能区間とは，手ぶれや失敗，取直しに関わる区間，もしくは，カメラの位置・フォーカスの調整や，意図した撮影区間の前後にあるカメラの撮影方向調整区間などである．これらは，いずれも無意味で無造作なカメラワークであり，不安定な区間として定義できる．また，これ以外に撮影開始の合図を行う際のカメラマンの手が入った区間や，撮影の行われていない区間などもある．

Girgensohn ら [48] は，ホームビデオ用の素材映像から編集に適した区間を選択する，

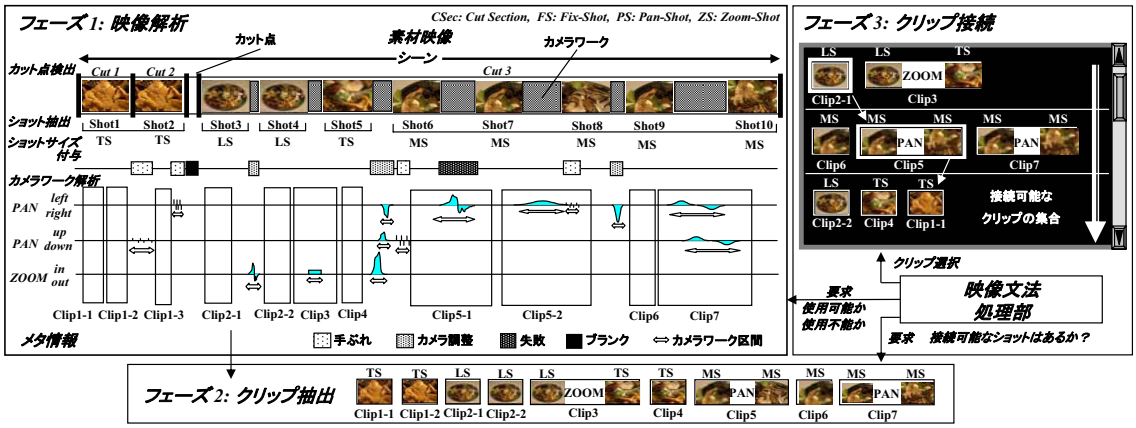


図 5.1　映像編集支援システム

半自動的な編集システムを提案しているが，使用不能な区間の判定については，明るさについての不適合度を用いるに止まっている．また，土橋ら [49] は，MPEG-2 の動きベクトルを用いてカメラワークの手ぶれ区間を推定しているが，手ぶれ以外の問題については言及してない．

本研究では，カット点とカメラワークの情報を用いて，使用不能区間を推定し，映像文法に従った使用可能なショット区間を自動的に推定する手法について提案する．

5.2　映像編集支援システム

7 章で実現する，映像文法に基づいた映像編集支援システムは，カメラマンが撮影した素材映像を受け取った時点から，映像を完成させるまでの作業，つまり，これまで編集者が行ってきた作業を，以下のような三つのフェーズに分けている．

フェーズ 1:　素材映像の解析と索引付け
フェーズ 2:　クリップ (5.3 で解説) の抽出 (切り出し)
フェーズ 3:　映像編集支援部 (クリップの接続)

図 5.1 は，映像文法を用いた映像の編集支援を行う上で，最低限の機能を有したシステムの概念図である．我々のアプローチでは，フェーズ 1，2 の完全自動化を目指し，フェーズ 3 では，5.3 で説明する接続可能なクリップの候補だけを編集者に一覧表示する．これにより，編集者は，非効率なフェーズ 1，2 の作業から解放され，フェーズ 3

内で，映像表現の要となるクリップの接続と，時間長の調整のみに作業を集中することが可能となる．

本論文で提案する使用可能なショット区間の自動抽出法は，フェーズ 1 の処理となる．ただし，フェーズ 1 中，ショットサイズの自動付与法に関しては，6 章，また，フェーズ 2，3 の実装法については，7 章にて提案する．ただし，7 では，複数のショットから構成されるシーンに相当する映像列を自動編集する手法について提案している．

5.3　映像用語と映像文法

ショットやカットなどを含め，海外から輸入された映像関連用語は国内で誤用されていたり，造語が定着している場合などがあり [26]，概念が統一されていない．本研究では，混乱を避けるため，抽出対象である使用可能・不能区間，ショット区間について定義を行う前に，文献 [26, 48, 50] に従って，カット点，カット区間，ショットを再度説明する．絶対的ショットサイズの名前は，図 2.6 において，方言的に異なる名前が使われていることがあるため，ここでは，再度，放送局で用いられている名前に基づいて説明を行う．また，定義を拡張したクリップの概念について述べ，最後にフォローについて述べる．

5.3.1　カット点とカット区間

通常，視聴者が見る映像は，カット点を区切りとして異なる場面が接続されている．瞬時に変化するこの変化点を本論文では「カット点」と呼ぶ．このカット点は，放送用として完成された映像 (以後，完パケ {PPP:Perfect Packaged Program} と呼ぶ) と「素材映像 (RVM:Raw Video Material)」とでは異なる意味を持っている（完パケと素材映像のカット点は国外でカットと呼ばれている）．完パケではショットを区切る点を意味するが，素材映像内のカット点は，カメラのスイッチをオン・オフしたことに相当する．このカット点とカット点で挟まれた素材映像上の区間を本稿では「カット区間 (CSec:Cut Section)」と呼ぶことにする（完パケのカット点は，5.3.2 で定義する確定ショットの開始点と終了点である）．

5.3.2　ショット

映像は実世界を四角い枠（フレーム）で切り出した記録物である．ショットはこのフレームを通して見える，構図や撮影対象に依存した時間方向にとぎれのない連続した区間であり，構図や撮影対象に着目した側面と，時間方向の区間に着目した側面を持つ．また，概念的に映像の最小単位と定義される（国内ではこれをカットと呼んでいる場合もあるが，ここではショットと呼ぶ）．このショットは，本論文中，複数の観点で用いられるため，便宜上，まず完パケの最小単位である編集後のショットを「確定ショット (CS:Confirmed Shot)」と呼ぶことにする（確定ショットは，国内で一般的にカットと呼ばれることが多く，海外ではショットと呼ばれる）．

概念的に，複数のショットを接続すると映像になるが，このとき，ショットを区切る変化点がカット点である．つまり，カット点で挟まれた映像区間をショットと解釈することもできる．この定義に従えば，第 1 に，素材映像上のカット区間，例えば，図 5.1 の Phase1 中 *CSec1,2* はショットである．しかし，これは撮影時，カメラマンが時間的に余裕を見て長めに撮影を行う慣習があるため，確定ショットとして使われるのは通常この一部分である．本論文では，ある確定ショットに対応する，確定寸前ではあるが，冗長な部分を含むショットを「冗長ショット (RS:Redundant Shot)」と呼ぶことにする．図 5.1 の *CSec1,2* は，この冗長ショットにあたる．

第 2 に，ショットをパンやズームなど，カメラワークの観点から捉える場合がある．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 (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って撮影対象の拡大縮小を起こすズームに大別される．カメラワークが起こっていない部分はフィックスと定義する．パンが行われている連続した区間は，パンショット (PS:Pan Shot)，またズームが行われている連続した区間は，ズームショット (ZS:Zoom Shot) と呼ばれる．これに対し，カメラを固定している連続した区間をフィックスショット (FS:Fix Shot) と呼ぶ．本論文では，パンショット，ズームショット，フィックスショットという三つのショットをカメラワークショット (CWS:Camera Work Shot) と呼ぶ．

$$PS, ZS, FS \in CWS$$

ただし，このショットの意味は，カット点で挟まれていることを前提としておらず，

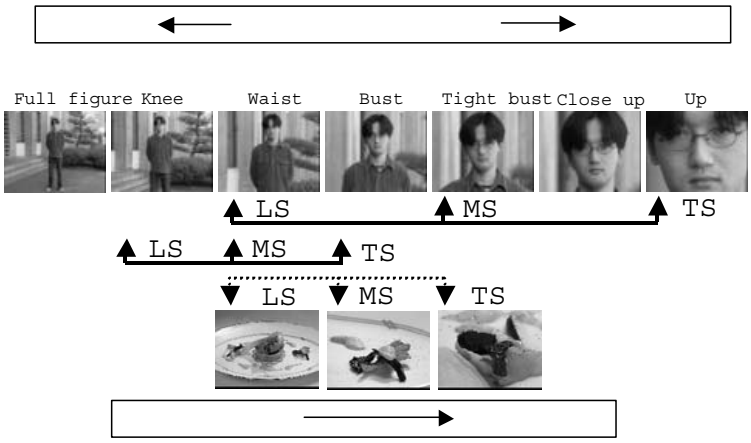


図 5.2　ショットサイズと相対的関係

パン，ズーム，フィックスいずれかの状態が連続している区間であるという観点で用いられている．また，冗長ショットは，カメラワークショットの一部を切り出したショットと見ることもできるため，カメラワークショットの部分集合となる．

5.3.3　ショットサイズ

図 5.2 には，人に関するショットサイズと人以外のショットサイズについて例が示されている．人に関するショットサイズでは，特に絶対名が存在し，本研究では，図 5.2 の人に関するショットサイズを絶対的ショットサイズとし，これを用いる．

2.5.2節で示したように，相対的なショットサイズは，タイトショット (TS:Tight Shot)，ミドルショット (MS:Middle Shot)，ルーズショット (LS:Loose Shot) に分類される．あるショットより被写体に近寄ったショットをタイトショット，引いたショットをルーズショット，両者の中間となるショットをミドルショットと呼んでいる．図 5.2 中，(α)，(β) の関係において，Waist shot は同じ絶対的ショットサイズであるにもかかわらず，相対的ショットサイズとしては位置づけが異なる．このように，相対的ショットサイズは三つのショットの内容を比較して，初めて定義できるショットの空間的分節概念である．

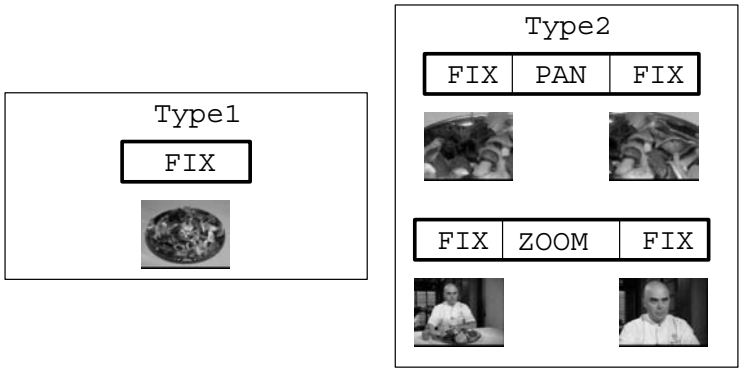


図 5.3　クリップの種類

5.3.4　クリップ

文献 [48] では，素材映像から切り出した映像の区間をクリップと呼んでいる．本論文で定義するクリップは，素材映像から物理的に切り出した区間として定義されるが，カット点を含まない連続した区間であるため，ショットの一種である．ただし，これまでのショットの概念に加え，映像文法に従うという観点が含まれている．

まず第一に，表 2.2 に抜粋した映像文法のうち，Rule(1-1) により，カメラワークに依存した，少なくとも 2 種類のクリップが考えられる (図 5.3)．Type1 は，カメラワークの観点に対応させた場合，フィックスショットと等価である．しかし，Type2 は，パンショットもしくはズームショットの前後に必ず 1 秒以上のフィックスショットを含んだショットとなる．

第二に，Type2 に含まれるパンショットやズームショットは表 2.2 の Rule(1-2)〜Rule(1-4) に従って，安定したカメラワークでなければならない．つまり，カメラワークショットの内部にある不安定な区間は使用不能区間とされ，クリップを構成できないという制約がある．

第三に，Type1，Type2 ともにフィックスショットを含んでいるが，フィックスショットは，通常，ショットサイズが対応付けられており，表 2.2 中，Rule(1-5) に従って時間長の制約を受ける．

ショットの概念には，時間長という制約はないため，一つのフィックスショットから複数のクリップを切り出すことが可能となる．例えば，図 5.1 中フェーズ 1 の FS1 は，一つのフィックスショットであるが，ショットサイズが TS であるため，この区間内で

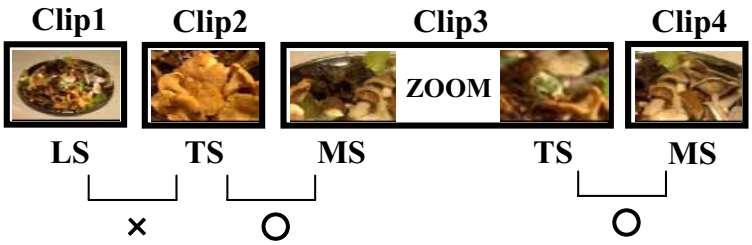


図 5.4 クリップの接続判定

切り出すクリップは基本的に Type1 で 2.5 秒の区間となる．FS1 が少なくとも 5 秒あれば Clip1-1,Clip1-2 のような同じショットサイズを持つ二つのクリップを確保できるのである．カメラマンが FS1 を一つのショットとして撮影したとしても，編集段階で FS1 から二つのクリップを取出すということは間違いではない．このように，カメラマンが意図したショットと，編集者が切り出すクリップが異なる場合がある．

最後に，クリップは，フェーズ 3 において接続が行われるが，表 2.3 に示す Rule(2-1) から Rule(2-5) などの映像文法を用いた図 5.4 に示すようなクリップどうしの接続判定が行われる（例えば，図 5.4 の Clip1 と Clip2 は Rule(2-1) に従い接続できない）．このようにして，一つのシーンが生成されるが，表 2.4 の構文的規則 Rule(3-2) により，シーンの最初と最後に位置するクリップは，時間長が 2 秒増やされる．また，編集の最終段階では，映像全体を時間内に収めるため，それぞのショットの時間を縮めるなどの調整が行われる場合がある．

このように，クリップは，本論文の映像編集支援システムで扱う映像文法に依存した物理的な最小単位であるが，フェーズ 3 にて時間長の調整，また切り出し位置を調整する場合もあるため，確定ショットではなく，冗長ショットの一種である．

$$Clip \in RS$$

5.3.5 フォロー

フォローとは，撮影対象を追いかけて撮影することであるが，大きく分けて以下の二つが想定できる．

1: カメラを固定し，パンやズームで対象を追跡
2: カメラマンがカメラを担いで対象を追跡

フォローを行っている場合のパンやズームには，これまでの映像文法をそのまま適用してクリップを検出することが困難である場合が多い．2:の場合は，カメラを担いで撮影しているため，カメラが絶えず動いており，映像的にはパンやズームが頻繁に起こっている状態と見た目で同じとなる．特に，今回扱っている料理番組用の素材映像では，調理シーンでフォローがよく使われている．このとき，対象はフレーム内で大きい領域を占め，なおかつ動いており，カメラマンはその対象を追いかけているため，カメラが動いているのか対象が動いているかについて判定することは困難である．今日のカメラワーク検出技術では，いずれの方法を用いても，本質的にこの問題に対処できない．

また，1:の場合でも，レポータが料理を食べながら話している場合，レポータをフレーム内に収めるよう若干のパンやズームを行う場合や，パンとズームを併用したパン・ズームにより，レポータから料理長へカメラを向け，そのまま料理長が話すようなものもある．この場合も，被写体を追いかけている影響で微妙なカメラワークが発生している．

このように，映像が絶えず動いていると，ショットの定義だけでなく，パンやズームも同様に定義があいまいになる．これは表 2.2 の Rule(1-1)〜(1-4) が要求するようなカメラワーク動作ではなく，無造作な動きとなるが，対象を追いかけている観点で，意味がないカメラワークとは言えない．また，レポータをフォローしているシーンでは，レポータが発話していたり，意味のある動作をしている場合がある．このようなシーンでは，表 2.2 の映像文法，Rule(1-1) や Rule(1-5) のように，カメラワークの前後を単純な時間間隔で切断できない．例えば，Rule(1-1) により，単純に 1 秒で切断したり，Rule(1-5) のように，6，4，2.5 秒の点で単純にクリップを切り出すことはできない．なぜなら，それぞれの時間位置でレポータが発話中である可能性があるからである．現在，フォロー内でクリップを切り出すのに有効な規則がないため，本研究ではフォローが撮影されるカット区間を特定するに留め，フォローが行われているカット区間内のクリップの抽出については今後の課題とする．また，ショットの抽出は編集者に委ねるため，フェーズ 1 においては使用可能区間に位置付ける．なお，このフォローが行われているカット区間を「フォロー (Follow)」と呼び，それ以外のカット区間は本論文内で「通常カット区間 (NCSec:Normal Cut Section)」と呼ぶ．

5.4　使用可能・不能区間とショット区間

ここで，使用可能・不能なカット区間について述べる．また図 5.1 を用い，カメラワークの手ぶれ区間や不安定な区間と使用可能・不能区間の関係について述べる．ただし，図 5.1 の PAN・ZOOM に関して，縦方向は変化量の大きさを表している．PAN については right:left と up:down に分け，それぞれ中央の線を 0 として上側:下側の大きさが変化量を表している．ZOOM の場合は in:out の倍率である．また，本節の最後に，本論文で抽出するショット区間について定義を行う．

5.4.1　使用可能・不能なカット区間

図 5.1 中，Phase1 は，あるシーン内に含まれる通常カット区間列の一例を示している．素材映像は,「カット区間 (CSec)」の集合であり，要素として「通常カット区間 (NCSec)」と「フォロー (Follow)」があることを前節で述べた．ただし，図中 CSec2 と CSec3 の間には，何も撮影が行われていないカット区間があり，その区間を「ブランク (Blank)」と呼ぶ．このブランクは放送用には使用されない区間であるため，使用不能区間と位置付ける．したがって，素材映像を解析する最初の課題は，素材映像をカット点で分離し，CSec を抽出して，NCSec，Follow，Blank に分類することである．また，CSec において，Follow は使用可能区間に，また Blank は使用不能区間にそれぞれ分類される．

$$NCSec, Follow, Blank \in CSec$$

本研究では，カット区間内のカメラワークの変化量を用いてこの判定を行う手法を提案する．

5.4.2　使用可能・不能区間と手ぶれ

通常カット区間に関しては，カメラワークショット内の手ぶれや不安定な区間を対象に，より詳しい使用可能・不能区間推定が行われる．まず，放送用に使用不能となる区間の第一要素は「手ぶれ (Camera shake)」である．図 5.1 中，フェーズ 1 の FS1 は，フィックスショットのみを含む通常カット区間であり，FS2 は FS1 の取直し（リテイク）である．しかし，FS2 には両端に手ぶれ区間がある．一つのフィックスショット

として見れば，手ぶれを含んでいるため，このフィックスショットを使用不能区間とすることもできるが，図中 Clip1-3 の範囲を 2.5 秒取ることができれば，ぎりぎり使用可能なクリップの候補として採用できる．しかし，これがシーンの最後のクリップとなった場合は，手ぶれ区間に掛かり，2 秒増やすことができず，表 2.4 の Rule(3-2) に従えないため，候補からはずれることになる．ただし，この場合は，FS2 が FS1 のリテイクであるため，Clip1-1 や Clip1-2 で代替が可能である．これに対し，Clip5-2 は，ズームショットの直後のフィックスショットに手ぶれを含むため，クリップとしては成立しない．映像文法，表 2.2 の Rule(1-1) に従えば，このズームショットの直後に，１秒のフィックスショットが確保できないため，このズームショットとフィックスショットの手ぶれ区間が使用不能区間となる．しかし，リテイクがないために，手ぶれの程度が低い場合は，Clip5-2 を採用する場合も想定できる．この場合は，手ぶれの区間検出だけでなく，どの程度の手ぶれであるかを示し，使用可能なクリップとしてのスコアを提示するといった対処法も考えられる．あるいは，画像処理によって手ぶれを除去する処理も考えられる．

5.4.3　使用可能・不能区間と不安定な区間

図 5.1 中，フェーズ 1 にあるパンショット PS1 は，左パンから右パンに急激な変化を起こしていると見ることができる．この PS1 は，意図して撮影しようとしたが，カメラマンが気に入らなかったのか，カメラを引き戻したパンショットの失敗（Failure）区間である．その意味でこの PS1 は，表 2.2 の Rule(1-2)〜Rule(1-4) に従い使用不能区間となる．こうした放送用には使用できないカメラワークは通常，急峻かつ不安定な動きをしていたり，上下反転するなどの特徴を持つ．しかし，Clip7 の PS4 は，滑らかな動きをしているが，上下反転をしているにもかかわらず，使用可能なカメラワークである．これは皿をなめるように円運動する特殊なパンを行っている区間であり，使用可能区間である．また，この使用不能区間となる PS1 の前後にある FS6 の手ぶれ区間以外の部分区間，また FS7 は Type1 のクリップとして利用できる可能性があるため，使用可能区間である．

次に，カメラマンは，スイッチを入れたまま複数の冗長ショットを想定して撮影を行ったり，そのリテイクを撮影する場合がある．図 5.1 のフェーズ 1 中 *CSec3* は，その

様子を示した通常カット区間である．すると，放送用映像として意図して撮影した区間とそうではない区間が存在することになり，意図して撮影しようとした素材ショットどうしの間に，無造作に撮影した区間が存在する．これらを識別する特徴は，カメラワークの変化量に現れる．

例えば，ZS1 は，FS3 と FS4 の撮影の間で無造作にズームレンズを操作した区間である．また，ZS3 は TS である FS5 を取終え，MS である FS6 を撮影するために，ズームと撮影位置の調整を行った区間である．そして PS3 は，FS8 の撮影後，FS9 を撮影するために，カメラの方向を動かした区間である．いずれも，カメラの調整を行った区間であり，意味のないカメラワークである．したがって，使用不能区間となる．

これ以外にも，カメラ調整には，カメラの感度調整や照明の調整，合図と考えられるカメラマンの手が横切る区間なども存在する（カメラワークの解析結果としては「カメラ調整」と同じような特徴となる）．これらは，いずれも急峻な変化など，カメラワークの不安定性が特徴となるため，本論文では，不安定性を基準にカメラワークショット中の使用不能な区間を推定する手法を提案する．

5.4.4　ショット区間

以上より，素材映像上の使用可能・不能区間について述べたが，クリップは，映像編集システムのフェーズ 3 で時間長や位置の調整が行われる関係上，本論文では，フェーズ 1 でクリップを切り出さず，使用可能な区間と使用不能な区間のリストを素材映像から抽出し，メタ情報としてフェーズ 2，3 へ引き渡す手法を採用している．基本的には，カメラワークショットに使用不能な手ぶれ区間や，急峻（不安定）な区間が含まれる場合，カメラワークショットは，その使用不能区間で分断されて扱われる．その分断されたそれぞれの区間を本論文では「ショット区間 (SS:Shot Section)」と呼んでいる．ただし，パンショットやズームショットは，分断しても，クリップの条件，表 2.2 の映像文法 Rule(1-4) を満たすような区間を確保できないため，使用不能区間が含まれるパンショットやズームショットは分断することなく使用不能区間と位置付ける．

例えば，図 5.1 の Phase1 中，FS2 は，一つのフィックスショットであるが，二つの手ぶれ区間があるため，二つの使用不能区間と一つの使用可能区間がショット区間のリストとして表 5.1 のように生成される．

表 5.1　使用可能・使用不能区間のリスト

Start(frame)	End(frame)	CWS	Useful/Useless	Kind
528	539	-	useless	Hand shake
540	610	FS	useful	-
611	619	-	useless	Hand shake

フィックスショットは使用不能区間を除去しても，残った区間がフィックスショットとして採用できる場合があるため，これを新たなフィックスショットとして用いる．

これまでの定義を総合すると，確定ショットから素材映像までの包含関係は以下のようになる．

$$CS \subseteq RS \subseteq SS \subseteq CWS \subseteq CSec \subseteq RVM$$

ただし，CS，RS は，5.3.2 で定義したように，CS が「確定ショット」，RS が「冗長ショット」であり，SS は，5.4.4 で定義したように「ショット区間」のことである．また，CWS は「カメラワークショット」，CSec は「カット区間」，RVM は「素材映像」であった．以上より，本論文で抽出する「ショット区間」SS は，表 2.2 の映像文法 Rule(1-1)，また，Rule(1-2)〜Rule(1-4) に従って抽出されるカメラワークショットを基盤とする連続した区間であり，映像文法を背景として撮影された素材映像に対し適用することで抽出が可能となる．

5.5　使用可能・不能区間の推定

図 5.5 は，使用可能・不能区間を特定し，それを索引情報として生成するまでの処理過程を示している．A は，ブランクを検出して区間リストを生成する処理である．B は，カット点検出を行い，カット区間のリストを生成する．C はカメラワークの解析部である．D は，A と B，C の結果をもとに，ブランクを使用不能区間とし，それ以外のカット区間に対して，使用可能区間のフォローであるか，通常カット区間であるかを判定する処理である．E は，C の結果と，D で生成されたフォロー以外の通常カット区間リストを受け取り，通常カット区間ごとにカメラワークの変化量を用いて，ショット区間を抽出する処理である．このショット区間には，映像文法の立場から使用可能なカ

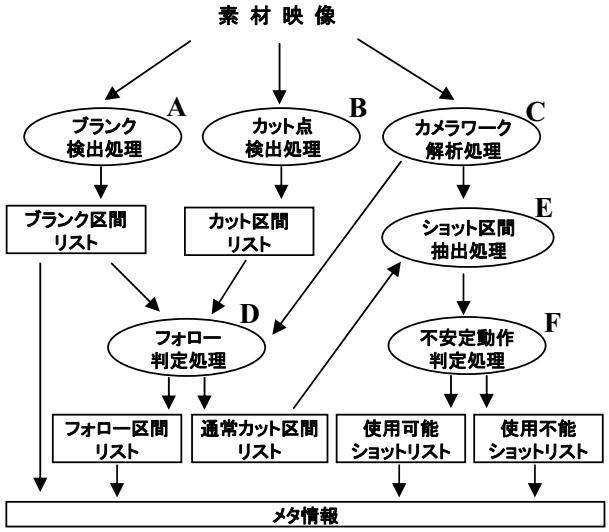


図 5.5　索引情報生成過程

メラワークショットの区間，使用不能な不安定なカメラワークショットの区間や手ぶれの区間などが含まれる．F は，E で得られたショット区間ごとに，手ぶれやカメラワークの不安定性を評価し，使用可能・不能なショット区間のリストを生成する．

5.5.1　素材映像に対するカット点検出

編集支援システムでは，素材映像が得られてから編集処理に至るまでの処理時間は短いことが求められる．カット点検出については，数多くの方法が報告されている [51] が，処理速度と精度の高さを考慮すれば方法論は限られてくる．我々は，フラッシュなどの瞬時的な変化に対応し，比較的高精度で実時間処理可能な方法として，ヒストグラムインタセクション [54] と，バッファリング法を併用したカット点検出法を提案している．バッファリング手法は，音声の研究領域で異なる話者の発話区間検出法として知られている GLR(Generalized Likelihood Ratio) [55] を簡易化した方法となっている．

カット点検出では，まず，$h'_{f,i}$ をフレーム f，クラス i のヒストグラムとして，これを求める．次に，正規化されたヒストグラム値 $h_{f,i}$ を，$h_{f,i} = h'_{f,i} / \sum_j h'_{f,j}$ $(i, j = 1 \cdots I : I = Q^3)$ として計算する．このとき，Q は R,G,B カラー空間のクラス数であり，実験では $Q = 8$ を用いているため，$I = 512$ となる．次に，フレーム a とフレーム b におけ

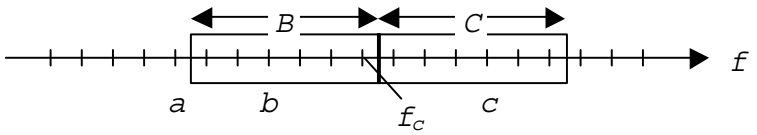


図 5.6　カット点 f_c の抽出

るヒストグラムインタセクション $HI(h_a, h_b)$ を式 (5.1) により求める．

$$HI(h_a, h_b) = \sum_{i=1}^{I} min(h_{a,i}, h_{b,i}) \tag{5.1}$$

図 5.6 は，バッファリング法によるカット点検出を図示したものである．まず，注目フレーム a に対して，幅 B と C を持つ二つの先読み区間を設定する．次に 2 つの先読み区間内に，任意のフレーム $b = a + m(m = 1 \cdots B)$ と $c = a + B + n(n = 1 \cdots C)$ を設定し，式 (5.2) を計算する．

$$Cut(a) = \min_b HI(h_a, h_b) - \max_c HI(h_a, h_c) \tag{5.2}$$

式 (5.2) の定式化は次の意味を持っている．

> フレーム a と b が同一区間に属し，フレーム a と c が，異なる区間に属している場合，フレーム a と b の類似度の最小値は，フレーム a と c の類似度の最大値をはるかに上まわっている．

従って，$Cut(a) > \theta_c > 0$ を満たすとき，カット点 f_c は $f_c = a + B$ として求められる．実験では，$\theta_c = 0.2$，$B, C = 6$ としている．

この手法を用いて，25 分の素材映像に適用した実験結果を表 5.2 に示す．基本的には，映像データから 1 フレームごとに RGB 画像が得られれば良いが，現時点では，カラー，352x240，フルフレームの MPEG-1 ファイルを作成し，フレーム精度で画像を復元できるツール [56] を用いて実験を行っている．

表 5.2 に示すとおり，再現率，適合率ともに高い値を得ることができている．ここでの「過剰検出」は素材映像特有のもので，カメラ調整による明るさの変化が二つ，カメラマンもしくはスタッフの手が横切った部分が三つである．この区間は使用不能区間検出で判定できるため，カット点検出としては間違いであるが，クリップの生成については問題とならない.「未検出」はカメラのスイッチが瞬時的に切れたようなもの

表 5.2　カット点検出の結果

正解検出	M	50
未検出	D	3
過剰検出	E	5
再現率 (%)	$M/(M+D)$	94.3%
適合率 (%)	$M/(M+E)$	90.9%

で，人が見ても見逃しやすいものが一つ，残り二つは同じ対象で，若干ショットサイズが異なるものであった．これはヒストグラムにあまり違いが得られなかったものであると推測できる．

5.5.2　素材映像に対するカメラワーク解析

映像編集支援システムでは，カット点検出と同様，カメラワークの解析にも速度が要求される．また，手ぶれ，カメラワークの安定性を問う場合，変化量についても精度が必要となる．これまで，カメラワークの解析法についてはいくつかの研究が行われている [38, 39, 57, 58]．我々は，処理速度とともに，手ぶれやカメラワークの安定性を充分に検証できる方法として投影法を用いた手法 [38] を採用している．

濃淡画像に変換された縦 h，横 w の画像フレーム f の位置 (i,j) にある画素値を $Gray(f,i,j)$ とすると，輝度投影量 P_X，P_Y は式 (5.3)，(5.4) のように定義できる．また，ブランク区間は投影量を用いて式 (5.5) により判定できる．

$$P_Y(f,i) = \frac{1}{h}\sum_{j=1}^{h} Gray(f,i,j) \tag{5.3}$$

$$P_X(f,j) = \frac{1}{w}\sum_{i=1}^{w} Gray(f,i,j) \tag{5.4}$$

$$Blank(f) = \frac{1}{w}\sum_{i=1}^{w} P_Y(f,i) < \theta_b. \tag{5.5}$$

ここで，2 フレーム間の縦横方向の投影距離は，それぞれ式 (5.6)，式 (5.7) として計算され，横方向のパン Pan_{lr} と縦方向のパン Pan_{ud} は，それぞれ式 (5.8)，式 (5.9) と

して定義できる．なお，δ_p は，$(\delta_p = -20, -19, \cdots, 19, 20)$ を用い，δ_p 中，最小距離が複数現れた場合は，前フレームと同符号を優先とし，また $|\delta_p|$ が最も小さいものを選択する．これは小さい動きを優先することに相当する．

$$D_{P_Y}(f, i, \delta_p) = |P_Y(f, i) - P_Y(f+1, i-\delta_p)| \tag{5.6}$$

$$D_{P_X}(f, j, \delta_p) = |P_X(f, j) - P_X(f+1, j-\delta_p)| \tag{5.7}$$

$$Pan_{lr}(f) = \arg\min_{\delta_p} \sum_{\substack{i=1+\delta_p \ (\delta_p \geq 0) \\ i=1 \ (\delta_p < 0)}}^{\substack{w \ (\delta_p \geq 0) \\ w-\delta_p \ (\delta_p < 0)}} D_{P_Y}(f, i, \delta_p) \tag{5.8}$$

$$Pan_{ud}(f) = \arg\min_{\delta_p} \sum_{\substack{j=1+\delta_p \ (\delta_p \geq 0) \\ j=1 \ (\delta_p < 0)}}^{\substack{h \ (\delta_p \geq 0) \\ h-\delta_p \ (\delta_p < 0)}} D_{P_X}(f, j, \delta_p) \tag{5.9}$$

ズーム量については，$\delta_z(\delta_z = -20, -19, \cdots, 19, 20)$ を変化させることにより，式 (5.10)，(5.11) によって得られる，拡大・縮小された投影量を用いる．

$$P_{ZY}(f, p, \delta_z) = P_Y(f, p\frac{w-2\delta_z}{w} + \delta_z) \tag{5.10}$$

$$P_{ZX}(f, q, \delta_z) = P_X(f, q\frac{h-2\delta_z}{h} + \delta_z) \tag{5.11}$$

放送用に撮影された映像のズームは非常にゆっくりしたものが多く，隣接するフレーム間ではズームが検出できない場合が多い．そこで，フレーム f から $|P_Y(f,p) - P_Y(f+n_c, p)| > \theta_C$ となる変化が検出された $f+n_c$ フレームを基準とし，$n_c \leq n \leq n_c + m (n \leq \theta_l)$ を範囲として，式 (5.10), (5.11) の距離を最も短くする δ_z 値によりズーム量を計算する．これを定式化したのが式 (5.12) である．ただし，n_c が θ_l を越える場合はズームはなかったものとして判定する．なお，$\theta_C = 140, m = 5, \theta_l = 15$ を用いる．

$$\begin{aligned}\hat{\delta}_z(f) = \arg\min_{\delta_z}\{\min_n(\min\{ & \\ & \sum_{\substack{p=1 \ (\delta_z \geq 0) \\ p=1+|\delta_z| \ (\delta_z < 0)}}^{\substack{w \ (\delta_z \geq 0) \\ w-|\delta_z| \ (\delta_z < 0)}} |P_{ZY}(f, p, \delta_z) - P_Y(f+n, p)|, \\ & \sum_{\substack{q=1 \ (\delta_z \geq 0) \\ q=1+|\delta_z| \ (\delta_z < 0)}}^{\substack{h \ (\delta_z \geq 0) \\ h-|\delta_z| \ (\delta_z < 0)}} |P_{ZX}(f, q, \delta_z) - P_X(f+n, q)|\})\}\end{aligned} \tag{5.12}$$

ただし，求められた$\hat{\delta}_z$は，フレームfからの距離が考慮されていない．例えば，同じ$\hat{\delta}_z$であっても，$f+n$のnが1の場合と10の場合では，後者の方がゆっくりとしたズームである．そこで、$\hat{\delta}'_z = \hat{\delta}_z/(1+\alpha n)$により，$f$から離れているほど値が小さくなるよう補正する．$\alpha$は，現在実験上最も良かった0.1を用いる．ズームの倍率は$(w - 2\hat{\delta}'_z(f))/w$にて計算できるが，拡大は1より大きい実数，また縮小は0より大きく1より小さい実数として計算されるため，正負の値を持つパン量と同列に比較ができない．そこで，パンと同列に扱うため，式(5.13)を用いて変換する．なお，パン量とのスケールの違いを合わせるため，$w_d = 600$を用いている．

$$Zoom(f) = w_d \cdot log(\frac{w - 2\hat{\delta}'_z(f)}{w}) \tag{5.13}$$

ブランクは式(5.5)に示したように投影量を用いて計算できるため，5.5のA，Cは同時に処理を行うことができる．

5.5.3　フォローの判定

映像文法に従った映像では，ショットの接続によって映像を表現することが主になるため，比較的フィックスが多い．また，プロが作成する映像では，カメラワークの多用を避け，ポイントとなるところにのみカメラワークを使用することが良いとされる．

従って，フォロー以外の区間では，カメラワークが検出される区間は全体から見れば一部である．これに対し，フォローは図5.7のように，比較的パンやズームが急峻かつ断続的に行われる場合が多い．特にカメラを担いだ撮影で，対象が大きく動いている区間では，特に顕著に現われる．そこで，カット区間集合中，任意のカット区間をcとして，断続性を表現する関数D'の値と，急峻な変化など，不安定性を表現する関数I'の値に基づき，通常カット区間とフォローが行われているカット区間の特徴を調べた．

断続性は，カット区間内にカメラワークがどの程度存在するかにより判定することができる．したがって，フレームkにおけるカメラワーク$(CW|Pan_{lr}, Pan_{ud}, Zoom)$の各変化量を$x_{CW}(k)$としたとき，カメラワークの存在の有無を表わす関数$f(x_{CW}(k))$

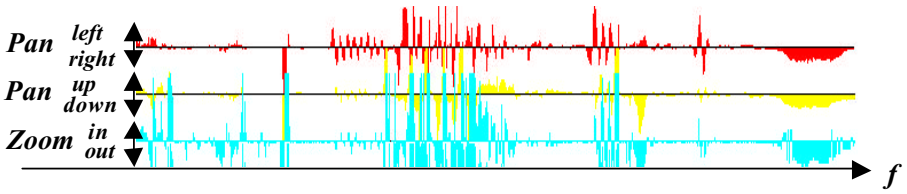


図 5.7　フォロー区間の特徴

を用いて $D(CW)$,$(CW|Pan_{lr}$，Pan_{ud}，$Zoom)$ を式 (5.15) のような密度によって表わすことができる．C は，カット区間 c 内の総フレーム数である．

$$f(x_{CW}(k)) = \begin{cases} 0 & (x_{CW}(k) = 0) \\ 1 & (x_{CW}(k) \neq 0) \end{cases} \tag{5.14}$$

$$D(x_{CW}) = \frac{1}{C}\sum_{k=1}^{C} f(x_{CW}(k)) \tag{5.15}$$

ただし，フォロー区間では，パンやズームすべてが検出されるため，実際には

$D'(Pan_{lr}, Pan_{ud}, Zoom) = \{D(Pan_{lr}) + D(Pan_{ud}) + D(Zoom)\}/3$

を用いる．

また，急峻な変化が断続的に起こっていることを検出するため，短い区間（窓）ごとの分散をカット区間内で積算し，カット区間長で正規化した式 (5.16) を不安定度の基本式とする．C は対象となるカット区間長，W は窓区間長を表し，$x_{CW}(j)$ はフレーム番号 j におけるカメラワーク $(CW|Pan_{lr}$，Pan_{ud}，$Zoom)$ の各変化量である．$\mu_{CW,i}$ は窓区間内の x_{CW} の平均値を表している．

$$I(CW) = \frac{1}{(C-W)W}\sum_{i=1}^{C-W}\sum_{j=i}^{W+i-1}(x_{CW}(j) - \mu_{CW,i})^2 \tag{5.16}$$

特に，フォローの区間では，図 5.7 のように，縦横方向のパンやズームが同時に検出されるため，横方向のパンを $Pan_{lr}(j)$，縦方向のパンを $Pan_{ud}(j)$，ズームを $Zoom(j)$ として，$I'(Pan_{lr}, Pan_{ud}, Zoom) = \{I(Pan_{lr}) + I(Pan_{ud}) + I(Zoom)\}/3$ を用いる．

図 5.8 は，表 5.3 の学習用素材映像から，フォロー区間と通常の区間を手動で抜き出し，D' を縦軸に，また I' を横軸に置いて値を描画したものである．

この図 5.8 から，D' のうち，フォローの場合はいずれも高い値を示していることがわかる．しかし，通常カット区間の D' 値は全体的に広がっており，フォローと通常カッ

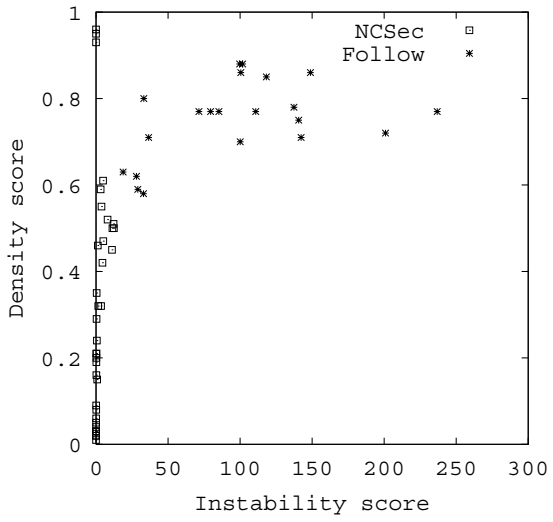


図 5.8　フォロー区間の特徴

表 5.3　実験環境

・計算機：	PentiumIII 800MHz, ノートパソコン
・映像：	毎日放送，あまからアベニュー素材映像 料理店名焼き肉はやし:14 分 46 秒, 料理店名焼き肉はなき:25 分 16 秒 (352x240,29.97frame/s)MPEG1,Quality:75%

トを分離する上では，有効とはいえない特徴であることがわかる．これに対し，I' の値に注目すると，通常カット区間とフォローを線形分離する境界があることがわかる．そこで，学習用素材映像から手動で抜き出した，学習用の通常カット区間群に対し，式 (5.16) の標準偏差 σ をあらかじめ求めておく．その後，判定すべきカット区間に対して式 (5.16) を計算し，図 5.8 に示すように，その値が平均 μ から $3 \cdot \sigma$ 内に入るものを通常カット区間，それ以外はフォローとして判定を行った．

5.5.4　ショット区間抽出処理

図 5.5 の E では，カメラワークの変化量を基盤として，通常カット区間内でまとまりのある区間を抽出し，ショット区間リストを生成する．正常もしくは急峻なカメラワー

クは，いずれも変化量が時系列上で連続している．しかし，手ぶれなどの区間は，カメラワークの変化量が不連続に存在している．両方の区間抽出が可能な方法として，ここではカメラワークの変化量が存在する 0 以外の値を持つフレームがあった場合，そのフレームを基点とし，30 フレーム先 (1 秒に相当) 先までを上限として，1 フレームずつカメラワーク変化量の有無を調べる．もし，存在すれば検出したそのフレームをまとまりのある区間に加え，検出したフレームを新たな基点として，30 フレーム先までを上限として調べる．これを繰り返し，次の 30 フレーム先までにカメラワーク変化量を検出できなければ，この時点で基点となっているフレームと最初の基点となったフレームまでをまとまりのある区間とし，これをショット区間と位置付ける．また，まとまりのある区間どうしの間にある，カメラワーク変化量が 0 の連続した区間は，フィックスショットであるため，これもショット区間と位置付ける．これにより，連続・不連続を問わずショット区間の抽出が可能となる．このショット区間検出は，パンの右左，上下，ズームの三つそれぞれ独立に行った．図 5.1 の白抜き矢印がその区間であり，素材映像上で重なる場合もある．このように，ショット区間リストには，正常なカメラワーク，急峻な動きをするカメラワーク，手ぶれが含まれることになる．

今回，ズーム検出については，ゆっくりとしたズームを検出できるように感度を上げているため，ズームに対して正負に振動する雑音成分が多く現れる．しかし，素材映像上の正常なズームは通常，変化量が一定であるため，ズームに関しては一定の値が続く区間についてのみ残し，変動が小刻みに激しい区間は除去した．また，パンの区間検出では，リスト中，１秒以内の連続した区間は手ぶれではなく雑音と考えこれも除去した．

5.5.5　使用可能・不能区間判定

図 5.5 の F では E で得られた区間リストに対し，使用可能区間と使用不能区間の識別を行う．使用不能区間については，手ぶれの区間と急峻な変化のある区間とを識別し，索引情報に付与する．これにより，手ぶれや急峻な変化の度合いが小さいと判定できる区間は，編集の段階で採用することも可能となる．そこで，この F では，手ぶれ区間，急峻な変化を伴なうカメラワークの区間，正常なカメラワーク区間を表 5.4 の特徴を用いて識別する．

表 5.4　カメラワーク判定の指標

カメラワーク	密度:D (Density)	不安定度:I (Instability)
手ぶれ	疎	–
急峻	密	大きい (短い区間内で)
安定	密	小さい (短い区間内で)

表 5.4 の密度とは，ある区間内にカメラワーク変化量が 0 以外の値を持つフレーム数がどの程度存在するかを表現している．手ぶれの場合，図 5.1 のフェーズ 1 中，FS2 にある手ぶれ区間の例のように，値が存在するフレームが少なく，それを疎であると表現している．これに対し，安定もしくは急峻なカメラワークショットは，値が連続して存在しているため，密であるとしている．この点から，手ぶれとそれ以外の安定もしくは急峻なカメラワークショットをこの密度により判別する．

また，安定なカメラワークショットは，隣り合うカメラワーク変化量の微分が小さく，短い区間内でその微分の総和が小さい値を示すものと考えられる．それに対し，急峻な区間では，微分が比較的大きく，短い区間内の微分の総和が大きくなると考えら得ることから，この特徴により，安定と急峻なカメラワークショットを識別する．

まず，手ぶれとそれ以外を識別するために，このショット区間に対して，式 (5.15) を適用した式 (5.17) を用いる．L は，ショット区間の区間長である．

$$D(CW) = \frac{1}{L}\sum_{k=1}^{L} f(x_{CW}(k)) \tag{5.17}$$

連続する 1 秒以内の微少区間は，区間検出の際に雑音として除外されているため，ショット区間で密度の小さなものは，手ぶれ区間である可能性が高い．そこで，$D(CW) < \theta_h$ を満たすショット区間は手ぶれ区間と判定される．

次に急峻な変化をする区間と正常な区間の識別については，式 (5.16) を区間リストに適用した式 (5.18) を用いる．

$$I(CW) = \frac{1}{(L-W)W}\sum_{i=1}^{L-W}\sum_{j=i}^{W+i-1}(x_{CW}(j) - \mu_{CW,i})^2 \tag{5.18}$$

この閾値に関しては，学習用の素材映像から急峻な変化をする区間と正常な区間を取り出し，それらの平均値で決定する．W については，なめらかな変化との違いをつ

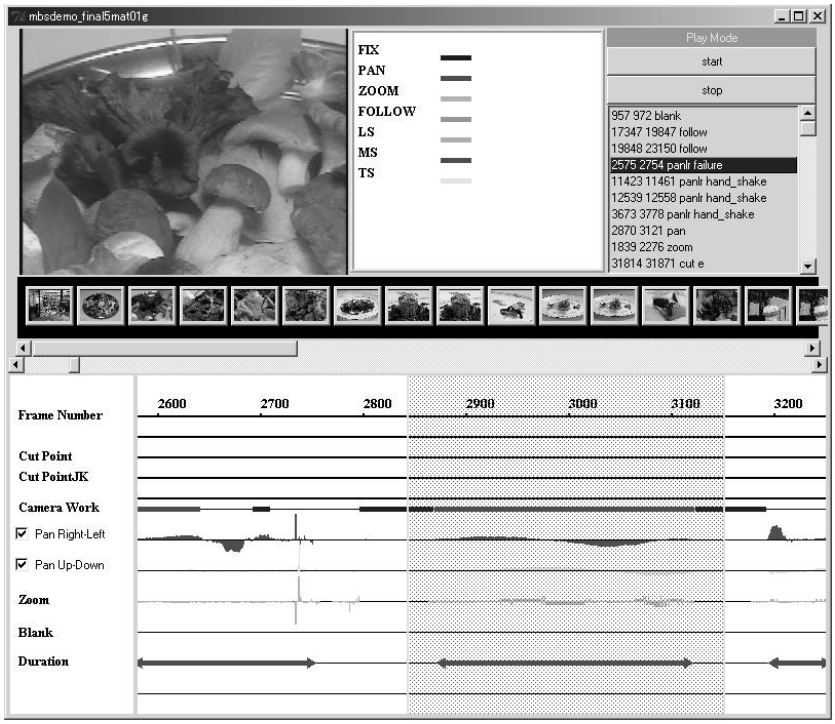


図 5.9　索引情報表示システム

ける上で，区間長が長すぎてはあまり意味がなく，短すぎても違いがでない．急峻な変化は 1 秒以内でも起こっているため，1 秒以内で予備実験を行った結果，10 程度が良かったことから，現在はこの値を用いている．

5.6　実験

使用可能・不能区間の判定処理が有効であるか否かを判断するため，プロのカメラマンが撮影した表 5.5 の素材映像 1 本に対し実験を行った．実験の評価に関しては，目視により作成した正解データと照合し，区間長の異なり，またずれが生じていても，一定量の重なりがあれば正解とした．カット区間の数はこの素材映像の場合 55 区間存在する．

MPEG-1 を用いているのは，速度を重視しているためである．また，カット点検出とカメラワーク解析の精度についても，MPEG-1 と比較して，解像度が高い MPEG-2 や MotionJPEG などと大きな違いがなかったためである．処理時間は PentiumIII800 を搭載したノートパソコン上でカット点検出，カメラワーク解析それぞれリアルタイ

表 5.5 実験環境

映像：	料理店名パスカルペニョ(352x240,29.97frame/s) MPEG-1,Quality:75%,25 分 27 秒,45763frame

表 5.6 使用可能・不能区間推定の実験結果.

	ブランク	カメラワーク				フィックス
		フォロー	手ぶれ	急峻	安定	
正解検出数	6	5	22	7	21	78
未検出数	0	0	5	1	2	8
過剰検出	0	0	14	0	2	5
再現率 (%)	100	100	81	88	91	91
適合率 (%)	100	100	61	100	91	94

ムで処理できる．また，使用可能・不能区間推定処理はカット点検出とカメラワーク解析の結果を用いて，全体で１分以内に終了する．

表 5.6 に実験の結果を示す．表 5.6 中，使用可能な区間はカット区間の「フォロー」，また「安定」に含まれるパンショットやズームショット，そして「フィックス」に含まれるフィックスショットなどのショット区間である．ただし，ショット区間としてのフィックスショットは，5.4.4 で述べたように，本来のフィックスショット中に使用不能区間がある場合，細分化されたものである．これらの再現率は「フォロー」が 100%,「安定」が 91%，そして未検出が二つあったものの「フィックス」は 91%であった．

また，使用不能区間は,「ブランク」,「手ぶれ」，カメラワークの「急峻」の３区間である．手ぶれの未検出区間が５つあるが，カメラワーク解析の精度よりも，まとまりのある区間の検出で失敗しているものが多く，改良の余地がある．また，区間の一致度については，前後 10 フレームの誤差を許せば，視察による正解とほとんど一致するが，ズームを行っている三つの区間では大幅に異なっていた．これは，ゆっくりとしたズーム区間の検出を重視すると，映像内で動くものがあるとき，それに過剰反応することで雑音が増え，区間検出を間違うことが原因である．

また，これらの結果は，図 5.9 に示す Tcl/Tk によって作成された索引情報表示システムで見ることができる．図 5.9 の中央から下は，横軸をフレームとし，図 5.1 のフェー

ズ 1 の様子と同じような表現法により，カット点を Cut Point の横方向ラインで，パンやズームを Pan right-left，Pan up-down，Zoom ラインで，また，ブランクは Blank ラインで，パンやズーム変化量のまとまりのある区間は Duration ラインで視覚的に確認することができる．右上のリストには，ショット区間が一覧されており，その区間をクリックすることで，対応する区間の映像，また各ラインが左方向にシフトしていくことで，同期して再生が行われる．

5.7　結言

映像の編集作業は，編集を行う前のショットの切り出し作業が多大な時間を浪費し，必ずしも人が介在する必要のない作業となる．この観点に着目し，本研究では，撮影の失敗や取り直しが含まれる素材映像を対象に，映像文法に従って使用不能区間を推定し，使用可能なショット区間の自動切り出しを可能とする，第三の問題「作業コスト・人材不足」の解決に貢献する手法を提案した．これを基に編集において重要なショットの区間を特定することが可能となった．この使用可能・不能区間推定法としては，映像文法に従い，カメラワークの速度量の変化を評価した．実験の結果，再現率の平均値で 91.8%，適合率で 95.0%を達成した．

第6章

相対的ショットサイズ自動付与による映像編集支援方式

6.1 緒言

西洋文法の基盤が品詞であり，文法記述には品詞がかかせないように，映像文法を基盤とする場合，相対的ショットサイズが最も基盤となる概念であり，最も基本的な索引情報である．相対的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．しかし，相対的ショットサイズは，相対的に決まる索引情報のため，画像から得られる特徴から直接的に索引情報を決定できない．

これらの判断は，プロのカメラマンが映像文法に従う編集を意識して撮影した素材映像について，編集者が行う．ただし，素材映像では，カメラのスイッチを OFF することなく，一つのカット区間内でのちに完パケで使う複数の異なるショットや，リテイクを撮影している．また，5 章に示したように，そのようなカット区間には，映像文法上，使用不能の区間も混在しており，使用可能な区間のみを選び出す作業は非常に労を要する作業である．つまり，編集者は，ほぼ連続する映像素材から，完パケに使用可能なショットを切り出すという作業が必要であり，さらに相対的関係から決まる相対的ショットサイズを素材映像のフィックス部ごとに判断しなければならない．

このように，素材映像が映像文法に従って撮影されているとしても，この労力を要する作業を自動化することが可能となれば，編集者の負担を減らすことができ，編集

図 6.1　素材映像の特徴

者は，ショットの接続作業という知的作業に集中することが可能となる．本研究では，映像文法に基づいて映像編集の作業を支援するシステムの実現を目的とし，映像編集支援システムの部分システムとして，映像文法に従って撮影した素材映像について相対的ショットサイズを自動的に付与する手法を提案する．

6.2　相対的ショットサイズ付与法の概要

6.2.1　素材映像の特徴

テレビのプロのカメラマンは，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 6.1 は素材映像の構造的な特徴を表わしている．

編集された映像では，カット点はショットを区切る転換点である．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．本章では，素材映像中のカット点で挟まれた区間をパート (Part) と呼ぶ．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，１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 6.1 のカメラ調整)．使用不能区間としては，この他にも実際には何も撮影されていないブランクの部分 (図 6.1 のブランク) や，撮影時に起こったカメラマンの手ぶれ (図 6.1 の手ぶれ)，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．従って，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．

6.2.2　編集作業とショット

編集作業は単純化すると，次の 2 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である．

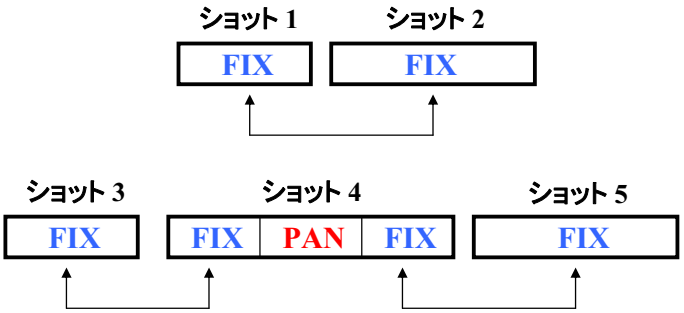


図 6.2　ショットの接続形態

図 6.2 は，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 6.2 のショット 1 とショット 2 は，双方が FIX ショットであるため，一般的なショットの接続例である．しかし，ショットのタイプにはカメラワークを含んだ Type2 のようなショットがある．この Type2 のショットの場合，単純にショットどうしを比較するのではなく，例えば，ショット 3 とショット 4 の接続では，ショット 4 のカメラワーク前部にある FIX 部が映像文法の接続規則の判定対象となり，ショット 4 とショット 5 の接続では，ショット 4 のカメラワーク後の FIX 部と，ショット 5 の FIX ショットの接続判定を行うことになる．つまり，素材映像上の FIX 部と，カメラワークの存在区間を把握することができる場合，素材映像の FIX 部に相対的ショットサイズを付与すれば良いことになる．

ただし，相対的ショットサイズは，古典的デクパージュを前提とした場合，もともとは同じ空間を連続しているように分節するという概念に基づいてカメラマンは撮影を行っている．その「連続」の印象を与えるため，古典的デクパージュに従うショットの接続では，基本的にショット内に存在するプロット（映像のフレーム枠内に存在し，内容伝達に関わるもの）が接続されるショットどうしで共有される．これは 2.2.4 の図 2.5 に示された古典的デクパージュの時空間分節に関する 15 の視点のうち，9 番目に相当し，空間的分節 (1) 空間連続性の保持に含まれる「包含関係」を示す規範的規則概念である．そのため，素材映像上の同一シーン内に含まれる映像区間の中で，特に FIX 区間どうしは，基本的に何らかの包含関係で結ばれる可能性が高い．

例えば，図 6.1 の素材映像は，一つのシーンに含まれる複数のカットから構成されており，そのカット中に存在する FIX 区間どうしは，包含関係を成す．図 6.1 におけるプロットとは，お皿のことである．ただし，無作為に選択した FIX 区間どうしが包含関係にあるとはいえない．また，図 6.1 のショット 3, テイク 1 は LS であり，他のすべてのショットを含む関係にあるが，ショット 2 のテイク 1 とショット 4 のテイク 1 は，お皿上，異なる場所を拡大しているため，ショット 2 のテイク 1 とショット 4 のテイク 1 どうしは包含関係にない．しかし，プロのカメラマンは，古典的デクパージュに偏る映像文法を背景として，同じ被写体について，少なくとも LS を撮影し，加えて，MS, TS を撮影するよう指導される．このため，LS が，被写体の全体像を撮影することから，同一シーン内の LS は，MS を含み，MS は，TS を含む関係がいずれかの FIX 部どうしで成立しており，直接的に包含関係がないショットがあるとしても，LS を親として，包含関係によるネットワークを形成している可能性がある．つまり，一つのシーンを特定できれば，そのシーン内の FIX 部についての包含関係を探索することで，ショットどうしの相対的な関係を決定できる．

6.2.3　相対的ショットサイズ索引付けの処理過程

相対的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．素材映像には，複数のシーンが撮影されていることがあるが，相対的ショットサイズは，同一シーン内の複数のショット間で相対的に決まるショットサイズであるため，素材映像内で，複数のシーンを構造化しておく必

要がある．ただし，シーンは，基本的に異なる場所を撮影するものであり，素材映像内に複数のシーンが含まれる場合でも，同一シーンは，時系列上混在もしくは交差することがほとんどなく，異なるシーンの転換位置の数も少なく，シーンを人手で構造化しても労力は少ないものと思われる．

本研究の関連研究には，シーンの色調が大きく異なる可能性が高いことから，色調の変化に着目してシーンの転換を自動検出する手法 [59] がある．正解率 92%，適合率 48%を達成しているが，過剰検出も多く，これらの研究の精度向上を行っても 100%の精度を常に得られる訳ではない．しかし，本研究の相対的ショットサイズ索引付けの処理では，シーンの構造化が正しく行われていない場合は，精度が大きく後退するため，シーンの構造化は正しく行われていることを前提としたい．もしシーンの数が少ないのであれば，100%の精度を持たないシーン転換点検出システムを用いずに，人手でシーンの構造を正しく決定し，逆にシーンが比較的多い場合でも，シーン転換点検出システムの検出結果を人手によって修正する方法を採用する．つまり，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するシステムを構築したり，逆に完全自動化システムの実現化を見送るのではなく，シーンの構造化で編集者の介在を許すことにより，現実的な編集支援システムを構築する立場を取る．

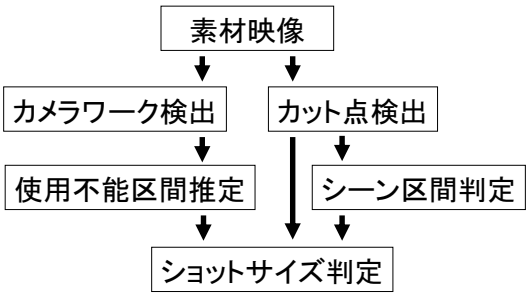


図 6.3　ショットサイズ索引付けシステム

前処理を含めた相対的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6.3に示す．相対的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必要がある．このカット点検出やカメラワーク検出，使用不能区間の推定手法については，5 章に示した手法を用いる．また

シーン転換点の解析手法は，文献 [59] を用いてシーン区間を判定する．ただし，本研究の実験は，これらの手法による前処理で索引情報が理想的に得られたことを想定し，相対的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相対的ショットサイズが決定できないような，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相対的ショットサイズの付与を行う．

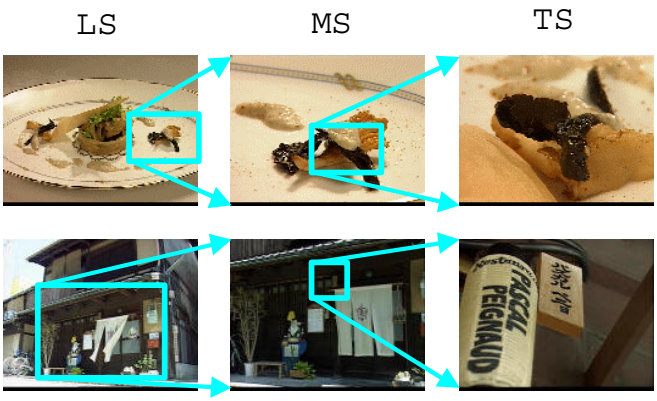


図 6.4　ショットサイズ

6.3　相対的ショットサイズの索引付け

6.3.1　相対的ショットサイズ付与における問題点

2.5.1 の図 2.6 に示したように，人については，絶対的ショットサイズが存在する [60]．絶対的ショットサイズは，対象が人であることが事前にわかっている場合，人を検出し，フレーム枠内に占めるその大きさを算出することによって，絶対的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，認識して大きさを算出する問題は画像認識の領域で極めて困難である課題とされている [61]．

例えば，図 6.4 の上にあるショットは，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 6.4 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 を見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相対的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，この相対的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．従って，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，相対的ショットサイズを付与する有効な方法論となる．

6.3.2　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，相対的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 6.5 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．

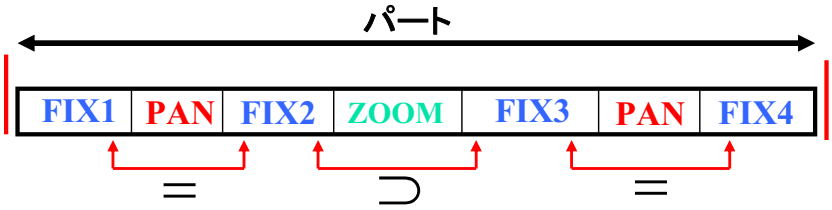


図 6.5　パート内でのショットの包含関係

この図で，PAN 前後では，ショットサイズが変化しないと言えるので，FIX1 と FIX2，FIX3 と FIX4 は同じショットサイズとなる．また ZOOM 前後では，ショットサイズが変化するので，FIX2 と FIX3 は包含関係にあることが判る．

6.3.3　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

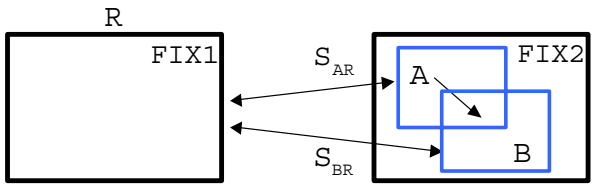


図 6.6　アクティブ探索

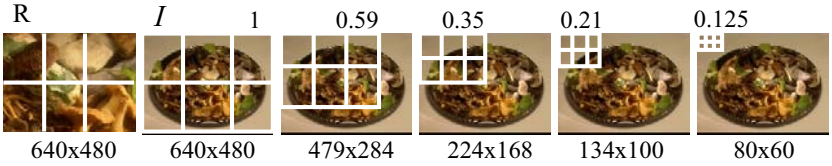


図 6.7　窓比に対応する画像サイズ

ト間では，カメラワークの情報を使わずに，ショットの FIX 部どうし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 [62] 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 6.6 は，アクティブ探索法の概念図である．この図 6.6 は，FIX2 の一部をズームアップして FIX1 が生成された場合を想定している．この場合，アクティブ探索法は，FIX1 を参照画像 R とし，これを縮小して探索対象画像 I である FIX2 上で最も類似度の高い部分を探索する．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，図 6.7 に示した $(1, 0.59, 0.35, 0.21, 0.125)$ の窓比で表される 5 種類とした．この 5 種類の窓比は，正解データの中から包含関係にある相対的ショットサイズの組み合わせを吟味し，ヒューリステックに決定した．アクティブ探索では，比較する二つのフレームを FIX1，FIX2 としたとき，FIX1 が FIX2 を含む，また FIX2 が FIX1 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閾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るとし，その比を包含の比率として類似度と共に記録する．

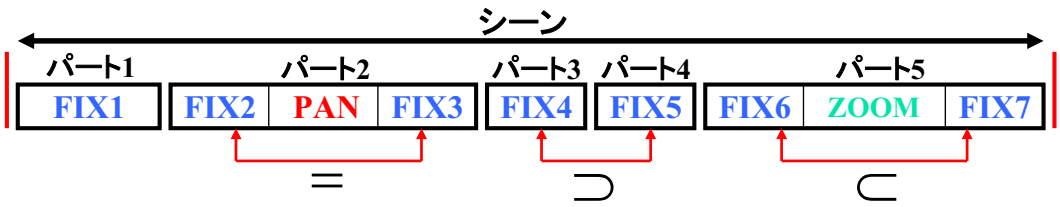


図 6.8 包含関係だけでショットサイズが決定できない例

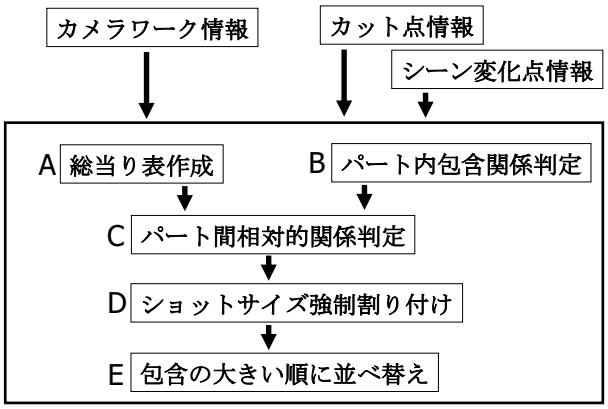


図 6.9 ショットサイズ自動付与部の処理過程

6.3.4 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 6.8 で，パート 1 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できない場合は，US(Unknown Shot) となってしまう．またパート 2 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．パート 3 とパート 4，またパート 5 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，隣り合うショットどうし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 6.9 に示す．

図 6.9 中，A のブロックでは，シーン内のすべての FIX 部どうしの包含関係を総当

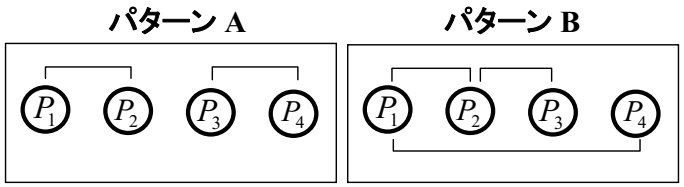


図 6.10　パート間の包含関係を判定する方法

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り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 6.10 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 番目のパートを表わしている．パターン B の場合，シーン内のすべての FIX が包含関係で結ばれ，ショットサイズを付与できる．しかし，パターン 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 どうしの総当りによる類似度計算を行う．その中で，最も高い類似度が閾値を越えていれば，その FIX 部どうし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．

図 6.10 の 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替える．E のブロックでは，包含の比率を 3 つに分割し，上位を LS，中位を MS，下位を TS とする．分割の閾値は，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 2.4 の Rule(3-1) を適用する．Rule(3-1) は，シーンに必ずマスターショット，つまり LS が 1 つは存在することを規定している．図 6.8 は 1 つのシーン内でのパートの関係を表わしており，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い FIX 部を LS とする．

6.4　ショットサイズ自動付与実験

表 6.1 に本研究の実験環境を示す.

表 6.1　実験環境

・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 料理店名パスカルペニョ 640x480,30frame/s Motion-JPEG,Quality:75% 25 分 27 秒,45763frame

表 6.2　ショットサイズ自動判定の結果

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	合計	総合	合計
	119	84	70.6	1

また，表 6.2 に実験結果を示す．表 6.2 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．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 6.2 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 6.2 中，Shot Size の未付与数は，6.3.4 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)%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共に高い値を持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 6.3 に示す．

表 6.3　窓比ごとの包含関係判定率

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

素材映像内には，ブランク区間など，同一シーン内のショットを含む区間に，そのシーンとは関係のない区間も存在する．これらの区間は，あるショットとの包含関係の判定において，関係のない区間として判定されるべきである．包含関係を判定する複数のショット間では，窓画像どうしの類似度を計算するが，この類似度の計算で，関係のない区間は，低い類似度となり，設定した閾値以下の類似度となることが正しい結果となる．しかし，この関係のない区間が閾値以上の類似度となる場合も想定できる．つまり，包含関係の判定には，同一シーン内のショットどうしを正しく判定した精度とともに，関係のない区間を関係のない区間として正しく判定する精度も考慮しなければならない．そこで，包含関係判定率は，包含関係の正解数と共に，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として，式 (6.1) のように定義した．

$$\text{包含関係判定率} = \frac{\text{包含関係の正解検出数} + \text{関係なし正解検出数}}{\text{包含関係・関係なしの正解総数}} \tag{6.1}$$

表 6.3 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場合の判定結果である．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 6.4 に図 6.9 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．

表 6.4 シーンごとの処理時間

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

この処理は，シーン内に含まれる FIX 部の包含関係を総当た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 2 つのフレーム間に対する処理を組み合わせ数分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，本手法による相対的ショットサイズ判定法は，ショットを切り出した上で判定を行っている訳ではない．もし，素材映像から使用可能なショットの区間を抽出し，その後に相対的ショットサイズを付与しようとした場合，画像としては，必ずしも包含関係にあるとは限らない場合があるため，画像処理を用いて相対的ショットサイズを決定することは非常に困難な課題となる．しかし，本手法により索引付けされた相対

的ショットサイズは，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ても，そのショットに含まれる FIX 部に関して相対的ショットサイズを調べるだけで良い．つまり，編集者は，切り出すショットの位置やショットの時間長を変更しても，相対的ショットサイズは再計算する必要はなく，ショットの切り出し位置を変更するなど，編集を再調整するような場合でも，柔軟な対応が行える編集支援システムが構築できる．

6.5　結言

本研究では，映像編集支援システムの一部として，相対的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絶対的なショットサイズを付与することができる．しかし，顔のようなパターンが比較的安定したプロット以外の一般的な物に関しては，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，画像の包含関係に基づく探索問題を適用することでこの問題を解決した．

映像文法では，相対的な関係を持つショットサイズに基づいてショットの接続を行う文法的規則があり，これに基づいて編集を行うことから，相対的ショットサイズの自動付与を行えば，編集者は知的な編集作業のみに集中することができ，第三の問題「作業コスト・人員不足」に貢献する．この観点から，本研究では，相対的ショットサイズの自動付与手法について提案した．

相対的ショットサイズの自動付与法としては，プロのカメラマンが撮影し，映像文法を背景として撮影された素材映像を対象とするが，カット点検出を行って得られる映像の区間は，編集上のショットに必ずしも対応しておらず，一つの断片に複数のショットに相当する部分が含まれる場合が多い．そこで，カメラワーク解析情報を用いて，FIX 区間の推定を行い，FIX 区間の代表フレームどうしの包含関係をアクティブ探索法を用いて判定し，その窓比から，相対的ショットサイズを決定する手法を採用した．この実験の結果，平均値で，70.6%の自動判定正解率を達成し，有効性を示した．

第7章

映像編集支援・自動編集方式

7.1 緒言

本研究では，本論文5章で示した，素材映像上の使用可能なショットの区間情報であるフィックスショットやカメラワークショットの区間情報，また，6章に示した，相対的ショットサイズの索引情報に基づき，映像編集支援システムの一つとして，映像文法に基づいて使用可能なショットを自動抽出し，映像文法の規則に従ったショット列候補を自動編集する方式を提案する．

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み合わ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する技法が先人により生み出された．その方法は，大衆が視覚的に，自然に理解できるように練り上げられた視覚的な普遍的規則の側面を持つ，映像の編集技法であり，その流れを汲む古典的デクパージュ（2章）に偏った規範的技法が本研究で扱う映像文法 [5] であった．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をす

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上で，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わかってきた[1]．そこで，本論文では，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．

7.2　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 [63] 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 [48] 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．

[1] 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．

Sundaram ら [64] は,「visual complexity」および「film syntax」という二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,「一つの会話は，最低三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，一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，属性値が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.

7.3　映像編集支援システム

7.3.1　映像編集支援システムで着目する映像文法

視覚的に見易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の長さで編集した場合，視聴者はどのショットが重要なのかを判断しづらく，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．本章で注目する映像文法を以下に抜粋する.

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，ズーム等は開始点と終了点を 1 秒以上フィックスさせて編集する．

例えば 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．

7.3.2　属性値

表 7.1　属性値の一覧

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

本論文 5 章，6 章の手法で素材映像に自動付与された索引情報，映像の文法的要素ならびに規則を考慮して，本研究で提案する映像編集支援システムの素材映像集に付与すべき属性値を検討した．その属性値を表 7.1 に示す．

属性 CutID, ShotID, CameraWork, StartFrame, EndFrame, ShotSize はそれぞれ 5 章と 6 章で述べた手法を用いて全てインデクシングされる．ただし，SceneID は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て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，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 7.1 中の例は図 7.1 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．例えば，図 7.1 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．

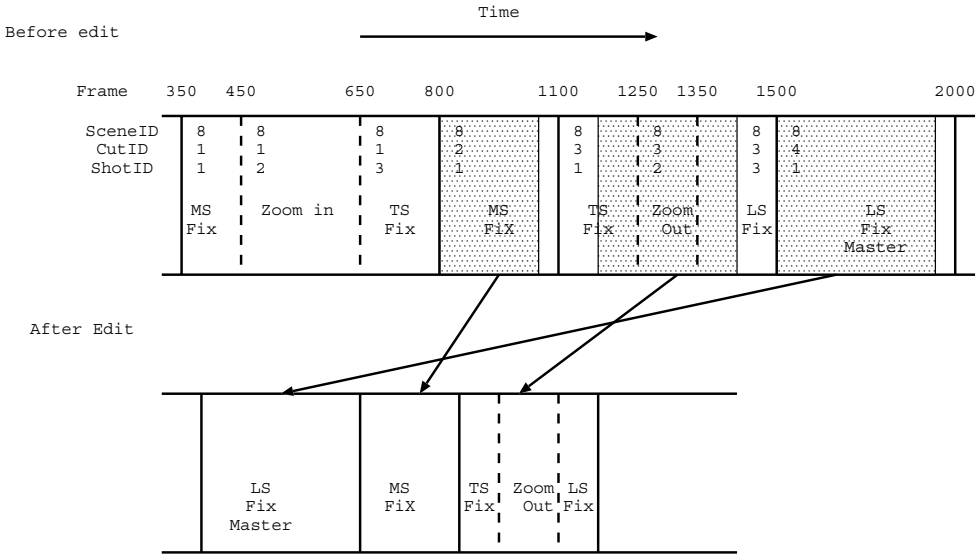


図 7.1 編集過程の概要

7.3.3 編集過程

図 7.1 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，規則 (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)）．ただし，MS は，もともと規則 (1) に基づいて，LS から TS，もしくは TS から LS への接続を可能にするために存在する中間的なショットである．また，シーンは，マスターショットとして LS から始めることが規則としてあるため，TS から LS へ接続するために MS を利用することよりも，LS から TS へ接続するために MS を用いることが優先される．このため，本研究では，MS からの接続候補に LS と TS が同時に存在する場合，MS から LS への接続より，MS から TS の接続がより高い優先度を持つので，TS が選ばれるようにしている．つまり，3 番目のショットは TS となる．

3 番目の TS の候補としては，CutID 1 で，MS からズームインした後の TS である ShotID 3 や，CutID 3 の最初の TS でありズームアウトの ShotID 2 が続く ShotID 1 がある．ここで，規則 (4) から導かれ，映像文法に従うショットの形式がこの選択を決定する．5.3.4 でも述べたように，映像文法に従い，素材映像から抽出可能なショット候補（クリップ）の形式には，フィックスショットのみの Type1 や，カメラワークの前後に 1 秒以上のフィックスショットを有する Type2 がある．この例では，CutID 1 の ShotID 3 は，TS の後にカットが存在するため，Type1 のショットとして抽出するしかない．一方，CutID 3 の ShotID 1 は，TS のみの Type1 クリップを抽出する方法以外に，CutID 3 の ShotID 2 のカメラワークショット, カメラワーク後のフィックスショットである ShotID 3 を含んだ Type2 のショットを抽出することも可能である．プロのカメラマンは，カメラワークを極力使用しないように指導される．そのプロのカメラマンが素材映像上に撮影したカメラワークを含む区間は，カメラマンが重視した区間である可能性が高い．本研究で使用する素材映像は，プロのカメラマンが撮影した映像であるため，本研究では，使用可能なカメラワークの区間があり，その前後に 1 秒以上のフィックスショットが存在する区間は優先的に選択することにした．このため，CutID 3, ShotID 1 から始まる Type2 の区間が選択される．この区間には，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，全てのシーンを編集する．

7.3.4　編集支援システムの構成

前向きプロダクションシステム

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 7.1 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 7.2 に示す．

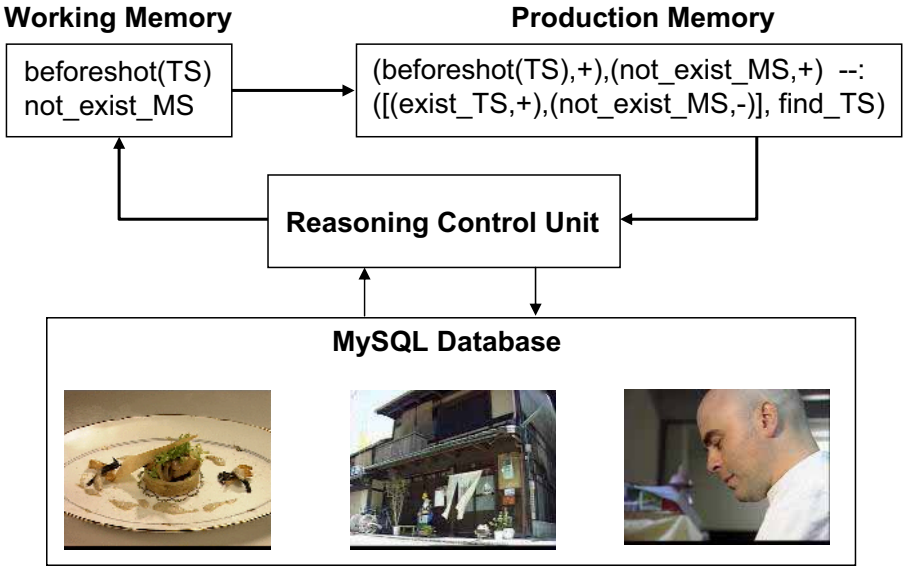


図 7.2　編集支援システムの概要

Production Memory はプロダクションルールとしての編集規則の集合，Working Memory は編集過程の各時点で定まる「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．

プロダクションルールの記述

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，7.3.1 で述べた編集規則に対応づけている．例えば，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」は，相対的ショットサイズが基本的に LS，MS，TS の関係で接続されることから，LS と TS がつなぐことができない接続関係に該当する．

この LS と TS を接続する上では，間に MS を挟むことが，滑らかで自然な接続を生み出すとするのが映像文法である．これは，古典的デクパージュの概念に偏るショット接続法となる．このため，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」等いくつかの細かいルールにわけられ，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)`

前向きプロダクションシステムでは，プロダクション規則を左辺から右辺に適用し,「認識」—「行動」のサイクルを実現する.「認識」は Working Memory で行われ，その「事実」に基づいて「行動」が行われる.「認識」とは Working Memory の状況 (condition) を問い合わせることであり，その「事実」に基づいて「行動」(action) が行われる．これらのプロダクションルールは，“if **condition** then **action**” が，以下のように表現されている．

condition --: action

condition

condition 部の例としては，`(beforeshot(LS),+)` がある．この condition の記述に現れる「＋」は，Working Memory 上に記憶された「事実」の存在状態として「存在」することを調べており，この例では，`beforeshot(LS)` という事実が存在すれば「真」，存在しなければ「偽」となる.「真」が成立すれば，action が「行動」として実行される．

action

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」を返した時にのみ，その「行動」が実行される．つまり，action 部は，次のように記述される．

(「行動」,「述語」)

((exist_MS,+),find_MS) の例では，find_MS が術語であり，(exist_MS,+) が「行動」である．例えば，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」として exist_MS が「行動」(exist_MS,+) として実行される．このとき,「行動」の記述に使用される「+」もしくは「-」は,「+」が Working Memory へ新たな「事実」を加えること，また「-」が Working Memory 内の操作対象となる「事実」を削除することを意味する．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．

このように，7.3.1 に示したシーンレベルの 4 つの編集規則を全てプロダクションルールで表わした場合，総数は 27 個となった．本研究では，この 27 個のルールを用いて自動編集を行う．

編集規則の競合時に用いる優先度

また，本システムでは編集規則に優先度をつけている．

一つは，相対的ショットサイズの接続に関する優先度である．例えば，TS の次には TS と MS が接続可能であるが，本研究では，TS から TS の接続より，TS から MS への接続の方に高い優先度を与えている．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 が連続してしまい相対的ショットサイズの変化がなくなり，単調な映像になってしまうからである．

もう一つはカメラワークに関する優先度である．図 7.1 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等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ーを用いて 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ーフェース部は Java で記述している．

7.4　映像文法による編集支援システムの実験

7.4.1　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 7.2 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．

表 7.2　実験対象

	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

7.4.2　素材映像の使用率による編集評価

素材映像は，一つ一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 7.3 に示す．編集に使用したショットの割合を Shot Rate，編集に使用したフレームの割合を Frame Rate 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 (7.1)，(7.2) で算出される．

$$Shot\ Rate(\%) = \frac{ES}{AS} \times 100 \tag{7.1}$$

$$Frame\ Rate(\%) = \frac{EF}{AF} \times 100 \tag{7.2}$$

表 7.3　カットとフレームにおける利用率

素材映像番号	Shot Rate (%)	Frame Rate (%)
素材映像 1	44.1	12.8
素材映像 2	72.5	11.9
素材映像 3	53.6	16.9
素材映像 4	41.0	15.3

表 7.3 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わかる．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 7.3 より，素材映像をかなり短縮できたことを示した．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．例えば，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．

7.4.3　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 (7.3) で表される．

$$Quality\ Rate(\%) \quad = \quad \frac{High}{High + Low} \times 100 \tag{7.3}$$

素材映像ごとの Quality Rate を算出した結果を 表 7.4 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等，システム異常が考えられる．しかし，表 7.4 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，見易い映像ができているといえる．

表 7.4　品質の評価値

素材映像番号	Quality Rate (%)
素材映像 1	68.2
素材映像 2	66.7
素材映像 3	78.6
素材映像 4	72.7

7.4.4　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，映像が保持する情報量として，画像の視覚的な複雑さをとりあげる．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．言い換えると，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 [64] 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．言い換えると，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 (7.4) で表される．

$$Compressibility(\%) \quad = \quad \frac{Compressed}{normal} \times 100 \tag{7.4}$$

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言える．LS, MS，TS ごとに平均をとったものを 表 7.5 に示す．カメラワークを含むものについては，カメラワークの全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 7.5　ショットサイズごとの圧縮率の平均

素材映像番号	LS (%)	MS (%)	TS (%)
素材映像 1	79.6	73.8	75.4
素材映像 2	93.4	80.4	74.6
素材映像 3	85.5	78.8	76.9
素材映像 4	88.7	88.0	91.6
total	87.2	83.4	82.6

表 7.5 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．

7.4.5　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．第一の指標は，編集に適したショットを選択したかどうかを示す．二番目に，1 つ 1 つのショットの継続時間の妥当性，三番目に内容の一貫性，四番目に映像の見易さ，五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．これらの五項目について，被験者から得た回答を表 7.6 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 7.6　被験者から得た回答結果

	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)
映像の見易さ	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)

表 7.6 より，mat1 では「ショットの選択」,「内容の一貫性」,「全体の長さ」が低いことがわかる．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き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 7.5 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」,「映像の見易さ」,「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 7.5 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，リズムのある見易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「映像の見易さ」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づ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，見易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．

7.5　結言

本研究では，5 章と 6 章の手法によって自動的に索引付けされた情報を用い，映像文法のショット接続ルールに従って自動的に接続可能なショット候補を自動選択・接続し，複数のショットから 1 つのシーンを生成する自動編集手法を提案した．自動編集システムは，映像文法をプロダクションルールとして記述し，前向きプロダクションシステムを用いて実現した．接続ルールを適用する場合，競合が起こった場合には，映像文法の概念から導かれる接続の優先度を用いて競合を解消する方法を示した．

本研究で用いる映像文法では，ショットの継ぎ目となる編集は意識されず，内容に没入可能な見やすい自然な接続が望ましく，各ショットの時間長によるリズムや接続順序を考慮した映像の見易さが評価の指標となる．主観評価実験によって，シーンの見易さが高い評価を得ており，有効性を示した．

映像の編集作業は，編集を行う前のショットの切り出し作業が多大な時間を浪費し，必ずしも人が介在する必要のない作業となる．この観点において，本研究は，作業コストや作業時間を軽減することから，第三の課題に貢献することになる．また，5 章と 6 章の手法と本研究で提案した手法の異なる具体化法としては，完全に自動編集とせず，編集者が選択したショットに対し，映像文法により接続可能な候補の一覧を提示する方法へ変更することも容易に可能である．この場合，編集者は，知的ではないショットの切り出し作業から解放され，編集作業にのみ集中することができる．このため，本研究の応用としては，コンテンツ産業を支援する映像編集支援技術として捉えることも可能である．

第8章

デジタルシューティングによる映像コンテンツ自動撮影方式

8.1 緒言

放送の多チャンネル化とディジタル技術の発展，またインタネットやブロードバンドの普及により，映像メディアに関して様々なサービスを展開できる基盤が整ってきた．しかし，放送局側では，コンテンツと人材の不足が原因となり，充分な放送コンテンツを用意することは困難である．この問題に対し，ディジタル技術を導入して，必ずしもプロのカメラマンや編集者を必要としない映像生成技術が注目されている．

プロの放送関係者を必要としない映像コンテンツとして，民間のスポーツの試合を配信するというサービスが考えられる．これまで，プロの選手が行うスポーツの試合は，収益が期待できる一般大衆を対象としてきたが，民間のスポーツ映像コンテンツに関しては，何らかの同好者や個人が潜在的なマーケットとして有力視されつつも，放送局が請け負うコストの面で割が合わず，放送コンテンツとして成り立たなかった映像コンテンツである．ここで，スポーツの試合に対し，映像生成の支援，さらには自動的に撮影・編集・配信を行うシステムが実現できれば，人件費や中継に必要となるコストを抑えることが可能となり，この問題に対応できると考えられる．

本研究では，最終的なシステム像として，内容の解説を可能とする同好者や個人向けの映像生成システムを目指している．その第一歩として，サッカーを対象とし，8.3で説明する自動撮影の一手法と考えられるディジタルカメラワークの実現法を提案する．

表 8.1　高画質映像の規格と解像度.

	規格	縦横比	解像度	画素数
1	SD D1:525i,D2:525p	9:16,3:4	720x480	約 35 万
2	HD D4:750p	9:16	1280x720	約 100 万
3	HD D3:1125i,D5:1125p	9:16	1920x1080	約 200 万
4	Next	9:16	3840x2160	約 800 万

8.2　サッカーに関する関連研究

8.2.1　サッカー映像コンテンツに関する問題点

スポーツ映像コンテンツとしては，全世界的に知名度の高い，サッカーが注目されている．しかし，サッカーに詳しくない素人にとっては，どこをどのように見るべきかがわかりにくいため，その面白さがわかりにくく，新たな視聴者やファンを獲得しにくいと言われている．特に，現状の放送映像では，スーパープレーに対する依存度が高く，また，玄人が注目するスーパープレーに至るまでのパスワークや，撮影フレームの外側で起こっている各選手どうしの駆け引きなどが，撮影範囲の限界によって必ずしも撮影されていないため，玄人にとっても見たい部分が見れない映像コンテンツとなっている．それゆえ，サッカーの内容を効果的に表現する手段が注目されている．

8.2.2　バーチャルスタジアム

ここで，表 8.1 のように，現在知られている映像の解像度に関する規格を示す．現在，最も普及している放送品質としては，表 8.1 の 1 に示した SD(Standard Definition) がある．これに対し，HD(High Definition) レベルの放送品質が普及しつつある．しかし，縦横比が 9:16 と撮影時の視野が広くなり，未撮影領域が減るものの，撮影できてない部分があることに代わりはなく，表 8.1 の 3 においても，画質が良くなるだけである．

以上のような現状に対し，サッカーの見せ方についてディジタル技術を導入したバーチャルスタジアムがある．ただし，バーチャルスタジアムについては，二つのアプローチが知られている．一つは，実際のスタジアムで観戦しているかのように見せる，巨大なパノラマスクリーンを用いた実写による 2 次元的システムである [65, 66]．2002 年の

図 8.1　3 台の HD カメラで実現されるサッカー映像の構図

FIFA ワールドカップにおいては，表 1 の 3 に相当する HD カメラ 3 台分の映像をシームレスに接続することで縦横比 9:48 の映像を実現し，通信衛生 (N-STAR) の Ka バンドを使用して，MPEG2,155Mbps による中継実験が行われた [66].

図 8.1 は，その Mega Vision システムによりサッカーコート全体を撮影した場合の構図である．現時点では，表 8.1 の 3 に示す解像度の HD カメラ 3 台分を統合することにより，5760x1080 の解像度を持つことになる．図中白枠は，SD(720x480 画素) 相当のフレームを示している．また，表 8.1 の 4 に示す次世代の高解像度カメラを三つ用いた場合，11520x2160 の解像度を持ち，SD の枠は黒枠相当になる．

二つ目は，仮想の視点を 3 次元復元し，自由な視点から観戦することができる，バーチャルカメラ [67] を用いた 3 次元的システムである [68, 69]．これらは，サッカー映像の見せ方について新たな基盤を提供するが，サッカーの試合の内容を効果的に提示する問題は，別の問題である．

8.2.3　自動撮影ロボット

サッカーコート全体を固定で撮影するのではなく，ある部分に着目して撮影する意味としては，視聴者に対し，どこを見るべきか提案していることに相当するものと考えられる．つまりカメラマンは，サッカーの試合の見方をガイドしているのである．

この撮影法では，撮影できない領域が存在するものの，こうしたプロのカメラマンの代替となる知的ロボット撮影システムが研究されている [70]．このようなシステムでは，カメラマンを模倣することが一つの目標となっており [71]，生成される映像は，カメラマンが撮影する映像と基本的に同じものとなる．そして，一度撮影が行われると，そのカメラワークを後で変更することが困難であるという問題がある．また，撮影環境としては，ライブ映像の撮影となるため，基本的に取り直しがきかない．

そこで，カメラマンには，試合の進行を予測しながら撮影するなど，人間にとっても失敗を伴う高度な撮影技術が必要であり，画像処理とカメラ駆動を含め，実時間以上の処理や予測技術が要求される．そのため，現在の画像処理技術では，人間と同程度のシステムを構築することは難易度が高い問題である．

これに対し，数多くのカメラを少しずつ離れた位置に配置し，イベントごとに最適な角度からの映像を提示するという方法も提案されている [72]．しかし，装置が複雑かつ大規模になる傾向があり，試合が行われる場所によっては設置が困難であることも考えられる．

8.2.4　放送映像の 2 次利用

映像の生成に関する研究としては，プロのカメラマンが撮影し，スイッチャーがカメラを切り替えることにより配信される映像を素材として，2 次利用する研究がある．例えば，構造化を目的として，ショットを分類したり [73, 74]，イベントの特定につながるゴールポスト [75] やライン [76]，ボール [77] の検出，選手の追跡 [78]，イベントを検出してメタ情報を付与し，内容検索を可能とすること [79]，また，イベントに着目し，ハイライトなどを自動抽出するもの [80, 81] や，要約映像を生成するもの [82, 83] などがある．このような 2 次利用においても，同好者もしくは個人向けに好きな映像区間を提示するなど，嗜好に対応した映像の生成手法は考えられる．しかし，これらのアプローチでは，プロのカメラマンが撮影した映像を加工できる範囲でしか処理できないという制約があり，自由度が少ない．

8.2.5　ディジタルシューティング

これに対し，ディジタルシューティングという自動撮影法が考えられる．例えば，表 8.1 の 3 によれば，HD は最大で SD の 6 個分の解像度を有し，次世代の 800 万画素品質では，SD の 24 個分の解像度を有することがわかる．このとき，画面の一部をクリッピングし，SD レベルの品質にマッピングすると映像が生成できる．つまり，クリッピングする方法によって，パン (クリッピングサイズを固定し，位置を連続的に上下左右へ移動)，ズーム (クリッピングサイズの連続的な変更)，カット (クリッピングの位置とサイズを離散的に大きく変更) を表現できる．

このようなディジタル処理によるパンやズームは，ディジタルカメラワーク [84] と呼ばれ，一つのショットを形成する．また，ショットどうしを接合するとき，クリッピングのサイズや位置が大きく変化しているときは，カメラをスイッチングしたことに相当し，これをディジタルスイッチングと呼んでいる．ディジタルシューティングとは，このディジタルカメラワークとディジタルスイッチングを用いて映像を生成することを言う．

8.2.6　ディジタルシューティングの利点

このディジタルシューティングによる映像生成では，一つの素材映像から任意のカメラワークやショットの接続によって映像を幾通りも生成することができる (One Source Multi-Production System)．つまり，テレビの撮影法を模倣できるだけでなく，取り直しに相当する処理が何度でも行え，また同好者や個人向けに嗜好を考慮した視聴者ごとに異なる映像を生成するなど，多くの可能性を秘めており，テレビ映像の 2 次利用よりは自由度が高く，効果的な映像提示法としての可能性が期待できる．また，選手個人の動きに注目し，選手の指導用映像を生成することや，素人または玄人向けの映像生成，ルールを解説するための映像を生成することなど，応用範囲は広いと考えられる．また，ディジタルシューティングでは，高い位置からの俯瞰映像を必要とするが，カメラは固定で良く複雑な制御は必要としない．サッカーコートがスタジアムでない場合でも，俯瞰映像に必要な高さの塔を設置できれば良い．また，バーチャルスタジアムのパノラマ映像をそのまま映像素材とすることもできる．本研究では，ディジタルシューティングのうち，ディジタルカメラワークの実現法を提案する．

8.3　サッカーに対するディジタルカメラワーク

放送用のサッカー映像の撮影では，ライブ性のため，取り直しがきかないことから，ボールの進行方向に大きな空間を持たせて急激な変化に対応するなどの撮影技術を用いている [71]．しかし，この点については，ディジタル化された映像に先読み処理を導入することで，遅延は発生するものの，構図的に最適なクリッピング位置を決定できる可能性がある．

また，サッカー撮影に熟練したカメラマンは，サッカーの試合進行に関係する情報を把握して撮影する高度な撮影技法を身につけている．例えば，できるかぎりアップで選手を撮影することを目標とした場合，ドリブルの得意な選手に対しては，パスなどの急激な動きがないと判断してアップで撮影し，あまりドリブルをせず，すぐにシュートやパスを行う選手に対してはロングで撮影する．これはアップで撮影した際，ボールが大きく移動すると，追い切れなかったり，追従に専念することで映像が見にくくなることを避けるなど，不要な意図を発生させない映像文法 [47] に極力従おうとするためであると考えられる．これ以外にも，パスが行われた際，ボールをフレーム内に収めつつも，パスを受け取ると思われる選手が中心となるよう，いち早くカメラをパンし，選手の移動方向の追従に専念した撮影法が用いられている．このような撮影技法は，選手個人の特徴に関する情報や高度な判断力が求められる．0 本来ならば，アップとロングを切り替えるタイミングも，試合進行を適切に見せる上で重要な問題であるが，このディジタルスイッチングを具体化する技術については，今後のディジタルシューティングを応用する研究で取り扱う．

ディジタルカメラワークでは，試合進行の何に着目するかによって，異なるフレームの軌跡やフレームサイズ，ズームのかけ方を変更することが可能である点が特徴であった．本研究では，まず，テレビの撮影法を模倣する立場から，その要素技術として，サッカーの撮影で基本情報となる，選手の動き，ボールの動きという二つの観点によるディジタルカメラワークの実現法に着目する．ただしサッカーでは，基本的にズームは撮影フレームのサイズ調整程度に使われるだけで，パンの制御法がディジタルカメラワークの中心課題である．そこで本稿では，ディジタルカメラワークのうち，選手の動き，またボールの動きに着目したパンの制御を提案する．この選手の動きとボールの動き，それぞれによる生成映像の違いを明確にするため，フレームサイズは，テレビ映像のフレームサイズと同程度の大きさで固定とした．8.4.1 の撮影条件下では，およそ 320x240 となる．

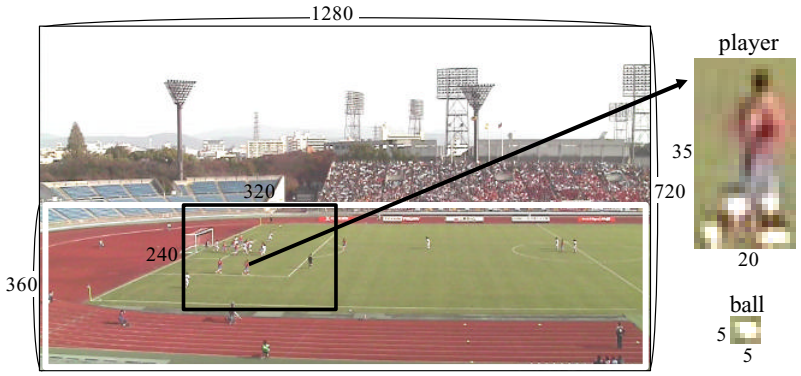


図 8.2　取得画像，選手，ボールのサイズ

図 8.3　京都西京極競技場での撮影位置

8.4　選手に着目したフレーム位置制御法

8.4.1　撮影環境

図 8.2 は，図 8.3 中，×印に示した京都西京極競技場のロイヤルボックス横の位置から撮影した映像である．ロイヤルボックスは，テレビ関係者の撮影に用いられるが，横の位置においても，ほぼテレビと同等の高さから俯瞰の撮影を行うことができる．HD カメラとしては，表 8.1 に示す 2 の解像度で撮影ができる Victor の HD カメラ (HD-GR1) を用いた．

本研究では，ディジタルカメラワークを実現する上で，できるかぎり高解像度を維持しつつ，サッカーコートの広い範囲の撮影を目標とし，入手することができた 1 台のカメラを最大限に活用する方法として，半コートを撮影することにした．この撮影環境では，図 8.2 の右に示したように，選手はおよそ 35x20 画素，ボールは 5x5 画素となる．ディジタルカメラワークの処理対象領域としては，図 8.2 の上部が意味のない領

図 8.4　背景画像

域となるため，図 8.2 の白枠に相当する領域 (360x1280) のみを処理対象領域とする．

8.4.2　高解像度映像に対する背景差分

ディジタルシューティングでは，基本的に素材となる高解像度映像をカメラ固定で撮影する．本研究では，半コートの撮影によって実験を行うが，図 8.1 のような映像素材は技術的に取得可能である [66]．このような映像素材に対し，背景差分法を用いれば，試合進行上，動いている選手やボールの部分だけを差分として取り出すことができる．本研究では，処理を開始する直前の映像区間を対象とし，420 frame(14 秒) 内の 4 frame ごとに取得した画像から，出現頻度の高い輝度値を背景画素の輝度値とする手法により，図 8.4 のような背景画像を生成した．背景差分法は，サッカーコートの色に依存しないため，高所からの俯瞰映像さえ撮影できれば，民間の試合で用いられる土のグランドに対しても同じ処理が可能である．また，背景情報をある程度の時間で更新しておけば，日照の変化などにも対応できる．

8.4.3　選手に着目した追跡

テレビと同じ撮影技法としては，試合進行の中心を撮影することが考えられる．背景差分映像では，選手やボールが差分情報として得られるが，試合進行中は，サッカーコート全体で選手が動いているため，差分情報の存在位置だけでは撮影位置を決定で

図 8.5　移動情報画像

きない．しかし，試合進行の中心となるボール近辺では，選手が激しく動き密集する傾向にあることから，サッカーコート中，一定時間内で選手群が大きく動いている場所を撮影位置として決定できる．図 8.5 の下部は，図 8.4 に示す背景画像と処理対象区間の映像（図 8.5 の上部）との背景差分画像について，時間方向に 30 frame を対象として作成した移動情報画像である．その領域は，ちょうど図 8.2 の frame を基準位置とした白枠に対応している．この例では，ゴール前にボールがあり，その周辺で選手が密集している様子が示されている．図 8.5 から，ゴール前の選手が密集している位置で差分により得られた点が密集していることがわかる．まず選手ごとの領域を抽出するために，2 値化された背景差分映像の各画像フレーム (以後 frame とする) に対して膨張・縮小処理を適用し，孤立点除去を行った後，時間的に連続する領域を動オブジェクトと定義する．

8.4.4　移動情報によるフレーム位置の決定

2 値化された k 番目の背景画像 f^k の中に存在する N 個の各動オブジェクトに

$O_k^1, O_k^2, O_k^3, \cdots, O_k^n, \cdots, O_k^{N-1}, O_k^N$

とラベルを付与する．選手の情報を座標単位で扱うため，(8.1) 式を用いてラベル番号毎に動オブジェクトの重心を求め，選手の座標 $\mathbf{P}^{(O_k^n)}(P_x^{(O_k^n)}, P_y^{(O_k^n)})$ とする．

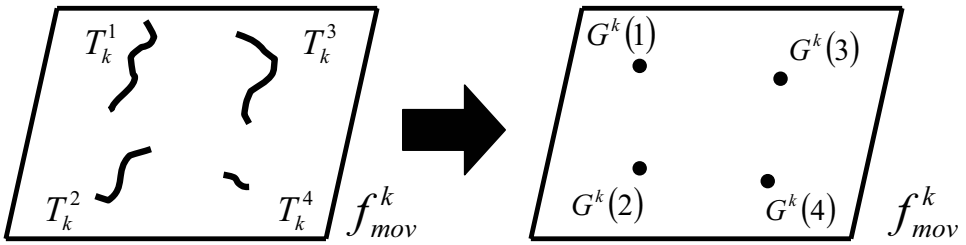


図 8.6　移動情報から重心を取りだす処理

$$P_x^{(O_k^n)} = \frac{\sum_{x \in O_k^n} x}{\sum_{x \in O_k^n} 1} \qquad P_y^{(O_k^n)} = \frac{\sum_{y \in O_k^n} y}{\sum_{y \in O_k^n} 1} \tag{8.1}$$

選手の移動情報を求めるために，f^k と $f^{k+l}(k \neq k+l : l = 1, 2, \cdots, L : L = 59)$ の 2 枚の 2 値画像において選手の座標を比較し，式 (8.2) に従って移動情報画像 f_{mov}^k に記述する．f_{mov}^k はあらかじめすべての画素の値を 0 に設定している．

$$f_{mov}^k(\mathbf{P}^{(O_{k+l}^n)}) = \begin{cases} 255 & (\mathbf{P}^{(O_k^n)} \neq \mathbf{P}^{(O_{k+l}^n)}) \\ 0 & (\mathbf{P}^{(O_k^n)} = \mathbf{P}^{(O_{k+l}^n)}) \end{cases} \tag{8.2}$$

f_{mov}^k では，オブジェクト m が移動しているときに点が書き込まれることになり，移動を続けると軌跡 T_k^n が得られる．図 8.6 は，図 8.5 を簡略化し，動オブジェクトの軌跡から重心を求める過程を示したものである．ただし，本手法では，個々の選手のラベルを軌跡に対応させる必要はない．複数の選手の動きが激しい領域を探索することが目的であるため，選手の軌跡が分断されたり，k 番目と $k+1$ 番目の frame で異なる選手の移動を同一選手の移動であると認識したとしても，その場所が移動の多い領域であることには相違ない．このため，同一選手の判定は行わずとも，移動の多い領域として探索することができる．

各軌跡へラベルを付与する際には，膨張処理を行い，軌跡をなめらかにすることでラベルを付けやすい状態にしておく．そのラベル毎に重心 $G^k(n) = (G_x^k(n), G_y^k(n))$ を求め，同一ラベル n の領域に含まれる画素の数 s_n だけ重心 $\mathbf{G}^k(n)$ に重み $W_k(n) = s_n$ を持たせる (図 8.6)．同一時間での軌跡であるため，曲線が長いほど大きな動きをしたと考えられることから $W_k(n)$ が大きいほど移動速度が大きくなる．

この状態で f^k_{mov} には k 番目から $k+L$ 番目の frame における選手の移動情報が記述されている．また映像の切り出し位置は最も移動の大きい場所と設定している．ここでは，f^k_{mov} における最も移動の大きい点を重心の密集している点と近似し，$G^k_{crowd}(n_k)$ として求めている．この点は重心の一つであり，n_k は (8.4) 式を満たす n である．これは，ある軌跡の重心であり，他の全ての軌跡の重心が最も密集している重心である．(8.4) 式は W'_k により軌跡間の距離が近くて軌跡の長さが長いほど値が大きくなるものである．

$$W'_k(n) = \frac{W_k(n)}{W_{k,total}},\quad W_{k,total} = \sum_{n\in N} W_k(n) \tag{8.3}$$

$$D(n,s) = \frac{W'_k(s)}{\{G^k_x(n) - G^k_x(s)\}^2 + \{G^k_y(n) - G^k_y(s)\}^2}$$

$$n_k = \arg\max_n \sum_{s\in N, s\neq n} D(n,s) \tag{8.4}$$

これにより映像のフレーム位置は (8.5) 式のように求められる．

$$G^k_{crowd}(n_k) = G^k(n_k) = (G^k_x(n_k), G^k_y(n_k)) \tag{8.5}$$

8.4.5　線形回帰分析によるフレームワーク

次に，$M+1$ frame 間の $\mathbf{G}^k_{crowd}(n_k)$ の移動情報からカメラワークを決定する．ただし，$\mathbf{G}^k_{crowd}(n_k)$ の軌跡は小刻みに動くため，そのままカメラワークの動きに反映させると映像が見にくいものとなる．そこで，複数 frame 間の $\mathbf{G}^k_{crowd}(n_k)$ を線形回帰直線で近似することで，自然なカメラワークを実現する．$\mathbf{G}^k_{crowd}(n_k)$ を k から $(k+M)$ frame の $M+1$ frame 間で解析し，座標を $(G_x^{(k+m)}, G_y^{(k+m)})(m = 0, 1, 2, \cdots, M : M = 59)$ とし，(8.6)(8.7) 式を用いて線形回帰直線 $y = \alpha x + \beta$ を決定する．

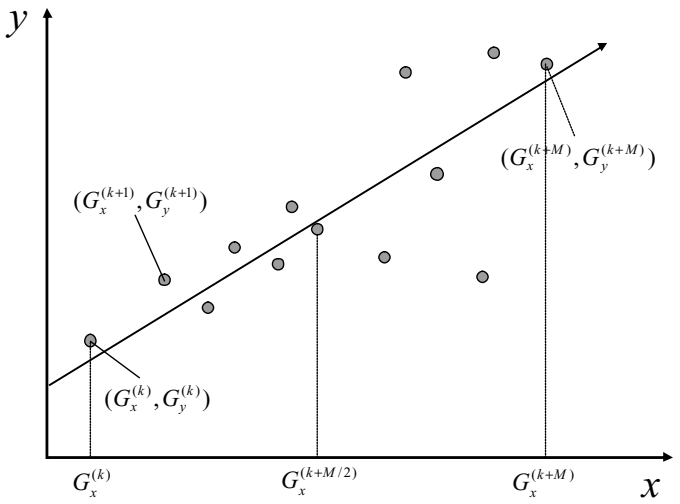


図 8.7　パンの軌跡を計算するための線形回帰直線

$$\begin{aligned}\bar{G}_x(n_k) &= \frac{1}{M+1}\sum_{m=0}^{M} G_x^{(k+m)}(n_k)\\ \bar{G}_y(n_k) &= \frac{1}{M+1}\sum_{m=0}^{M} G_y^{(k+m)}(n_k)\end{aligned} \tag{8.6}$$

$$\begin{aligned}\alpha &= \frac{\sum_{l=0}^{M}(G_x^{(k+l)}(n_k)-\bar{G}_x(n_k))(G_y^{(k+l)}(n_k)-\bar{G}_y(n_k))}{\sum_{l=0}^{M}(G_x^{(k+l)}(n_k)-\bar{G}_x(n_k))^2}\\ \beta &= \bar{G}_y(n_k)-\alpha\bar{G}_x(n_k)\end{aligned} \tag{8.7}$$

この線形回帰直線は，$(G_x^{(k+m)}, G_y^{(k+m)})$ を 2 次元平面上に表したとき，図 8.7 のような直線を示す．これは $M+1$ frame 間で移動の大きな点を逃さずに撮影できる直線である．ただし $(m = 0, 1, 2, \cdots, M)$ とする．

この回帰直線に沿って切り出し部分を移動させる．$M+1$ frame ごとに独立してカメラワークを決定すると，前後の関係を無視してしまうためカメラワークが不自然になってしまう．そこで f^k から f^{k+M} までの処理を行い, f^k から $f^{k+M/2}$ までのカメラワークを決定する．次に同様にして $f^{k+M/2+1}$ から $f^{k+M/2+M}$ までの処理を行い，$f^{k+M/2+1}$ から f^{k+M} までのカメラワークを決定する．このように常に処理 frame を重ねることで,

次の frame を考慮したカメラワークとなるため，自然な動きを実現することができる．また，1 frame 目を始点とし，$M/2+1$ frame 目を終点とした．この間は $M/2+1$ frame であるので，1 frame あたり $(G_x^k - G_x^{k+M/2})/(M/2+1)$，$(G_y^k - G_y^{k+M/2})/(M/2+1)$ で x と y を変化させてカメラワークを決定した．

8.5　ボール情報を用いたデジタルカメラワーク

映像文法に基づいて，サッカーの撮影技法を考慮した場合，試合進行を視聴者へ正確に伝えるようフレームサイズと撮影位置を決定し，余計な意図を発生させないよう滑らかにこの試合進行を追従する必要がある．しかし，ボールは，ドリブルやパスでは小刻みに激しく動き，ロングパスやシュートでは大きく動く．このように，サッカー映像でのディジタルカメラワークでは，小刻みな動きについてはボールがフレームからはずれない程度に動きを押さえ，ボールが大きく動く時には素早く追従する手法が求められる．

そこで本章では，これらの問題の解決手法を提案する．ただし今回入手した素材映像は，表 8.1 の 2 レベルの映像であり，図 8.2 の右下へ示したように，ボールが非常に小さく，ボール検出が困難であった．そこで，今回の実験では，ボールに着目したディジタルカメラワークの有効性にのみ焦点を当てるため，ボール情報は事前に手動で付与した．

8.5.1　小刻みなボールの移動に反応しないカメラワーク

小刻みなボールの動きにフレームが影響を受けないように図 8.8 のように白枠の切り出しフレームだけでなく黒の内枠を導入する．

図 8.8　フレーム (白枠) と内枠 (黒枠)

図 8.9 の t_1 から t_2 の時点でボールの座標がこの内枠の外に出るとき，フレームの中心を図 8.9 の t_3 のように，そのボールの座標まで移動させる．逆に内枠の中にボールが存在する時はフレームを静止させる．このことにより，内枠の中でボールが細かな動きを行っても，フレームがボールの細かな動きを追従することはなく，ぶれの少ない映像が生成できる．

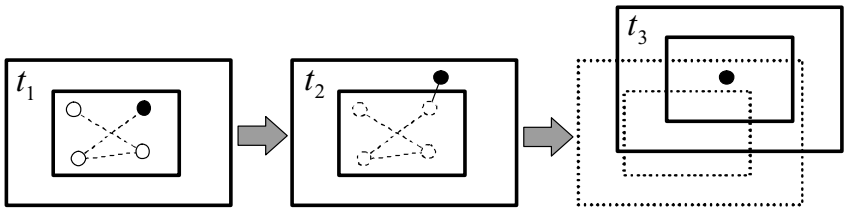


図 8.9　内側に枠を設定した時のフレーム移動

8.5.2　ボールの速度ベクトルを用いたフレームの移動

内枠を設定することにより，次のフレームでボールの座標が内枠の中にあるような細かいボールの動きに対する反応は防ぐことができた．しかし，この内枠は大きさや位置が固定であるため，ボールの大きな移動に対しては，フレームはボールの座標が内枠を出る瞬間まで移動せず，図 8.10 の左側のように，内枠の外に出た t_5 で急激に移動する．この動きによりカメラワークは急激な速度で移動してしまう．

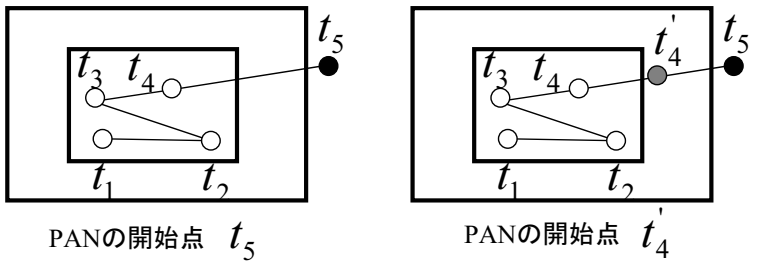


図 8.10　ボールの移動を考慮したカメラワーク

この急激な移動を防ぐためには，大きな移動に対してはあらかじめその移動を予測してフレーム移動を行うような方法が必要となる．この問題に対して前後のフレームでのボールの移動ベクトルの加重平均 $V = (V_x, V_y)$ を求め，これを用いて次の時刻の

ボールの座標を予測し，その座標を用いてフレーム移動の判断を行う方法が考えられる．そこで，図 8.10 右側において，時刻 t_4 での座標 (x_{t_4}, y_{t_4}) に対して，V を加えることにより $t'_4(x_{t_4} + V_x, y_{t_4} + V_y)$ を取得し，この t'_4 を用いてフレーム移動の判断を行う．この結果フレームは図 8.10 左の場合より早い時刻で移動を始める．これにより，ボールの大きな移動に対して早い段階から追従し，自然なカメラワークを実現することができる．

8.6 評価実験

8.6.1 AHP を用いた主観評価法

映像の評価については，主観評価となる．従来このような主観評価法の基本技法としては SD 法 (Semantic Differential Method) が良く知られている．しかし，SD 法では主観評価の項目が独立に評価されるため，複数の評価項目がいずれも良い評価を得たとき，どの評価を重視すべきか判定できない．

このような問題に対し，一対比較法として，サーストン法 (Thurston Method) や AHP(Analytic Hierarchy Process) [85, 86] 等が知られている．特に AHP は人の主観判断を取り扱う事に適している．AHP とは，意志決定を，問題・評価基準・代替案という階層構造として捉え，階層ごとに一対比較を行った上で，代替案のどれが好ましいかを決定する手法である．この評価基準の階層で一対比較によって得られる重みにより，どの基準が重視されているかを判定することができる．また，代替案を優位性の順に並べ替えることも可能であるが，評価基準の重みが似ている同好者どうしをグループ化し，グループごとに代替案の評価を見ることで，同好者ごとに代替案の順位を見ることができる．このことから，多角的な評価が可能である．

本研究では，今後の同好者や個人化された映像の評価も考慮に入れて，AHP 法を用いて実験映像の評価を試みる．

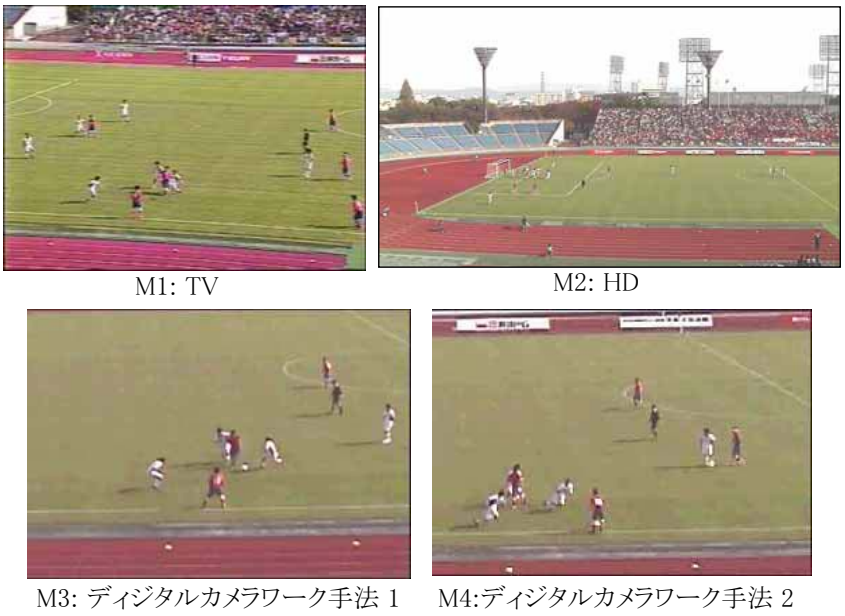


図 8.11　評価実験用映像

8.6.2　実験環境

実験用の映像としては，8.4.1 に示した条件下で，第 83 回全国高校サッカー選手権大会京都府大会決勝，京都朝鮮高校対桂高校の試合を撮影した．また，この試合については，ロイヤルボックスにて TV 用の撮影が行われていたため，放送された映像を入手し，これを 640x480,30fps の AVI ファイルとして生成した．これを M1 と呼ぶ．また，HD カメラで撮影した映像を 1280x720,30fps の AVI ファイルとして生成した．これを M2 と呼ぶ．また，撮影した映像からオフラインにて 8.4 で述べた手法による実験映像を M3(図 8.11:method 1)，8.5 で述べた手法による実験映像を M4(図 8.11:method 2) として，320x240 の固定フレームによる切り出しを行い，これを 640x480,30fps，SD レベルでのディジタルカメラワークにより自動映像生成を行った．図 8.11 は，M1〜M4 の映像の例を示している．

HD の映像を評価用に加えた理由は，試合進行のガイドになると位置づけた，試合の一部を見せる表示法と，HD 映像のように，全体を固定で見せる表示法の評価上の違いを判定するためである．

8.6.3　AHP 法による評価実験

評価実験の映像は図 8.11 に示した４種類であり，各種類ごとに 6 本の映像を用意し，被験者六人に対してそれぞれ種類ごとにランダムに選択した 2 種類の映像を見せて評価した．ただし，被験者のうち 4 名はサッカー経験者であり，2 名はサッカーに興味はあるが，サッカーについての知識がないサッカー未経験者である．また，現時点ではサッカー映像に対する充分な嗜好の分類調査を行っていないため，サッカー経験者とサッカー未経験者の分類は，必ずしも同好者としてのグループ分けとはなっていない．本研究では，まずサッカーの経験者と未経験者でどのような評価の違いが現れるかについて焦点をあてた評価実験を行う．

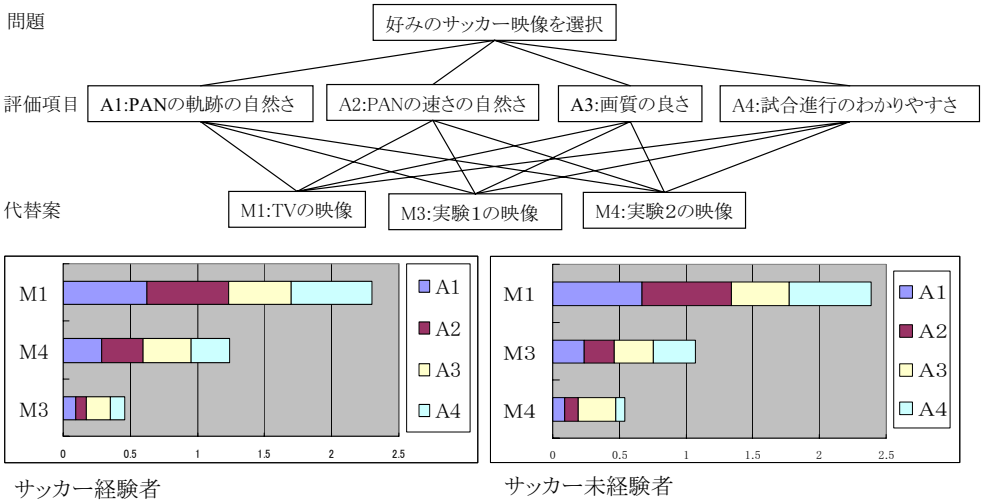


図 8.12　カメラワークを基準とした評価実験

「みやすさ」による評価実験

今回生成された映像の評価としては,「みやすさ」が評価の基盤となる．そこで，AHP の中間層について，パンの軌跡の自然さ (A1) と速度の自然さ (A2)，また解像度の見劣りがあるかどうかを見極めるため，画質の良さ (A3) を導入し，最後に試合進行のわかりやすさ (A4) を加えて評価要素とした．この観点から，テレビ (M1) と二つの実験手法 (M3,M4) を対象として，カメラワークを基準とした評価実験を行った．図 8.12 に実験結果を示す．HD 映像である M2 を評価の対象から省いているが，これは，カメラ

ワークを基準とした評価であるため，カメラワークの存在しない M2 は，比較対象から除外した．

評価の結果，中間層の嗜好の重みについては，サッカー経験者で A4(0.50), A1(0.20), A3(0.18), A2(0.12)，未経験で A4(0.50), A3(0.28), A1(0.16), A2(0.06) となり，いずれも試合進行のわかりやすさが最も重視されつつも，サッカー経験者ではパンの軌跡，未経験者では画質が次に重視されるという違いが現れた．また，AHP 全体の評価結果としては，サッカーの経験・未経験を問わず提案手法よりテレビが優位となり，次に M4, 最後が M3 となった．また，中間層の評価要素についてはサッカーの経験・未経験による大きな違いはないが，M4 の映像に対し，未経験者より経験者の A1,A2 に対する評価が高くなっている．これらの結果に対しては，嗜好に関するより詳しい評価法を必要とするため，今後の課題とする．

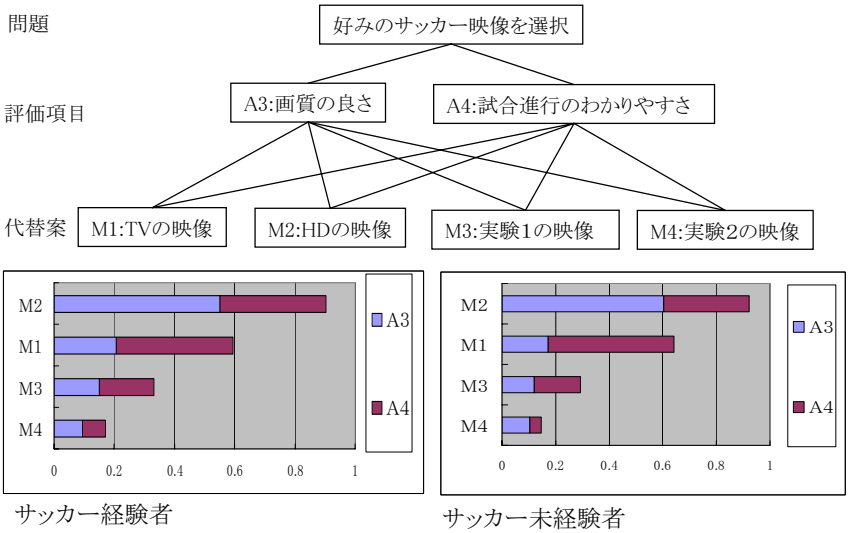


図 8.13　画質と試合進行を基準とした評価実験

「試合のわかりやすさ」と「画質」による評価実験

また，カメラワークを用いて視聴者に一部を見せる表示法とサッカーコートを固定カメラで広く見せる表示法について，試合のわかりやすさと画質を評価項目として HD 映像 (M2) を加えた評価実験の結果を図 8.13 に示す．ただし，中間層の嗜好の重みについてはサッカー経験者で A4(0.79),A3(0.21) 未経験で A4(0.65),A3(0.35) となり，図

8.12 の結果同様，試合進行が最も重要とされるが，サッカー経験者は，画質よりも試合進行を重視していることがわかる．また，評価結果としては，サッカーの経験・未経験を問わず，HD の映像が高い評価を得ている．しかし，中間層の評価要素の貢献度を見ると，いずれの場合も試合進行のわかりやすさでテレビが優位となっており，特に未経験者のテレビに対する評価が高いことが示された．

この結果より，いずれの評価実験においても最も重視された試合進行のわかりやすさという点においては，サッカーコートを広く見せることよりも，カメラワークを用いた撮影が優位であることを示しているものと考えられる．これは，8.3 で示したように，プロのカメラマンが試合進行をガイドする高度な撮影技法を用いているからであると思われる．

本研究で示した手法は，プロのカメラマンが用いる高度な撮影技法を取り入れるには至っておらず，それらの高度な撮影技法の分析と実装法が今後の課題となる．

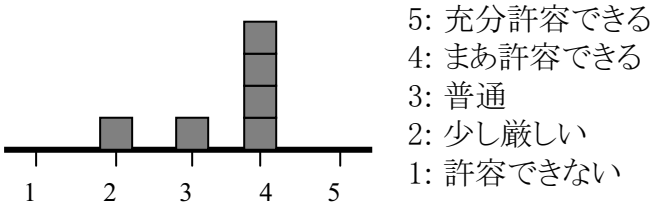


図 8.14 アンケート

8.6.4 生成映像の許容度アンケート

ここで図 8.14 は，ディジタルカメラワークで自動生成した映像が，必ずしもプロのカメラマンを必要としないサッカー映像を視聴するという観点において，受け入れられるかどうかについてアンケートを行った結果を示している．ただし，この許容度アンケートについては，8.4 と 8.5 の手法で生成した両方の映像を含めて５段階で評価してもらった．その結果，許容度アンケートの最高評価５「十分許容できる」の評価は得られなかったが，４の「まあ許容できる」に評価が集まった．これより，実験手法はまだテレビやサッカーコート全体を見せる表示法より評価が低いものの，必ずしもプロのカメラマンを必要としないサッカー映像として，許容できるレベルには到達しているものと考えられる．

8.7　結言

本研究では，サッカー映像を対象として，デジタルシューティングの観点からデジタルカメラワークを実現する手法について提案を行い，映像自動撮影技術の一手法を示した．これは，第二の問題「コスト・技術不足で実現できないコンテンツ」の解決に貢献する．また，大衆向けの映像コンテンツではないが，同好者小集団から見て映像コンテンツのチャネル数を増やす点では，第一の問題「コンテンツ不足」の解決にも貢献する．また，必ずしも人が介在する必要がないため，第三の問題「作業コスト・人員不足」の解決にも貢献する映像コンテンツ自動生成支援技術となる．デジタルカメラワークでは，高解像度の映像の一部を切り出して映像を自動生成する際，何らかの対象の動きに基づいて仮想のフレーム枠の動きを操作する必要がある．本研究では，小刻みに動くボールを対象としながらも，映像文法に従う安定した仮想カメラワークを実現している．被験者による主観評価を行ったところ，十分許容できる映像ではないが，許容できるとする評価が得られ，その有効性を示した．

第9章

映像の構文に依存したライブ映像の二次コンテンツ自動生成方式

9.1 緒言

近年，WWW などの大規模マルチメディア空間や，携帯電話などのモバイルマルチメディア空間が拡大，普及している．さらに，映像や音声などのマルチメディアコンテンツを，家庭のコンピュータや携帯電話などに実時間配信することが可能となっており，不足するコンテンツと制作者の人材不足を補うディジタル技術や，必ずしもプロの製作者を必要としない，もしくは補助するコンテンツ生成技術，またその生成支援技術などが注目されている．コンテンツ不足を補い制作コストを抑え，必ずしもプロの制作者を必要とない対象として，スポーツ映像が注目されており，特にプロの試合だけでなく，民間のスポーツを対象として，地域性の高いコンテンツを生成するためのディジタル技術も注目されている．

本研究が対象とする野球についても，高校野球を対象とした，スコアボード付中継映像のインタネット配信実験が行われている [87]．すでに EPG などの番組情報付コンテンツの配信が始まっており，プロ野球の試合内容に関するインデクス情報が将来付与されることを前提として，ユーザの好みに合わせたダイジェストを見ることのできる，インタラクティブ TV システムも提案されている [88]．また，人手で更新される WWW 上で公開されているスコアボード情報を利用した，ダイジェストシステムも提案されている [89].

このように，多くの人が場所を問わず，自分の好みに合った情報を入手できるサー

ビスが提供されようとしている．しかし，以上のシステムは，いずれも人手，もしくは既に何等かの方法で，インデクス情報が付与されたことを前提としている．近年の多チャネル化や，サービス形態の多様化によるインデクス情報の増大を考えると，人手によりインデクス情報を付与することは現実的でない．

また，映像コンテンツの場合，特に商品価値が高いものは，アーカイブ化された過去の映像コンテンツではなく，同時性を有する，現在もしくはある出来事が起こった直後の映像コンテンツであり，実況中継などのライブコンテンツもその一つである．このような，実況中継を前提とした映像コンテンツの生成では，実況中継中，あるいは実況中継を完了した直後に配信を始める必要があるので，予め時間をかけてインデクス情報を付与し，構造化しておくことができないという問題を解決する必要がある．

以上の点から，本研究では，スポーツ実況中継，特に野球の実況中継映像を対象として，自動的にインデクス情報を付与して，ハイライトシーンを検出することを目的としている．特に，本研究が着目するサービスの形態は，外出中の野球ファンに対し，野球中継中のハイライトシーンを速報として，リアルタイムに配信するシステムであり，人手によるインデクス情報を用いないシステムである．このようなシステムを実現するためには，インデクス情報を自動付与するだけでなく，映像をディジタル化しながら並行して，コンテンツ解析を行うオンライン処理，かつすべての処理がリアルタイム内に収まるよう，高精度で高速に処理する手法が必要となる．

そこで，野球の実況中継映像を対象に，オンライン処理を前提としたシステムの実現を主眼に置き，実時間音声認識技術と実時間画像解析技術を統合した，野球映像中のハイライトシーン配信システムの部分システムとして，PC シーンを高速で高精度に自動検出する手法を提案する．

9.2　関連研究

映像を構造化するには，音声，言語，画像など，マルチモーダル情報ストリームの協調的処理が有効であることが示されている [90]．本研究は，実時間向けの画像処理とアナウンサーの実時間音声認識を協調させるアプローチを用いているため，この部類に入る [91]．画像処理のみを基本として，ディジタル化された後の野球映像にインデクスを付与したり，それをもとに構造化する研究としては，カメラワークを用いてシー

ンを特定する研究 [92] や，その情報をもとに，ホームランシーンのカメラワークをテンプレートとして，DP マッチングにより他のホームランシーンを検出する手法 [93]，差分画像から得られる動作情報を用いて，投球やスイングを DP マッチングで検出し，インデクス情報を付与する手法 [94, 95]，映像中のテロップを解析する手法 [96, 97] がある．また，音声認識のみによりヒットなどのインデクスを付与する手法 [98] などが提案されている．この他にも，人手によるメタ情報の付与作業を支援する技術を背景として，検索を目的としたデータマイニングによるシーン検索システムも提案されている [99]．しかし，いずれもオフライン処理を前提としているため，ライブ映像に対応できない．また，クローズドキャプションなどの言語情報を用いる手法 [100] もあるが，訓練された人手による作業が前提となる．

本研究の対象である，投球ショット (本研究では放送用語として PC[Pitcher and Catcher] ショットと呼ぶ) に対して，そのテンプレートとカメラワークを抽出して映像を構造化する手法 [101] も提案されているが，PCS の検出を高精度化するために，マスキングする領域は手動で指定しているため，最適性に疑問が生じる．

この他にも，ショットの分類を目的として，野球の意味モデルをベイジアンネットで表現する手法 [102]，最大エントロピー法を用いる手法 [103] また，ハイライト抽出やダイジェスト作成を目的として，最大エントロピー法を用いる手法 [104] や，HMM を用いる手法 [105] などがある．このように，野球に限定しても，２次的なコンテンツ生成や生成支援・応用システムに対する関心の高さが伺える．

9.3　ハイライトシーン配信システムの概要

9.3.1　野球中継映像の撮影に関する背景

野球におけるハイライトシーンを定義する上での意味的な最小単位は，投手の投球に始まり，次の投球が始まるまでの区間として概念的に容易に構造化できる．ただし，野球映像の撮影法としては，スタジアムで観戦しているかのように，ある俯瞰の位置から Long Shot に相当する位置のまま撮影し続けることで，編集のない映像として提示することも可能である．その場合は，常に連続している印象を与えるが，野球の試合構造に関係のない映像の表現スタイルとなる．この点については，8 章の 8.6.3 で示

図 9.1　ショットの種類

されたように，試合進行のわかりやすさという観点において，素人を含めた一般大衆向けには，スタジアムを広く一定の視点から見せるよりも，テレビ放送の技法が好まれる傾向があった．ただし，サッカーの試合内容は，素人にとってわかりにくいため，テレビ放送の技法が好まれる傾向を示したとも考えられる．放送局の映像制作者は，テレビの技法が,「余計なお世話」の一種に成り得ることを意識しつつ,「優れたお世話」を目指す立場から，試合内容をより良く視聴者へ伝える表現技法を意識して断片の組織法を練り上げている．

古典的ハリウッドの流れを強く継承するテレビ放送において，カメラの異なる配置からの映像やアングルの変化を用い，空間的断絶を伴う映像は，視聴者に対し，連続性の欠如や試合進行の自然な解釈を阻害する可能性がある．このため，視覚的断絶を意識させない自然な接続によって，試合内容が自然に連続しているように見せる技法の適用が意識される観点においては，古典的デクパージュの概念に基づいて映像は組織されていると見なすことができる．ただし，野球中継映像では，この役目を担うのは編集者ではなく，複数のカメラを切り替える権限を持つスイッチャーの役割である．

9.3.2　野球中継映像の絶対的ショットサイズ

プロ野球中継の場合，球場に配置されるカメラの位置はほぼ決められており，我々は，それらのカメラを用いて撮影されるショットを図9.1のように分類している．図9.1のようなショットのカテゴリーは，試合進行を自然に表現するための視覚的断片の材料として選ばれた絶対的ショットサイズであると考えることができる．図9.1中の記号は，

それぞれPCS(Picher and Catcher Shot：ピッチャーの投球場面でバックスクリーンから撮影されているショット)，LS(Long Shot:グランドの様子を撮影している高い位置からのショット)，MS1(Middle Shot 1:グランドに配置され，選手の動きを追うショット)，MS2(Middle Shot 2:LSより低い位置から撮影されるショット)，TS(Tight Shot:選手の顔などを撮影するショット)，FS(Full Shot:球場全体を撮影するショット)，AS(Audience Shot:観客を撮影するショット)，OS(Other shot:その他のショット)を表している．本研究では，これらのショット集合からPCSのみを検出することに主眼を置くため，PCS以外のショットはNPCS(Non-PCS)と呼ぶ．

また図9.1のショットは，複数台のカメラとカメラのスイッチングを行う中継車によって撮影が行われる．しかし，カメラの台数は撮影・中継コスト等に関連して一定とは限らず，コストを抑える必要がある場合は，１台のカメラで，図9.1の複数の絶対的ショットサイズの撮影を担うこともある．このため，カメラと絶対的ショットサイズの種類は基本的に対応しておらず，カメラの数が減るほどその傾向は強くなる．つまり，カメラが特定されても絶対的ショットサイズはカメラと一対一の関係として必ずしも固定的に特定できない．このため，メタ情報の付与は基本的に人手を必要とする．また，この撮影環境を構築する業者と放送局は，基本的に別会社となる場合が多く，放送局は，中継車から送られてくる映像の放映権について契約を行っている．このため，メタ情報を用いたシステムを構築する場合，メタ情報は新たな取引の対象となり，恒常的なコストを必要とすることになる．

我々は，コストの削減や新たなビジネスの対象となるメタ情報の自動取得や，人手によるメタ情報付与作業の支援技術に着目しており，音声や画像の解析による問題解決手法に着手している．本研究では，中継車から送られてくる映像を対象としたPCSの自動検出に焦点を当てる．

9.3.3　PCSと構文的なPCシーン

本研究で取り扱うハイライトシーンは，投手が投げたボールをバッターが打つことにより発生するヒットやホームランなどのイベントに限定する．野球中継映像では，試合内容の進行を視覚的にも理解しやすくするための工夫が行われており，試合構造を反映した構造を持つ．野球の試合進行は，投手の投球に始まり，次の投球が行われるま

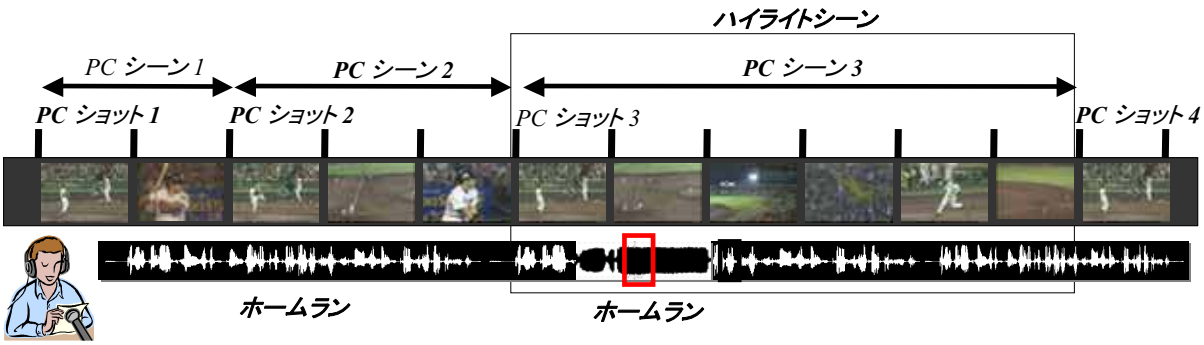


図 9.2　PCS とハイライトシーンの関係

での時空間に対応するイベントの繰り返しであると考えられ，これを「PC シーン」と呼ぶ．この PC シーンは，空間断絶による内容伝達の阻害要因を考慮しながら，複数のショットによる分節構造を用いて表現される．その野球中継映像に適用された表現法としては，図 9.2 のように，PC シーンに含まれるショット列の構造として，必ず PCS で始まる形式が用いられている．ここで，構文を，分節構造を持つ表現体において，意味的な最小単位の接続系列が持つ構造的パターンとすれば，PC シーンは，図 9.1 に示した野球映像で最小単位となる絶対的ショットサイズの分類に基づいて，PCS で始まるという一つの構文を形成していると見ることができる．

9.3.4　PCS とマスターショット

古典的デクパージュでは，表 2.4 の Rule(3-1) のように，シーンの冒頭をマスターショットで始める規範規則があった．マスターショットは，普通，広角 (Loosely Frame) で撮影され，あるシーンで起こっているすべてのアクションをカヴァーするショットのことであり，古典的デクパージュの「連続性」表現を成功させるために欠かせない手段である [8]．古典的デクパージュでは，まず基本として，マスターショットがあり，シーンは最悪のケースとして，分節を行わない場合，マスターショットだけで構成しても良い．概念的には，このマスターショットを機軸として，異なるカメラ・アングルや異なるショットサイズによる映像を挿入し，シーン内の焦点を宛てる対象を操作することになるが，その焦点の当て方がわからない場合はマスターショットのみで済ませる，また，焦点を当てる途中で，焦点の当て方がわからなくなった場合は，マスターショットに戻る，等の観点で分節構造が決定される．野球中継映像などの，ストーリーがあら

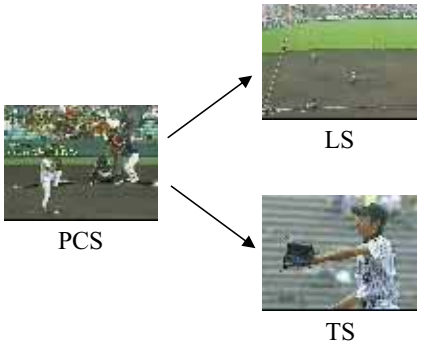


図 9.3　PCS に関する野球映像独自の映像文法

かじめわかってない映像の撮影では，その場の判断でショットの挿入の仕方を決定するため，失敗も発生する．この点で，マスターショットは起点として特に重要である．

マスターショットは，本来，シーン内の空間的関係が掌握できるフレーム枠であることが指針とされるため，その観点を適用すれば，PCS より，図 9.1 の LS や FS などが，適していると考えられる．しかし，野球の試合進行を，どのようにして視聴者へ正確に伝えるのか，また大衆に広く受け入れられる表現はどのような提示法が良いかについて検討された結果，野球の試合進行の中心となる PCS がマスターショットとして選ばれたのである．そして，構造表現を明確にするためか，PCS は三振などバッターがボールを打つイベントが起きない場合でも，一投球に対し，一つの PCS が対応するよう，次の投球が始まるまでに，別のショットが挿入される指針が採用されている．このような背景から，基本的に，PC シーンは複数のショットで構成され，PCS が PC シーンの冒頭にマスターショットとして用いられる構文的構造が現れる．

9.3.5　映像ジャンルに依存した映像文法

大衆向けの放送映像には，内容を効率良く視覚的に伝達するための工夫が施され，視聴者が内容理解を自然に行えるよう，接続の善し悪しが検討されており，映像のジャンルごとに独自の映像文法が存在すると言われている．

例えば，PC シーンは PCS で始まる以外に，PCS に続くショット遷移について文法的規範が導入されている．例えば，バッターがボールを打った場合は，基本的に LS が挿入 (接続) される (図 9.3)．LS は，バッターが打ったボールがカメラのフレーム内に入るよう，グランドの様子を上から撮影し，カメラワークによってボールを追いかけ

るショットである．このように挿入することによって，バッターが打ったという意図をショットの接続という映像表現を用いて伝えているのである．もし，PCS でバッターがボールを打ったにもかかからず，図 9.3 の下向きの接続のように，ここで TS が接続されると，ボールの行方がどうなったかを知ることができず視聴者は混乱する．

以上のように，PCS は，PC シーンの構文的構造を決定するだけでなく，文法記述の機軸ともなっているため，野球中継映像において，PCS を判定することが構造化において重要な役割を果たす．本研究では，この慣習化した構文的構造を持つ映像の表現スタイルに基づいて，プロが制作した映像の二次利用としての映像コンテンツ自動生成手法に頂点をあてる．また，本研究では，この構文的構造による PC シーンが，ホームランなどのイベントにおいて，ピッチャーが投げ，バッターが打ち，スタンドに入ったという一連の構造と一致するため，ハイライトシーンの抽出単位として，PC シーンの抽出を行う手法に着目する．本研究の目的は，ホームランなどのイベントが起きた際，即時的にハイライトシーン映像を外出先のファンへ送信することであり，その実現手法として，オンライン処理向きの高速で高精度な PCS の判定法について提案する．

9.3.6　PCS の判定

PCS を抽出するためには，まずショットの切り替えを検出する必要がある．本研究で扱う野球映像では，ショットの切り替えにカットだけでなく，ディゾルブも含まれるため，これらのショットの切り替えを高精度に効率良く検出する方法が望まれる．本研究では，MPEG １の映像に対し，高精度かつ実時間の 1/30〜1/50 でカットやディゾルブを検出できる手法 [106] を用いてショットの切り替え点を検出した [107]．検出されたショット切り替え点で挟まれた区間を，本報告ではショット区間と呼ぶ．PCS は，カメラを完全に固定してはいないが，構図は安定している．そこで，PCS の判定を行う場合，ショット区間から代表フレーム（先頭の１フレーム）を取り出し，１枚のフレーム画像が PCS に属するものであるかどうかを判定すればよい．

この手法を用いることで，大量の画像データを処理する必要がなくなり，計算量を大幅に減らすことが可能となる．ショットの切り替え判定が実時間の 1/30〜1/50 であるため，PCS の判定が高速であれば，ディジタル化の処理を含めてもリアルタイム処理が可能となる．

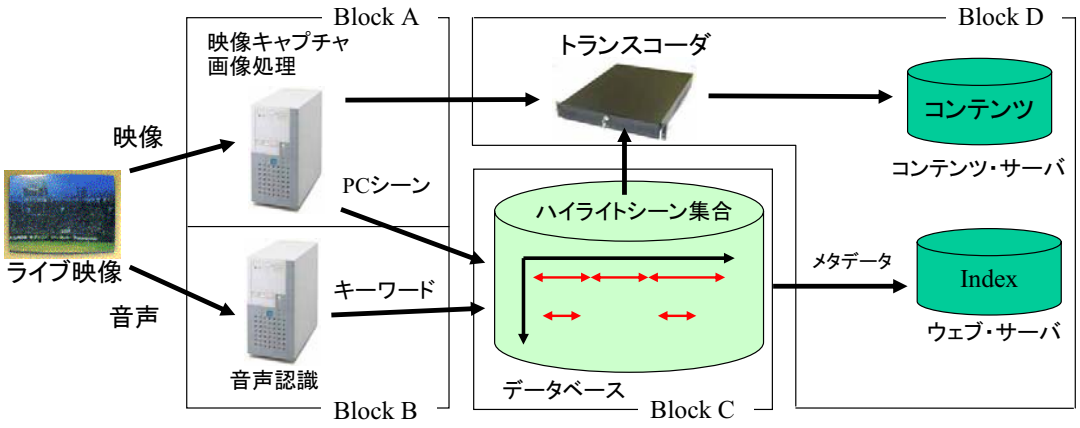


図 9.4　ハイライトシーン検出システムの概略

9.3.7　ハイライトシーン検出システム

本研究で扱うハイライトシーン検出システムの全体像を図 9.4 に示す．図 9.4 において配信用映像生成部 (Block D) では，入力されたライブ映像をモバイル配信用の映像フォーマットに変換する．メタデータ生成部 (Block A,B) では，ハイライトシーンの映像検索に用いるメタデータを生成する．メタデータ生成の流れは，まず，ライブ映像と中継音声の同期をとり，それぞれ別々の計算機に取り込む．次に，映像解析用計算機 (Block A) では，部分的にディジタル化が終了した入力映像を順次的に処理し，ショットの切り替えが起こった直後の画像フレームを対象として PCS の判定を行う．PCS と判定された場合は，次に現れる PCS を待って PC シーンを切り出し，各シーンにおける始終端の時刻情報をメタデータとして XML 形式でデータベース (Block C) に出力する．

一方，音声解析用 PC(Block B) では，入力された中継音声から無音区間を抽出して音声を分割する．分割された音声区間に対して，音声認識を実行しキーワードを検出する．検出されたキーワードを，その始終端の時間情報とともに XML 形式でデータベースに出力する．これらの XML ファイルにより，ハイライトと判定された映像の区間情報に基づいて，モバイル配信用に変換された映像の区間を携帯端末に配信する．

つまり，ハイライトシーンとは，PC シーンのうち，音声認識と音声のパワーを用いてイベントと判定されたキーワードを含む区間となる。画像解析により PC シーンの区間を決定し，音声認識により PC シーンの中からハイライトシーンを特定するため，

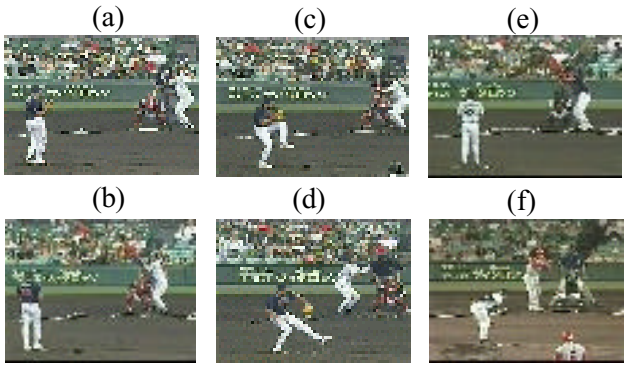


図 9.5　PCS の変動要素

画像・音声の協調システムとなっている．このためシステムの性能は，PCS をいかに精度よく検出できるかに依存する．本研究は，図 9.4 中，Block A を実現する．

9.4　PCS の判定法

9.4.1　PCS の変動要素

PCS は，安定した画像であると思われるが，実際にはデーゲーム，ナイトゲームの要因以外にも，様々な変動要因がある．図 9.5 の (a) は典型的な PCS であり，(b)〜(f) は，PCS の画像上，変動する要素を示したものである．PCS は，固定されたカメラで撮影されているが，実際の野球映像では，(b) のように，カメラの撮影方向が多少左右にずれる．これは，右バッターや左バッターが PCS の中へ適切に入るよう，カメラを左右に動かすことができるようになっているためである．この点においては，画像のずれに頑健な特徴量が望まれる．また，図 9.5(c) のように，ピッチャーの位置やフォームが変化したり，(d) のように，バッターの位置が異なる場合もある．この他にも，PCS の判定においては，色情報を用いる方法が考えられる．しかし，(e) のように，選手のユニホームの色が変化したり，(f) のように，天候や照明の影響によって，全体の色が変化するため，必ずしも安定した特徴量を得ることができない．

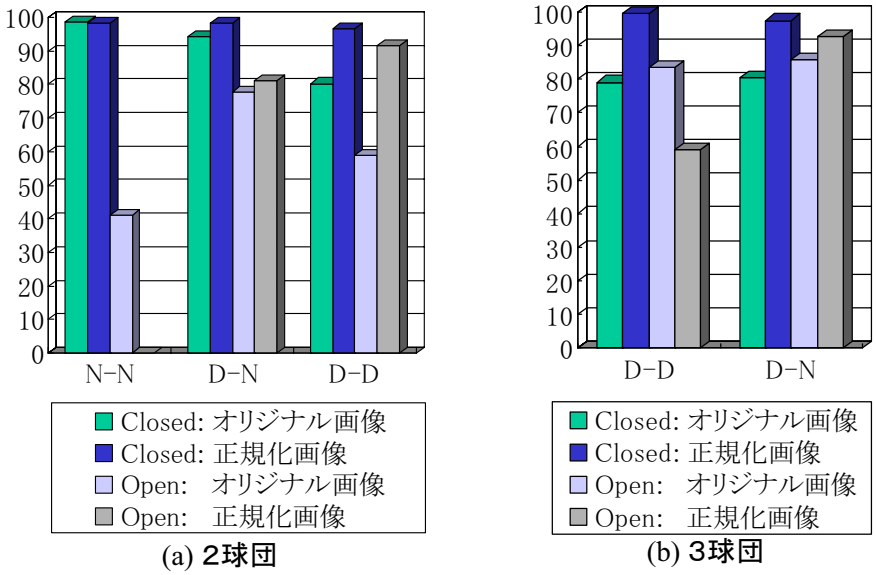


図 9.6　ヒストグラム法による PCS の検出結果

9.4.2　ヒストグラム法による実験

このような構図がある程度定まった画像のシフトや選手の位置関係の変動に強く，また画像を正規化することで色や照明の変化に強い特徴として輝度ヒストグラムが考えられる．画像の正規化とは，元画像の画素 i の輝度値を $x_i = (0, 1, \cdots, 255)$ とするとき，正規化された画像の画素 i の輝度値を x_i^N として，$x_i^N = (x_i - \mu_{x_i})/\rho_{x_i} + 127$ と定義される．ここで，μ_{x_i} とは，x_i の平均値であり，ρ_{x_i} とは，x_i の標準偏差である．

図 9.6 は，学習データに用いたデータのみで PCS の判定を行う Closed な実験と，学習データには用いなかったデータを使って PCS の判定を行う Open な実験の結果を示したものである．また，図 9.6 中，Original とは，画像の正規化を行わなかった場合，また Normalized は，画像の正規化を行った場合を示す．図 9.6(a) は，ユニホーム色を２球団に限定し，デーゲーム (D) とナイトゲーム (N) の組合せで実験した結果である．N-N はナイトゲームのみのデータ，D-N は，デーゲームとナイトゲームを混ぜたデータ，D-D は，デーゲームのみのデータを用いたことを示す．また，図 9.6(b) は，ユニホーム色を３球団に限定して PCS の学習を行い，判定実験を行った結果である．ここで，図 9.6(b) に図 9.6(a) のような N-N の組み合わせによる実験結果が無いのは，入手することのできた映像データの組み合わせ上，N-N の組み合わせが実現できなかったためである．この N-N の組み合わせによる判定結果はないが，いずれも Closed な実験

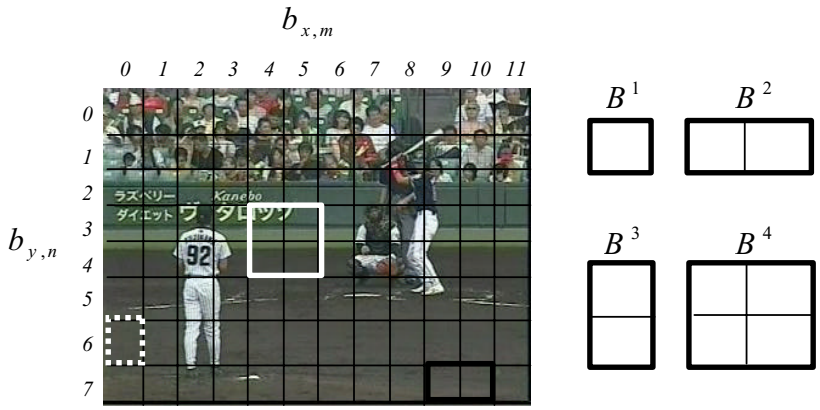


図 9.7　ブロックと領域の種類

に対しては，非正規化で 80%以上の精度を示し，正規化すると精度が上がる．しかし，Open な実験にすると基本的に精度が悪くなり，正規化した方が良い場合もあれば，正規化をすることで，悪くなる場合もあり，精度が 0%になる例もあった．これらの不安定要因を吟味し，新たな処理を単純に追加して PCS を精度良く判定しようとすれば，計算量を増大させてしまうことになる．

このように，構図が安定している割に安定して高い精度が得られないことがわかる．これは，不安定な要因をすべて特徴量に反映させていることに問題があると考えられる．そこで本研究では，精度を保ちつつ計算量が少なく，また若干の平行移動に強い上に，色の変化に左右されない特徴量を画像中から自動的に検出し，判定を行う手法を提案する．

9.4.3　特徴量のマイニング

図 9.5 に示すような PCS の代表フレーム画像（以後，PCS 画像と呼ぶ）の構図は，ピッチャーとバッターで特徴付けられるが，ポーズの変動やユニホームの変化などにより，特徴量としては必ずしも安定していない．より安定した特徴としては，グランドや，観客席とグランドの間の壁などが想定できる．しかし，人手による安定領域の決定は信頼性がない．また，同一種類より多種類の特徴量により PCS を決定づける方法が考えられる．そこで，安定した領域を抽出するために，学習データから特徴のマイニングを行い，PCS 判定の識別率の高い安定した特徴の種類とその特徴に対応した

領域を選択し，組合せることで判定処理を行うアプローチを提案する．

まず，PCS 画像を図 9.7 のようにブロックに分割し，図 9.7 右に示す B^1，B^2，B^3，B^4 といった四つのブロックの組合せブロック内で輝度値の平均値，分散，さらに分散の対数を特徴量として着目する．これは次の理由からである．

(1) グランドや壁の領域は平坦な輝度値を持っているため分散値が小さく，

(2) 色の大きく異なる部分が含まれていても図柄が安定しており，

(3)PCS 画像のみで構成したオンセット（PCS 画像集合）内では分散が安定している．

(4) ブロック内の輝度に大きな差がある場合でも，対数を取ることによって，分散値の桁，つまり指数が安定している．

(4) の特徴量を用いる理由を述べる．まず，もし領域内の輝度値が大きく 2 極化する場合，構図は安定しているにも関わらず，その分散は大きくなる．例えば，図 9.7 の白枠で囲まれた領域の上部には，壁に白い大きな文字が存在しているが，このブロックで壁の輝度と文字の輝度は 2 極化しており，それぞれの輝度値で安定した集団を形成している場合が想定される．このような場合，その輝度の分散は，ある大きな値で安定している可能性がある．この安定性は，輝度の分散を指数表現した場合，指数の安定性として表現されると考えることができる．分散の対数を取ることで，文字の変化や文字位置の変化などによる分散値のばらつきをある程度抑制でき，指数の安定性を得ることができる．この意味においては，観客席の特徴を表現できる可能性もある．

次に，安定度という指標に注目すると，ブロックから得られる 3 種類の特徴量は，PCS 画像集合中，どの PCS 画像においても，それぞれよく似た値であることが望ましい．したがって，それぞれの特徴量の分散値が PCS 画像集合内で小さいブロックを，PCS として判定する際の良い特徴領域と考えることができる．そこで，学習用の野球映像から PCS を取り出し，それぞれの代表フレームとして取り出した PCS 画像の集合を学習データとしたとき，この学習データから得られるすべてのブロックの特徴量に対して，その PCS 画像集合内で分散値を求め，この分散値を昇順に並べて上位に位置するものが良い特徴量を示すブロックであると判定する．このような方法で特徴量を選択することにより，PCS 画像判定処理で用いる比較演算を大幅に減らすことが可能となる．

9.4.4　PCS 判定の学習アルゴリズム

学習は，二つのステップに分かれる．１つは，今回用いた三つの特徴量に基づき，特徴量の分散が安定した領域を選択することである．しかし，この段階では，分散の安定した特徴領域を用いることがPCS の判定精度にどう影響するかわからない．そこで，学習セットを対象としたClosed の実験により，異なる特徴量の最適な組合せを決定する．

ステップ１では，ショット区間の代表フレーム画像を図 9.7 に示すように，x 軸を 12 に，y 軸を 8 の領域に分割する．また，このブロックを図 9.7 の右のように組合せた四つの領域タイプ B^1〜B^4 を設定する．これら四つの領域タイプは，フレーム内でブロック単位に移動する．

そこで，これら四つの領域タイプに対して，フレーム内で存在する位置を領域番号 i として記述する．いま，あるフレーム f において，領域タイプ t ，領域番号 i の領域 B_i^t から得られる平均 $M_{f,t,i}$ ，分散 $V_{f,t,i}$ は，次の式 (9.1),(9.2) によって求められ，対数分散 $LV_{f,t,i}$ は $logV_{f,t,i}$ として求められる．ここで，$Gray_f(x,y)$ は，フレーム f の位置 (x,y) における濃淡値，また $|B|$ は，ブロック内の画素数を表している．

$$M_{f,t,i} = \frac{1}{|B_i^t|} \sum_{x,y \in B_i^t} Gray_f(x,y) \tag{9.1}$$

$$V_{f,t,i} = \frac{1}{|B_i^t|} \sum_{x,y \in B_i^t} (Gray_f(x,y) - M_{f,t,i})^2 \tag{9.2}$$

最後に，特徴の安定した領域を求めるために，各特徴量について安定している順に並べた領域リストを生成する．領域リストとは，領域タイプ t と領域番号 i の組み合わせにより得られるすべての可能なパターンを成す領域のリストである．ただし，ここでは，特徴量として平均値をとりあげ領域安定性を計算する方法について述べる．学習に用いるPCS 画像フレーム f の枚数を N としたとき，各PCS 画像上で，同じ領域 B_i^t において平均値の分散を，$V_{M,t,i} = mean_f(M_{f,t,i} - mean_f(M_{f,t,i}))^2$ として求める．

次に，$V_{M,t,i}$ をすべての t と i に対して昇順に並べたとき，順位 n に位置する分散を $V_{M,t,i}^n$ とする．すると，1 番目に位置する分散は，$V_{M,t,i}^1$ となり，これに対応する領域 $B_{i_M}^{t_M}$ をブロック平均値における最安定領域とする．なぜなら，この領域は，学習データにおいて，その平均値が最も変化しない領域として考えられるからである．言い換

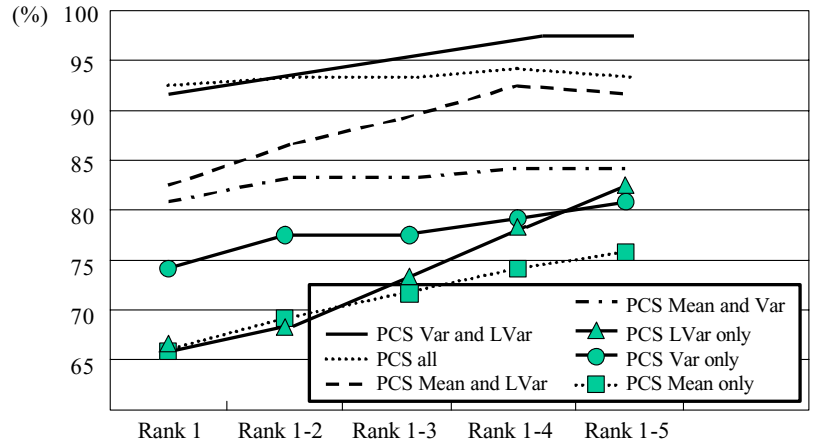


図 9.8　ステップ 2 の学習実験結果

えると，同じ値を取り続けると考えられるからである．このとき，フレーム f 上の領域 $B_{i_M}^{t_M}$ における平均値を M_{f,t_M,i_M} とすると，f に関してこの値の最大値を式 (9.3)，最小値を式 (9.4) として保存する．この最大値と最小値は，自動決定された閾値となる．

$$M_{t_M,i_M}^{max} = \max_f(M_{f,t_M,i_M}) \tag{9.3}$$

$$M_{t_M,i_M}^{min} = \min_f(M_{f,t_M,i_M}) \tag{9.4}$$

同様にして，分散の分散 $V_{V,t,i}$，対数分散の分散 $V_{LV,t,i}$ を計算し，これを基準として昇順に並べ，それぞれ 1 番目に位置する分散の分散 V_{V,t_V,i_V}^1 と対数の分散 $V_{LV,t_{LV},i_{LV}}^1$，また最大値と最小値として，$V_{t_V,i_V}^{max}, V_{t_V,i_V}^{min}, LV_{t_{LV},i_{LV}}^{max}, LV_{t_{LV},i_{LV}}^{min}$ が得られる．

PCS の判定は，野球映像の各ショットから代表フレームとして取り出した画像フレームを f としたとき，

$$\begin{aligned}(M_{t_M,i_M}^{min} \le M_{f,t_M,i_M} \le M_{t_M,i_M}^{max})\ and\\ (V_{t_V,i_V}^{min} \le V_{f,t_V,i_V} \le V_{t_V,i_V}^{max})\ and\\ (LV_{t_{LV},i_{LV}}^{min} \le LV_{f,t_{LV},i_{LV}} \le LV_{t_{LV},i_{LV}}^{max})\end{aligned} \tag{9.5}$$

を満たす代表フレーム画像を PCS 画像として判定し，その PCS 画像が含まれるショットを PCS と判定する．

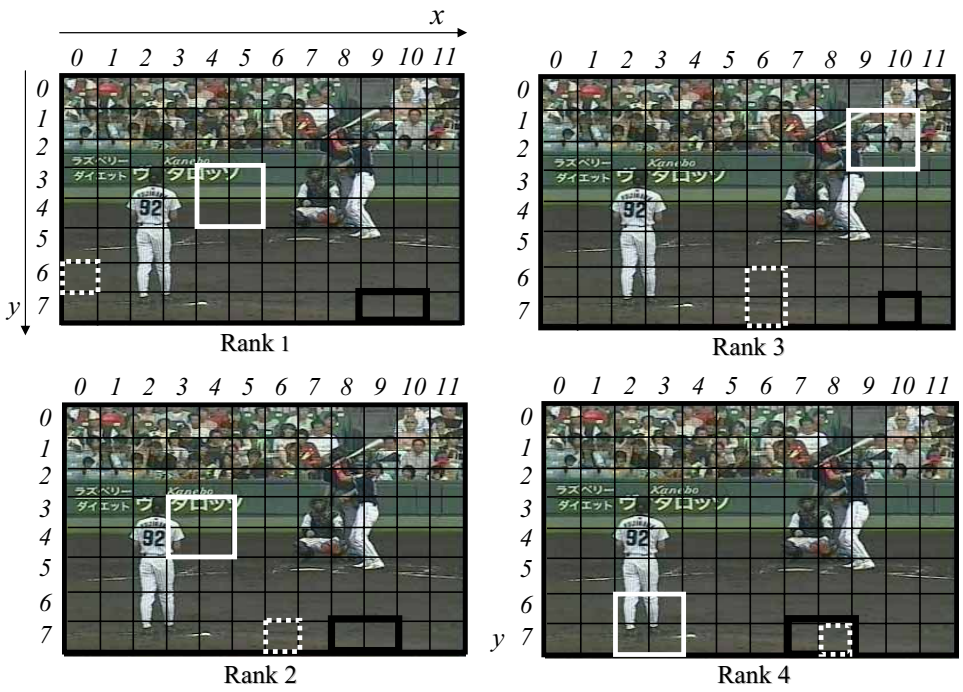


図 9.9　選択された領域

ステップ２では，ステップ１で得られた特徴量の昇順リストに基づき，リストの上位に位置する特徴量の組合せを生成し，Closed な実験を行って PC 判定の精度を求める．この精度とは，9.5 で定義する $Fmeasure$ を示す．この判定精度を降順に並べたとき，最も上位にある組合せが判定に最適な特徴の組合せとなる．図 9.8 に，ステップ２での実験結果を示す．図中，横軸は組合せ方を示している．例えば，Rank1 は，各特徴の Rank1 のみを組合せた方法，また Rank1-2 は，各特徴の Rank1 と 2 を組合せた方法である．この結果より，ブロックの分散と対数分散を Rank1-5 まで用いた特徴が，ステップ２の Closed 実験の中では最も良い値を示した．三つの特徴量を用いた場合も比較的高い値を示しているが，ブロックの輝度平均値を取り除いた分散と対数分散の組合せの方が良い結果を示している．これは，輝度平均もある程度は判定に貢献するが，逆に輝度情報に左右されるためであると考えられる．分散と分散の対数はブロック内の輝度の拡散度に依存するため，色や照明などの変化に大きな影響を受けなかったものと考えられる．

ここで，参考として，図 9.9 に各特徴量の安定性によって選択された領域を上位 Rank1 ～Rank4 の順に示す．図 9.7 の 3 種類の領域 (白点線枠，黒枠，白枠) は，１の実験に

おいて V^1_{M,t_M,i_M}，V^1_{V,t_V,i_V}，$V^1_{LV,t_{LV},i_{LV}}$ に対応する領域である．白の破線で囲まれた B^1 タイプのものが V^1_{M,t_M,i_M}，黒の枠で囲まれた B^2 タイプのものが V^1_{V,t_V,i_V}，白の枠で囲まれた B^4 タイプのものが $V^1_{LV,t_{LV},i_{LV}}$ にそれぞれ対応する領域である．

ブロックの平均値と分散に対応する領域は，画像全体のずれが起こっても輝度値の変化が起こりにくい領域が選択され，また輝度値は一様である部分が選択されていることがわかる．対数分散に対応する領域は，壁の輝度値と白字の輝度値の差が大きく，単純にブロックの分散を計算しただけでは分散値が大きくなると考えられる．しかし，壁の輝度値と白字の輝度値それぞれが安定しているので，PCS 画像集合内の同じ領域を比較すれば，いずれも分散値が近い値を示し，分散の対数が安定している領域が選ばれていると考えられる．

9.5 PCS判定実験

9.5.1 実験条件

実験に用いた映像は，甲子園球場を対象として，中継車から送信された映像を取り込んだ 320x240，29.97fps の MPEG1 映像 10 本であり，人手の情報が付与されていない，それぞれ 2 時間程度の素材映像である．各映像には，100〜250 程度の PCS が含まれている．また，さらなる高速化を目的として，MPEG1 から得られた画像 320x240 を 72x48 へダウンサンプリングを行い，カラー画像を濃淡画像へ変換している．このとき，最小ブロック の縦横の画素数は，それぞれ 6 である．

PCS 判定実験において本研究の手法の精度を検証するために，下記の式 (9.6) で示される Fmeasure を用いる．

$$Fmeasure = \frac{2 \cdot \text{再現率} \cdot \text{適合率}}{(\text{再現率} + \text{適合率})} \tag{9.6}$$

$Fmeasure$ は，二つの指標を統合する評価法であり，本研究の場合，その二つの指標として再現率と適合率を用いる．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

表 9.1　特徴の組合せと Fmeasure のランキング

教師データ	1:N-N			1:D-N			1:D-D			2:D-N			2:D-D		
Features	M	V	LV	M	V	LV	M	V	LV	M	V	LV	M	V	LV
Rank 1	0-0	1-4	1-4	0-0	1-4	1-4	0-0	1-6	1-6	0-0	1-4	1-4	0-0	1-4	1-4
Fmesure	96.8%			95.8%			97.5%			96.2%			96.5%		
Rank 2	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5	0-0	1-5	1-5
Fmesure	95.9%			95.0%			97.3%			95.1%			95.2%		
Rank 3	0-0	1-6	1-6	0-0	1-6	1-6	0-0	1-4	1-4	0-0	1-6	1-6	0-0	1-6	1-6
Fmesure	95.3%			94.9%			97.2%			95.0%			95.1%		
Rank 4	0-0	1-3	1-3	0-0	1-3	1-3	0-0	1-7	1-7	0-0	1-3	1-3	0-0	1-3	1-3
Fmesure	93.7%			94.9%			95.5%			94.5%			94.6%		

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．式 (9.6) では，再現率と適合率の両方が 100%に達したとき，Fmeasure は 100%となる．本研究では，未検出がなく，過剰検出のないシステムが理想的であるため，$Fmeasure$ の精度が 100%に近いことが望ましい．

学習データの選択法としては，球団ごとの色の違いや昼夜の違いがどのように影響するかを判別するため，グループ 1:阪神 vs 広島のみ（２球団），グループ 2:阪神 vs 広島と阪神 vs ヤクルトの組合せ（３球団）という二つのグループに分け，それぞれデーゲーム (D) とナイトゲーム (N) の組合せによる実験を行った．ただし，提案手法の特徴として，少ない学習データで高い精度が得られることを示すために，学習データには，二つの映像を用いて残り八つの映像を評価する．グループ 1 の２球団が含まれる学習データでは，入手した映像データの組み合わせ上，二つの映像を選択する組み合わせは N-N，D-N，D-D が可能であり，これをそれぞれ 1:N-N，1:D-N，1:D-D と表記する．つまり，例えば「1:N-N」という表記は,「阪神 vs 広島戦」のナイトゲーム（N) を２本用いた実験であることを示す．また，グループ 2 の３球団が含まれる映像データでは，入手した映像データの組み合わせ上，二つの映像を選択する組み合わせは D-N，D-D が可能であり，これをそれぞれ 2:D-N，2:D-D と表記する．

表 9.1 に Open な PCS 判定実験についての結果を示す．

表 9.2　ステップ 1 における各実験ごとの PCS 画像教師データ数

	1:N-N	1:D-N	1:D-D	2:D-N	2:D-D
PC	277	362	365	402	458

9.5.2　実験結果

表 9.1 は，平均，分散，分散の対数，それぞれの特徴量を用いて高い安定性を示した領域順位に従い，領域順位の高いものを複数組み合わせて PCS 判定実験を行い，Fmeasure による判定精度の高かった上位 1 位（Rank1) から 4 位 (Rank4) までの実験結果を抜粋したものである．表 9.1 で，M，V，LV は，それぞれ平均，分散，分散の対数による特徴量であることを示す．また，Rank 1〜Rank 4 それぞれの行に示された，例えば「0-0 1-4 1-4」という表記は，M，V，LV の特徴量の組み合わせ方を示している．「0-0 1-4 1-4」の場合，9.4.4 で説明したように，ステップ 1 で得られるブロックタイプ t とブロック番号 i を伴う特徴量の順位 V^n_{M,t_M,i_M}，V^n_{V,t_V,i_V}，$V^n_{LV,t_{LV},i_{LV}}$ において，M は上位 $n=0$〜0，V は上位 $n=1$〜4，LV は上位 $n=1$〜4 の領域特徴量の組み合わせを用いて PCS 判定を行うことを示す．ただし，M の上位 $n=0$〜0 とは，M の特徴量を使わないことを示す．表 9.1 縦軸は横軸の特徴量の組合せ方に対応する Fmeasure に基づいた判定精度のランキングを示している．

また，表 9.2 に各実験ごとの学習ステップ 1 で教師データとした PCS 画像の数を示す．表 9.1 より，最も高い値を示したのは，学習データにデーゲームのデータのみを用いた，V^n_{V,t_V,i_V}，$V^n_{LV,t_{LV},i_{LV}}$ の上位 $n=1$〜6 を組合せた場合である．しかし，この特徴量の組合せは，他の学習データの組合せではいずれも 3 位に位置するのに対し，V^n_{V,t_V,i_V}，$V^n_{LV,t_{LV},i_{LV}}$ の $n=1$〜4 や $n=1$〜5 は，最高値の 97.5%に近い精度でどの学習データの組合せにおいても高い精度を保持している．

表 9.1 の V，LV の組み合わせにおいて，Rank 1 に多く出現する上位 $n=1$〜4 位の領域特徴と，Rank 2 に出現する上位 $n=1$〜5 位の領域特徴の組合せを抜粋した Fmeasure の棒グラフを図 9.10，またその数値を表 9.3 に示す．本研究では，学習データの違いに影響されにくい特徴を良い特徴と位置づけるため，ここでは表 9.1 の Rank 1 に安定して位置する V^n_{V,t_V,i_V}，$V^n_{LV,t_{LV},i_{LV}}$ の上位 = 1〜4 の領域特徴の組合せを最も優れた特徴と定め，その中で最も高い精度 97.2%をシステム上の最も高い精度と位置づける．

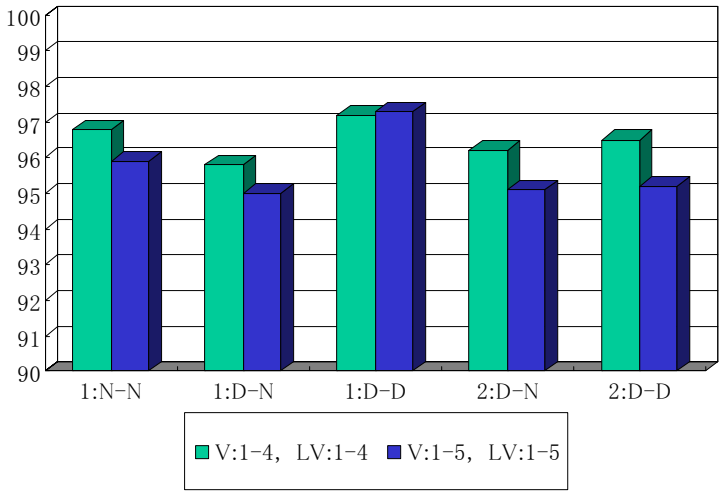


図 9.10　PCS 判定結果

表 9.3　実験結果

	1:N-N	1:D-N	1:D-D	2:D-N	2:D-D
V:1-4,LV:1-4	96.8%	95.8%	97.2%	96.2%	96.5%
V:1-5,LV:1-5	95.9%	95.0%	97.3%	95.1%	95.2%

以上の結果より，特徴に分散と分散の対数を用いることがPCS の判別に有効であるという結果が得られた．また，いずれも学習データに依存せず，映像 2 本分という少ない学習データで高い精度が得られることも示しており，提案手法の有効性を示すことができた．ただし，本研究において甲子園球場の PCS 画像から得られた特徴量は，他球場でも有効であるとはかぎらない．しかし，本研究で提案した手法は，比較的構図が安定していると思われる画像を対象として，安定要素を探索することに着目した手法である．また，表 9.3 より，学習データの選定について，昼夜，チームカラー等の画像特徴に注意を払わなくても，いずれも高い精度を示すため，システムの初期調整が容易になると思われる．つまり，表 9.2 に示した程度の比較的少ない教師データを用いて，検出精度を高くする特徴量をマイニングするため，一つの球場で得られた特徴を他球場に流用するよりも，むしろ各球場ごとに特徴量をマイニングすることに適した手法と考えられる．

実験のグループ 1 における処理時間を計測した．Xeon3.0MHz，1CPU のパソコン上

で，8 特徴（Rank1-4:L1,L2,L3,L4,LV1,LV2,LV3,LV4）を用いた場合の PCS 判定の平均処理時間は，最大で 8ms であった．このように，ブロックの輝度平均，分散，対数分散を求める計算量は非常に小さく，判定に用いる特徴数を 8 個までに限定することができるため，本手法は精度を高く保ちつつ，計算量を最小限に抑える手法であることがわかる．

9.6 結言

本研究では，即時的価値が高い映像コンテンツとされるプロのスポーツ中継映像のうち，野球中継映像に着目し，常に全ての中継映像を見ることのできない環境に置かれる外出中のファンへ，速報として，スポーツ中継映像のハイライトシーンを自動的に携帯端末へ配信するための，実時間ハイライトシーン自動抽出法に着目し，この実時間ハイライトシーン自動抽出法の画像処理部に焦点を当てて研究を行った．

ハイライトシーンの決定を行う手法としては，野球中継映像の構文的構造に着目し，ピッチャーとキャッチャーが同時に映る PCS を高速に安定して検出する手法に焦点を当てた．提案手法は，PCS の抽出において，どういう特徴が有効であるかをブロックの位置，形状を考えて発見する手法となっている．PCS 検出では，まず学習ステップ１でいくつかのブロック形状について輝度平均，ブロック内の輝度分散と，その分散の対数値を用い，それらの値が安定なブロックを求めた．次に，ブロックと特徴量の組合せをマイニングし，分散と対数分散によるブロックの上位 1〜4 位を用いたものが学習データに左右されにくく，高い精度を示す特徴量の組合せであることを示した．その特徴量の組合せにより，最大で 97.2%の検出精度が得られた．以上より，映像に対するメタデータ付与に関しては，充分に高い精度が得られていると考えられる．

本研究の提案手法と，音声認識を併用した手法 [108] により，オンライン処理でホームラン等のハイライトシーンが起こった直後に PC シーンを配信するプロトタイプシステムを実現しており，Net.Liferium 2003 にて実証展示を行った．本研究により，新たなサービスを提供する意味で，コンテンツ不足に応える点から第一の問題「コンテンツ不足」の解決に貢献し，必ずしも人が介在する必要がなく，第三の問題「作業コスト・人員不足」の解決にも貢献する映像コンテンツ自動生成支援技術を実現したことになる．

第10章

結論

本論文では，映像コンテンツ業界が抱える課題について，人材不足，人材育成，コンテンツ不足の軽減，新しい映像コンテンツの提供を可能とする技術の確立を目的として，映像撮影訓練，映像編集支援，ディジタルシューティング，ハイライト映像実時間配信という四つのテーマについて研究を行った．それらは，まず撮影訓練について，高時間分解能高速カメラワーク解析法（第3章），訓練指向オンライン単一ショット映像撮影支援方式（第4章）であり，編集支援について，索引情報を付与する，使用可能・不能区間推定による映像編集支援方式（第5章），ショットサイズ自動付与による映像編集支援方式（第6章），そして映像編集支援・自動編集方式（第7章）である．これらの背景となる映像文法については，第2章において概念を説明した．また，特にコンテンツ不足を補い，新しい映像コンテンツの提供を可能とする枠組みのディジタルシューティングについて，デジタルシューティングによる映像コンテンツ自動撮影方式（8章），最後にハイライト映像実時間配信として，映像の構文に依存したライブ映像の二次コンテンツ自動生成方式である．

第3章では，映像撮影訓練をオンライン処理として実現し，訓練対象者のカメラワーク動作の問題点を指摘するために，高時間分解能の高速カメラワーク解析法を提案した．従来法である輝度投影相関法では，ズームの解析精度が悪く，カメラワークの分類法やオンライン処理に向けての速度の改善を必要とした．これに対し，提案法では，二分化テンソルヒストグラム法をズームの解析に適用することにより，ズームの解析精度が向上しただけでなく，カメラワークの分類法として，ベクトルを基盤とした方法を導入することにより，全体的な精度の向上を実現し，速度の面でも改善し，撮影訓

練システム全体の処理として，ほぼ30fps を実現することが可能となった．第 4 章では，第 3 章のカメラワーク解析システムを用い，撮影者に映像文法に従う基本的なショットの撮影法を強制的に課題として与えることにより，提案システムを用いずに概念だけを説明して訓練を行った場合より，すべての課題を克服するよう撮影者の撮影技能が向上することが可能となった．

第 5 章では，映像編集支援システムの部分システムとして，映像文法に従った撮影が行われつつも，一つの連続した素材映像の中にあり，完パケで使用可能となるショットを自動的に切り出すための索引情報を映像文法の概念を利用して抽出する手法を提案した．この手法により，これまで煩雑で労を要する作業を自動化することが可能となった．第 6 章では，映像編集支援システムの部分システムとして，映像文法の基本となる相対的ショットサイズを自動付与する手法を提案した．相対的ショットサイズは，相対的にしか決まらない問題はあるものの，映像文法に従って撮影された映像であれば，シーン内のショットは包含関係で関係付けられるという背景知識を用いることにより，その自動付与が可能となった．第 7 章では，第 5 章と第 6 章の索引情報を用いて，映像文法に従い，シーン内のショットの接続候補を自動編集する方式を提案した．この提案方式により，編集者は，その候補群から良いものを選択し，その候補に修正を加えるなどの新しい編集環境を提供することができ，編集作業を軽減し，知的作業にだけ集中することが可能となった．

第 8 章では，ディジタルシューティングという新しいコンテンツ自動生成技術の枠組みを提案し，サッカー映像を対象として，ディジタルカメラワークによる映像の生成法を提案した．放送局の現場の意見としては，コンテンツ不足を解消する一つの観点として，必ずしもプロ並の品質を持つ必要がないコンテンツも存在するという見解がある．その必ずしもプロの制作者が関わらなくてもよい映像の候補は，民間のスポーツ活動を撮影した映像コンテンツであり，本研究では，サッカーを対象とし，自動撮影・編集によるシステムを提案した．この自動生成された映像は，プロの撮影には及ばないが，受け入れられるとの評価を得ており，民間のスポーツ活動を自動撮影し，新しい映像コンテンツとして提供する基盤技術を実現することが可能となった．

第 9 章では，プロの撮影した映像の中でも，即時的価値の高い野球中継の映像コンテンツから速報コンテンツとしてハイライト映像を外出中のファンに提供することを目的として，その速報性を実現するために，実時間で映像をキャプチャしながら，高

速で高精度なピッチャー・キャッチャーショットを検出する手法を提案した．これにより，オンライン処理で野球映像の投球からバッターの行動を反映した一続きのシーンを自動構造化することが可能となった．

以上のように，本論文では，撮影・編集支援技術により，これまで提案システムが担う役割を担当していた人材を必要としなくなる点で，人材不足に対応し，また，人を介さずに映像文法に従った撮影法を修練させる方法を提供しているため，人材育成にも貢献している．また，映像コンテンツの自動撮影・自動生成における基盤技術を提供したことにより，コンテンツ不足の解消に貢献し，人を介さず新たなサービスを提供する観点から，人材不足の更なる負担とはならないままに，コンテンツ不足を解消することに貢献し得ると思われる．

最後に，本研究の更なる応用について，展望する．

10.1　本研究の応用と展望

10.1.1　映像撮影・編集指南システム

映像文化を含め、日本のコンテンツ産業には次期戦略産業としての期待がある．しかし，映像制作に対する無知や，労働集約的な印象が映像コンテンツ産業・映像文化の活性化阻害要因となっており，製作現場でも高度な撮影・編集技法の伝承・教育不足の問題を抱えている．一方映像撮影機器が一般家庭に普及し，素人作品を公開する基盤も整いつつあるが，編集の困難性，あるいは技法不足から，一般人が自由に映像制作を展開するには到っていない．本論文では，3 章と 4 章において，オンライン型の訓練指向単一ショット映像撮影ナビゲーションシステムを提案した．しかし，撮影の全体的傾向や癖を見抜き，きめ細かい問題点の指摘や助言などは行えていない．なぜ提示したショットの型が必要であるかという，映像文法を背景とする編集を考慮した撮影概念を撮影者に定着させるにも至っていない．このような複数回撮影したショットなどの映像から，撮影者の癖や問題点の傾向を自動分析し，適切な問題点の指摘や助言を行って，きめ細かい指導と問題点を克服するための特別な訓練メニューなどを提案し，ガイドを示すのは，オフライン型の解析・提示システムの役割であり，オンライン型のシステムと連係すれば，より高い学習効果が得られるものと期待される．

また，映像文法には，本論文で説明したような単一ショットに関する映像文法だけでなく，少なくともシーンを撮影する際のシーンレベルや構文レベルの映像文法もある．5 章，6 章，7 章では，シーンレベルの映像文法を用いた映像編集支援システムの一種を提案した．撮影支援システムや編集支援システムの応用として，オンライン・オフライン型のシステムを統合し，高次の撮影・編集概念を組み込むことで，より高次の映像撮影・編集を学習するシステムや撮影・編集を指南するシステムが展望される．

10.1.2　ビデオカメラの高機能化とデジタルシューティングの応用

映像コンテンツの生成支援技術を市場で有用なものとするためには，大衆が映像コンテンツを制作する習慣が増え，それらの諸技術を必要とする基盤環境が整備されなくてはならない．そのような環境を実現する方法として，大衆に普及したビデオカメラをより便利なものにするアプローチが考えられる．本研究における，映像撮影訓練システムは，映像撮影の学習用のシステムとして提案したが，この機能をより高め，多角的に撮影をサポートすることで，映像コンテンツの制作を，民間レベルまでより活性化させる，その方向性に関する第一歩と言える．

また，ディジタルシューティングの特徴は，高解像度の映像を固定的に撮影し，その一部をトリミングすることで，仮想的なカメラワークや，ショットの切り替えを模擬する擬似的な撮影である．一方，これまでの観光などでは，記念として写真を撮ることが一つの趣向であったが，近年，映像を撮る傾向も高まっている．しかし，現地で映像を撮影しても，ビデオカメラを持った者の視点でしか映像は撮影できない．もし，映画的な俯瞰映像や，観光地の天候の良い日や，季節のごとの特色のある映像を自分の作品の中に取り込むことができれば，映像作品としての質が高まり，創作意欲が湧き，工夫された映像コンテンツが盛んに制作される起爆剤となる可能性がある．その際，観光地の映像は，観光地が用意した固定の映像ではなく，ディジタルシューティングによって，映像購入者の趣向が反映できる映像を提供する環境が整えられれば，より独自の作品を作成する意欲につながる可能性もある．

現代においては，無線通信技術が進展しており，現地に固定的に設置された複数の高解像度ビデオカメラとアドホックに映像をやりとりする環境も技術的には可能となるものと思われる．つまり，観光地に設置された高解像度のビデオカメラを見つけた

場合，その無線エリア内で，ビデオカメラと通信し，好みの映像をディジタルシューティングによって，現地で購入・取得し，持ち帰るということが可能となる．このとき，手持ちの撮影を支援するだけでなく，ディジタルシューティングを行う際の支援を映像文法により行う方法も考えられ，ビデオカメラに多角的なアドバイス機能，支援機能，通信機能を統合することで，映像コンテンツ制作環境の幅を広げることが可能となる．

10.1.3 創作支援と順列芸術

映像文法の研究を行っている点で，よく指摘される点が，芸術性との関係である．映像編集支援システムでは，決まりきった映像の接続法についてのみ，支援を行う印象がつきまとう．しかし，この断片の接続による順列表現の創作支援概念が存在する．Moles は，順列芸術 [110] という構想を示しており，本論文で示した編集支援システムの自動編集について，映像文法の概念を拡張していくことで，創作支援法を実現する可能性が考えられる．

順列芸術の観点を述べる前に，脳と芸術の関係を述べた経学者の Ramachandran の視点を述べる．Ramachandran によれば，芸術家には，共感覚者が多いとされる．その共感覚とは，例えば視覚的に数字を見ると色を感じる特殊感覚である．これは，色を感じる大脳皮質の部位と数字を認識する大脳皮質の部位が隣接しており，その大脳皮質の領域で，クロス配線などのミス配線が起きていることを原因とする説がある．この共感覚は，遺伝することが知られている．脳の角回という部位は，意味の処理に関与しているが，この部位は，聴覚と視覚と触覚の部位が隣接しており，この角回を損傷すると，隠喩が理解できなくなり，言葉の表面的な字義通りの意味しか理解できなくなる症例が報告されている．Ramachandran は，この領域について，人類の祖先が樹上生活をしていた際，樹の上を飛び敵から逃げ，獲物を見つける生活が，聴覚，視覚，触覚の連合強化を必要とし，この領域のクロス配線を起こし，優位に立った系統が今日のヒトであるとの可能性を述べている [109]．つまり，共感覚を脳のいずれかの部位で起こした者に芸術家が多いとすれば，芸術的能力の一端は，同種族が結び付けなかった情報を結びつける能力が高いことによるものとする仮説が成り立つ．つまり，思いもよらない結びつけを行う作業を支援することが一つの創作支援法と考えられる．

映像編集の観点で考えれば，撮影したショット群について，そのショットの数が多ければ多いほど，その組み合わせ方は膨大となり，編集に携わる者の思考の中では思いもつかない接続候補が存在する可能性が高まる．そして，その自動生成された候補の中で，より良い表現と思えるパターンが存在する可能性もある．このような思いもよらぬパターンを提示し，選択したパターンから，新たなアルゴリズムを生成し，人間と人間の創作活動を支援するコンピュータの在り方を論じる概念がある．それが Moles の順列芸術（art permutationnel）の構想の一部として存在する [110]．Moles は，美的知覚の情報理論の観点から,「コンピュータが，新たなアルゴリズムの実現と，その順列的応用の実現を組織的に関係づける芸術をもたらす」と考え，そこに導かれる新たな芸術領域として，いわゆる順列芸術を置いている．その順列芸術において，芸術家は，芸術アルゴリズムとして，感覚的要素の記号コードから構成される候補と，それらを組み合わせる一連の法則をプログラム上に設定し，コンピュータがその組み合わせの巨大な可能領域から引き出されるすべての作品を実現していくことになる [111]．そして，その中で偶然現れた順列表現が芸術家を喚起し，触発されて新たな芸術アルゴリズムの導入を試みることにより，芸術家とコンピュータが相互に影響を与え，相互に発展を遂げる好循環の創作活動が可能となる．

進化人類学者の Deacon は，人間の言語能力は，突然創発されたものではなく，共進化により，発達したものであるという仮説の立場にある [112].「共進化」とは，進化論における概念であり,「複数の種が相互に影響を与えながら環境への適応力を高める方向に進化すること」という定義である．たとえば被捕食種は，捕食種に捕らわれないよう進化し，捕食種はその進化に対応すべく自らを進化させる．このような選択的圧力を及ぼし合い，共に進化する関係を共進化と呼ぶのである．Deacon は，言語と脳は，共進化の関係によって成長・発達したという仮説を唱えている．これは，原始言語から始まり，脳の発達が，より複雑な言葉を可能とし，複雑な言葉がさらに脳を発達させるという共進化である．Deacon は言語と脳の共進化の議論として，文法能力が生得的ではありえないと論じる．もし，この仮説の通りだとすれば，順列芸術が暗示する人間とコンピュータの関係は，その表現法を共進化的に向上させる可能性を秘めている．本研究で示した映像編集支援システムにおける，自動編集の概念において，映像文法を動的に成長させる方式を導入することができれば，順列芸術のような創作支援法の一つとして実現できる可能性がある．

10.1.4 映像による概念辞書

映像コンテンツは，物語やニュース，話題や生活情報，文化・スポーツ活動などを伝えるだけのものではない．ソヴィエト・モンタージュ理論が目論んだ視覚的情報伝達言語としての技法に着目し，それらを契機とした応用を考えることができる．

例えば，百聞は一見にしかずという諺があるように，文章や言葉で説明するよりも，視覚的に伝えた方が分かりやすい情報がある．そのような観点に着目して，製品の説明書を映像として提供するための技術が研究されている．また，伝達したい内容によっては，言葉や文化の違う民族どうしで情報を伝え合う手法として，映像として伝えることにより，言葉や文化の壁を越えた視覚的共通言語となる可能性を秘めている．ソヴィエト・モンタージュ理論は，その可能性を探求したが，その方法論は，未完のまま過去の遺物となっている．その伝えたい内容を，視覚化する際，映像化する優れた支援技術が存在すれば，思ったことを言葉でなく，映像として伝える，視覚的意味表示装置という道具となりうる．さらには，ニュアンスがわかりにくい表現について，映像コンテンツとして表現する語学学習支援コンテンツにもなりうる．これは，語学学習の分野で映像を用いた学習コンテンツとして進展しつつある．

また，言語の自動翻訳技術は，意味変換の根底にある問題をうまく処理できない問題に多数直面しており，その問題を軽減する別角度の方法として，映像による情報伝達方法が考えられるかもしれない．一方，電子辞書が広汎しつつあるが，それらの辞書も，文字情報だけでなく，映像による概念提示法が有効になる可能性がある．

そのような環境を実現するためには，コンピュータのシステムとして，映像を簡易的に生み出すための支援技術が数多く必要になると思われる．音声言語から文字の発明，印刷技術の開発による言語に関する道具が数多く発明されたように，映像の視覚化支援技術により，映像による概念表現技法が当然となる時代も来るかもしれない．そのためには，まだ混沌として整理が十分行われていない，映像の性質を集約し，体系化し，応用することが必要となる．

10.2　映像文法のゆくえ

10.2.1　文法学の流れと映像文法

Aristoteles 以来，弁論や詩作をより効果的に伝達するために，言語表現を明瞭にする多数派の共通性を反映した「標準」を知り，規範を定めて言語を考察する方法が文法の礎となった．そして，文法学の一派は，多数派が受け入れた慣用であり，先人の慣用を検討した結果であり，多数派が用いるがゆえに何らかの特性を持ち，多数派が共有することで自然な対話ができ，多数派の一員となるための，お手本としての規範文法となる．ただし，規範における「正しい」,「正しくない」は,「正読，正しいとは何かを知るための学」となるインプット型の文法観，また,「正話，実践，より表現を高めるための術」となるアウトプット型の文法観へ偏りを見せる．

中世以来，正しく読み，正しく書くアウトプット型の規範的文法は，長きに渡り継承されてきたが，紀元前の古き文法では，むしろ文学的観点に焦点があてられ，雄弁家を育てるためのすぐれた表現法を体得するためのノウハウを集成した学問であった．その点において，現代の科学的言語研究における文法観は，言語の理を追求するためのインプット型に偏る系統が多く，焦点を絞った研究が行われていると言えるが，一般大衆に言葉や文法の謎を知るための学問として科学的言語研究の印象を強く与えている点で閉塞的であるとも言える．

一方，語学学習の分野では，長きに渡り実用的に規範文法のアウトプット型の方法が伝承されており，映像文法の一部としての映画文法は，規範文法を参考にして体系化が進めれた経緯が明らかとなっている．しかし，規範文法は，科学的言語研究が進めるように，文法や意味の正しい理解による，言語表現の「なぜ」には十分な説明を行えていない．その点で，語学学習の分野では，インプット型の説明が必要とされており，科学的言語学に期待がかけられている．

映像は，まだ 100 年程度の歴史が浅いメディアである．20 世紀初頭から中頃にかけて，映画言語，映画文法の理論が研究されたが，20 世紀前半は，古来の言語学が対象としてこなかった意味への本格的な探求の契機となる現代言語学の祖，Saussure の言語学が始まったばかりであり，言語とは何かについて，その定義が曖昧なまま，混沌の中での議論が進められた．言語学は，映画言語・文法の活発な議論が終息した 20 世紀中頃以降，生成文法が出現し，意味から自律した文法のみの研究により，それまで

の言語学であまり進展させられなかった統語論を進展させる契機となった．また，その生成文法の意味に関する扱いに不満を持つ系統が融合し，20 世紀末にかけて文法論と意味論を統合的に扱い，認知言語学が勢力を拡大しつつあり，脳科学の後押しもあって，人間の認識機構を研究対象に含む，意味の研究を進展させつつある．また，生成文法や認知言語学は，いずれも言語表現の「なぜ」を問う，インプット型に偏るが，選択体系機能文法 [113] のように，文法の機能の面に着目し，アウトプット型の研究を進める動きもある．つまり，21 世紀以降は，文法の観点からすれば，偏りを持ったインプット型，アウトプット型，それぞれの研究成果を取り込みながら，失われた文法観を再検討し，再融合する方法論が模索される可能性がある．なぜならば，言語学の課題に，言葉の生成の謎を問う次世代の課題が残されているからである．

このような状況は，映像の持つ文法的観点，意味の観点を研究する点においても，言語・文法学の研究成果を取り込む基盤が整いつつあることを示していると見ることができる．また，失われた文法観にこそ，映像文法を役立てる一つの側面がある．それは，映像の正しい生成という観点だけでなく，映像生成における雄弁家を育てるためのノウハウを集成した教育プログラムである．そして，それらの雄弁性をより創造的に発展させる，もしくは創造活動を支援する道具を生み出すことである．この観点から言えば，現代の文法観が偏りがちな，守らなければならない，規則ではなく，創作の活動を促す文法観に再度目を向け，それらを進展させるための方法論の創出に目を向ける機会が訪れていると見ることができる．

10.2.2　意味伝達に直接関わる映像文法の例

図 10.1 は，空間の自然な連続性を目論む点で，古典的デクパージュの一種であるが，カメラワークが直接意味に関係する規則の例である．図 10.1 にはカメラワークの前後に FIX 部がある映像文法で示した Type2 のショットである Shot1 と，FIX 部のみの Type1 のショットを接続する例である．ただし，Shot1 では，建物の全体が撮影された LS の状態から建物の入り口へ ZOOM-IN が行われており，次に接続する Shot2 に関する文法である．この場合，一般的には ZOOM-IN が建物の中に入るというコードとして我々の意識下に慣習化されており，次に接続される Shot2 は，上側の Shot2 にあるような建物の中のショットを接続すれば違和感はないが，下側の Shot2 のように，建物の外を撮

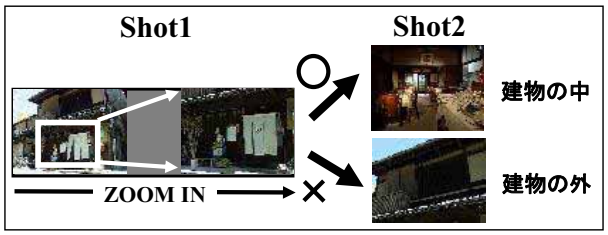


図 10.1　意味を直接伝達するカメラワークと映像文法

影したショットが接続されると，その接続の意味をどう解釈するかが曖昧になり，継ぎ目が浮き彫りになるため，視聴者にとって誤解や混乱が引き起きる可能性が高くなる．このように，図 10.1 の×に示した接続では，送り手の計画的な意図を担わせて撮影を工夫しないかぎり，空間の連続性について自然さが失われ，編集の存在を意識させてしまい，映像への没入状態から視聴者を現実へ引き戻す，悪しき要因と捉えられる．

ただし，この例で注目すべき点は，ZOOM-IN というカメラワーク自体に「入る」という意味の伝達機能があるということである．つまり，古典的デクパージュの方法に従った概念であっても，意味伝達の脇役のような，連続性という空間的関係の構成機能だけでなく，カメラワーク自体が意味の伝達に直接関与する例があるという事実があり，それに付随して，後続のショットの接続に制限を与える規則が存在することである．このような例は，どれほどあるのだろうか．

ソヴィエト・モンタージュ理論のような構成的意味生成の方法，古典的デクパージュのような連続体の連続性を維持する分節の方法など，映像を組織する抽象概念は，これだけなのだろうか．もしくは，まだ未開拓の方法論があるのだろうか．現在用いられれている手法を整理すれば，新たな体系的観点でその技法を捉え，映像の制作や教育，自動化技術に生かせるかもしれない．また，先人の見本的技法をまねるだけでは，古来の失われた文学的側面としての，より表現力を高めるためのノウハウ的文法観が失われたままになる．まだ眠っている映像の特性をつむぎ出す概念の探求は，これからの課題である．

10.2.3　人材育成と伝承

日本映画の巨匠，黒澤明監督は，モンタージュについて，きわめて非凡な才能を見せる巨匠として有名である [50]．黒澤は，晩年まで、映画産業の発展には科学が必要で

あるとしきりに訴えていたとされる．黒澤は次のように述べている.「今日,（ＴＶ隆盛に対し）映画の斜陽化は世界的な現象のように云われているが，その中で，アメリカ映画が隆盛を取り返しつつある理由は何か．アメリカ映画のバック・ボーンには，アメリカ映画芸術科学アカデミーという組織があり，映画は科学と密接に結びついた芸術である，という確固たる認識の上に立っているからだ」[114]．ハリウッドは，映画産業をただ芸術や娯楽として位置づけたのではなく，もとより，芸術・産業 (技術)・学術の融合を前提として発展してきたのである．

ただし，黒澤の想いに対し，学術の観点から逆に黒澤の功罪を指摘する声がある．日本のアニメーション業界を牽引する富野由悠季も，映画の文法を映像の原則と呼んで，映像制作の背後に存在する映像の性質についてアニメーション制作の立場から述べている [115]．富野は，映像の原則が現代の制作者の知識として不足していると嘆く．富野の視点によれば，黒澤監督や小津監督など，日本の巨匠が活躍した時代に，映像の持つ性質を熟知し，職人芸的に監督が映像の制作を進めたために，学術的に継承が可能な資料が残らず，次世代の映像制作者が育たなかったと指摘している．また，テレビの現場でも，特に，現代の若手制作者は映像文法を意識せず，制作する映像について，視聴者への配慮に欠けるとの熟年制作者の嘆きの声や，若手の映像表現に嘆く声がある．放送局との映像文法に基づく応用に関して行われた共同研究も，その観点が事の始まりであった．つまり，古典的デクパージュの方法は，少なくともテレビの世界で実用的に扱われているが，その配慮は薄れており，巨匠の編み出した高度な技法は，教育プログラムの形として十分に伝承されず，失われていることになる．

一方，編集行為が映像独自の表現法を見出すモンタージュの観点や，さらにソヴィエト・モンタージュの方法は，単に古い技法というだけななく，古典的デクパージュの中で慣習化された技法よりも，その扱いに高度な資質を要求されるため，その扱いの困難さから，避けられ，現代の映像作成法としては，伝承されにくい技法とも言える．また，ソヴィエト・モンタージュ理論を発展させた Eisenstein は，独自の映画理論を未完のまま数多く残しており，その膨大な資料から，新たな映像の技法が見出される可能性も残されている．

以上のように，映像の組織法は，古い手法としても，学術的に伝承可能な体系として整理されていない状況があり，新しい手法についても，まだ優れた制作者の直感に頼る側面が強い．また，古き手法の中で未完の手法が残されており，そのような手法

を掘り起こし，現代風に応用するための，教育プログラムとしての学術的体系化の試みの中で，映像文法は，文法学の動向を取り込みつつ，映像制作での雄弁家を育成するために，映像コンテンツ産業に貢献すべき基礎を築く役割を担っていると思われる．

謝辞

本研究を行うに当たって，常日頃より御指導を頂き，この研究を行う機会を与えて下さった有木康雄教授に深く感謝致します．さらに，本研究に関して，御意見及び御指導を下さった上原邦昭教授，研究に協力していただいた天野美紀氏に心から感謝致します．また，本研究で用いた映像文法の契機を与えていただいた，下條真司教授，毎日放送の春藤憲司氏，塚田清志氏，ならびに濱口伸氏，清瀬基氏，毎日放送の関係諸氏に深く感謝致します．

参考文献

[1] 稲蔭正彦: “ユビキタス社会のライフスタイルに適応するディジタルシネマ”, 映情学誌, **59**, 2 pp.208-209 (2005)

[2] 経済産業省商務情報政策局監修, ” デジタルコンテンツ白書 2004“, 財団法人デジタルコンテンツ協会

[3] 坂井滋和: “ディジタル時代の人材育成”, 映情学誌, **59**, 2 pp.210-213 (2005)

[4] 熊野雅仁:“映像の言語と文法 (1)–デジタルコンテンツ時代の人材育成–”，龍谷理工ジャーナル，No.48, 第 17 巻 2 号，pp.9-13 (2005)

[5] M.Kumano, Y.Ariki, K.Shunto, K.Tsukada: “Video Editing Support System Based on Video Content Analysis”, Proc. of Asian Conference on Computer Vision (ACCV) pp.628-633 (2002).

[6] 熊野雅仁:“映像の言語と文法 (3)–モンタージュ・編集・意味作用–”，龍谷理工ジャーナル，No.50, 第 18 巻 2 号，pp.12-22 (2006)

[7] 熊野雅仁:“映像の言語と文法 (7)–映画文法・時空間の分節とデクパージュと文学–”，龍谷理工ジャーナル，No.54, 第 20 巻 1 号，(2008) 掲載決定済

[8] スティーブ・ブランドフォード，バリー・キース・グラント，ジム・ヒリアー，フィルム・スタディーズ事典–映画・映像用語のすべて，フィルムアート社 (2004)

[9] ジェイムス・モナコ, “映画の教科書”, フィルムアート社 (1983)

[10] Noel Burch，Theory of Film Practice, Princeton (1981)

[11] Michael, Ian. 1970. English Grammatical Categories and the Tradition to 1800. Cambridge: Cambridge University Press.

[12] 宮脇正孝，“ディオニュシオス・トラクスの文法 (*Téchnē grammatikḗ*) について”，ASTERISK, Vol.X，No.4，pp.243-250 (2001)

[13] Ｊ・オーモン, Ａ・ベルガラ, Ｍ・マリー, Ｍ・ヴェルネ, “映画理論講義”, 勁草書房 (2000)

[14] Raymond Spottiswoode, ”A Grammar of the Film“, University of california press

[15] レイモンド・スポティスウッド, ” 映畫の文法 : 映畫技巧の分析 “, 映畫評論社 (1936)

[16] 酒井優子, “言語伝達説と言語認識説の系譜”, リーベル出版 (2002)

[17] 宮脇正孝, “Anselm Bayly の品詞分類の変遷について”, Asterisk, Vol.III, No.7, pp.367-379 (1994)

[18] 熊野雅仁:“映像の言語と文法 (4)–文法とは–”, 龍谷理工ジャーナル, No.51, 第 18 巻 2 号, pp.1-12 (2006)

[19] 宮脇正孝, “クインティリアヌスの文法概念”ASTERISK, Vol.X, No.3, pp.175-180 (2001)

[20] ジェフ・ウィリアムズ, 英語教師のための–機能文法入門, リーベル出版 (2002)

[21] Roy Thommpson, “Grammar of the shot”, Focal Press (1998)

[22] Dan Ablan, [digital] CINEMATOGRAPHY & DIRECTING 日本語版-3DCG クリエータのための映画撮影術と監督術-, 株式会社ボーンデジタル (2003)

[23] 純丘曜彰, “エンターテイメント映画の文法–ヒットを約束する脚本からカメラワークまで–”, フィルムアート社

[24] ジョージ・レイコフ, 認知意味論–言語から見た人間の心, 紀伊国屋書店 (1993)

[25] 伊藤康児, “カテゴリーの研究–Rosch を中心とする研究の概観–”, 名古屋大學教育學部紀要 教育心理学科, Vol.27, pp.17-32 (1980)

[26] 日本映画・テレビ編集協会編, “図解映像編集の秘訣”, 玄光社 MOOK (1999)

[27] 熊野雅仁, 有木康雄, 上原邦昭: “実時間カメラワーク評価に基づく単一ショット訓練指向型オンライン映像撮影ナビゲーションシステム–映像文法を背景とした映像撮影学習システムに向けて–”, 映情学誌, **61**, 8, pp.1159-1167 (2007)

[28] http://www-nlpir.nist.gov/projects/trecvid/

[29] Paul Over, Tzveta Ianeva, Wessel Kraaij, and Alan F. Smeaton: “TRECVID 2005 - An Overview”, In Proc. of TRECVID 2005, 2005. NIST, USA (2006)

[30] B.K.P Horn and B.G. Schunck:“Determining optical flow”, Artificial Intelligence, **17**, pp.185-203 (1981)

[31] 井宮淳: “動画像理解の数理”, 情報処理学会研究報告 CVIM, 5, pp.137-150 (2006)

[32] 興梠正克, 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, **J82-D-II**, 7, pp.1161-1170 (1999)

[33] Michael A. Smith, Takeo Kanade: ”Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques”, In Proc. of CVPR1997, pp.775-781 (1997)

[34] Ruggero Milanese, Frederic Deguillaume, Alain Jacot-Descombes: “Video Segmentation and Camera Motion Characterization Using Compressed Data”, In Proc. of SPIE, vol.3229, Multimedia Strage and Archiving Systems II, pp.79-89 (1997)

[35] Rong Jin, Yanjun Qi, Alexander Humptmann: “A Probabilistic Model for Camera Zoom Detection”, In Proc. of ICPR2002, pp.859-862 (2002)

[36] Ralph Ewerth, Martin Schwalb, Paul Tessmann, Bernd Freisleben: “Estimation of Arbitrary Camera Motion in MPEG Videos”, In Proc. of ICPR2004, pp.512-515 (2004)

[37] Bernd Jahne: Spatio-Temporal Image Processing: Theory and Scientific Applications. Number 751 in Lecture Notes in Computer Science. Springer (1993)

[38] 阿久津明人，外村佳伸: “投影法を用いた映像の解析方法と映像ハンドリング”, 信学論, **J79-D-2**, 5, pp.675-686 (1996)

[39] 長坂晃朗, 宮武孝文: “輝度投影相関を用いた実時間ビデオモザイク” 信学論, **J84-D-2**, 10, pp.1572-1580 (1999)

[40] K.Uehara, M.Amano, Y.Ariki, M.Kumano: “ Video Shooting Navigation System by Real-Time Useful Shot Discrimination Based on Video Grammar ”, Proc. of ICME2004 (International Conference on Multimedia and Expo), CD-ROM, 2004.

[41] Chong-Wah Ngo, Ting-Chuen Pong, Hong-Jiang Zhang, Roloand T. Chin: “Motion Characterization by Temporal Slices Analysis”, In Proc. of CVPR2002, pp.768-773, (2002)

[42] B.Barry and G.Davenport: “Documenting life: Videography and common sense” In Proceedings of IEEE International Conference on Multimedia. (ICME 2003), Baltimore, MD pp.1-4 (2003).

[43] B.Adams and S.Venkatesh: “Situated Event Bootstrapping and Capture Guidance for Automated Home Movie Authoring”, In ACM International Conference on Multimedia, Singapore, pp.754-763 (2005).

[44] B.Adams and S.Venkatesh: “Director in your Pocket–Holistic help for the Hapless Home Videographer–”, In ACM International Conference on Multimedia, New York, pp.460-463 (2004).

[45] 熊野雅仁，有木康雄，上原邦昭: “輝度投影相関と二分化テンソルヒストグラムを用いたオンライン処理向けカメラワーク解析法の精度向上–訓練指向型オンライン映像撮影ナビゲーションシステム”, 映情学誌, **61**, 8, pp.1150-1158 (2007)

[46] 坪見博之，苧阪直行:“視覚的注意のトップダウン制御の脳内表現”, 心理学評論, **49**, 2 pp.321-340 (2006)

[47] 熊野雅仁, 有木康雄, 春藤憲司，塚田清志：“映像文法に基づいた映像編集支援システムのための使用可能なショット区間の自動抽出‘ ’; 映情学誌, **57**，7，pp.829–839 (2003)

[48] Andreas Girgensohn and John Borecxk, “A Semi-automatic Approach to Home Video Editing,” Proc. of UIST ’00, ACM Press, pp.81–89, (2000).

[49] 土橋 健太郎，小舘 亮之，富永 英義: “手ぶれを考慮したカメラワーク検出に関する検討” 信学総大，D-12-56，pp.223，(2001):

[50] Arijon, D., Grammar of Film Language, Hollywood, CA., Silman-James Press (1976) (邦訳: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))

[51] 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.63-70,2001-07.

[52] 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”, 電子情報通信学会論文誌, Vol.J81-D-II,No.7,pp.1564-1575

[53] 金子敏充，堀修: ” ゆう度比検定を用いた MPEG ビットストリームからの動画像カット検出手法”: 電子情報通信学会論文誌, Vol.J82-D-II,No.3,pp.361-370

[54] M.J.Swain and K.H.Ballard: “Color indexing”, IJCV, 7, pp.11-32, (1991).

[55] Liu,D.,and Kubala,F.:Fast Speaker Change Detection for Broadcast News Transcription and Indexing,Erurospeech99, pp.1031-1034, (1999).

[56] http://avs.kddlabs.co.jp/mpeg/mpfs/index.html

[57] Michael A.Smith,Takeo Kanade: ”Video Skimming and Characterzation through the Combination of Image and Language Understanding Techniques”,CVPR1997,pp.775-781, (1997).

[58] 興梠 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, J82-D-II, 7, pp.1161-1170, (1999).

[59] 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.

[60] 永田, 徳平, 山口, 山本, 熊野, 有木, 春藤, 塚田: “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”, 信学総大,D-12-138, p.305, (2001).

[61] 柳井啓司: “一般物体認識の現状と今後”, 電子情報通信学会技術研究報告, PRMU, 106(229), pp.121-134 (2006-9)

[62] 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－”, 電子情報通信学会論文誌 D-II,Vol.J81,No.9,pp.2035-2042,1998.

[63] Tzi-cker Chiueh and Tulika Mitra: “Zodiac: A History-Based Interactive Video Authoring System,” Proc. of ACM Multimedia ’98, ACM Press, pp.435–443, 1998.

[64] Hari Sundaram and Shih-Fu Chang: “Condensing Computable Scenes Using Visual Complexity and Film Syntax Analysis,” Proc. of ICME 2001, pp.389–392, 2001.

[65] 田中, 鈴木他: “高精細度映像 (WHD:Wide/Double HD) 伝送システム”, 信学論, J84-D-2, 6, pp.1094-1101 (2001.06)

[66] http://www.megavision.co.jp/

[67] 林正樹: “バーチャルカメラ”, 信学誌, 81, 3, pp.244-246 (1998.03)

[68] 大田友一，北原格，斉藤英雄，秋道慎志，尾野徹，金出武雄: “仮想化現実技術による自由視点映像スタジアムの構築”, 映情学技告, 25, 76, pp.57-62 (2001.11)

[69] Takayoshi Koyama, Itaru Kitahara, Yuichi Ohta: “Live Mixed-Reality 3D Video in Soccer Stadium”, ISMAR 2003: pp.178-187.

[70] 加藤大一郎: “新しい番組制作支援技術 知的ロボットカメラと放送番組への応用”, NHK 技研 R & D,No.48, pp.34-47 (1998)

[71] 加藤, 山田他: “被写体を追尾撮影時の放送カメラマンのカメラワーク特性分析 (＜論文小特集＞生体計測応用)”, テレビ学誌, 50, 12 pp.1941-1948 (1996.12)

[72] 松本圭介，須藤智，斉藤英雄，小沢慎治: “サッカーシーンにおけるボール追跡に基づく最適視点決定システム”, 信学技告 PRMU2000, 6 pp.29-36 (2000.06)

[73] P. Xu, L. Xie, S-F. Chang, A. Divakaran, A. Vetro, and H. Sun: “Algorithms and system for segmentation and structure analysis in soccer video”, Proc. IEEE Int’l. Conf. on Mult. and Expo (ICME), pp.928–931 (2001)

[74] L. Xie, S-F. Chang, A. Divakaran, and H. Sun: “Structure analysis of soccer video with Hidden Markov Models”, Proc. IEEE Int’l. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), (2002)

[75] Kongwah Wan, Xin Yan, Xinguo Yu, Changsheng Xu: “Real-time goal-mouth detection in MPEG soccer video”, ACM Multimedia 2003: pp.311-314.

[76] Hyunwoo Kim, Ki-Sang Hong: “Soccer Video Mosaicing Using Self-Calibration and Line Tracking”. ICPR 2000: pp.1592-1595.

[77] Xinguo Yu, Changsheng Xu, Hon Wai Leong, Qi Tian, Qing Tang, Kong Wah Wan: “Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video”. ACM Multimedia 2003, pp.11-20.

[78] Okihisa Utsumi, Koichi Miura, Ichiro Ide, Shuichi Sakai, Hidehiko Tanaka: “An object detection method for describing soccer games from video”, Proc. 2002 IEEE Intl. Conf. on Multimedia and Expo (ICME2002), vol.1, pp.45-48 (Aug. 2002)

[79] A. Woudstra, D. D. Velthausz, H. J. G. de Poot, F. Moelaert El-Hadidy, Willem Jonker, Maurice A. W. Houtsma, R. G. Heller, J. N. H. Heemskerk: “Modeling and Retrieving Audiovisual Information: A Soccer Video Retrieval System”, Multimedia Information Systems 1998: pp.161-173.

[80] Dennis Yow,Boon-Lock Yeo,Minerva Yeung,Bede Liu: “ANALYSIS AND PRESENTATION OF SOCCER HIGHLIGHTS FROM DIGITAL VIDEO”, ACCV’95,pp.499-503.

[81] Jurgen Assfalg, Marco Bertini, Carlo Colombo, Alberto Del Bimbo, Walter Nunziati: “Automatic Interpretation of Soccer Video for Highlights Extraction and Annotation”, SAC 2003: pp.769-773

[82] A. Ekin, A. M. Tekalp, and R. Mehrotra: “Automatic soccer video analysis and summarization”, IEEE Trans. on Image Processing, 12, 7 pp.796-807 (2003)

[83] Ngoc Thanh Nguyen, Tuoung Cong Thang, Tae Meon Bae, Yong Man Ro: “Soccer Video Summarization System Based on Hidden Markov Model with Multiple MPEG-7 Descriptors”, CISST 2003: pp.673-678.

[84] 大西正輝，泉正夫，福永邦雄: “ディジタルカメラワークを用いた自動映像生成”, 画像の認識・理解シンポジウム, MIRU2000, pp.I-331-I-336, Jul, (2000)

[85] Saaty, T.L.: “The Analytic HierarchyProcess”, McGraw-Hill,1980.

[86] 木下栄蔵: “AHP の理論と実際”, 日科技連出版社

[87] 知念賢一, 吉田豊一，山口英，香取啓志: “全国高等学校野球選手権大会 Internet ライブ中継実験報告”，映情学誌, 51, 6, pp.925–930 (1997)

[88] 橋本隆子，白田由香利，真野博子，飯沢篤志: “TV 受信端末におけるダイジェスト視聴システム”，情報処理学論, 41, SIG3, pp.71–84 (2000)

[89] 田中清，阿久津明人，外村桂伸，秦泉寺浩史: “見たいシーンを見逃さないライブ中継 LiveWatch”，信学技報, パターン認識・メディア理解, PRMU2002-30, pp.51–56 (2002).

[90] 馬場口登: “メディア理解による映像メディアの構造化”, 信学技報, パターン認識・メディア理解, PRMU1999-42, pp.39–46 (1999)

[91] 重森 猛, 金子 剛志, 緒方 淳, 藤本 雅清, 有木 康雄, 塚田 清志, 濱口 伸, 清瀬 基: “音響・言語適応処理を用いたスポーツ実況中継音声の認識 〜ハイライトシーン検出への応用〜”, 信学技報, 音声研究会, SP2003-166, pp.3–40 (2003)

[92] 遠藤斉，片岡良治: “カメラモーションに基づく類似動画像検索”，信学技報, データ工学, DE, 99, 202, pp.147–152 (1999)

[93] 片岡良治，遠藤斉: “サンプルマッチングによる動画インデクシング支援方式”，信学技報, 画像工学, IE, 99, 179, pp.59–66 (1999)

[94] 舘山公一，川嶋稔夫，青木由直: “野球中継におけるシーン検索”，第３回知能情報メディアシンポジウム論文集：pp.195–202（1997）

[95] 舘山公一，川嶋稔夫，青木由直: “動作スポッティングによるシーン検索, 情報処理学会研究報告”, CVIM, コンピュータビジョンとイメージメディア, 97, 70, pp.115–122 (1997)

[96] Zhang, Dongqing and Shih-Fu Chang: “Event Detection in Baseball Video Using Superimposed Caption Recognition”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM ’02), pp.1–6 (2002)

[97] 広部一弥, 牛尼剛聡，酒井宏治，孫魯英，渡邊豊英: “イベントと状況変化の依存関係に基づいた野球中継のインデキシング支援”，情報処理学会研究報告. データベース・システム研究会報告, 98, 34, pp.87–94 (1998)

[98] Yong Rui, Anoop Gupta, Alex Acero: “Automatically Extracting Highlights for TV Baseball Programs”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM ’00), pp.105–115 (2000)

[99] 金山智一，瀧本裕一，小西修: “データマイニング法による映像情報の内容検索”，情報処理学会，研究報告，人文科学とコンピュータ, CH, 99, 59, pp.35–42 (1999)

[100] 新田直子, 馬場口登, 北橋忠宏: “言語と画像の情報統合によるスポーツ映像からの人物・アクション・イベント抽出”, 電子情報通信学会技術研究報告, パターン認識・メディア理解, PRMU1999-256, 99, 709, pp75–82 (2000)

[101] 山本拓, 佐藤宏介, 千原國宏: “野球中継映像における各種プレイシーンの自動検索/編集システム”, 2000 年電子情報通信学会総合大会講演論文集情報・システム 2, D12-77, pp.247 (2000)

[102] Huang-Chia, Chung-Lin Huang: “A Semantic Network Modeling For Understanding Baseball Video”, Proc. of IEEE Int’l Conf. on Acoustics,Speech,and Signal Processing(ICASSP ’03), V, pp.820–823 (2003)

[103] Wei Hua, Mei Han and Yihong Gong: “Baseball Scene Classification Using Multimedia Features”, Proceedings of the 2002 IEEE International Conference on Multimedia and Expo(ICME ’02), I, pp.821–824 (2002)

[104] Mei Han, Wei Hua, W. Xu and Yihong Gong: “An Integrated Baseball Digest System Using Maximum Entropy Method”, Proceedings of the 8th International ACM Conference on Multimedia(ACM MM ’02), pp.347–350 (2002)

[105] Peng Cheng, Mei Han, and Yihong Gong: “Extract Highlights From Baseball Game Video With Hidden Markov Models”, Proceedings of IEEE International Conference on Image Processing (ICIP ’02) I, pp.609–612 (2002)

[106] 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, J81, D-II, 7, pp.1564–1575 (1998)

[107] http://w3-mcgav.lab.kdd.co.jp/mpeg/mpfs/indexe.html

[108] 重森 猛，金子 剛志，緒方 淳，藤本 雅清，有木 康雄，塚田 清志，濱口 伸，清瀬 基: “ 音響・言語適応処理を用いたスポーツ実況中継音声の認識 〜ハイライトシーン検出への応用〜 ”, 電子情報通信学会，音声研究会，SP2003-166，pp.33-40，2003-01.

[109] 熊野雅仁:“映像の言語と文法 (6)–言語・文法の自律と言語・文法能力獲得へのミラーニューロン・音楽の関与–’，龍谷理工ジャーナル，No.53, 第 19 巻 2 号，pp.8-22 (2007)

[110] Moles, Abraham A., Art et Ordinateur, blusson (1971)

[111] 吉積健, ” メディア時代の芸術 “, 勁草書房

[112] テレンス・W・ディーコン，ヒトはいかにして人となったか，新曜社 (1999)

[113] M.A.K. ハリデー，機能文法概説–ハリデー理論への誘い–，くろしお出版 (2001)

[114] 黒澤明, “蝦蟇の油–自伝のようなもの–”, 岩波書店, pp.312–314 (1990)

[115] 富野由悠季:“映像の原則”, キネマ旬報社

関連論文

学術論文

[1] 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，Vol.J85-D-I, No.7，59–602，2002.

[2] 熊野雅仁, 有木康雄, 春藤憲司，塚田清志: “映像文法に基づいた映像編集支援システムのための使用可能なショット区間の自動抽出 ”，映像情報メディア学会誌 Vol.57，No.7，829–839，2003.

[3] 天野美紀, 上原邦昭, 熊野雅仁, 有木康雄, 下條真司，春藤憲司，塚田清志: “映像文法に基づく映像編集支援システム ” 情報処理学会論文誌，第 44 巻，第 3 号，915–924，2003.

[4] 熊野雅仁，有木康雄，塚田清志: “ボールと選手に着目したディジタルカメラワークの実現法–ディジタルシューティングによるサッカー解説映像生成システムに向けて–”，映像情報メディア学会誌，Vol.59，No.2，271–278，2005.

[5] 熊野雅仁，有木康雄，塚田清志: “野球中継のハイライトシーン実時間配信を目的とした特徴のマイニングによるＰＣシーンの自動検出”，映像情報メディア学会誌，Vol.59，No.1，77–84，2005.

[6] 熊野雅仁, 有木康雄，上原邦昭: “輝度投影相関と二分化テンソルヒストグラムを併用したオンライン処理向けカメラワーク解析法の精度向上–訓練指向型オンライン映像撮影ナビゲーションシステム–” 映像情報メディア学会誌，Vol.61，No.8，1159–1167，2007.

[7] 熊野雅仁, 有木康雄, 上原邦昭: “実時間カメラワーク評価に基づく単一ショット訓練指向型オンライン映像撮影ナビゲーションシステム–映像文法を背景とした映像撮影学習システムに向けて–” 映像情報メディア学会誌, Vol.61, No.8, 1150–1158, 2007.

国際会議

[1] M.Kumano, Y.Ariki, K.Shunto and K.Tsukada: “Video Editing Support System Based on Video Content Analysis”, Proc. of Asian Conference on Computer Vision (ACCV’02), 628–633, 2002.

[2] M.Kumano and Y.Ariki: “Automatic Useful Shot Extraction for a Video Editing Support System”, Proc. of Machine Vision Applications (MVA’02), 310–313, 2002.

[3] M.Kumano, Y.Ariki, M.Amano, K.Uehara, K.Shunto and K.Tsukada: “Video Editing Support System Based on Video Grammar and Content Analysis”, Proc. of 16th Inter’l Conf. on Pattern Recognition (ICPR’02), 1031–1036, 2002.

[4] M.Kumano, Y.Ariki, K.Shunto and K.Tsukada: “Automatic Shot Size Indexing for a Video Editing Support System”, Proc. of 3rd-International Workshop on Content-Based Multimedia Indexing (CBMI’03), 57–62, 2003.

[5] K.Uehara, M.Amano, Y.Ariki, M.Kumano: ”Video Shooting Navigation System by Real-Time Useful Shot Discrimination Based on Video Grammar”, Proc. of International Conference on Multimedia and Expo (ICME’04), CD-ROM, 2004.

[6] M.Kumano, Y,Ariki, K.Tsukada: “A Method of Digital Camera Work Focused on Players and a Ball –Toward Automatic Contents Production System of Commentary Soccer Video by Digital Shooting – ”, Proc. of Pacific-Rim Conference on Multimedia (PCM’04) III, 466–473, 2004.

[7] M.Kumano, Y.Ariki, K.Tsukada, S.Hamaguchi and H.Kiyose: “Automatic Extraction of PC Scenes Based on Feature Mining for a Real Time Delivery System

of Baseball Highlight Scenes", Proc. of International Conference on Multimedia and Expo (ICME'04), CD-ROM 23–25, 2004.

[8] M.Kumano, K.Uehara, Y.Ariki: "Online Training-Oriented Video Shooting Navigation System based on Real-Time Camerawork Evaluation", Proc. of International Conference on Multimedia and Expo (ICME'06), CD-ROM, 2006.

商業誌論文

[1] 熊野雅仁: "野球中継のハイライトシーン配信システム-画像特徴のマイニングによるＰＣシーン実時間自動検出法-", 画像ラボ，Vol.16, No.8, pp.6-13 (2005)

解説

[1] 熊野雅仁: 映像の言語と文法 (1)–デジタルコンテンツ時代の人材育成–，龍谷理工ジャーナル，No.48, 第 17 巻 2 号，pp.9-13 (2005)

[2] 熊野雅仁: 映像の言語と文法 (2)–言語・非言語・ソシュール言語学・動物の言語–，龍谷理工ジャーナル，No.49, 第 17 巻 3 号，pp.7-12 (2005)

[3] 熊野雅仁: 映像の言語と文法 (3)–モンタージュ・編集・意味作用–，龍谷理工ジャーナル，No.50, 第 18 巻 1 号，pp.12-22 (2006)

[4] 熊野雅仁: 映像の言語と文法 (4)–文法とは–，龍谷理工ジャーナル，No.51, 第 18 巻 2 号，pp.1-12 (2006)

[5] 熊野雅仁: 映像の言語と文法 (5)–脳・記憶・意識・知の根底から見た言語–，龍谷理工ジャーナル，No.52, 第 19 巻 2 号，pp.5-16 (2007)

[6] 熊野雅仁: 映像の言語と文法 (6)–言語・文法の自律と言語・文法能力獲得へのミラーニューロン・音楽の関与–，龍谷理工ジャーナル，No.53, 第 19 巻 2 号，pp.8-22 (2007)

[7] 熊野雅仁: 映像の言語と文法 (7)–映画文法・時空間の分節とデクパージュと文学–, 龍谷理工ジャーナル，No.54, 第 20 巻 1 号 (2008) 掲載決定

学術講演

[1] 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), pp.31-pp38, 2001-09. (電子情報通信学会オフィス研究会　オフィス研究会賞　2001)

[2] 熊野雅仁, 有木康雄: “映像編集支援システムのための使用可能ショット自動抽出”, 電子情報通信学会技術研究報告, パターン認識とメディア理解研究会, PRMU2002-31, pp.1-8, 2002.6.

[3] 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ショットサイズ自動付与”，画像の認識・理解シンポジウム MIRU2002，II，pp.215-222，2002-7.

[4] 熊野雅仁，神崎伸夫，藤本雅清，有木康雄，塚田清志，濱口伸，清瀬基: “野球中継のハイライトシーン実時間配信を目的としたＰＣシーンの自動検出” 電子情報通信学会, パターン認識 ・メディア理解, PRMU2003-18,pp.27-34,2003/07

[5] 熊野雅仁，岩本健，有木康雄，塚田清志: “ボールと選手に着目したディジタルカメラワークの実現法–ディジタルシューティングによるサッカー解説映像生成システムに向けて– ”, 画像の認識理解シンポジウム (MIRU2004)，SUP-C1-12，— 341- — 346，2004-07.

[6] 熊野雅仁，天野美紀，有木康雄，上原邦昭: “映像文法に基づいた実時間使用可能ショット識別による撮影ナビゲーションシステム ”, 電子情報通信学会技術研究報告, PRMU, パターン認識・メディア理解, Vol.104, No.369, pp.1-6, 2004-10

[7] 熊野雅仁, 天野美紀, 有木康雄, 上原邦昭, 春藤憲司, 塚田清志: “映像文法基盤の訓練指向型単一ショット撮影ナビゲーションシステム ”, 第 2 回デジタルコンテンツシンポジウム, CD-ROM, 2006-06

[8] 熊野雅仁: “映像指南カメラ”, 第 1 回 BIZ-NET 研究会, 2007-06

[9] 熊野雅仁: “映像文法を背景とした訓練指向オンライン単一ショット映像撮影ナビゲーションカメラの試作開発”, 映像情報メディア学会，コンシューマエレクトロニクス研究会 2007-10

[10] 熊野雅仁: “映像文法を背景とした映像撮影・編集支援技術”, 京都産学公連携フォーラム 2007, 2007-11

[11] 熊野雅仁: “映像文法を背景とした撮影・編集支援技術の提案”, 滋賀県産学官ニーズ・シーズプラザ, 2008-1