映像文法に基づく
自動撮影システムに関する研究

平成16年度

井 上　亮 文

目 次

図 目 次

表 目 次

第1章　序論

1.1　はじめに

20 世紀では，映像は映画とテレビを通して供給されてきた．最古の映像メディアである映画は，テレビの登場と普及により一時期落ち込みもあったが，今なお繁栄を続けており，現在でも多くの作品が上映され続けている．この映画が映画館に行かなくては見られなかったのに対し，テレビは各家庭へと普及し，現在では最も影響力のある映像メディアとなった．ほとんどの映像はテレビを通して供給されてきたといってもよい．

そして現在，21 世紀を迎え，ディジタル多チャンネル時代に突入した．従来のテレビや映画に加えて，BS デジタル放送，2003 年から地上波デジタル放送も開始され，チャンネル数が飛躍的に増加した．放送業界に限らず，インターネット，ゲーム，携帯端末などあらゆるメディアで映像が配信されるようになっている．

また，映像の用途も広がった．かつての映像の用途は，映画やテレビが主流の時代では娯楽，記録，ニュースといった用途がほとんどであったが，現在では企業が DVD などのパッケージメディアを通じて自社や製品の紹介映像を配布することも珍しくない．テレビ会議も家庭にまで普及し，大学ではインターネットを通じて授業の映像を中継する遠隔講義が始まるなど，コミュニケーション用途でも映像の果たす役割が重要になってきている．

このように映像の供給先，用途ともに急速に拡大する一方で，肝心の映像をどのように作っていくかが課題になっている．この課題は，何も放送業界に限ったことではない．撮影には依然としてカメラの台数と同じだけのカメラマンを用意する必要があり，さらにその編集には膨大な時間を要する．そこで，このような負担やコストを軽減するため，撮影を自動化しようという試みがなされている．

1.2　本論文の目的

従来の典型的な自動撮影システムでは，移動物体の追跡など，被写体の変化にどのように対応するかに重点が置かれてきた．従ってそのカメラワークは “被写体を捉え続ける” という基本的なタスクの遂行を重視したものになる．しかし，そのような映像は，我々が普段目にしている映画やテレビの映像と比べて単調であったり，時に機械的で見づらいものであったりする．今後，自動撮影技術が普及していくためには，そのようなタスク重視型のカメラワークから一歩進んで，“どのように撮影すればよいか” という，映像の見やすさ，面白さといった視点に立った演出志向のカメラワークが必要になる．

この演出に関しては，映画やテレビの撮影現場では，映像の意図を効果的に伝えるための知識が存在する．この知識の集大成を映像文法 [1] と呼ぶ．本論文では，複数台のカメラを映像文法に基づいて協調動作させ，効果的に演出された映像を自動的に撮影するシステムの実現を目的とする．ここで，すべての撮影対象を演出可能なシステムはあらゆる演出用カメラワークを用意する必要があり現実的とはいえない．本研究では撮影対象が大き

く分けて (1) 次に何が起こるかを判断することができない場合（シナリオの無いシーン），(2) 次に何が起こるかを事前に判断することができる場合（シナリオのあるシーン），に分類できることに着目した．そして，それぞれに該当する具体的な撮影対象を設定し，その技術課題を解決していくアプローチを取った．

1.3　本研究の概要

1.3.1　対面会議の自動撮影

まず，シナリオの無いシーンの例として対面会議を取り上げた．会議では次に誰が発言するのか分からないため，この研究では映像文法に基づいた演出用カメラワークをリアルタイムに生成・実行することに焦点を置いている．

通常我々は会議をする際，円卓もしくは四角形の机を囲んで議論することが多い．このような会議を複数のカメラで撮影する場合，発言する参加者の変化に応じてカメラの映像を切替える（スイッチングする）必要があるが，その方法によっては映像に急激な変化が生じ，視聴者が混乱したり，非常に見づらい映像になってしまう．本研究では，映像文法を “正確で分かりやすい” 映像を制作するための技法としてとらえ，人物の位置関係を明確にする映像理論であるイマジナリーラインに注目した．この映像理論に基づいて複数台のカメラを協調制御し，参加者同士の対話シーンを見やすく演出する撮影手法を提案する．手動でスイッチングを行った映像との比較実験を通じて提案手法の映像表現における有効性を確認する [2, 3, 4] .

1.3.2　オーケストラ演奏の自動撮影

次に，シナリオのあるシーンの例としてオーケストラ演奏を取り上げた．オーケストラでは楽譜に，“いつ”，“どの楽器が”，“どのような音を演奏するか” が記述されているため，この研究ではシナリオの情報をもとにして映像文法に基づいた演出用カメラワークを自動で生成することに焦点を置いている．

オーケストラの撮影では，用意できるカメラの台数に比べて被写体となる楽器の数が多い上，カメラを設置できる場所にも制限がある．そのため，カメラワークが事前に適切に計画されていないと，編集段階で必要なショットが撮影されていない，似たようなショットばかり撮影している，といった状況が発生し，効果的な映像を編集することができない．本研究では，映像文法を “バラエティに富んだショット” を撮影するための技法としてとらえ，被写体の種類と構図の変化に着目した．そして，複数台のカメラが協調し，なるべく多くの被写体を様々な構図で撮影するようなカメラワークを楽譜から自動的に生成

する手法を提案する．別の手法で計画されたカメラワークとの比較実験を通じて，本方式で計画されるカメラワークが映像表現の向上に一定の効果があることを示す [5, 6, 7].

1.4　本論文の構成

本論文は，以下の 5 章で構成されている．

第 1 章では，本研究の目的および概要について述べた．

続く第 2 章では，本研究の背景と位置づけについて述べる．まず背景として，映像制作の成り立ちと，その分類について言及する．次に，その分類に基づいて関連研究を整理する．最後に，それら関連研究との比較から本研究の位置づけを明確にする．

第 3 章では，シナリオが無い，その場の状況に応じて進行が決定する場面の撮影について議論する．数人の参加者が一地点に集まって議論する対面会議を撮影対象とし，映画の撮影技法を考慮しながらこれを見やすい映像に編集するための手法について述べる．

第 4 章では，シナリオが存在する，進行があらかじめ決定している場面の撮影について議論する．オーケストラ演奏を対象とし，楽譜をシナリオとして利用して限られた台数のカメラを効果的に被写体に割り当てる手法について述べる．

最後の第 5 章は，結論として本研究を総括するとともに，今後の展望について言及する．

第2章　研究の背景と位置づけ

2.1　はじめに

本章では，研究の背景と位置づけについて述べる．本研究は，自動撮影に “映像文法” を組み込み，シーンを魅力的・効果的に演出した映像を自動的に生成するという視点に立って行われた．ここで言う魅力的・効果的な映像とは次のような条件を満たす映像のことである．

見飽きない映像　視聴者を映像に惹きつけ，興味を持たせる映像であること．

分かりやすい映像　シーンの状況が理解しやすく，誤解を生じさせない映像であること．

これを実現するためには，単に映像を “撮る” だけでなく，“作る” ことが必要になる．そこで，まずは研究の背景として，現在の映像コンテンツがどのように制作されているかを概観，分類する．次に，この分類をもとに関連研究の動向を述べる．最後に本研究が目指す自動撮影システムの特徴を整理し，関連研究での対応状況を挙げながら本研究の位置づけを行う．

2.2　映像制作

2.2.1　映像の構成

映画やテレビのように編集された映像は，概念的に図 2.1 のような階層構造を形成している．編集された映像 (Video) は最上層にあたり，シークエンスの接続により構成される．シークエンスはシーン，シーンは映像の最小単位であるショットの接続により構成される．

ショットは映像の最小単位であり，あるカメラのスタートボタンを押してから留めるまでの間に撮影された，連続した映像の一区切りである．主人公やその話し相手のアップなどがショットに相当する．シーンは “場面” と定義され，単一の場所や時間を扱ったいくつかのショットで構成される．同じ部屋での会話などがシーンに相当する．シークエンスは “エピソード（挿話）” と定義され，シーンよりもストーリーにまとまりをもったものである．一般的な書物にたとえると，ショットは文章，シーンは段落，シークエンスは章，映像が書物そのものになる．

このように，普段我々が目にする映像は，多くの素材となる映像をつなぎ合わせることで構成されている．1 つ 1 つの素材は，映っている事実以外に何の意味も持たない．映像制作とはその事実の断片をつなぎ合わせて，意味を持ったまとまりのある映像を作り上げることだといえる．

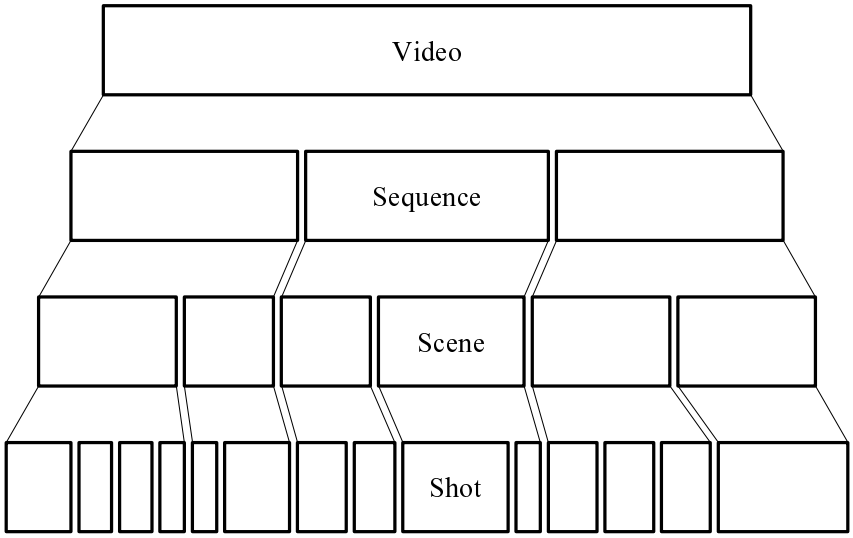


図 2.1: 編集された映像の構造

映像制作の現場では，映画の誕生以降 100 年に渡って，どうすれば制作側の意図することを効果的に視聴者へ伝えることができるかが試行錯誤されてきた．その規則の集大成を映像文法と呼ぶ．この映像文法に関しては，20 世紀を代表する映画監督であるヒッチコックがインタビューの中で以下のように述べている [8].

> 「わたしは映画の数々の小さな断片しか撮らない．その無数の断片を組み合わせると一本の映画になるわけだが，その編集をきちんとできるのはわたしだけで，ほかの人間には絶対できないように撮るわけなんだよ．撮影中にわたしの頭のなかですっかり編集ができあがっているから，わたしの指示なしには勝手に編集することが不可能なんだ.」

本研究は，限られた専門家によって利用されてきた映像文法を利用することで，自動的に撮影される映像をより効果的なものにすることを目的とするものである．

2.2.2　フェーズによる分類

映像制作は，大きく分けて “計画”，“実行”，“編集” の 3 つのフェーズに分類することが出来る [9]．表 2.1 にその手順と詳細を示す．

計画フェーズは，“何を”，“どのように” 撮影するかを決定する，実際の作業に入る前の準備的な段階である．撮影方法は現地での撮影，スタジオ収録，コンピュータグラフィックス合成 [10] などの手法から選択され，映像のイメージを検討しつつ構成を決定する．その際，テレビ局などの専門家集団では，カメラの位置や使用するレンズ，三脚，照明器具

表 2.1: 映像コンテンツ制作のフェーズ

フェーズ	計画	実行	編集
出力	構成表，台本	映像・音声素材	完成コンテンツ
作業内容	絵コンテ	V ロケ，スタジオ制作，電子映像制作，効果音，選曲	映像編集，MA 処理
制作機材	ワープロ，作画ツール	カメラ，VTR，作画装置，照明・音響装置	編集装置，特殊効果装置，VTR，MA 装置
メンバ	脚本家，映像デザイナ，プロデューサ	出演者，カメラマン，照明，音声，美術他	編集マン，ミキサ

の種類や量まで検討する．この段階は番組の流れを示す構成表と具体的な映像や音声のイメージを表す絵コンテや台本という，いわば映像の設計図を作成する段階である．

実行フェーズは，カメラなどの撮影機材を駆使して映像や音声を収録する段階である．通常この作業には多くの人手を要する．スタジオ制作では美術スタッフ，照明，カメラマン，ミキサ，スイッチャ，出演者，ディレクタ，タイムキーパなどが参加する．各担当者は感性と技能で完成形をイメージしながら映像や音声が収録されていく．この工程ではカメラやマイクをはじめとする各種の制作機器を使って必要な映像音声素材が収録される．通常この作業は何度も試行錯誤を繰り返し，その中から最も良くできたものを取捨選択していくため，ここで収録される映像・音声素材は，完成したコンテンツの十数倍になることも多い．

編集フェーズは，実行フェーズで撮影・収音された映像・音声素材を編集加工し，コンテンツとして完成させる段階である．この工程は映像・音声の編集作業と，音入れの MA（マルチトラック・オーディオ）処理，映像に文字を重畳する処理が行われる．まず，複数の素材映像の中から必要な部分を選択し，それをつなぎ合わせて 1 本のストリームにする．その際，映像と映像のつなぎ目にフェードやワイプといった光学的特殊効果を付加したりする．次に，この編集された映像を参照しつつ，コメントや BGM，効果音などを重ね合わせていく．最後に，出演者の名前や映像の注釈文字を重畳してコンテンツが完成する．編集作業は映像コンテンツの質を決める重要な作業であり，編集者の技量が問われる箇所でもある．

一般的に認識されている追尾機能のような自動撮影は，この中でも実行フェーズに相当する．しかし効果的な映像制作のためには，実行に際して個々のカメラをいつ・どのように制御するか（計画），複数のカメラ映像をどのように切替えるか（編集）が重要になるといえる．

表 2.2: 撮影対象の分類

イベント型	特徴	その場の状況に応じて進行
	状況理解	画像処理・音声認識など
	例	会議・講義・スポーツ中継など
ストーリー型	特徴	ある程度決まった流れに沿って進行
	状況理解	シナリオなどの事前知識
	例	演劇・コンサート・結婚式など

2.2.3　撮影対象による分類

撮影対象は，その進行方法の違いから，表 2.2 に示す 2 種類に分類することができる．1 つは講義やスポーツのように，その場で次に何が発生するのかわからないものであり，本論文ではこれをイベント型シーンと呼ぶ．もう 1 つはドラマやコンサートのように，ある程度事前に決まった流れに沿って進行するものであり，ストーリー型シーンと呼ぶ．

イベント型の撮影対象は，例えば会議で誰が発言したかという現場で発生する事象（イベント）に基づいて撮影方法が変わる．このようなイベントはあらかじめ予測することは困難であり，人物の発言や表情，行動を画像処理や音声認識などを用いることでリアルタイム認識し，それに応じてカメラワークを変更していく必要がある．

これに対し，ストーリー型の撮影対象にはほとんどの場合シナリオが存在し，プロのカメラマンによる撮影においても計画の段階でこのシナリオが重要な役割を果たしている [11]．シナリオとはシーンのどこで何が起こるかといった動作や状況の変化などのイベントが時間軸に沿って記述されているものであり，これを利用することであらかじめシーンの状況を把握することができる．

2.3　関連研究

本節では，映像制作の関連研究をフェーズや撮影対象ごとに分類して紹介する．ここで，映像制作は多くのプロセスから成り立っている．また，撮影するシーンには様々なものがあり，各々の映像的特徴や視聴者の目的も様々である．従ってあらゆるシーンを自動的に撮影可能なシステムを実現するには，あらゆる撮影規則を用意する必要があり現実的ではない．

多くの研究は，図 2.2 に示すように，2.2.2 節で述べたフェーズを限定したり，1 つの撮影対象に特化するアプローチがとられている．本研究が想定する “〜の自動撮影システム” に関する研究は，撮影対象を 1 つに限定し，それに必要な技術をトータルで提供する

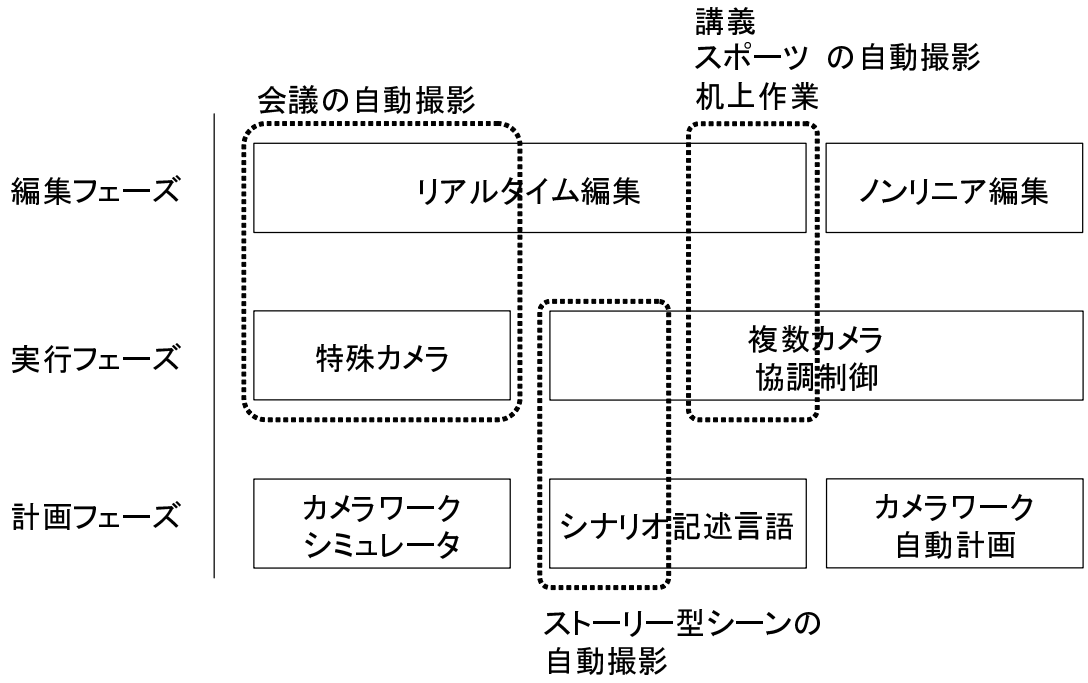


図 2.2: 関連研究の分類

アプローチであるといえる．

2.3.1　シナリオ記述言語

シナリオ記述言語は，ストーリー型シーンの撮影に必要なシナリオをどのように記述するかを定義するものである．現在シナリオを記述する場合は独自の仕様によってアナログ的に書かれていることが多い．この仕様を統一することで，ユーザ間でのシナリオ共有やシステムからの利用が可能となる．

その一例である TVML(TV program Making Language) はテレビ番組を記述できるテキストベースの言語で，NHK 放送技術研究所が開発したものである [12, 13, 14]．この TVML で書いた番組台本（TVML 台本）は，ソフトウェアとして提供されている TVML プレイヤーで即座にテレビ番組として再生することができる．ユーザーはエディターで TVML 台本を書くだけで，自分だけのテレビ番組をパソコン上で簡単に制作することが可能となる．TVML ではテレビ番組を作るのに必要な次の機能を持っている．

- CG のスタジオセットに CG の小道具，キャストを自由に配置できる．
- CG のキャストを台本の記述に従って会話させたり，動かすことができる．

```
set: assign(name=studio)
set: openmodel(name=studio, filename="studio.iv")
set: change(name =studio)
character: casting(name=Mary)
character: bindmodel(name=Mary, modelname=MARY)
camera:closeup(what = Mary)
super:on(type = text, text = "Mary")
character:bow(name = Mary)
character:talk(name = Mary, text = "こんにちは")
```

図 2.3: TVML スクリプトの例

- CG 内のカメラワークを自在に制御できる．
- テキストや画像をスーパーインポーズできる．

TVML は，実際のテレビ番組の制作現場で用いられている番組台本の中で採用されている記述法を手本とし，誰でも簡単に使いこなすことのできる言語になるようにデザインされている．このため TVML 台本では，コンピュータプログラミング言語にある条件分岐やループなどは一切なく，時間の流れに従って何のイベントがどのように行われるかを単純に列挙した形になっている．TVML 台本は，1 行があるひとつのイベントに対応する．TVML プレイヤーは 1 行分のイベントを実行し，そのイベントが終了したら次の行に記述されたイベントを実行する．その書式は次のようになる．

```
event_type:command_name(arg1 = data1, arg2 = data2...)
```

event_type は，CG 内のどの対象を制御するかを指定するもので，表 2.3 に示す 12 種類が存在する（TVML ver.1.1）．command_name は，event_type で指定した対象をどのように制御するかを決定するものである．

例として，スタジオセットに CG キャラクターの Mary を登場させ，カメラを Mary にクローズアップし，“Mary” という文字をスーパーし，Mary におじぎをさせて “こんにちは” としゃべらせるスクリプトは図 2.3 のように記述する．この結果は図 2.4 のように再生される．

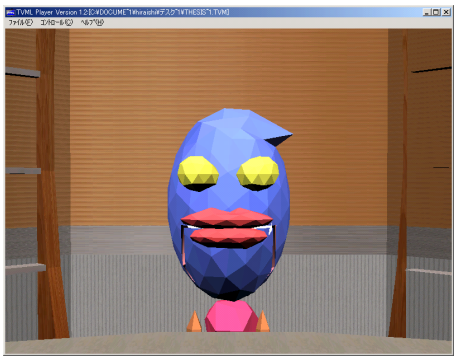
2.3.2　カメラワークシミュレータ

シナリオだけでは実際の映像のイメージが想像しにくい．結果として，当初考えていたものと，実際に収録したものとのイメージがかけ離れたものになり，何度も撮りなおしを

表 2.3: イベントタイプとコマンド例

イベントタイプ	機能	コマンド例
character	CG キャラクタ関係	casting（CG キャラクタに名前をつける） bindmodel（モデルをバインドする） talk（セリフをしゃべる） sit（座る） bow（お辞儀する）
camera	CG カメラ関係	assign（カメラに名前をつける） switch（指定カメラにスイッチングする） movement（カメラを指定位置に動かす） twoshot（2 つの対象物をツーショットにする） closeup（対象物をクローズアップする） catch（対象物をフォローする）
set	CG スタジオセット関係	assign（セットに名前をつける） openmodel（セットのモデルをオープンする） change（セットをチェンジする）
prop	CG 小道具関係	position（小道具の配置）
light	CG 照明関係	model（照明の作りこみ）
movie	動画再生	play（ムービーファイル再生）
title	静止文字情報・静止画	display（静止情報表示）
super	スーパーインポーズ	on（スーパー表示）
sound	音声再生	play（オーディオファイル再生）
naration	ナレーション	talk（セリフをしゃべる）
video	ビデオエフェクト	switcher（ビデオスイッチャーを制御する）
cgenv	CG エフェクト	shadow（CG の影をつける）

カメラをMaryにクローズアップ
camera:closeup(what = Mary)



Maryという文字をスーパーする

super:on(type = text, text = "Mary")

Maryがおじぎをする
character:bow(name = Mary)

「こんにちは」としゃべる
character:talk(name = Mary,
text = "こんにちは")

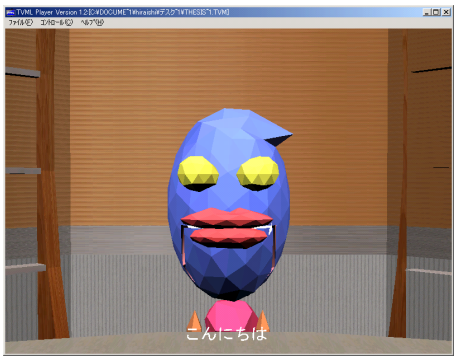


図 2.4: TVML プレイヤーと再生の様子

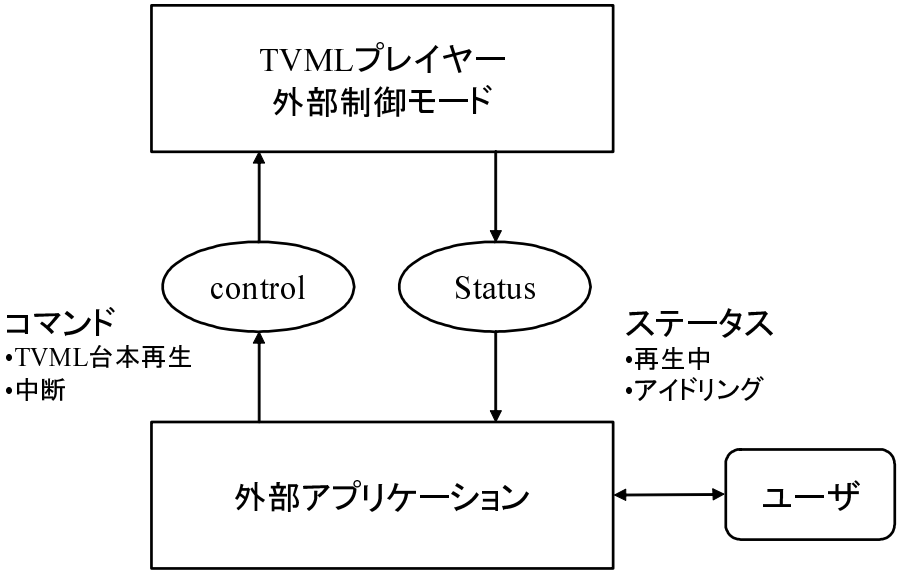


図 2.5: TVML 外部制御モード

することもしばしばである．カメラワークを仮想空間でシミュレーションすることによって，実際の撮影で得られる映像のイメージの把握が容易になる．これは，カメラのトレーニングツールや，計画段階でどのようなショットを撮影すべきかを検討するのに非常に有効である．このような視点の移動に制約が無い仮想空間におけるカメラワークの研究は数多くなされている [15, 16, 17, 18]．本節ではこのシミュレータとして，2.3.1 節の TVML を応用したものについて述べる．

TVML プレイヤーには，その機能を外部から制御することができる “外部制御モード” が用意されている [19]．図 2.5 にその仕組みを示す．通常，TVML プレイヤーは完全インタープリター動作のため，1 行のスクリプトを読み込むと即座にこれを構文解析し実行する．外部制御モードでは，起動中の TVML プレイヤーに対して外部のアプリケーションから非同期に任意のスクリプトを送信したり，逆に TVML プレイヤーの実行状態（ステータス）を得ることができる．例えば，外部からカメラ操作に関するスクリプトを送信することで，TVML プレイヤーに現在表示されている画面の視点を任意の位置に変更することが可能になる．

牧野らはこの外部制御機能を用いて，実際にカメラマンが使っているカメラ雲台と TVML とを連携させたカメラワークシミュレータを開発した [20]（図 2.6）．ユーザは雲台のレバーを上下左右させることで，TVML プレイヤーにカメラの操作に対応する TVML スクリプトを動的に送信することができる．結果として，TVML プレイヤー上のカメラの向きやズームをインタラクティブに制御することができる．

図 2.6: TVML 外部制御モードを用いたカメラワークシミュレータ

2.3.3　カメラワークの自動計画

“いつ”，“何を”，“どのカメラで”，“どのように撮影するか” というカメラワークを決定するのは，たとえシナリオがあったとしても時間がかかる作業とされている．その上，このカメラワークの出来不出来が映像コンテンツの完成度を大きく左右する．そこで，適切なカメラワークを自動的に計画することが期待されている．

道家らは TVML を用いて，番組に必要な情報を入力するだけで自動的にテレビ番組を制作する手法を提案している [21]．図 2.7 に TVML を用いて人間が番組制作を行う場合と，コンピュータが自動的に番組を制作する場合との比較を示す．人間が番組制作を行う場合，番組に必要な情報をもとに，人間が TVML の言語仕様に基づいてテレビ番組の台本を記述する（図 2.7(a)）．これに対して “人間が TVML 台本を記述する” 部分を “コンピュータ” に置き換えることができれば，人間はコンピュータに対して番組に必要な情報を与えるだけで，自動的にテレビ番組を生成することが可能となる（図 2.7(b)）．

図 2.8 に TVML を用いた自動番組制作の流れを示す．まず，ユーザは出演者のセリフや番組で使用する映像素材などを含む番組の “内容” データと，セットや出演者，カメラの画割といった番組の “見せ方” データをシステムに入力する．これを受け取る番組構成部は実世界でのディレクターに相当し，“内容” データから得られる番組構成をもとに，各制作モジュールに指示を行う．プレート，照明生成など各制作モジュールは，用意された TVML スクリプトのテンプレートの中から適切なものを選択し，その一部を書き換えて番組の部品（TVML スクリプトの断片）を生成する．番組構成部はこれらスクリプトを

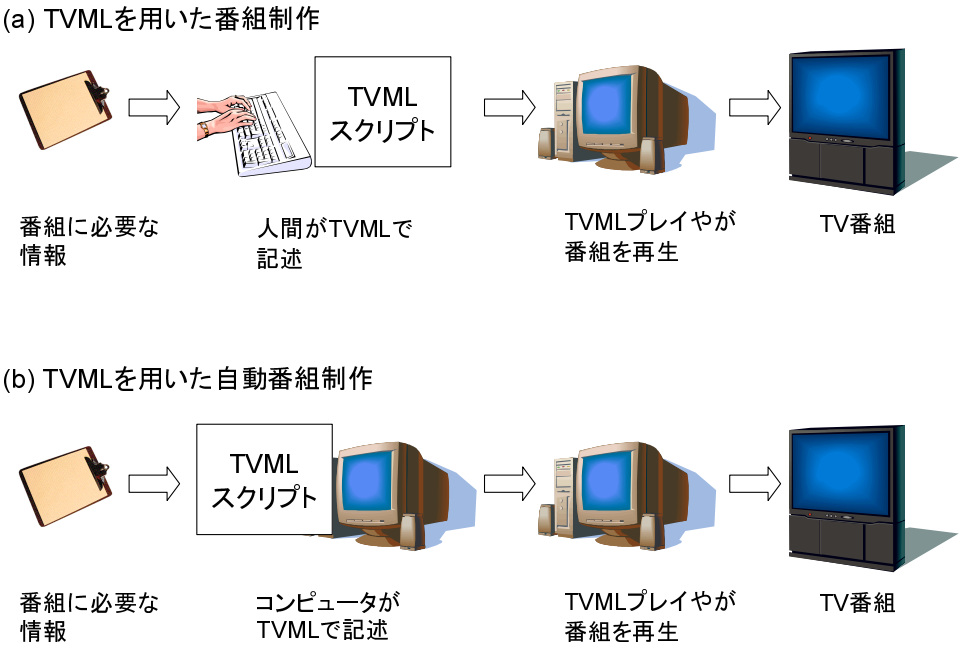


図 2.7: TVML を用いた番組制作の比較

統合し，番組（オンエアに用いる TVML スクリプト）を生成する．この手法を用いて，ニュース番組を自動的に生成するシステムを開発している．

2.3.4　特殊カメラ

通常の撮影では，ユーザは自分の持っているカメラ 1 台でしか撮影できない．しかし映像の完成度を高めるには，複数のカメラで様々な地点から撮影する必要がある．そこで離れた場所から容易に操作できたり，自動的に被写体を追跡する特殊なカメラが必要になる．

NHK 放送技術研究所ではプロのカメラマンの技術を反映した知的ロボットカメラを開発している [22, 23]．プロのカメラマンは，カメラの操作に関して熟練した技量を持ち，パン・チルト・ズームどれをとっても一般のユーザとは異なるノウハウを持つ．我々一般の撮影者による映像と比較すると，その品質には大きな差が出てしまう．加藤らは知的ロボットに放送品質の映像を撮影させるために，プロのカメラマンが被写体を追跡する際にカメラをどのように操作するかを細かく分析した [24, 25, 26]．その結果，次のような特性を明らかにしている．

(1) 画面内での被写体の位置は，被写体の速度よりサイズとの関係が深い．

(2) 画面内での被写体位置の広がりは，被写体のサイズが大きいほど，また被写体の速度が速いほど大きい値を示す．

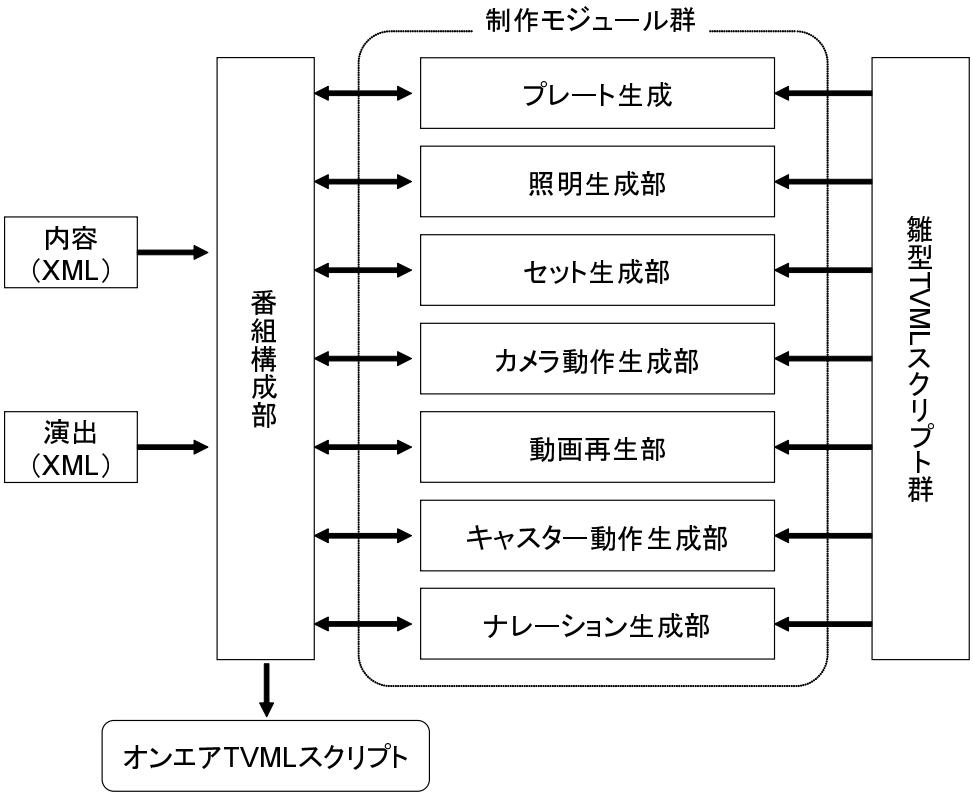


図 2.8: TVML を用いた自動番組制作の流れ

(3) カメラマンは，最適と考えている画面内での位置に対して被写体が誤差を持っていても，画面内での急激な被写体位置の修正を行わない．

(4) カメラマンが被写体の状況を判断し，カメラを操作するまでに要する応答時間は，被写体サイズ，速度には依存しておらず，最短で 200ms から 400ms 程度の値をとる．

郷らは，カメラの画角の変化量に対してパン・チルトの角速度を一定にするカメラ制御手法を提案した [27]．我々がハンディカメラを用いる場合，ズームアウト時とズームイン時でパン・チルト量を柔軟に変化させる．ズームイン時には少し小さめにカメラを動かすということを経験的に知っている．しかしネットワーク経由で遠隔のカメラを制御する場合，ズームによって画角が変化した場合でも，パン・チルトの角速度は一定である．したがって，カメラ操作に慣れるまでは，ズームインしたときにカメラを動かしすぎてしまい，見たい場所の映像をうまく表示することが難しい．アルファベット群の中から指定された文字を探し出すタスクを課した結果，ズーム量に応じてパン・チルト角速度を補正した場合は，補正しない場合に比べて短時間でタスクを完了できることを示している．

2.3.5　複数カメラの協調制御

ロボットカメラを導入したとしても，それぞれがバラバラに撮影をしていては良い映像は撮影できない．例えばあるカメラが追跡しきれなくなった被写体を次のカメラが引き継ぐといったように，カメラ同士が通信をして情報を共有し，協調しながら撮影を行う必要がある．

この代表的な例として，松山らが提唱する分散協調視覚プロジェクトがある [28]．このプロジェクトでは，多自由度の雲台を備えたカメラに高度な実時間画像処理機能を搭載した能動視覚エージェントを実現し，それらエージェントが有線・無線ネットワークで互いに通信しながらシーンの撮影を行う．中でも，複数エージェントで移動対象の追跡を行うシステムは入退室管理などの広域監視システム [29]，対話型遠隔会議・講義システム [30] などの応用システム実現のための重要な基盤技術の 1 つとして多くの研究がなされている．

分散協調視覚プロジェクトではその例として，能動視覚エージェント（AVA: Active Vision Agent）によって実時間対象追跡を行う際に，(1) 各エージェントにおけるシーン中の観測可能領域，(2) 追跡対象の移動軌跡，に関する知識をエージェント間で共有し，対象追跡能力を向上させる手法を提案している [31]．この手法により，複数のカメラが役割を適宜変更しながら，共通の対象を効率的に追跡することに成功している．浮田らはこの成果を発展させ，複数の対象を実時間で追跡することを目的にエージェント同士が協調するための三層構成アーキテクチャも実現している [32]．このアークテクチャを図 2.9 に示す．

Intra-AVA 層　最下層．能動視覚エージェントの各機能である視覚，行動，通信の各モジュールとダイナミックメモリ [33] で構成されている．各モジュールはダイナミックメモリを介してインタラクションを行い，その結果として一つのエージェントの動作が発現する．

Intra-Agency 層　中間層．同時に同一対象を追跡する能動視覚エージェントをエージェンシと呼ぶこととし，この層はそのエージェンシを組織するエージェントによって構成される．同一エージェンシ内のメンバは対象検出結果を交換し，追跡対象の同定を行う．

Inter-Agency 層　最上位層．システム内に存在するすべてのエージェンシによって構成されている．複数の対象を継続的に追跡するには，対象の移動やメンバであるエージェントの能力を考慮し，エージェンシ間でメンバを交換する必要がある．こうした動的なエージェンシの再構成を実行するため，エージェンシ間で追跡対象とメンバの情報が交換され

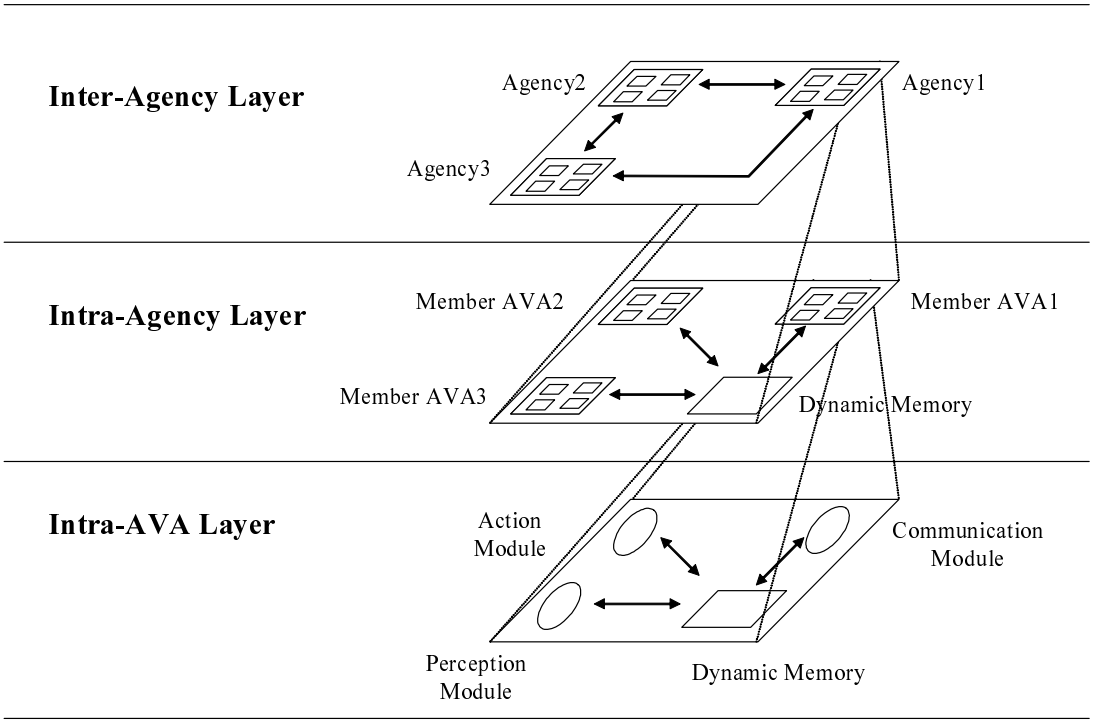


図 2.9: 分散協調視覚における対象追跡システムのアーキテクチャ

る．

また，冷水らは複数のカメラを連携させてあたかも 1 つのカメラであるように動作する Union-Camera を提案している [34]．この Union-Camera では，3 台のカメラを適宜切替えることにより，撮影可能な範囲を仮想的に拡張している．例えばユーザが “135 度の方向を見たい” という要求を出すと，要求された 135 度の地点はどのカメラを何度パンさせた地点なのかを計算する．これにより，360 度全方位を撮影可能な仮想パノラマカメラとして動作させている．

2.3.6　ノンリニア編集

ノンリニア編集とは，実行フェーズで収録された素材となる映像をいったんハードディスク等に記録して，コンピュータ上行う編集のことである．編集は完成までに多くの試行錯誤を伴うため膨大な時間がかかる作業である．そのため，この編集作業を支援する研究が数多く行われている．

Chiueh らは，編集過程の履歴を木構造で表現することにより編集のやり直し作業を容易にした対話型ビデオオーサリングシステムを構築している [35]．また Girgensohn らは，

素材映像から編集に使用可能なショットを切り出してユーザに提示する半自動的な編集システムを提案している [36].

これらの研究は編集作業を支援するものには違いないが，接続される素材映像の前後関係は考慮していない．接続の仕方は無限に存在するが，特に映像中における被写体の大きさはその際の重要な指標となる．熊野らは，素材となる映像からショットを自動的に抽出した上で，対象を遠くから撮影したルーズショット，近くで撮影したミドルショット，接近して撮影したタイトショットの 3 種類に自動的に分類する手法を提案している [37]．天野らはこの成果をもとに，設定したルールに沿って適切な映像を自動的に選択してスイッチングを行うシステムを実現している [38].

上記の研究が時間的に連続した素材映像を接続することを目標にしたのに対し，森山 [39] や Sundaram ら [40] のように，映像の不要な部分を削除して接続する研究も行われている．これらは時間的に離散した断片を接続しつつ，元の映像が持っていた内容をできるだけ損なわないで要約することを目的としたものである．他に要約の対象としては講義 [41]，料理 [42, 43] などが研究されている．

秦らは，カメラ操作のメタファを用いて，現在見ているシーンを別視点から撮影したシーンの検索手法を提案している [44, 45, 46]．まず，ファインダに映る被写体を観察しながら，興味ある被写体を探して指定する．次に，この時の撮影時刻と撮影範囲を問い合わせ情報として，同様のメタデータを有する他の視点からの映像を時空間上で検索する．その際，被写体の映り具合を考慮することで，ユーザの好みに応じた映像を選択表示することが出来る．

住吉らは，映像や音声以外にも台本，絵コンテ，字幕，読み原稿，撮影情報，編集情報，さらには調査段階で得られた資料まで含めた情報を番組情報として統合してデータベース化した DTPP(Desk-Top Program Production) を提案している [47]．番組の意味的な流れ（起承転結）にしたがって各種情報を管理するだけでなく，このシステムを使って編集作業をすると編集過程の情報がメタデータ化され，制作ノウハウなど多様な知見も蓄積することが出来る．

市村らは，運動会のように同じイベントに参加した複数撮影者の映像をサーバに集め，インターネットで映像編集できる Web システムを提案している [48]．サーバ上に集められた映像は自動で時間同期処理を施される．編集の際にはこの時間同期を利用して，ある映像クリップの前後につながる他の撮影者の映像クリップを一覧表示することができる．これにより映像素材の交換が容易になり，自分が撮影できなかったシーンを取り込むことや，プロフェッショナルの映像技法に沿った多彩な編集が可能となる．

2.3.7　リアルタイム編集

ノンリニア編集とは異なり，複数のカメラからの映像入力を蓄積することなくマトリックススイッチャなどを用いてリアルタイムに切替え，1 本の映像ストリームに編集していくことであり，生放送番組や実況中継などがこれに相当する．

関連研究としては遠隔会議，遠隔講義，スポーツ中継システムの一部として実現されていることが多い．これらに関しては次節以降で述べる．

2.3.8　会議の自動撮影

会議記録の 1 つの流れとして，会議室や参加者全体を効率的に撮影するためのパノラマカメラに関連した研究が行なわれている [49, 50, 51] ．

Lee らはパノラマカメラと 4 チャンネルの音声入力を持った Portable Meeting Recorder と呼ばれる小型デバイスを開発している [52]．会議記録が終了すると，MPEG2 ビデオと音声入力のデータを解析してメタデータを生成してデータベースに蓄積する．この結果をもとに，会議に参加できなかったユーザが後から自由に閲覧したり，任意の場面へのアクセスが可能となる．

Rui らは 1300 × 1030 の高解像度の映像を秒間 11 フレームで撮影可能なパノラマカメラを用いて小規模な会議を撮影する際に，どのようなインタフェースが好ましいかを実験している [53]．その結果，参加者全員を映したパノラマビューを使用したいという被験者が多かったこと，会議の雰囲気伝達に肯定的な意見が多かったこと，カメラの自動制御については人によって意見が分かれたことを明らかにしている．

これらの研究は，記録再生の方法に重点を置いたものであり，実行フェーズ（特殊カメラ）と，編集フェーズ（ノンリニア編集）をカバーするものということができる．

もう 1 つの流れとして，遠隔会議への適用を目的に，映像をリアルタイムに演出しながら撮影する研究が行われている．従来の典型的な会議映像は，一定位置に固定されたカメラから参加者を撮影する．この映像は変化に乏しく平面的であるとされている [54]．これに対しテレビや映画では画面に映る対象を次々と切替えていくことで構成されている．そこで図 2.10 にあるような会議空間に首振りカメラを設置し，参加者の発言に応じて撮影する参加者を自動的に切替える方式が検討されている．

井上らはテレビ番組のカメラワークの知識を用いてこの切替えを行う手法を提案している [55, 56]．この研究によると，切替えには話者が交代する時，同一の話者が長時間発言する時の 2 種類があるという．前者はより重要な人物を映すため，後者は同一の構図が続いて単調な映像にならないよう視聴者の関心を維持するために行われる．また，テレビ番組においてどのショットからどのショットへ切替えられたかについて統計を取り，遷移確

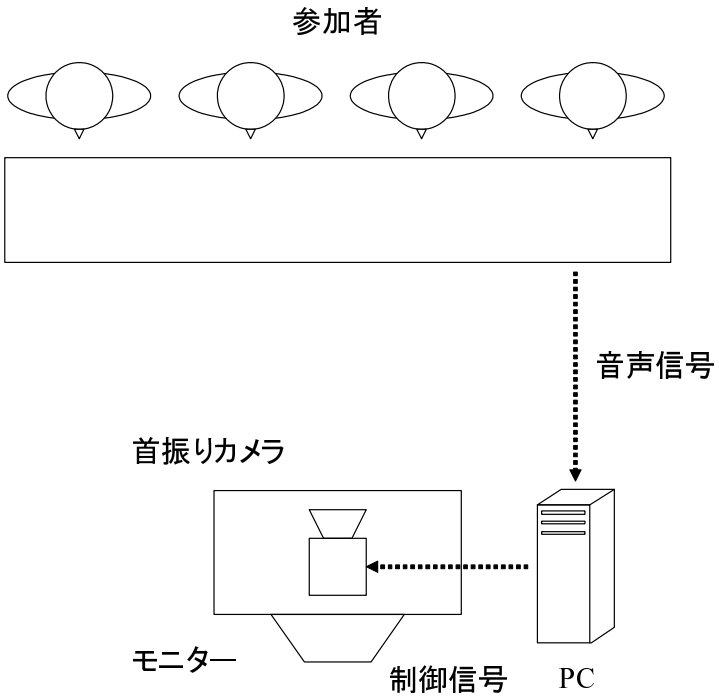


図 2.10: 映像演出 TV 会議システムの構成

率行列として定義している．この行列を用いてカメラの向きやズーム値を自動的に制御している．

大西らはこの遷移確率モデルに音源定位と画像処理を導入した自動撮影手法を提案している [57]．まず，マイクロホンアレーにより人物の発声位置を推定する．次に，活発に動作などを行っている映像上の人物領域を抽出し，両者の情報を統合することで注目すべき領域を決定し，カメラを制御する．これにより話者だけでなく身振りの大きな参加者を認識し，より豊かなノンバーバル情報の伝達を可能にしている．両研究ともにプロトタイプを実装し，既存の固定カメラによる映像との比較を行っており，会議空間の状況，雰囲気の伝達，視聴者の関心の維持に一定の効果があることを示している．

これら研究は，カメラを制御すると同時に，ショットを切替えてリアルタイムに映像を編集している．よって実行フェーズ（特殊カメラ）と編集フェーズ（リアルタイム編集）をカバーするものということができる．

2.3.9　講義の自動撮影

講義ですべての対象を同時に撮影しようとすると，個々の対象が画面内で小さくなりすぎ，文字を認識できなかったり，臨場感の低い映像になって学生の学習意欲を維持するこ

とができない．よって撮影対象の数だけカメラを用意し，それぞれの対象を個別に適切に撮影するという方法が主流である．この方法では，複数の映像から講義の状況に応じた映像を選択するという作業が新たに必要になる．現状では，この作業は講師自身，あるいは専用のオペレータによって手動で行われている．講義の自動撮影は，この適切な対象をどのように選択するかに重点が置かれているため，実行フェーズ（複数カメラの制御），編集フェーズ（リアルタイム編集）をカバーするものといえる．

大西らは，板書主体の講義を自動的に撮影する手法を提案している [58, 59, 60]．講義の状況を理解するために，固定カメラによって撮影した講義映像から講師と黒板の板書に関する情報を抽出し，それらの情報を用いて講師の行動推定を行う．次に，講師の行動に基づいて各カメラにおいて撮影領域を決定し，複数のカメラ位置から映像を取得する．最後に得られた複数の映像をそれぞれ評価することにより，現在の講義状況を最も効果的に表している映像を選択している．

亀田らは動物体の観測位置に応じて講義状況を把握し，撮影対象を自動的に選択する手法を提案している．この手法を適用したシステムを構築し，音声システム，プレゼンテーションシステムとともに高速ネットワークで接続し，実際の遠隔講義に適用している [61]．講師の位置は画像処理によって検出され，映像内に常に入るように切替えを行っているが，その誤差を小さくするため講師ごとの滞在位置の嗜好性に注目している．

先山らはあらかじめ定義した講義状態の遷移系列により講義状況を表し，その講義状況と対応する送信映像との関係をルール化した自動撮影手法を提案している [62]．講義状態は “講師が移動しながら説明をしている”，“講師が板書をしている”，“受講者が質問している” など 10 種類に分類され，その状態がどのように遷移してきたかで “受講者”，“黒板”，“スライド”，“講師” の 4 つの中から撮影する対象を選択する．どの対象を選択するかのルールは有限状態オートマトンとして記述している．

一方，宮崎らは送信対象の決定において，ユーザの要求を反映することに注目している [63]．受講者は，講義状況に応じてどの対象を見たいかという要求を映像化ルールとして登録しておく．しかし教室内に設置されるカメラの数は，その講義の受講者の数より少ない．また，1 台のカメラで撮影できる対象は 1 つに限られる．そこで，出来る限り多くの受講者要求を満足するカメラワークを制約充足問題として実時間で計算する手法を提案している．

2.3.10　スポーツの自動撮影

井口らはアメリカンフットボールを対象に，複数のカメラによる自動撮影と自動スイッチングを行なうシステムを提案している [64, 65]．アメリカンフットボールでは，プレイの内容により選手が密集したり分散するなど，画面内での急激な選手分布の変化が起こ

る．井口らはこの選手分布を 1 台の固定カメラの画像処理によって抽出して撮影すべき領域を決定し，別に用意された 3 台のアクティブカメラに適切に割り振ることを実現している．この研究は，実行フェーズ（複数カメラの制御）と編集フェーズ（リアルタイム編集）をカバーするものといえる．

金出らは 2001 年の Super Bowl において，Eye Vision と呼ばれる中継技術をデモンストレーションした [66]．この時使われた映像は，映画 “マトリックス” で有名になった，俳優の周囲をぐるりと一回転する演出（フローモーション法）であり，これをライブ中継で実現している．この実現のために，競技場の周囲に 33 台ものロボットカメラを設置し，それらの映像をフレーム単位で切り替える処理を行っている．この技術は後にアイスホッケー中継の他，日本の野球中継にも導入されている．

上記の研究が実在のカメラの視点のみを用いているのに対して，多数のカメラで撮影されたビデオを 3 次元モデル化し，自由に視点を変えたり実在しない視点を作る手段が提案されている [67, 68]．通常カメラは競技の邪魔になるためフィールド内に設置することはできない．しかしこの自由視点技術が実現すれば，サッカーであればゴールキーパーやペナルティーキックを蹴る選手の視点からの映像を自由に見ることができる．カーネギーメロン大学ロボティクス研究所の “3D Room” では，49 台のカメラを部屋の 5 面（前後左右と頭上）に埋め込み，ダンスやバスケットボールなど様々な場面を 3 次元モデル化している．また，大田らのシステムでは競技場の周囲に 18 台のカメラを配置し，これらの映像から自由な位置からの競技中継映像の合成を実現している [69, 70]．

金出らはこの技術を “Virtualized Reality”（仮想化された現実）として提唱している [71]．Virtual Reality のように現実空間を仮想化するのではなく，3 次元モデル化という仮想化技術を経て，現実以上の意味を持った現実空間を再構築するものである．

金出らの両研究は，実行フェーズ（複数カメラの制御）と，編集フェーズにおいてより高度な画像処理を用いたものということができる．

2.3.11　机上作業の自動撮影

尾関らは，科学実験のような机の上で行う作業（机上作業）を対象にしたプレゼンテーションの自動撮影方式を提案している [72]．テレビ番組の分析より，机上作業に必要なカメラワークが “撮影対象＝注目対象物＋注目すべき状態” と定義している．この注目すべき対象を認識するために，音声認識で “このように”，“こうやって” などの指示語，磁気センサーで “対象物を指す” といった指示動作を取得している [73]．また，注目すべき状態として “外観”，“動き”，“周辺関係” の 3 種類に分類し，これを実現するための可変枠制御方式を提案している [74]．この注目に基づいて自動編集した映像は，熟練者が編集した映像と並んで高い評価を得ている [75]．この研究では，実行フェーズ（複数カメラの制御）と編集フェーズ（リアルタイム編集）をカバーするものといえる．

2.3.12　シナリオのあるシーンの自動撮影

ストーリー型シーンには，進行に関してシステムに事前に何らかの知識を与えることができる．その事前知識としては次の 2 つが考えられる．

シナリオ　動的シーンの 3 次元的な状況を詳細に規定したもの．人物の位置，行動などが記述されている．

ストーリーボード　最終的な映像の構成を 2 次元的に規定したもの．各ショットでの画面構成や，その撮影時のカメラ制御の種類，映像効果などが記述されている．

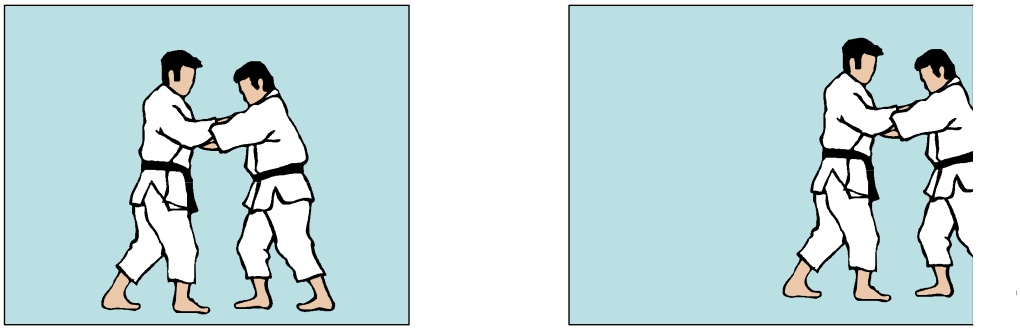
ここで，一般的に事前知識と実際のシーン状況とは完全に一致しない．このため，計画で得られたカメラワークをそのままの形で実シーンに適用すると，ストーリーボードで想定していた映像と実際に投影される映像との間には次の 2 種類のズレが生じてしまう．

幾何学的ズレ　画面構成上のズレ．図 2.11 では，本来はカメラの中央で人物を捉えているはずが，何らかの理由で人物の停止位置が事前知識であるストーリーボードでの位置とずれてしまい，画面の隅で捉えてしまっている．

時間的ズレ　カメラワークのタイミングのズレ．カメラ切替えのタイミングがシーンの状況に対して早すぎる/遅すぎるなどが原因となる．図 2.12(a) において，理想的な計画では，アップの映像からスイッチング後には投げを打つ動作を撮影するはずであった．しかし，実際のシーンでは何らかの影響で演技が計画より早く進行し，図 2.12(b) が示すようにスイッチング後には投げが終わった状態になってしまっている．

Bobick らは幾何学的ズレに対し，二段階で認識処理を行う手法を提案している [76]．最初の段階ではいつ，誰が，どんな動作をするかというシナリオと簡単な認識処理により，シーンの近似的なモデルを生成する．次に，この近似モデルから得られる情報をもとにして，より高次な認識処理を施してカメラの位置を補正している．この手法を取り入れたプロトタイプを実装し，料理番組の撮影に適用している．

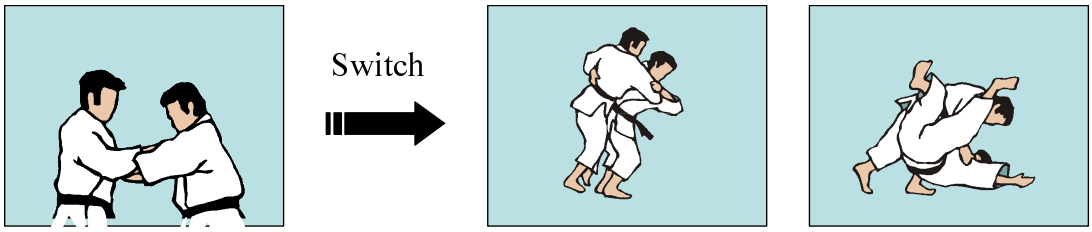
田中らはドラマのようにシナリオとストーリーボードが存在するシーンを想定し，シナリオを入力としてカメラワークを決定する計画システムと，その計画に基づいて実際にカメラを制御する実行システムの 2 段構成による自動撮影システムを提案している [77, 78]．その中で，先述のズレを修正するため，“人物が手を挙げた”，“人物が画面内に登場した”といったシーン中で発生するイベント情報をカメラワークに埋め込み，イベントが検出されたときに対応するカメラワークを実行して計画との同期を取る仕組みを実現している．



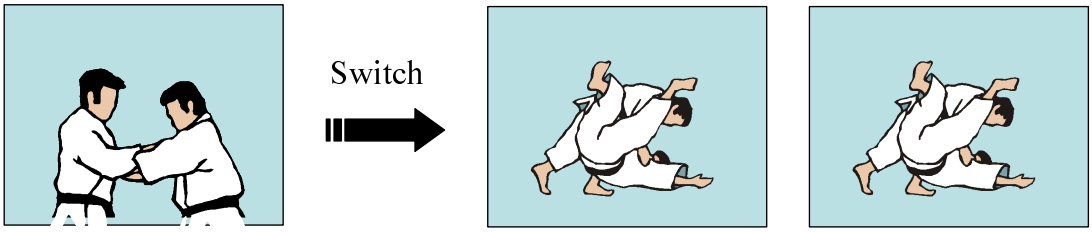
(a) 理想的なケース．対象を画面の中心に捉らえている．

(b) 実際のケース．中心から外れ，画面外に飛び出ている．

図 2.11: 計画と実シーンの間の幾何学的ズレ



(a) 計画段階での理想的な流れ．スイッチング後に投げを打つ直前の様子を捉えている．



(b) 撮影段階での実際の流れ．動作が速すぎ，スイッチング後には投げ終わってしまっている．

図 2.12: 計画と実シーンの間の時間的ズレ

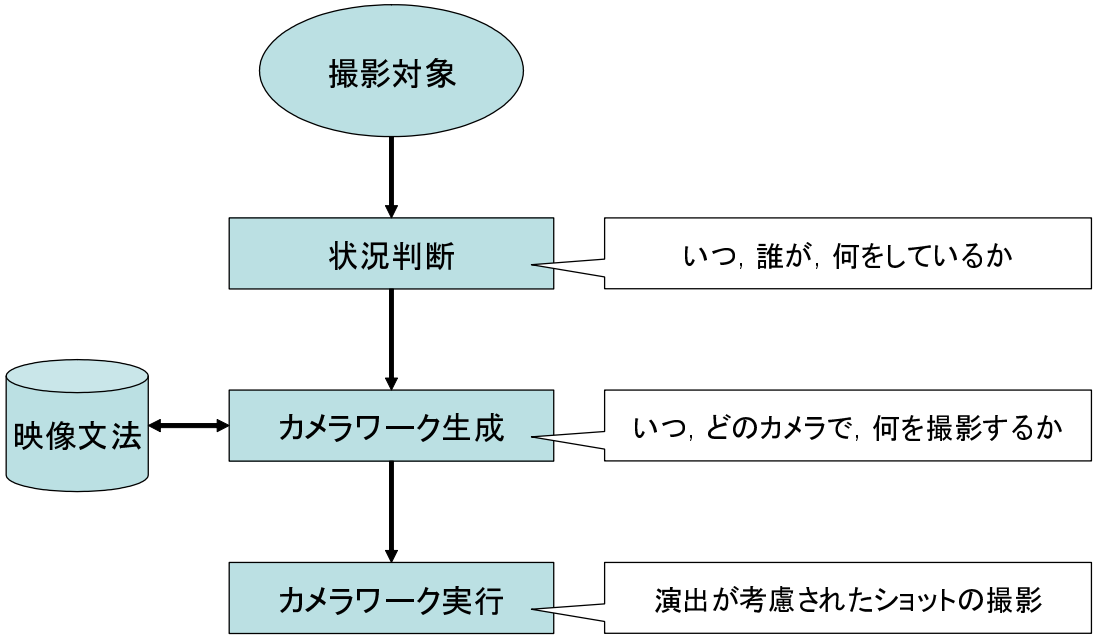


図 2.13: 映像文法に基づく自動撮影システム

この撮影計画・実行システムは実際のカメラやスイッチャーを用いて実装され，簡単なサンプルシーンの撮影を計画通り実行可能なことを示している．

これらの研究は，入力した事前知識に基づいて撮影機材を制御していることから計画フェーズ（シナリオ記述）と実行フェーズ（複数カメラの制御）をカバーするといえる．

2.4　本研究の位置づけ

本節ではまず，本論文が提案する自動撮影システムについて述べる．次に，一連の研究アプローチを述べながら，この概念とどう合致するのかについて触れる．また，それぞれの研究において，関連研究の対応状況をまとめ，本研究の位置づけを行う．

2.4.1　映像文法に基づく自動撮影システム

本研究では，映像文法を用いて複数台のカメラを協調制御し，シーンを自動で，かつ効果的に映像化することを目的とする．本研究の概念を図 2.13 に示す．第 3 章，第 4 章で述べる一連の研究は，いずれもこの概念に基づいたシステムの具体例として行った．

本研究が想定するシステムでは，何らかの撮影対象と，それを撮影する複数台のカメラが用意されている．まず，システムはこの撮影対象において，“いつ，誰が，何をしているか” という状況を判断する．

次に，取得した状況をもとにカメラワークを生成する．この決定の基準として映像文法が考慮される．一般に映像文法とは，効果的な映像を制作するための技法全般を指す．従ってその中には，照明の方法，撮影のアングル，カメラの操作方法，編集の方法など様々なものが存在するが，本研究では “いつ，どのカメラで，何（どのようなショット）を撮影するか” というカメラワークの決定基準を映像文法に求める．この実現には，各カメラ制御や被写体捕捉の精度も必要になるが，2.3.4 節で挙げた既存の研究で取り上げられているため対象とはしない．

最後に，決定したカメラワークをもとに，複数台のカメラを動作させ，実際の撮影・収録を行う．ここで収録されたショットは演出が考慮されているものなので，スイッチングなどの編集を経ることで効果的な映像を生成することができる．

2.4.2　イベント型シーンの自動撮影

本研究のアプローチ

第 3 章で述べる研究の位置づけを図 2.14 に示す．本研究では，撮影対象として対面会議を扱う．対面会議は，参加者の発言によって次の進行が決定していく．このような発言の順番は通常予測不可能であり，イベント型シーンの 1 つであるといえる．

対面会議は通常テーブルを囲む座席配置で行われる．このレイアウトでは 1 台のカメラで全員を適切に撮影することは困難であるが，複数カメラの映像を切替えるスイッチングワークにも高度な技術が必要となる．間違ったカメラの映像へスイッチングを行うと映像に急激な変化が生じ，視聴者が混乱してしまう．そこで本研究では，映像文法を “シーンの状況が理解しやすく，誤解を生じさせない映像” を制作するための技法としてとらえ，そのようなカメラワークをリアルタイムに生成する部分に主眼を置いた．

まず撮影対象の状況判断をリアルタイムで行うために，会議の状況を (1) 中心的な人物がいる場合，(2) 中心的な人物がいない場合，の 2 種類に分類し，これを参加者の発言の推移から判断する．

次に，会議空間に中心的な人物がいると判断した場合に，その人物同士の対話の様子を見やすく，かつ効果的に演出するようなカメラワークを決定する．この時考慮される映像理論はイマジナリーラインとカメラの三角形配置である．本来は概念的なものであるイマジナリーラインを会議空間に一意に設定し，このイマジナリーラインと三角形配置を基準にして複数台のカメラを協調させ，参加者を撮影するのに最適なカメラを決定する．ここで決定したカメラの映像をリアルタイムでスイッチングし，1 本の演出された映像を自動で生成する．

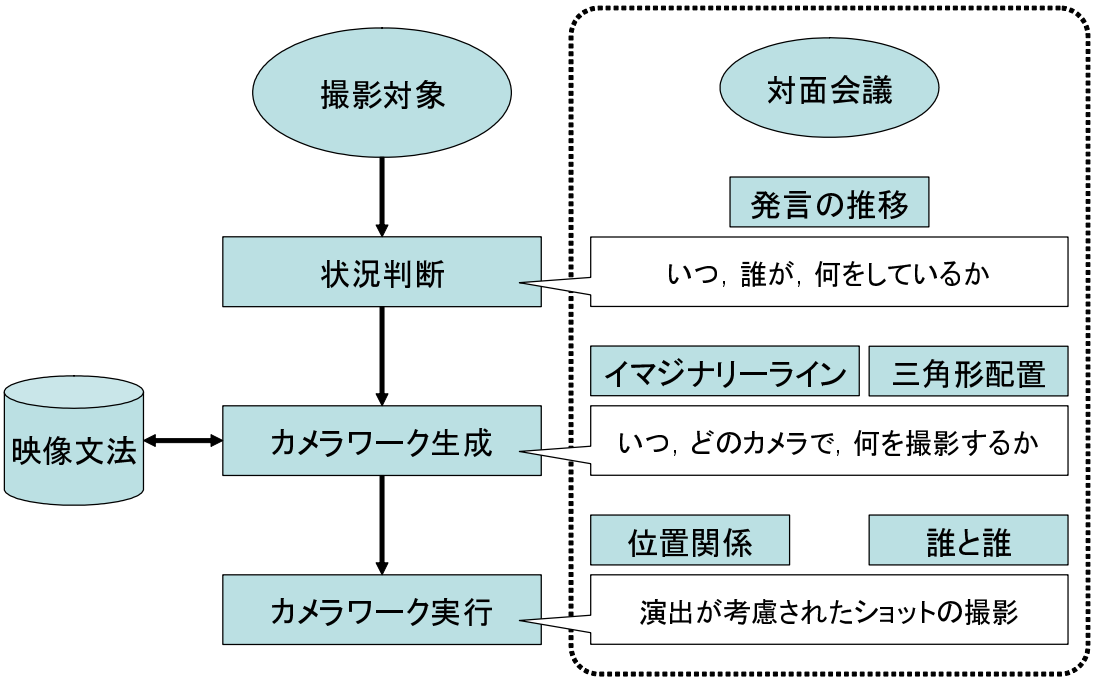


図 2.14: 第 3 章の位置づけ

関連研究の対応状況

図 2.2 に本研究をマッピングすると図 2.16 のようにまとめることができる．

イベント型のシーンで複数のカメラを協調制御するという面では，2.3.5 節で述べた分散協調視覚の諸研究 [31, 32] と目的を同じくしている．しかしそれらの多くは追跡対象をいずれかのカメラの画面内に捕捉し続けることのみを目的としている．その対象を “どのように映すか” という，映像の見栄えや演出に関する議論は十分に行われていない．

見やすい映像を演出するという面では，2.3.6 節で述べた熊野ら [37] や市村ら [48] の研究が該当する．しかしこれらはいずれも蓄積された素材映像を扱うノンリニア編集を対象としている．本研究は会議映像の特性上，記録以外にもテレビ会議の中継に利用することを想定し，リアルタイムで編集を行う．

加藤らの知的ロボットカメラ [24] も，見やすい映像を撮影するための研究に該当する．しかしこのカメラはカメラマン単独の知識に従って動作するものである．本研究は，ディレクターや映画監督など，シーン全体を把握した立場からの演出を試みる．

リアルタイムで編集するという面から見ると，2.3.8 節で述べた井上らの TV 番組の知識に基づく撮影 [56] が該当する．この研究では図 2.10 からも分かるように参加者を横 1 列に並べ，その前方に設置した 1 台のカメラの向きを制御する．しかし，本研究で想定している小規模な対面会議の多くは，参加者同士で机を囲む円卓型の環境で行われる．1 台のカメラですべての参加者を適切に撮影することは困難である．

2.3.9 節の講義の自動撮影に関する研究では，撮影対象が講師・黒板・スライド・生徒など，各対象に 1 台ずつ専用のカメラを割り当てることが多い．つまり講義における演出は被写体の切替えであり，やはり “どのカメラから撮影するか” という視点がない．これは，各カメラがあらかじめ撮影する対象を限定している尾関らの机上作業の自動撮影 [74] にもいえる．

2.4.3　ストーリー型シーンの自動撮影

本研究のアプローチ

第 4 章で述べる研究の位置づけを図 2.15 に示す．本研究では，撮影対象としてオーケストラ演奏を扱う．オーケストラ演奏の楽譜には，撮影に必要な “いつ”，“どの楽器が”，“どのような音を演奏するか” が記述されており，ストーリー型シーンの 1 つである．ストーリー型シーンでは，撮影対象の状況判断をこのシナリオから行うことができる．

オーケストラ演奏では，シナリオの役割を果たす楽譜が存在する．しかし，用意できるカメラの台数に比べて被写体の数が多いうえ，カメラを設置できる位置にも制限がある場合には，映像に関する知識のないユーザが適切なカメラワークを計画することは困難である．その結果として，必要なショットが撮影されていなかったり，別々のカメラで似たようなショットを撮影したりしていて，効果的な編集を行うことができないという問題が発生する．そこで本研究では，編集時に発生する様々なショットの要求に対応するため，映像文法を “バラエティに富んだショット” を撮影するための技法としてとらえ，そのようなカメラワークをシナリオから自動的に生成する部分に主眼を置いた．

オーケストラはストーリー型シーンのため，シナリオから撮影対象の状況を判断することができる．そこで楽譜からどのような情報をシナリオとして記述すればよいのかを定義する．

次に，シナリオから取得した状況から，被写体の候補となる演奏パートを抽出する．そして抽出された候補に対して，映像文法に基づく優先度を計算し，限られた台数のカメラに重要な被写体を割り当てていく．この優先度は被写体の種類，類似度，構図，カメラからの映り具合に基づいて計算する．最終的に，複数台のカメラが協調して，なるべく多くの被写体を様々な構図で撮影するカメラワークが生成される．

関連研究の対応状況

図 2.2 に本研究をマッピングすると図 2.16 のようにまとめることができる．

複数カメラでストーリー型シーンを撮影するという点で，2.3.12 節で挙げた田中ら [77] や Bobick の研究 [76] が該当する．しかしこの研究ではストーリーボードに相当する事前

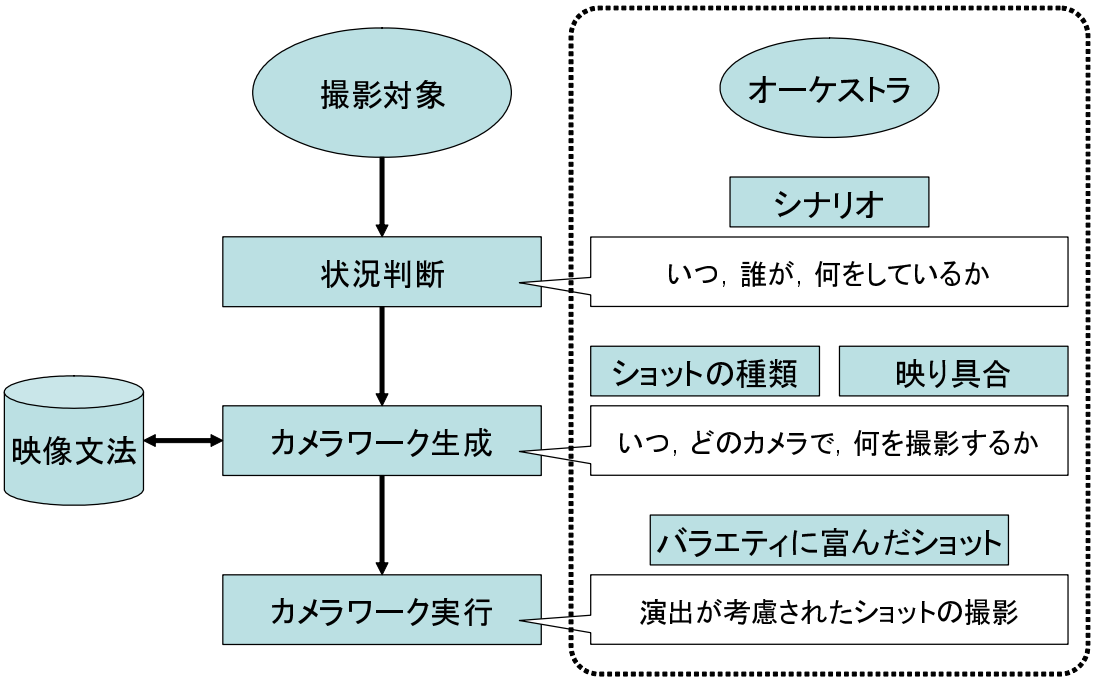


図 2.15: 第 4 章の位置づけ

知識は手動で入力するなど，カメラワークを自動計画はその手法の提案にとどまっており実現はされていない．

シナリオを扱うという点では，2.3.1 節の TVML が該当する．しかし TVML はシナリオを記述する言語仕様である．本研究でもこの TVML をシステムに入力するシナリオの記述方式として採用する．しかし本研究は TVML で記述されるカメラワークの自動計画を目的としており，その提案の範囲が異なる．

カメラワークを自動計画するという面で，2.3.3 節の道家らの研究 [21] が該当する．道家らは自動計画の例としてニュース番組を取り上げているが，ニュース番組での被写体はキャスターとニュース用映像クリップ程度であり，カメラワークの計画もそれほど複雑ではない．2.3.9 節で述べた講義の各研究も多くても 5 種類程度であり，それぞれに 1 台ずつカメラを割り当てることも可能なレベルである．これに対し本研究が対象とするオーケストラ演奏ではカメラの台数より多くの被写体が存在し，その取捨選択を迫られるなど複雑である．

2.5　まとめ

以上，本章では本研究の背景とその位置づけについて述べた．研究の背景として 2.2 節では，映像コンテンツができあがるまでの詳細とその分類について述べた．2.3 節では 2.2

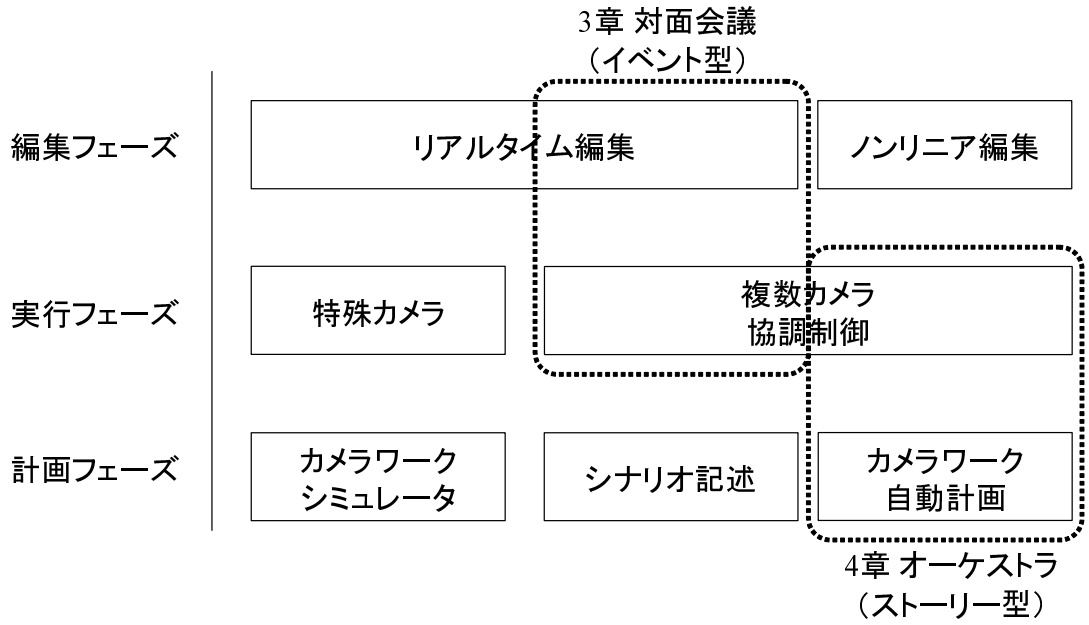


図 2.16: 関連研究との比較

節の分類をもとに各関連研究を紹介した.

これらの関連研究を踏まえた上で，2.4 節では，本論文が目指す自動撮影システムの全体像と，一連の研究との対応付けを行った．この位置づけを踏まえたうえで，次章以降，本研究における新規提案を含む本論を展開することにしたい.

第3章　対面会議の自動撮影

3.1　はじめに

本章では，次に発生する事象が予測困難であり，カメラワークを事前に計画することが困難なイベント型シーンの例として，対面会議を取り上げる．高橋によると，会議はその性質から伝達会議，調整会議，決定会議，創造会議の 4 種類に分類することができる [79]．会議を記録する場合，伝達会議であれば伝達事項が正確に伝わればよい．しかしそれ以外の会議では，どのようにして決定事項が導かれたのかを知るためなど，会議室の参加者の様子を映像で効果的に伝達できる必要がある．これを踏まえて本研究では，調整・決定・創造会議のような，企業において最も一般的な会議 [80] を撮影の対象とする．

従来の会議シーンは，パノラマカメラ [81] や，横一列に並べた参加者を固定カメラで撮影することが多かった [56, 57]．これに対して本研究で想定する調整・決定・創造会議は，参加者が机を囲む円卓型の座席配置で行うことが多い．そのような配置の全参加者を確実に撮影するには 1 台のカメラでは困難である．また，映画やテレビ番組，仮想空間 [16, 18] のようにカメラが撮影空間を自由に動き回ると参加者の注意がそがれてしまう．よって複数台のカメラをあらかじめ固定した位置に設置する必要がある．

このような条件下で映像に変化を与える演出方法として考えられるのは，カメラの映像を切替えるスイッチングと，撮影領域の決定であるズームに限定される．しかし，スイッチング方法には高度な撮影知識を要する．印象の異なるショットをスイッチングすれば映像に躍動感を出すことができるが，一方で映像の急激が変化が位置関係や方向感覚の混乱につながり見づらい映像となってしまう [63]．また，ズームアップしたショットは参加者の表情をよく映せる半面，カメラの視野が狭くなって空間的情報が減少する．その結果，やはり視聴者がどこから誰を見ているのかが分からなくなることがある．

そこで本研究では，このようなカメラの設置位置や操作が制限された撮影環境の中で，効果的にスイッチングを行う自動撮影手法を紹介する．本研究では，映像文法を “正確かつ分かりやすい映像” を制作するための技法としてとらえる．そのためにイマジナリーラインとカメラの三角形配置という映像理論に注目し，これに基づいて複数台のカメラを協調動作させ，参加者同士が対話をしているシーンを自動で効果的に撮影する手法を提案する．プロトタイプシステムを実装して評価を行い，本手法の有効性を確認する．

以下，3.2 節では本章で用いる映像理論について，3.3 節では提案手法について，3.4 節ではプロトタイプの実装について述べる．3.5 節および 3.6 節で実験と結果に対する考察を述べ，3.7 節をまとめとする．

3.2　映像理論

映像文法とは，映像を効果的に視聴者へ伝えるための技法である．この “効果的” という部分には，“白熱した”，“緊迫した” といった雰囲気の伝達方法など様々な方法が考え

られる．本研究ではスイッチング時に発生する見づらさに着目し，映像文法を “シーンの状況が理解しやすく，誤解を生じない” 映像を制作する技法として解釈する．本節では，この目標を実現するのに必要と思われる映像理論について述べる．

3.2.1　イマジナリーラインとカメラの三角形配置

シーンの中心となる 2 人の人物には，2 人の間に交わされる目線の方向に基づいた，相手に関心を示す概念的な直線が流れている．この直線はイマジナリーラインと呼ばれる．そしてこのイマジナリーラインを底辺として三角形を形作る位置にカメラを配置するのがカメラの三角形配置である（図 3.1）．

図 3.1 におけるカメラ 1 から 3 では，人物 A を画面の左側でとらえるのに対し，カメラ 4 では反対の右側でとらえる（図 3.2）．ここでカメラ 1，2，3 は，イマジナリーラインで分割された空間においてすべて同じ側に属している．このようにイマジナリーラインを越えないカメラの映像でスイッチングを行うことで，各人物の位置関係を明確にすることが可能となる．

また三角形配置されたカメラのうち，イマジナリーラインと平行な底辺を持つカメラ 2 台（底辺カメラ）からは，2 人のうちどちらか一方の人物の視点に偏ったショットが得られる．たとえばカメラ 2 はカメラ 1 より人物 A の表情をよりよく映せる．一方頂角に位置するカメラ（頂角カメラ）からは，2 人を均等に映したマスターショット的な映像が得られる．この 3 つのショットを切替えるとイマジナリーライン上にいる人物を強調できるといわれている．

対面会議シーンの撮影にイマジナリーラインを考慮することで，スイッチング時における参加者の位置関係が変化しない安定した映像を提供でき，会議室の状況を正確に伝達することができると考えられる．また，三角形に配置されたカメラの映像を用いてスイッチングを行うことで，誰と誰が話をしているかを強調することができると考えられる．

3.2.2　映画の分析

人物を撮影する際，イマジナリーラインは必ずしも越えてはいけないものではない．あえてこれを越えることで場面転換や時間経過を表現し，映像がより興味深く演出される場合もある [82]．そこで，実際の対面会議シーンの映像におけるイマジナリーラインの扱いを確認するための分析を行った．その対象として “12 人の怒れる男” という映画を取り上げた．この作品は 12 人の陪審員が殺人事件について審理を続けるというストーリーであり，ほぼ 1 つの部屋の対面会議シーンだけで作品全体 117 分を構成している．よって分析対象として最適であると考えた．調査したのは，作品の中で 2 人の人物の対話シーンにおけるスイッチング 253 回である．

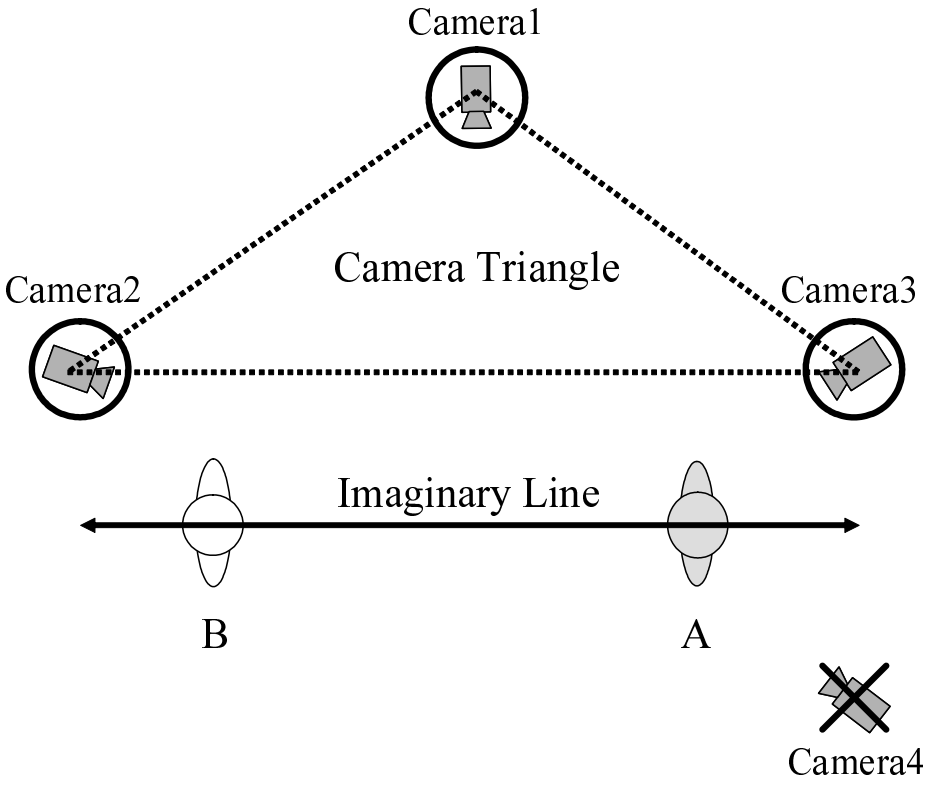


図 3.1: イマジナリーラインとカメラの三角形配置

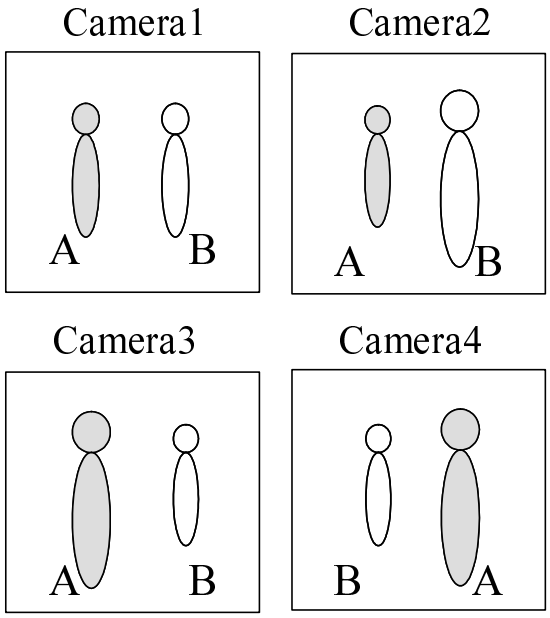


図 3.2: 各カメラの視点

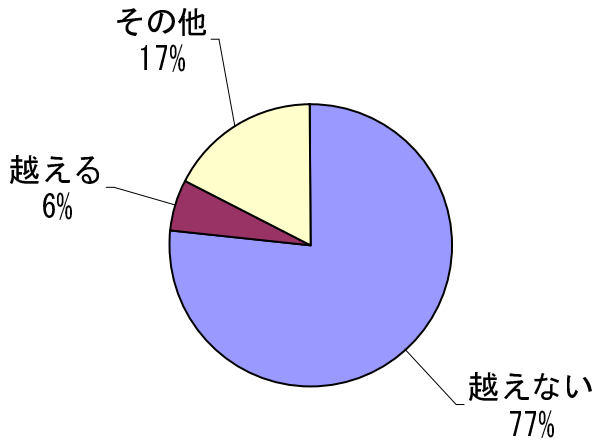


図 3.3: スイッチングとイマジナリーラインの関係

分析の結果を図 3.3 に示す．全体の 77% である 194 回はイマジナリーラインを越えないことが分かった．基本的に参加者の位置関係を重視した演出が必要なことが確認できる．

イマジナリーラインを超えるスイッチング（6%，15 回）は，カメラマンがイマジナリーライン上の位置までゆっくりと移動し，その後ラインを超える位置にあるカメラの映像へスイッチングするようにしていた．これは映画やテレビのように，カメラが撮影環境を自由に動けることが前提になる．本研究ではカメラが固定されていることを想定しているため，このような演出カメラワークは対象外となる．なお，その他（17%，44 回）は人物以外のショットへの切替えとその復帰であった．

次にショットの種類を調べたところ，ズームアップショットは視線で他の参加者の発言を促す場面や，発言をするまでの表情変化を伝える場面で利用されていた．ズームアップしたショットは迫力のある映像を作り出せる半面，位置関係がわかりづらくなるという欠点があり，会話の意味的内容と関連付けて注意深く用いられていると考えられる．

ほとんどのショットは，画面内に複数の参加者を映した構図であった．代表的なものとして会議室全体を撮影したショット，発言者とその相手を斜め後方から同時に映す “肩越しショット” が多用されていた．

3.3　提案方式

本節では，対面会議シーンに対してイマジナリーラインとカメラの三角形配置を組み込んだ自動撮影手法について述べる．提案手法の流れを以下に示す（図 3.4）．

(1) 円卓型座席配置の会議空間と，複数台のカメラが設置されている．

(2) 誰が，いつ発言したかを取得する．

(3) 参加者の発言の推移から会議の状況を判断する．

(4) 必要に応じて適切な位置，時間帯にイマジナリーラインを設定する．

(5) 参加者を効果的に撮影するのに最適なカメラを決定する．

(6) カメラの映像をスイッチングする．

(7) 位置関係，誰と誰が話しているかを強調した映像が生成される．

本研究では，映像文法を “正確で分かりやすい映像” を編集するための技法としてとらえる．そして，会議の進行に応じて，参加者を撮影するカメラをイマジナリーラインと三角形配置に基づいて逐次決定する．それらの映像をスイッチングすることで，位置関係と話者を強調した映像を自動的に撮影する．

ここで，イマジナリーラインは文字通り概念的なものであり，明確な実体や定義が存在するわけではない．実際の映像制作の現場では人物の表情，視線，発言内容などを組合わせてカメラマンが主観的に判断している．本研究ではこのイマジナリーラインをシステムで利用するため，会議空間中に一意に設定する方法を提案する．

なお，撮影にはカメラの位置や姿勢といったパラメータを取得するキャリブレーションや，被写体を適切に撮影するための画像認識技術が必要となる．これらに関しては既存の研究 [83, 22] で扱われているため本研究では対象外とし，事前にすべてのパラメータや撮影領域を設定しておくものとする．

3.3.1　撮影環境の設計

まず，参加者の人数を n 人，撮影に必要となるカメラを m 台とし，その数的関係から求められる撮影環境の基本設計方針について考える．なお，以降で扱うカメラは，パン・チルト・ズームの範囲といった各種パラメータがすべて同一のものであるとする．

$m = 1$ の時は，どのようなカメラを用意するかで映像が変わってくる．固定カメラを用いる場合は，常に参加者全員を画面内に映すようなショットで撮影を続ける．そのような

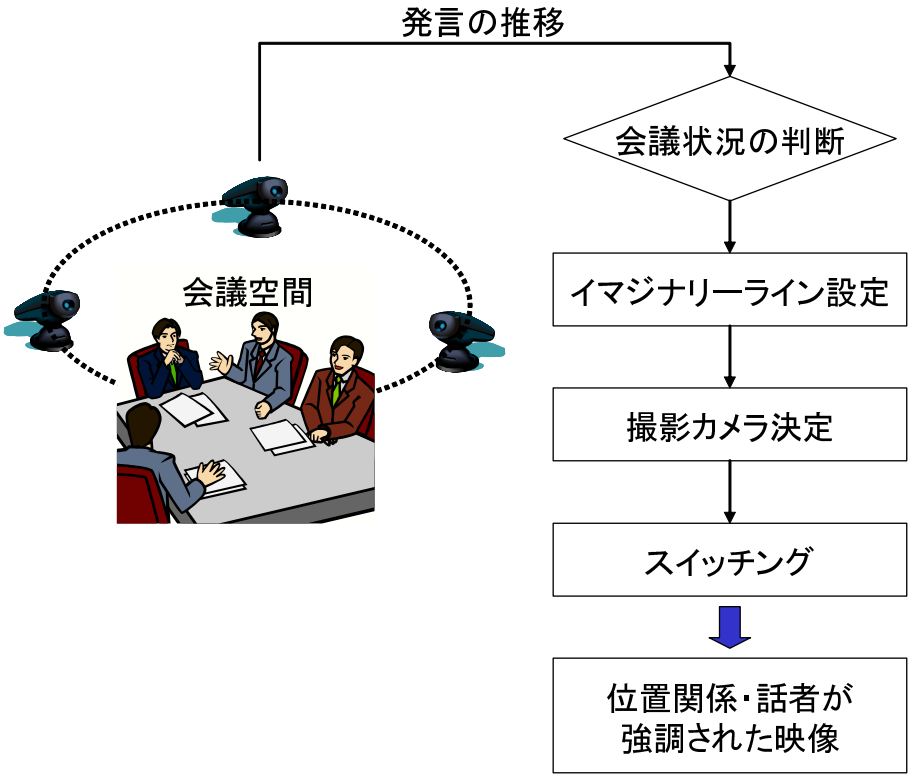


図 3.4: 提案手法の概要

同一角度からの同一内容を提示し続ける映像は最も平面的でつまらない映像とされている．首振りカメラの場合は，カメラの向きやズームを制御することでショットを変化させることができる．しかし本研究が対象としているのは円卓型座席配置であり，そのような配置では何人かの参加者は背中からしか撮影できない．これに対し，机の中心部分に 360 度回転可能なカメラを設置すれば，参加者全員を前方から撮影することができる．しかしこの場合，隣り合って着席する参加者 2 人は同時に撮影することはできても，正面に座る参加者との会話を同時に撮影することはできない．

$n > m$ の時，つまりある程度の台数のカメラを用意できる場合は，ショットも複数確保できる上，設置位置によっては参加者全員を前方から撮影することができる．また，正面に位置する参加者同士を同時に撮影するショットも実現可能になる．これを切替えることである程度演出された映像を生成することができる．しかし人数に比べてカメラの台数が少ない場合，1 台のカメラにいくつものショットが割り当てられる．その結果，パン・チルト値やズーム値の補正に必要な時間が多くなり，必要なショットを撮影するまでの時間が長くなる．会議では発言をする参加者が次々と代わるため，そのような遅延はカメラワークに大きく影響してしまう．

一方，参加者の人数に対して同程度，もしくはそれ以上の台数のカメラを用意すること

ができれば，1 台のカメラに割り当てられるショットの数も少なくなる．パン・チルト値やズーム値の補正回数も減り，会話の変動に迅速に対応できるようになる．よって本研究では，参加者と同等，もしくはそれ以上の台数のカメラを，参加者の周囲に固定して設置することとする．

3.3.2　会議状況の分類

次に，会議がどのような状況にあるときにイマジナリーラインを設定して演出用カメラワークを生成するかを議論する．会議の状況は参加者の発言の様子から次の 2 つに分類できる．

(1) 注目が集まる参加者がいる時

(2) 注目が集まる参加者がいない時

(1) に相当する状況として，一定時間内に特定の 2 人で対話を繰り返している状況が挙げられる．このような 2 人には，自然と周囲の注目が集まる．また，2 人以上でも，テーブルの一辺に並んでいるなど座席配置が直線に近い，もしくは局所的であるような場合も同様に注目が集まるといえる．

(2) に相当する状況として，各参加者が持ち回りで意見を述べたり，ほぼ全員が次々と発言をしている状況が挙げられる．この状況は，会議としては盛り上がってはいるものの，参加者の間で役割が均等に分担されているため，周囲の注目が分散する．また，テーブルの四隅に座った 4 人のように，座席配置が分散した少人数では，周囲の注目は位置関係によって分散してしまう．

本研究では，(1) の状況にある参加者を演出して撮影し，位置関係が明確で，誰と誰が話しているかを強調した映像の自動撮影を目指す．会議という特性から複数の参加者が同時に長時間発言をすることは少ないことに注目し，参加者同士の発言がどのように推移したかで (1) の状態にある参加者を特定し，イマジナリーラインとカメラの三角形配置に基づいて演出を行うこととする．

3.3.3　2 人におけるイマジナリーラインの設定

提案手法では，各参加者の発言をマイクから音声情報として検出し，現在の発言者に関する情報を蓄積していく．そして，現時点から過去にさかのぼって x 回の発言が特定の参加者 2 人だけで行われたとき，その 2 人を 3.3.2 節 (1) の状態にある “注目が集まる参加者” であると判断し，該当する参加者の座標上を通る直線を 1 本設定する．この直線をイマジナリーラインとして定義する．この時必要となる各参加者の座席位置やカメラの設置

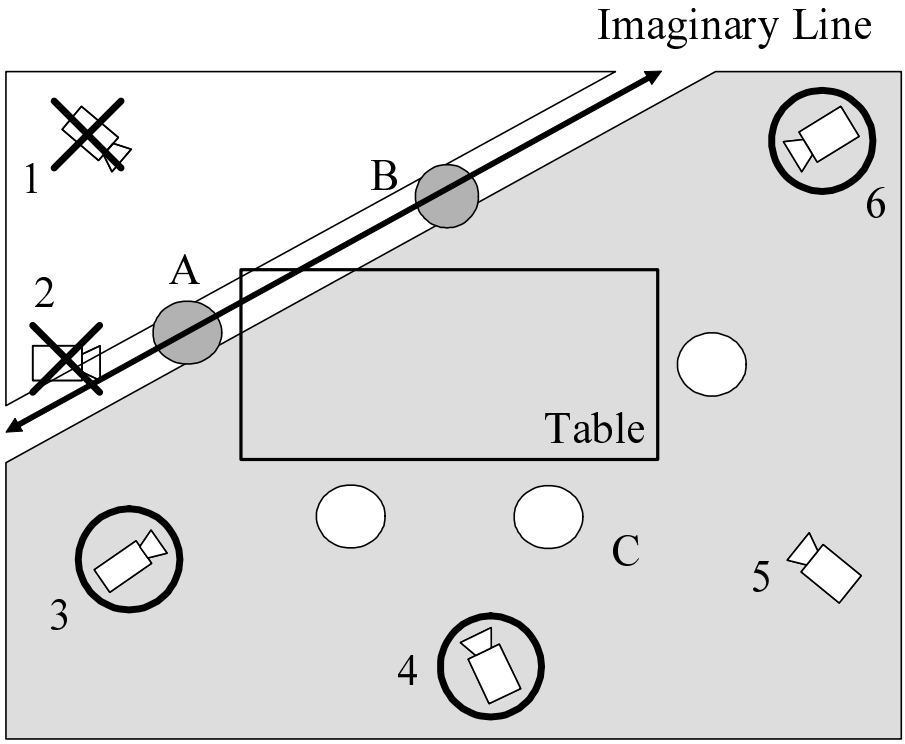


図 3.5: 2 人の対話におけるイマジナリーラインの設定

位置などは，あらかじめシステムが取得しておく．図 3.5 では，$A \rightarrow B \rightarrow A \rightarrow \cdots$ のように参加者 A と B の間で複数回発言が交換された時に設定されたイマジナリーラインの位置を示している．

設定に要する発言数 x は，会議の規模や，演出したい話題の重要度に応じて多少の変動がある．参加者数が少ないと，特定の 2 人の間で発言が繰り返される確率が高くなるため，このような場合は x を大きくして条件を厳しくする必要がある．また，重要な話題だけを演出したい場合も x を大きくする．これにより，2 人が何度も発言を繰り返す場合のみ演出することが可能になる．

この x が大きすぎれば条件が厳しく，イマジナリーラインが設定されないこともある．逆に小さすぎれば，不必要にイマジナリーラインが設定され，視聴者の混乱を招く．よって x は，比較的小さな値を中心にして，± 1 または 2 程度の狭い範囲に収まるものと考えられる．この x の適切な値は，実際の会議の規模や議題の内容に応じて適宜求める必要がある．本研究では，この値を 3.4.3 節の予備実験によって求めることとする．

3.3.4　撮影カメラの決定

提案手法では，(1) 位置関係が明確で視聴者を混乱させない，(2) 誰と誰が話をしているかを強調する，映像を自動生成することを目的とする．

(1) を実現するには，イマジナリーラインを越えない範囲でカメラを決定すればよい．しかし，(2) を実現するには，単にカメラを三角形に決定するだけでは不十分である．一般に，会話シーンのほとんどはエスタブリッシュショットと肩越しショットで構成される．エスタブリッシュショットはその場の環境が客観的にわかるようなショットで，状況判断の認識に効果がある．肩越しショットは 2 人のうち一方の斜め後ろから肩越しに相手の表情を映すショットで，対話の場面を効果的に表現するのに重要な役割を占める．つまり，各カメラにこれらのショットになるべく近いものを撮影させ，それらをスイッチングする必要がある．

イマジナリーラインが設定されている場合は，その直線によって会議空間が 2 つに分割できる．そこでまずそれぞれの空間に存在し，イマジナリーライン上の 2 人を撮影可能なカメラの台数をカウントし，台数が多いほうの空間を選択する．次に，イマジナリーライン上の参加者を撮影するカメラ（撮影カメラ）を三角形になるように決定する．この決定には各参加者とカメラの座標情報を利用する．エスタブリッシュショットを撮影する三角形の頂点カメラは，参加者の間に設定されたイマジナリーラインの垂直二等分線に最も近い位置のカメラを選択する．肩越しショットを撮影する三角形の底辺カメラ 2 台は，イマジナリーライン上のそれぞれの参加者に最も近いカメラを選択する．例として図 3.5 の A と B の対話では，カメラ 3，4，6 を撮影カメラとして決定している．

提案手法で選択される撮影カメラ 3 台はすべてイマジナリーラインを超えない位置に分布する．よってこれらの映像をスイッチングしても位置関係が反転しない分かりやすい映像を提供できる．

また，本研究では 3.3.1 節に示したように，性能が同等なカメラを参加者の周囲に固定して設置するため，各カメラのショットの内容を設置位置で決定することができる．各参加者からの距離が均等な位置にあるカメラを三角形の頂点とすることで，中心的な人物を均等に概観するエスタブリッシュショットに最も近いショットを撮影することができる．イマジナリーライン上の参加者に最も近いカメラ 2 台は，参加者の人数に対してカメラの台数が多めに設置されているならば，参加者の背後に位置する可能性が高い．これを底辺カメラとすれば肩越しショットに最も近いショットを撮影することができる．この 3 つのカメラのショットをスイッチングすることで，位置関係を混乱させること無く，誰と誰が話しているのかを強調することができる．

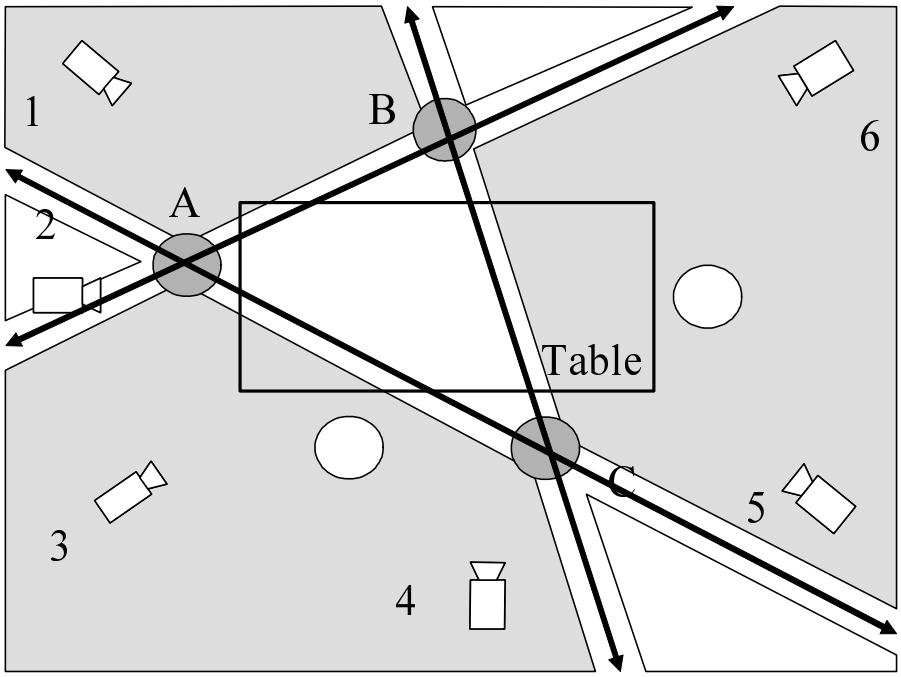


図 3.6: 複数人におけるイマジナリーラインの設定

3.3.5　複数人におけるイマジナリーラインの設定

実際の会議では，特定の複数人で発言を繰り返すこともしばしば存在する．3.3.3 節の概念をそのまま適用すると会議空間上には複数本のイマジナリーラインが存在することになる．そのすべてをいずれも越えないような位置に設置されたカメラ 3 台は，図 3.6 のようにその条件が厳しく存在しない可能性もある．たとえあったとしても，設置位置が近いカメラが多くなる．そのようなカメラから得られるショットは構図に大きな違いが無くなり，演出効果が期待できない．

このような場合，イマジナリーラインが設定されている複数の参加者をグループ化して 1 人の参加者とみなすことで，イマジナリーラインを 1 本に簡略化する．図 3.7 に 5 人の参加者のうち 3 人が発言を繰り返す例を示す．A と B の間にイマジナリーラインが設定されている状況では，この 2 人を 1 人の参加者とみなす．A，B のうちどちらか一方の発言を $(A|B)$ と表すと，$(A|B) \rightarrow C \rightarrow (A|B) \rightarrow \cdots$ のように AB と C の間で発言が複数回繰り返されたとき，A と B の間に設定されたイマジナリーラインを，AB の重心座標と C の座標との間に新たに設定しなおす．こうすることで，ある程度位置関係を保ったまま，多様な構図のショットを確保することができる．

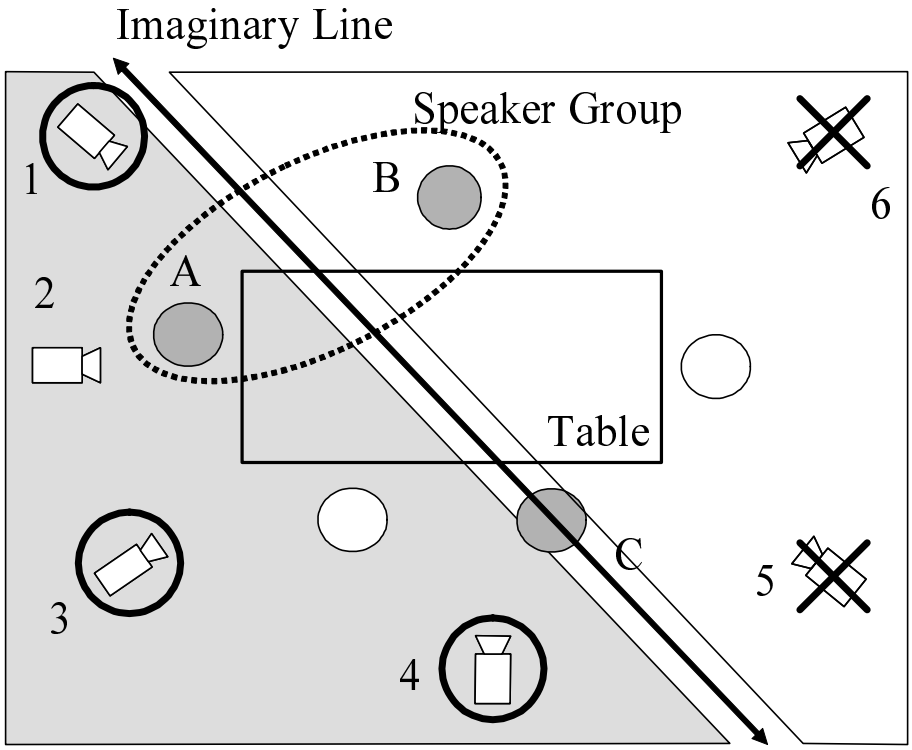


図 3.7: 話者のグループ化によるイマジナリーラインの設定

3.3.6　イマジナリーラインの解除

イマジナリーラインは中心的な人物の間に設定されるものである．よってシーンの中心的役割が他の参加者に移ったり，より多くの参加者の間で役割が分担されるようになった場合は，一度設定したイマジナリーラインを解除する必要がある．

本研究では，2 人，複数いずれの場合でも，設定されたイマジナリーラインと関係の無い参加者の発言が複数回割り込まれたときにこれを解除する．また，1 人が長時間発言を続けた場合にもこれを解除する．

解除に必要な割り込み回数 y も，設定に必要な発言数 x と同様に，会議の規模によって変動がある．総人数が多い場合は，イマジナリーラインに無関係な参加者の人数も多くなり，割り込まれる確率も高くなる．このような状態では，イマジナリーラインが実際にはまだ存在するにもかかわらず即座に解除される可能性がある．よって y を大きめに設定する必要がある．逆に少人数の会議では，y は小さくなる．

この y の適切な値も，実際の会議の規模や議題の内容に応じて適宜求める必要がある．これについても 3.4.3 節の予備実験にて述べる．

表 3.1: ショットの分類

ショット	説明
全景	参加者全員を映す
発言者	発言者とその周囲の参加者を映す
演出 1	底辺カメラから対話・議論中の参加者を映す
演出 2	頂角カメラから対話・議論中の参加者を映す

表 3.2: 1 ショットの持続時間と出現確率

持続時間（秒）	出現確率
2.5	55%
7.5	30%
12.5	10%
17.5	5%

3.3.7　スイッチング

映像を演出するために，撮影カメラのショットをスイッチングして 1 本の映像ストリームをリアルタイムに生成する．スイッチングに用いる 4 種類のショットを表 3.1 に示す．会議が行われている最中は，会議空間全体を映すショット（全景）と，発言者を前方から写すショット（発言者）が常に用意されている．イマジナリーラインが設定された場合は，そのラインに関わる参加者を演出するため，新たに底辺カメラ 2 台のショット（演出 1）と頂角カメラからのショット（演出 2）が用意される．

1 ショットの持続時間は，発言者が交代したときと，同一ショットが長時間続き単調な映像になるのを避けるときの 2 パターンがある．特に後者は，持続時間の長いショットほどその出現回数が減少することが分かっている [56]．本研究もこれに習い，3.2.2 節の映画の分析から，ショットの持続時間とその出現確率の関係を表 3.2 のように設定した．

これら表 3.1 と表 3.2 のデータをもとに，会議の進行と並行してスイッチングを行う．シーンの中心となる参加者が存在せずイマジナリーラインが設定されない場合は，全景ショットと発言者ショットの間でスイッチングを行う．対話・議論が発生し中心的な参加者がいる場合は，それら参加者たちの位置関係を明確にするために，底辺カメラと頂角カメラによる 3 つの演出ショットでスイッチングを行う．

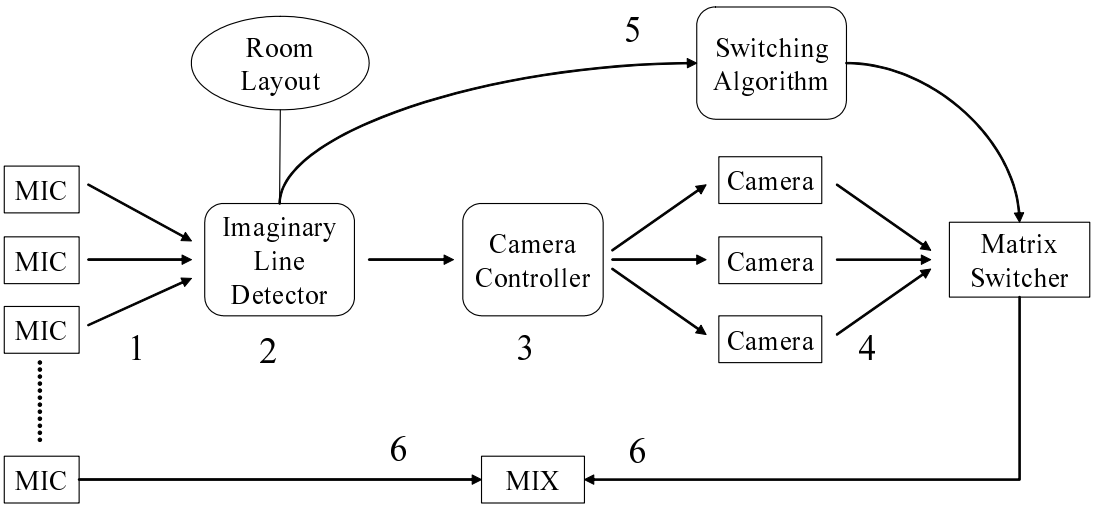


図 3.8: プロトタイプによる撮影の流れ

3.4　実装

提案手法に基づいて，対面会議シーンを自動撮影するプロトタイプシステムを構築した．プロトタイプにおける撮影の流れは次のステップからなる（図 3.8）．

(1) イマジナリーラインの設定に必要となる発言者の特定は，参加者がマイクに向かって話すことでシステムに認識させる．サンプリングレートは 8 kHz で 0.5 秒毎に入力信号の平均エネルギーを求め，閾値以上の入力が連続した時点で発言者と判定する．

(2) (1) で検出した発言者と会議空間のレイアウトをもとに，イマジナリーラインの設定・解除を行う．そしてその有無に応じて撮影カメラを決定する．会議空間のレイアウト情報はあらかじめ取得し，システムに入力しておく．

(3) イマジナリーラインが設定されている場合，撮影カメラは該当する参加者を映すようレンズの方向などが制御される．

(4) それぞれのカメラの映像出力はマトリックススイッチャへ入力されている．このスイッチャを制御することにより，任意の 1 つを出力映像として選択できる．

(5) 3.3.7 節のスイッチングアルゴリズムに基づいてスイッチャを適切に制御する．

(6) 各カメラの入力映像が 1 本の映像ストリームとして出力される．この映像は会議空間全体の音声とミックスされる．

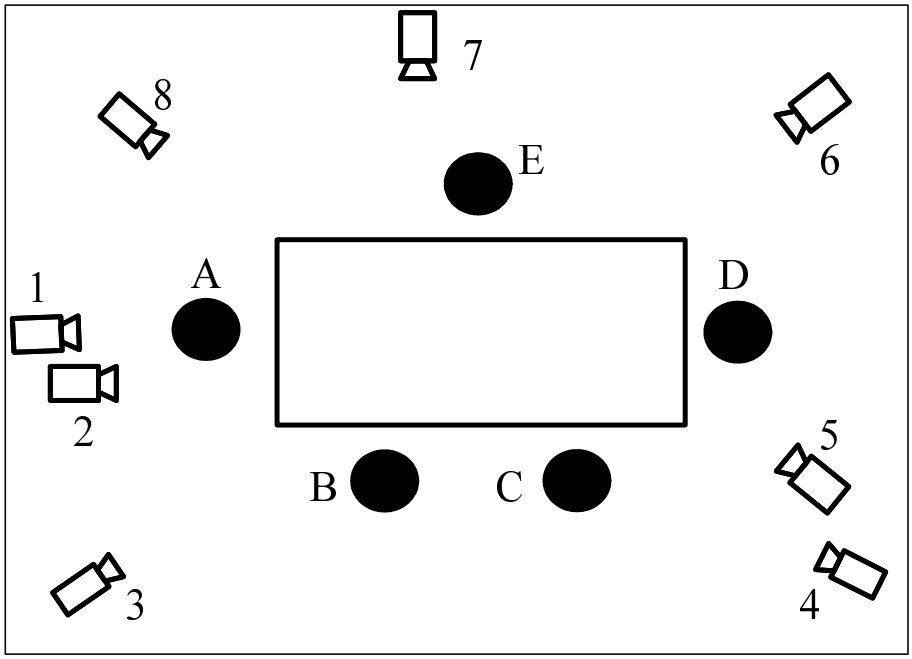


図 3.9: 会議空間のレイアウト

3.4.1　実装環境

イマジナリーライン検出部およびカメラ選択･制御部の実装にはWindows2000，Pentium III 600MHz の PC を使用した．本システムでは，各カメラのステータス（パン・チルト・ズーム値がいくらか，動作中か等）はすべてネットワークを介して自由に取得可能である [84, 85, 86]．システムの各モジュールはすべて Java 言語で実装した．使用したカメラは全て Canon 社製の VC-CI である．この雲台付きカメラ VC-C1 は RS-232C シリアルポート経由で PC から制御可能であり，首振り位置のプリセットメモリ機能を保持している．マトリックススイッチャは IMAGENICS 社製 SW-1010F を用いた．

3.4.2　システム構成

プロトタイプの構成を図 3.9 に示す．プロトタイプでは一例として 5 人の参加者，8 台のカメラで設計した．会議空間は小さいため，参加者同士の距離も近い．よって位置関係による中心的役割の分散は発生しないとした．

パンやチルトといったカメラの動作途中の映像が頻繁に登場すると，視聴者は映像に強い違和感を覚えてしまい，本研究の目指す評価が正しく得られない可能性がある．そこで，できるだけこのような動作をしないように，参加者人数に対してカメラの設置位置を

表 3.3: 各カメラのショット

Camera	shot 1	shot 2	shot 3
1	ABCDE	–	–
2	BCD	CDE	–
3	CDE	AE	–
4	ABE	–	–
5	ADE	ACE	–
6	ABC	BCD	ABE
7	BC	–	–
8	ABC	BCD	CDE

8 カ所と多めにした．この数的関係は，3.3.1 節で本研究が定めた撮影環境の設計方針に合致している．

各カメラのショットを表 3.3 に，2 者間でイマジナリーラインが設定された場合の撮影カメラを表 3.4 に示す．カメラ 1 が会議空間全体の撮影専用なのを除いて，他のカメラは 1〜3 通りのショットを保持し，常時 2 人もしくは 3 人を画面内におさめている．例として BC の対話時のショットを図 3.10 に示す．イマジナリーラインの位置に基づきカメラ 2，6，7 が選択されている．

スイッチングアルゴリズムがいかに働いているかは，ユーザに提供される GUI を通して確認する事ができる．それぞれ，2 人の間でイマジナリーラインが設定去れている場合（図 3.11），3 人の間でイマジナリーラインが設定されている場合（図 3.12）を示している．また，この GUI を用いて，カメラや参加者の絶対座標，参加人数，カメラの台数といった会議空間のレイアウト情報を調整できる．

なお，最も離れた A と D の対話時には，撮影空間の広さの限界から 1 台しか選択できなかった．またシステムがイマジナリーラインを認識する前にスイッチングが起こる場合も見られた．これら実装上の問題は今後の課題とする．

3.4.3　予備実験

プロトタイプシステムにおいて，イマジナリーラインの設定および解除に必要な発言回数を求めるための予備実験を行った．

5 人の被験者で 10 分間の会議を行った際，設定に必要な発言回数 x，解除に必要な割り込み回数 y を変えながら，いつ，どこにイマジナリーラインが設定されるかを記録した．この結果と，会話内容の分析と参加者の顔の向きからカメラマンが判断した理想的なイマ

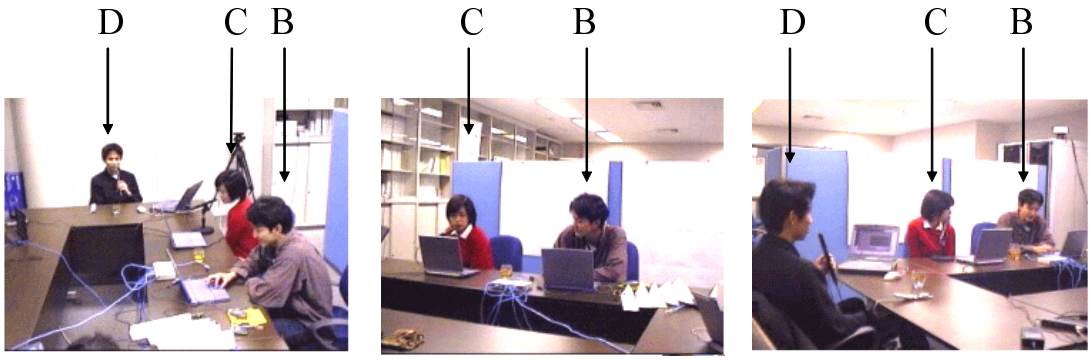


図 3.10: プロトタイプにおけるスイッチング例

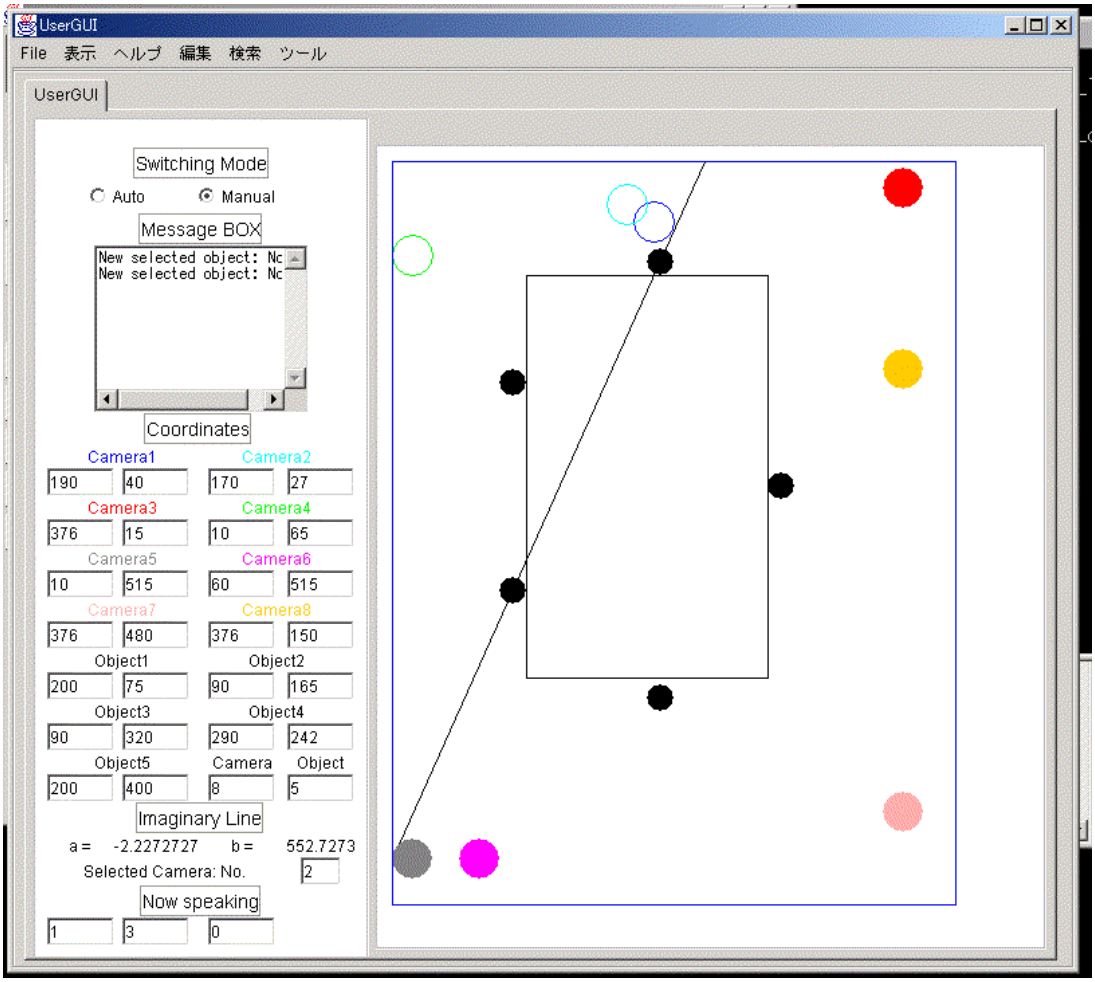


図 3.11: システム動作画面（2 人の間のイマジナリーライン）

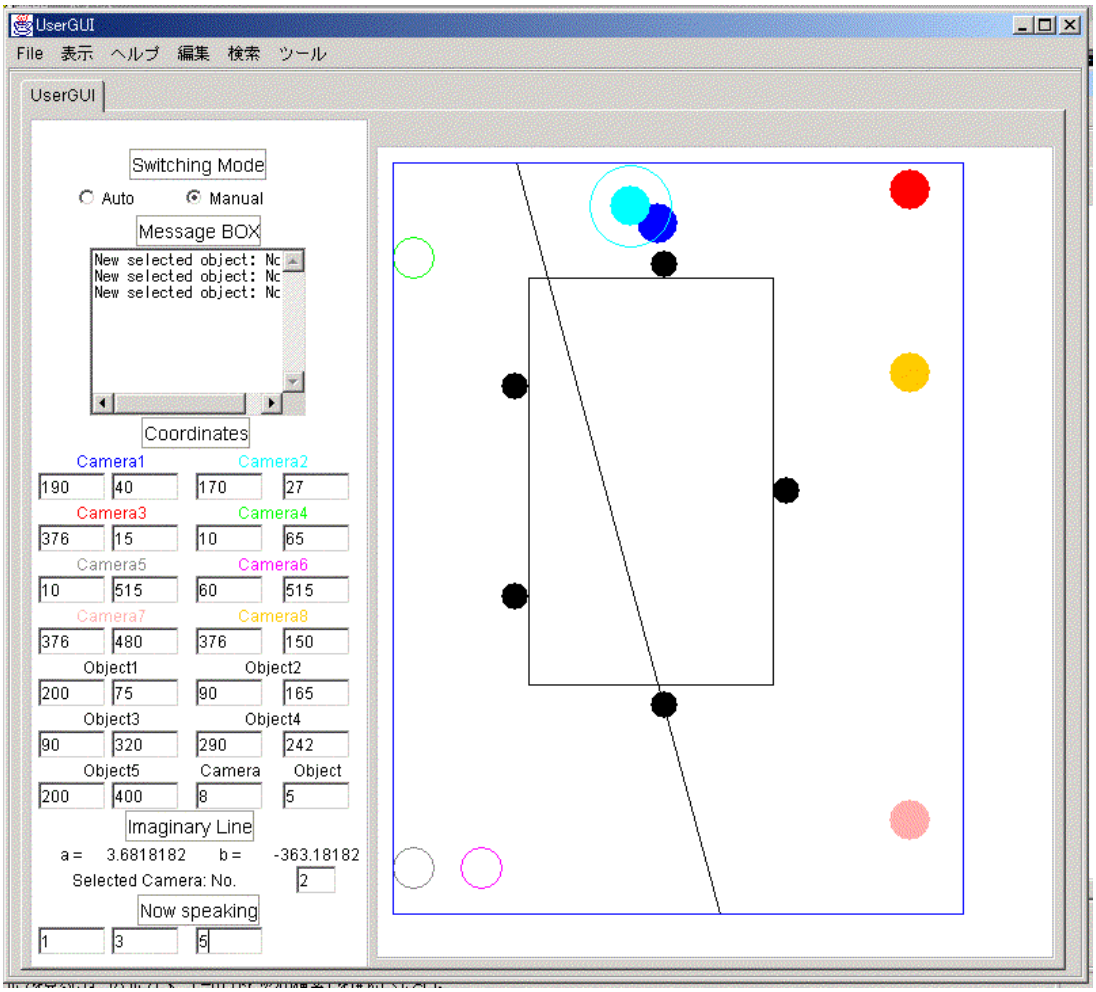


図 3.12: システム動作画面（3 人の間のイマジナリーライン）

表 3.4: 2 者間対話における撮影カメラ

参加者＼相手	B	C	D	E
A	4,6,8	5,6,8	5	3,4,6
B	–	2,6,7	2,6,8	4,6
C	–	–	3,6,8	2,3,8
D	–	–	–	3,5,8

ジナリーラインの発生時刻・位置とを比較した．ここで，$x = 2$ では常に会議空間上のどこかにイマジナリーラインが設定されてしまう．また，$y = 1$ では複数人の議論に対応することができないため，これらの値は除外した．

その結果，$x \geq 4$ では条件が厳しくなり，必要なイマジナリーラインがほとんど設定されなかった．また，$y \geq 3$ にすると一度設定したイマジナリーラインがなかなか解除されなかった．これらの値では，会議室の状況の変化に十分対応できないといえる．よって本研究では $x = 3$，$y = 2$ として以降の評価実験を進めていくこととする．

3.5　評価実験

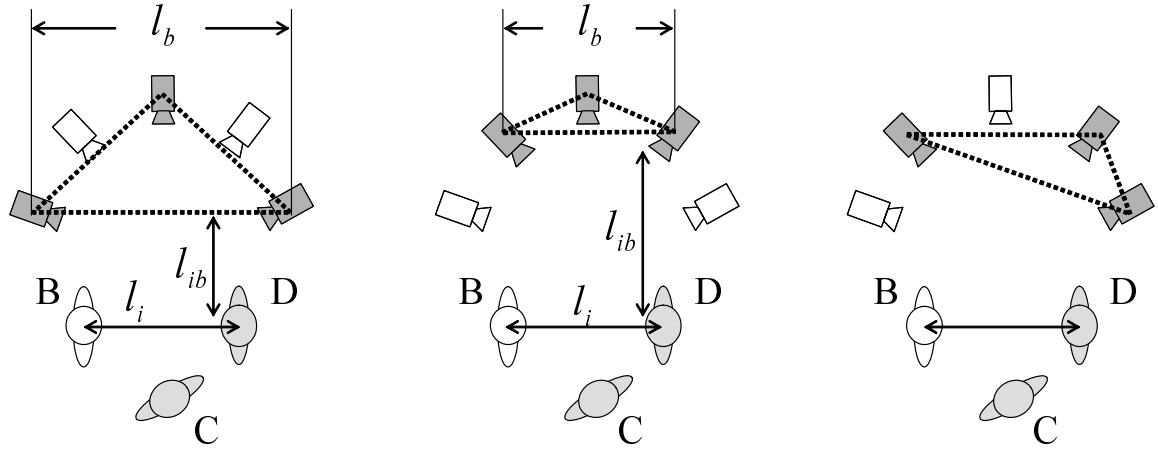
3.5.1　イマジナリーライン検出方法の評価

提案手法でどの程度イマジナリーラインを検出できたかを評価するため，図 3.9 のレイアウトによる対面会議を 10 分間撮影した．このとき，表 3.3 のすべてのショットを用意しておき，撮影後に会話内容の分析と参加者の顔の向きから理想的なイマジナリーラインの発生時刻・位置を 1 秒ごとに手動で決定し，提案手法で検出したイマジナリーラインの位置・時刻と比較を行った．また，両者のショットをそれぞれのイマジナリーラインをもとに編集し，映像表現に与える影響についても分析した．

3.5.2　撮影カメラ決定方法の評価

撮影カメラ決定方法の違いによる影響を調べるため，図 3.9 の B，C，D による対話・議論を 3 つのカメラ配置で撮影した．図 3.13 にこの配置を示す．

(a) は提案手法で決定した配置である．与えられたカメラの中から，頂角カメラはイマジナリーラインが設定されている参加者から均等に遠い位置にあるものを，底辺カメラはイマジナリーライン上の参加者それぞれに近いものを選択した．イマジナリーラインが



(a) 提案手法　(b) 小さな三角形　(c) いびつな三角形

図 3.13: 比較されるカメラ配置

設定されている参加者間の距離を l_i，底辺カメラ間の距離を l_b，イマジナリーラインと三角形の底辺までの距離を l_{ib} とすると，この配置では l_b が l_i に比べて十分に長い．(b) は，頂角カメラは (a) と同様に選択するが，底辺カメラはイマジナリーラインから 1 台遠い位置にあるものを選択した場合である．結果として (a) と比べて l_{ib} が長く，l_b と l_i の長さに差がない配置となった．(c) は他の 2 つに比べていびつな配置であり，頂角カメラを一方の参加者よりにずらし，底辺カメラも一方はイマジナリーラインから 1 台遠い位置にあるものを選択した．

次にそれぞれの映像（約 5 分）を大学生の被験者 16 人に見比べてもらい，1 分ごとにどの映像が好ましいかを順位付けするように指示した．1 位に選ばれていた場合には 3 点，2 位の場合は 2 点，3 位ならば 1 点として点数をつけ，順位付けの理由も簡潔に記述してもらった．スイッチングのタイミングは 3 つの映像すべてで共通とした．

3.5.3　映像の主観評価

本研究で意図した，位置関係と人物の対話を強調する効果が実際に映像に現れているかどうかを確認するため，プロトタイプで自動撮影した映像を大学生の被験者 16 人に見てもらい，アンケートに 5 段階で評価してもらった．

実際の映画やテレビでは，本研究が対象とした位置関係や対話の強調以外にも多くのカメラワークが存在する．よってこれらの映像とを単純に比較することはできない．そこで大学生の被験者 1 人にプロトタイプを使ってもらい，手動で撮影カメラとショット持続時

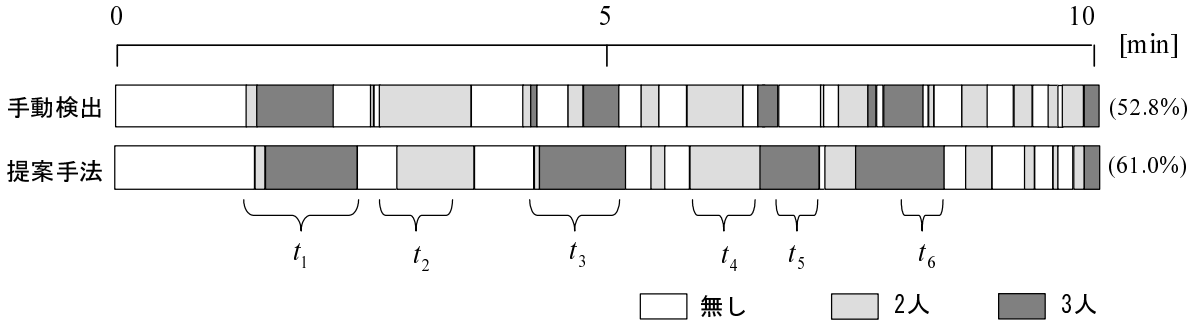


図 3.14: イマジナリーライン検出のタイムチャート

間を決定した映像を用いた．この比較により，一般の人に撮影を依頼する場合と比べてどの程度見やすい映像が自動撮影可能かを評価した．

アンケートの質問項目を表 3.5 に示す．イマジナリーラインの効果を比較するため，映像から伝わる位置関係に関するの質問項目を用意した（項目 4，5，6）．また，カメラの三角形配置の効果を比較するために，参加者の映り具合に関する質問を用意した（項目 1，2，8，9，11）．残りの項目は映像演出用のスイッチングに関するものである（項目 3，7，10，12）．

また，アンケート以外にも気になった点やシステムへの要求など自由なコメントを記入してもらった．

3.6　結果および考察

3.6.1　検出精度の影響

結果をタイムチャートとして図 3.14 に示す．濃色の部分はイマジナリーラインが設定されていた時間帯を表しており，設定場所については表記していない．右横の括弧内の数値は映像全体においてイマジナリーラインが設定されていた時間の占める割合である．

イマジナリーラインを正確な位置･時刻に設定できた割合を表すカバー率 P は，図 3.15(a) における手動による設定時間 T_i と，提案手法による位置・時刻が手動設定のそれと一致した時間 T_c を用いて次のように定義する．

$$P = \frac{T_c}{T_i} \times 100(\%) \tag{3.1}$$

ここで，提案手法の設定時間の合計は手動よりも約 50 秒長く，冗長な設定が含まれている．例として図 3.15(b) ではカバー率が 100%であるが，冗長な設定が多いこともわかる．そこで，いかに無駄なく正確な位置・時刻にイマジナリーラインを設定できたかを有

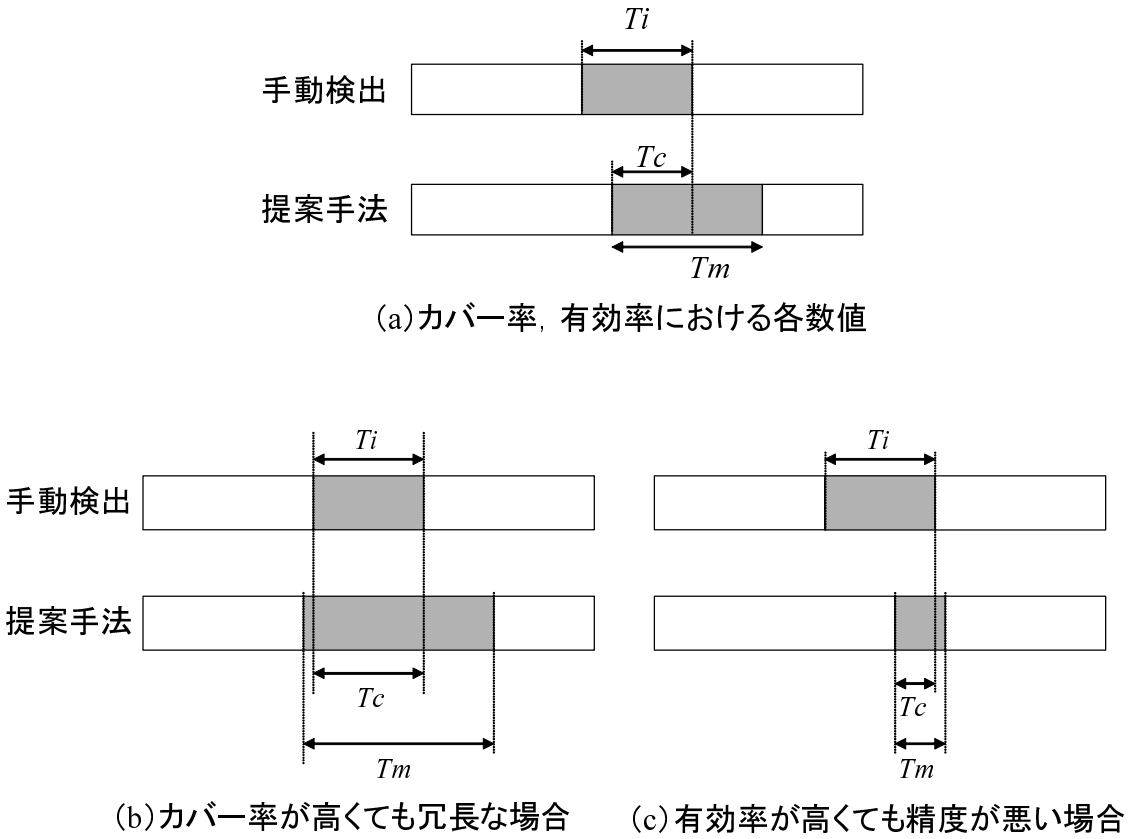


図 3.15: カバー率 P と有効率 E の定義

効率 E として定義する．この E は，図 3.15(a) における提案手法の設定時間 T_m と，先ほどの T_c を用いて次のように表すことができる．

$$E = \frac{T_c}{T_m} \times 100(\%) \tag{3.2}$$

ただし，有効率だけでも設定の正しさを評価できない．例として図 3.15(c) では有効率が 100%に近いにもかかわらず，正しい設定はほとんどできていない．よってカバー率，有効率の 2 つを同時に見ていく必要がある．

この実験を通してのカバー率 P は 70%であった．内訳は，2 人の対話時が 57%，3 人の議論時が 93%となった．一方，有効率 E は 61%であり，同様に内訳は 2 人の場合に 67%，3 人の場合は 56%となった．結果より，2 人の場合はカバー率は低いが有効率は 3 人に比べて良い．逆に 3 人の場合はカバー率は高いが有効率が低く，正確な位置と時間帯に設定はできたものの，それと同程度不必要な設定も存在していた．

この原因は設定と解除にかかる遅延である．図 3.14 中 t_1 の前半部分では，B が D に向かって直接話しかけ，D はその話を黙って聞いていた．手動ではこの時点で BD 間にイマ

ジナリーラインを設定したが，提案手法は $B \rightarrow D \rightarrow B$ と 3 ステップの発言を必要とするため，D と B の発言を待つ必要がある．実際，手動で設定された時間が 7 秒であったのに対して，提案手法では 5 秒遅れて設定されたため 2 秒間しか一致しなかった．2 人の対話時にカバー率が低いのは，このような少ない発言数で設定される場合に対応が遅れる影響が大きい．

また，t_1 後半ではしばらく 3 人の議論が進行したが，うち 1 人が途中から全員に向かって話を始めた．手動ではこの時点で解除されたが，提案手法は無関係な参加者の発言を基準にしているためこれを認識できず，解除が 13 秒遅れた．同様の例は t_3 前半，t_5 にも見られ，冗長部分が増えて有効率が低下した原因となっている．このように解除遅延は聞き手の変化が主な原因であるが，人数が 5 人だと議論中以外の参加者は 2 人と少なく発言頻度が低いことも影響がある．逆に 2 人の間のイマジナリーラインは，他の参加者が 3 人いるため解除されやすく有効率が高くなったと考えられる．

この遅延は 2 カ所で，編集映像の違和感につながった．t_3 前半では BCD の議論に表 3.3 のカメラ 2（shot 1），6（shot 2），8（shot 2）が選択される（図 3.9 および表 3.3 参照）．その後，t_3 中間部において手動では AC 間にイマジナリーラインが設定されたが，提案手法ではこれを認識できず，以前のイマジナリーラインが解除されなかった．先ほどの 3 つのショットには A を映すものは 1 つもないため，A が重要な発言をしているにもかかわらず声だけが聞こえる映像となってしまった．同様の状況は t_6 でも発生した．

しかしそれ以外の部分では，映像表現への影響は少なかった．t_1，t_2，t_4 のように，より演出が効果を持つ長時間の対話・議論のカバー率は 80% を超えている．t_2 では設定遅延が目立つが，これは $A \rightarrow D \rightarrow A$ という流れの中で D の発言が長時間続いたことによる．この間，提案手法では D を含むショットと全景ショットの 2 種類を切替えるため，特に違和感のない映像となった．t_1 後半や t_5 で目立つ解除遅延では，手動でイマジナリーラインが解除された後に発言を続けたのが，直前までその設定に関与していた参加者であったため，t_2 と同様に違和感のない映像を生成することができた．

3.6.2　カメラ配置の影響

映像全体を通しての平均得点は (a)2.43，(b)1.15，(c)2.43 となり，(a) と (c) が高く評価された．

提案手法の配置である (a) は，開始から 3 分までの B と D が積極的に発言した時間帯に常に評価が高かった．その理由として “表情がよく見える” というものがあった．(a) の底辺カメラは BD 間のイマジナリーラインの近くに位置し，それぞれの主観的視点に近い肩越しショットが得られる．このため BD の対話時に “話しかけている”，“話を聞いている” 表情をより正面から撮影できたためだと考えられる．

一方，(b) は全体を通して評価が低く，“変化に乏しい”，“横顔が多い” などのコメントがあった．(b) の底辺カメラの設置位置は頂角カメラに近く，スイッチングを行っても各ショットにあまり差異が見られず，平面的な映像になってしまった．提案手法はイマジナリーラインに近いものを底辺カメラとして選択するため，基本的には与えられた環境の中から最も肩越しショットに近いものを撮影できる．しかし，l_b が l_i と比べて差がない，もしくは短いようなカメラしか選択できない状況では肩越しショットに適した視点を確保することが難しい．この場合は位置関係の演出を取りやめ，ズームなどで別の演出を試みるほうが良いと思われる．

(c) の得点は，D と C が B に対して積極的に発言し，B が聞き役にまわっていた時間帯（3 - 5 分）で評価が高かった．発言頻度の極端な偏りにより D が重要人物と認識され，頂角カメラも D の主観的視点に偏った構図が評価されたようだ．底辺の大きな三角形状にカメラを配置する提案手法は，我々の想定した双方が発言を繰り返すシーンには適しているといえるが，発言頻度と配置に関して会議の種類や進行方法を含めて新たに検討してみる価値があると思われる．

3.6.3　アンケート回答結果の分析

アンケートの結果を表 3.5 に示す．各質問は “まったくあてはまらない”，“あまりあてはまらない”，“どちらともいえない”，“ややあてはまる”，“かなりあてはまる” の 5 段階にそれぞれ 1 点から 5 点を与え，映像別に各質問に対する平均得点を求めた．さらに，人手，提案手法のそれぞれの評点に有意差があるか確認するため，Wilcoxon の符号付順位検定により p 値を求めた．表中の人手・提案手法の各項目の値は評価値の平均得点である．

提案手法による映像は，有意水準 5%で検定を行ったところ，6 項目で有意であると評価された．項目 1 および 2 の結果から，人手よりも議論や会話の様子が分かりやすかった．また項目 4，5，6 からは，参加者の位置関係が明確で違和感の少ない映像であるという評価も得た．提案手法によるイマジナリーラインの検出と撮影カメラの決定が演出効果として現れたのが分かる．

これに対し，項目 8，9，10，11，12 では人手の映像が上回ったか，差がほとんど見られなかった．

項目 10 および 12 はスイッチングのタイミングに関するものである．本研究では，対話の推移しか利用せず，“落ち着いた”，“白熱した” といった会話の意味的内容までは踏み込んでいない．よって眺めの持続時間が望ましい落ち着いた状態においても，強制的に次のショットへ移行してしまうケースが見られた．また，プロトタイプでは特に意味の無い発言にも反応してしまう．人手による映像はカメラマンが会話内容を判断したため評価が高くなったものと考えられる．

表 3.5: 比較実験におけるアンケートの評価結果

No	質問項目	人手	提案手法	Wilcoxon 符号付順位検定 p 値
1	議論の流れがつかめた	2.81	3.94	**0.0017
2	だれとだれが会話しているかがよく分かった	2.69	4.06	**0.0017
3	映像に退屈しなかった	2.94	3.81	**0.0068
4	人物の位置関係がよくつかめた	3.13	3.88	*0.0107
5	カメラの切替に違和感を感じなかった	2.50	3.25	*0.0244
6	その場の状況が分かりやすかった	3.19	3.81	*0.0269
7	見やすい映像だった	3.00	3.44	0.1309
8	話し手がよく分かった	4.19	4.00	0.4375
9	画面上の人物の表情や身ぶりがよく分かった	3.69	3.50	0.4961
10	見たい映像に切り替わっていた	3.31	3.25	0.8438
11	画面上の人物の存在感があった	3.38	3.38	0.9999
12	切替えのタイミングは適切だった	3.06	3.06	0.9999

(N=16; **:$p < 0.01$, *:$p < 0.05$)

項目 8，9，11 は参加者 1 人の映り具合に関するものであった．プロトタイプは発言の意味的内容を理解しないため，人手に比べてスイッチングの回数が多くなる．その結果，参加者をじっくりと撮影することなく次々とショットが変わってしまったためだと考えられる．

また，アンケート項目以外のコメントには “映像が跳ねている” ように感じるという意見があった．図 3.10 の各ショットは B と C の画面上の位置がずれている．このため，これらショットを接続すると参加者が振動しているように見えてしまうことがあった．ショット接続の際に人物の視線や位置を一定に保つなど，新たな映像理論を取り入れることで改善できると考えられる．

3.7　まとめ

本章ではイベント型シーンの例として，映像文法に基づく対面会議の自動撮影手法を提案した．概念的で実体のないイマジナリーラインをシステムから利用するために，会議状況の分類，発言の推移に基づくイマジナリーラインの具体的な設定方法，エスタブリッシュショットと肩越しショットを撮影するためのカメラ決定方法を提案した．評価用映像では，提案方式のイマジナリーラインのカバー率は 70%，有効率は 61%であり，この検出ロスが映像の違和感につながる箇所は少なかった．撮影カメラ決定方法は，参加者が発言を繰り返すシーンにおいて効果があり，人手で撮影カメラを選択した場合と比較して位

置関係のわかりやすい映像を自動生成することができた．

一方，発話の推移だけを利用しているため，少ない発言数でのイマジナリーラインの設定と，聞き手が途中で変化した場合の解除には対応できず，冗長な部分が多いこともわかった．今後，顔や視線の向き，会話の意味的内容の認識などを組み合わせるなどして改善していく必要がある．また，会議では開始直後は発言数が少なく中盤には多くというように，時間に応じて進行方法が変動する．この変動に合わせてイマジナリーラインの設定に必要な発言数も適宜変化させるなどしていく必要があるだろう．

本研究での演出は位置関係に重点を置いたが，これ以外にも会議シーンを効果的に撮影する技法は数多く存在する．例えば緊張感や迫力を強調するには，適宜ズームアップしたショットを挿入していく必要があるだろう．また，ショットの切替え時に次のショットに音声をずれ込ませる “ずり上げ” や “ずり下げ” と呼ばれる技法を用いれば，発言に余韻を持たせたり，人物間の上下関係を強調できるため，会議記録に検討する余地は大きいと考えられる．

第4章　オーケストラ演奏の自動撮影

4.1　はじめに

本章では，シナリオのあるストーリー型シーン [87, 88] の自動撮影を試みる．ストーリー型における映像制作のプロセスは，“いつ”，“何を”，“どのカメラで撮影するか” というカメラワークを決める計画フェーズ，その計画に従ってカメラを制御し撮影を行う実行フェーズ，撮影された映像を編集する編集フェーズの 3 つに分類することができる．実行フェーズではシナリオと実際のシーンの時間的・空間的ズレの修正 [77]，編集フェーズでは特定場面の検索機能 [46] などが必要となるが，良い映像を作るには，編集フェーズで必要とされる映像素材を確実に提供できるカメラワークを計画することが重要になる．

しかし，シーンに多くの被写体がいたり，カメラの台数や設置位置に制限がある中では，カメラワークを 1 つ 1 つ決定していくことは，映像知識に乏しいユーザには大きな負担である．このカメラワークが適切に計画されていないと，必要なショットが撮影されていなかったり，別々のカメラで似たようなショットを撮影したりして，編集フェーズで効果的な編集を行うことができないといった問題がある．

そこで本研究では，シナリオからカメラワークを自動的に計画することを目的とする．提案手法では撮影対象としてオーケストラ演奏を想定し，シナリオである楽譜から被写体の候補を抽出する．次に，編集時に必要となるショットをできるだけ確保するために，映像文法を “バラエティに富んだショット” を撮影するための技法としてとらえる．そして抽出された候補に対し，被写体と構図の種類に基づいて優先度をつけ，限られた台数のカメラに効率よく割り当てる．カメラワークを自動的に生成，確認可能な仮想空間ベースのプロトタイプを実装して評価を行い，本手法の有効性を確認する．

以下，4.2 節では本章が撮影対象とするオーケストラについて，4.3 節では提案するカメラワーク計画方式について，4.4 節では実装したプロトタイプについて，4.5 節ではプロトタイプを用いた実験方法について，4.6 節では実験の結果および考察を述べ，4.7 節を本章のまとめとする．

4.2　撮影対象

本研究ではシナリオを “いつ，どこで，誰が，何をするといった撮影対象に関する情報が時間軸に沿って書かれているもの” と定義する．また，カメラワークを “いつ，何を，どのカメラで撮影するのか決定すること”，と定義する．本章で述べるシステムは，このシナリオを入力として，シナリオの開始から終了に至るまでのカメラワークを自動的に出力する．

ここで，カメラワークの計画方法は，目的とする映像の特徴によって変わってくる．表 4.1 にその分類を示す．映像全体を通して同じ被写体を集中的に撮影したり，同時刻に別々のカメラで同じ被写体を撮影するようなカメラワークは，ドラマのように特定の被写体

表 4.1: カメラワーク計画方法の分類

	撮影方法	映像の特徴
映像全体の構成	同じ被写体	主役がいる映像（ドラマ等）
	異なる被写体	主役がいない映像（紹介ビデオ等）
同時刻の各カメラ	同じ被写体	主役がいる映像
	異なる被写体	主役がいない映像
スイッチング前後	似た構図	変化の少ない安定感のある映像
	異なる構図	変化に富んだ躍動感のある映像

（主役）を中心に編集をするような映像に適している．逆に映像全体を通して異なる被写体をまんべんなく撮影したり，同時刻に別々のカメラで異なる被写体を撮影するようなカメラワークは，紹介ビデオのように主役が存在せず，被写体全体を網羅するような映像に適している．また，スイッチングの前後で各カメラが常に似たような構図のショットを撮影するカメラワークは，変化が少ない安定感のある映像を編集するのに適している．逆に各カメラで異なるショットを撮影するカメラワークは，変化に富んだダイナミックな映像を編集するのに適している．

本節では，具体的な撮影対象であるオーケストラの特徴を述べ，これを撮影するためのカメラワークに求められる要件について議論する．

4.2.1　オーケストラ

オーケストラとは，さまざまな種類の楽器を指揮者のもとで大合奏すること，またはその団体を指す．オーケストラの代表的な楽器編成と略称を図 4.1 に，舞台上の配置を図 4.2 に示す．各楽器は木管楽器，金管楽器，打楽器，高弦楽器，低弦楽器という 5 つのグループから成り立ち，これらは演奏形態の違い*から管打楽器，弦楽器にまとめられ，指揮者を含めて全体となる [89, 90, 91]．このように各楽器はグループ毎にまとめられた位置に配置される．

また，オーケストラには各楽器がどのような演奏をするかが音符で表現された楽譜（以下スコア）が存在する．これは “どこで，誰が，何をする” というイベントが時間軸に沿って書かれたシナリオとしての性質を持っている．

*管楽器と打楽器は 1 つの旋律を 1 人で演奏するのに対し，弦楽器は同じ旋律を複数人で演奏する特徴がある．

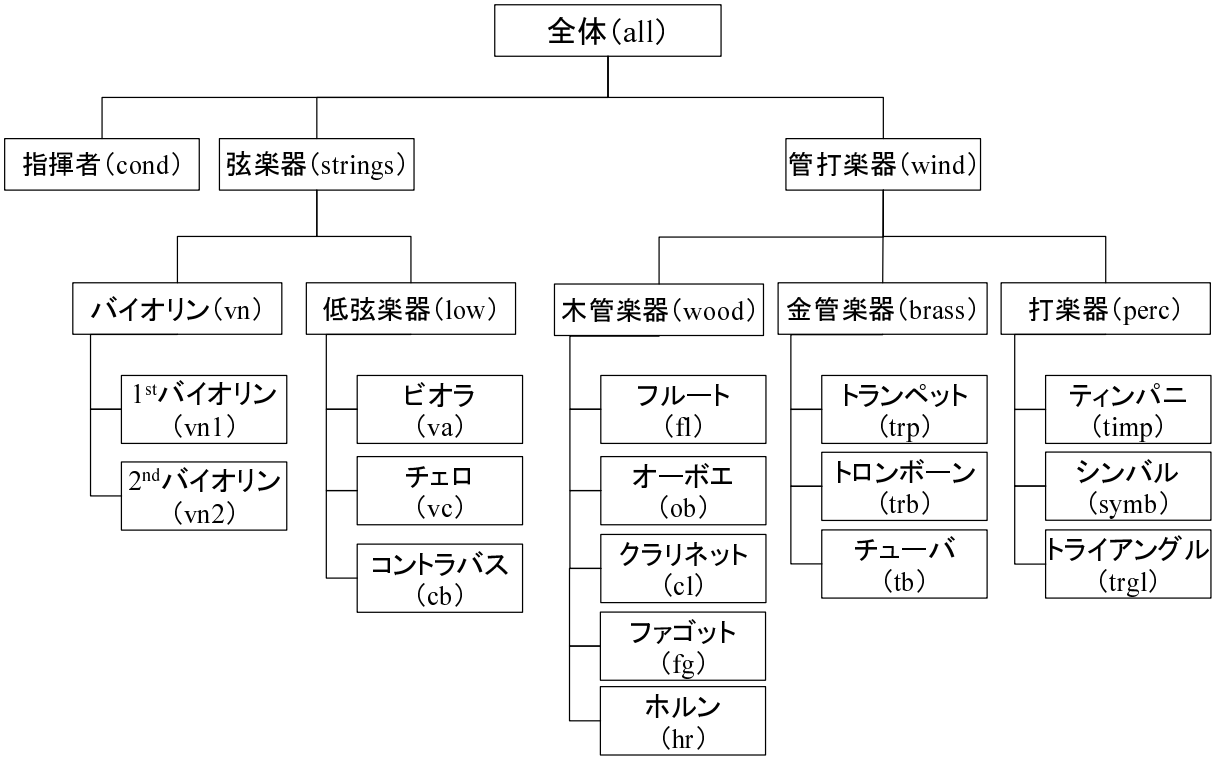


図 4.1: オーケストラの編成例

4.2.2　定性的分析

オーケストラ演奏の映像は，演奏者の記念や団体の紹介用として撮影されることが多く，ドラマのように 1 つの被写体を集中して撮影するよりも，映像全体を通してできるだけ多くの楽器を撮影することが期待されている．

また，同時にいくつもの楽器が演奏するため，同じ時間帯でも人によって注目する楽器が異なる．このようなシーンのカメラワークは，同時刻で各カメラが別々の被写体を撮影し，編集時におけるユーザの要求にできるだけ応えられるようなカメラワークが望ましい．

さらに，会場によってはカメラの設置位置が制限されており，常に同じカメラ配置で撮影できるとは限らない．ある配置では正面から撮影できる演奏者は，別の配置では真横からのショットになってしまうこともある．このような設置位置の変動にも対応できる必要がある．

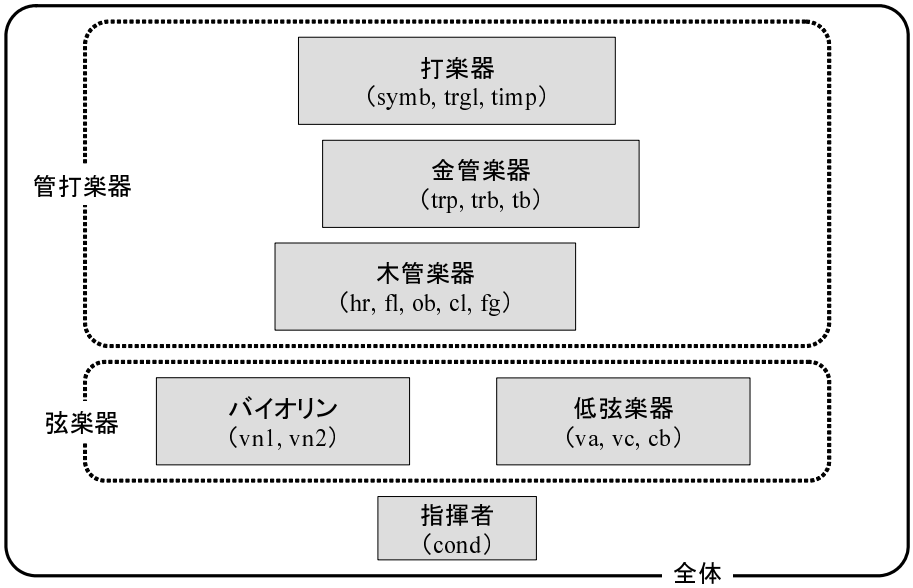


図 4.2: オーケストラの舞台における配置例

4.2.3　映像分析

オーケストラ演奏に適したカメラワークに関して，定性的分析で得られる以外の知見を得るため，プロのカメラマンが撮影・編集した映像 3 本を分析した．いずれの映像もマイスタージンガー前奏曲の演奏シーン約 10 分（223 小節）を撮影したものである．

映像分析の一般的な方法として，(1) ショットの内容に応じた分類，(2) ショット切替えのタイミング，(3) どの種類のショットからどの種類のショットへ移るかという遷移，の 3 点がある [56]．本研究ではこれに加え，(4) シナリオであるスコアとショットの関連性，についても分析を行った．

その結果，ショットの内容は図 4.3 に示す 4 種類に分類できた．“指揮者ショット” は演奏を指揮する指揮者を単独で映したものである．“メロディー楽器単独ショット” はフレーズの中心となるメロディーを演奏する楽器 1 種類を映したものである．“組合せショット” は音を出している楽器（演奏楽器）を複数まとめて映したものである．“全体ショット” はオーケストラ全体を映したものである．

次に，ショットの切替えは，メロディーのまとまりであるフレーズ単位で行なっていることがわかった．図 4.4 にフレーズの例を示す．小節はスコアの定期的な時間の区切りであり，これを基準にするとメロディーの途中でショットが切り替わってしまうことがある．これに対しフレーズは，文節における節ないし文に相当する音楽的に意味を持ったまとまりであり，人間の心理に安定感を与える [92]．その境界で映像を切替えることで無理のない映像にしていると考えられる．

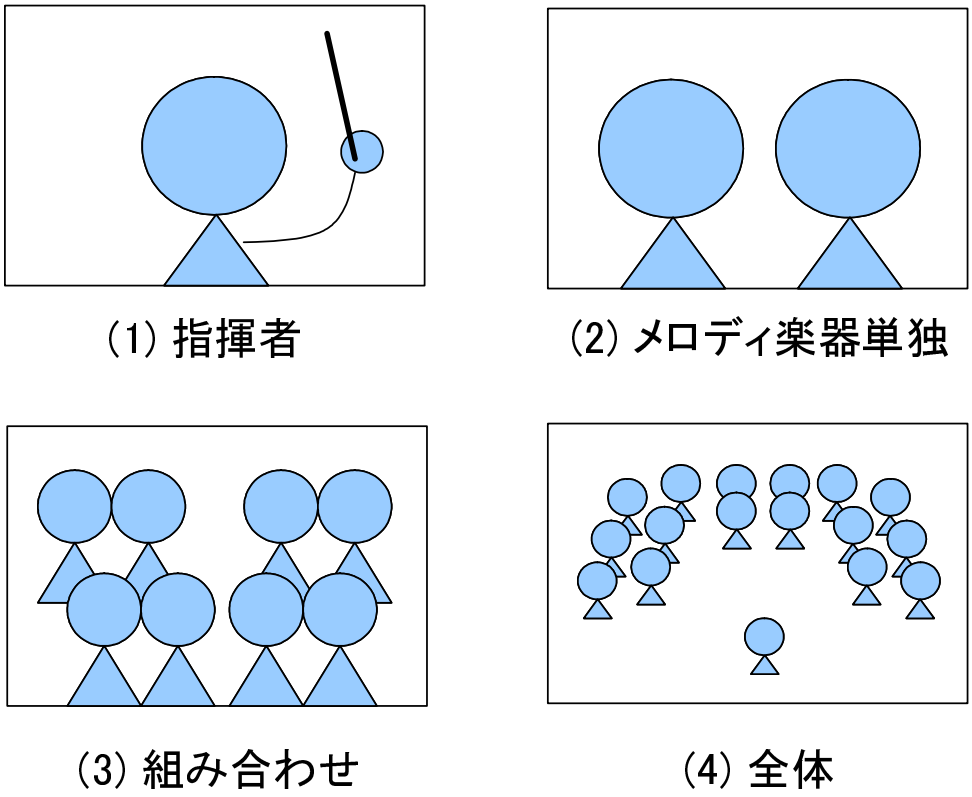


図 4.3: オーケストラ映像におけるショット分類

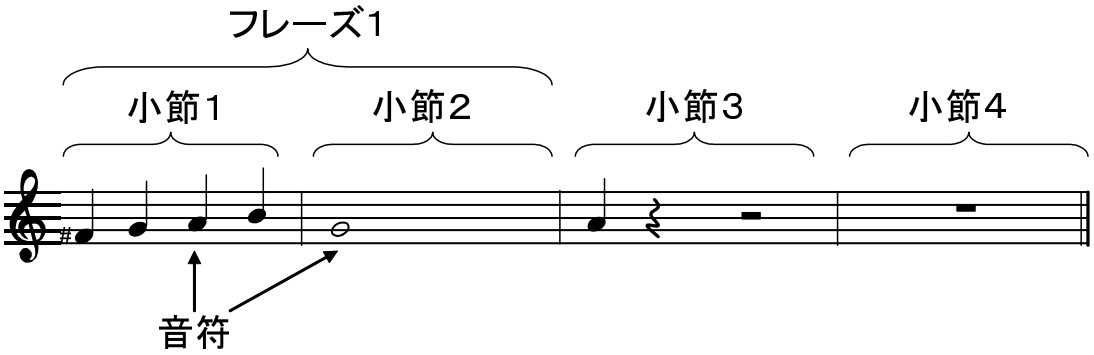


図 4.4: フレーズの一例

ショットの遷移では，切替えの前後でカメラのズーム値（ショットサイズ）に変化をつける傾向があった．一般には，ショットサイズを一定に保ったままスイッチングすると安定した映像が得られる．逆に構図を変化させると躍動感のある映像が得られる [93]．オーケストラ演奏では，アップショットとワイドショットを交互にしたりするなど，ショットの印象を変えるようにしていた．

最後に，スコアとショットは，演奏楽器の数と関連があることがわかった．ごく一部の楽器だけが演奏している “ソロ” 状態（全楽器数の 2 割未満）では，ほとんどがメロディー楽器ショットであり，稀に指揮者のショットが挿入されていた．いくつかの楽器が演奏している状態では，メロディー楽器ショット，組合せショットが多く存在した．ほとんどの楽器が演奏している “大合奏” 状態（全楽器数の 6 割以上）では，メロディ楽器ショット，組合せショットに加えて，指揮者または全体ショットが多く存在した．

4.2.4　要求されるカメラワーク

以上の分析をまとめると，オーケストラ演奏に求められるカメラワークの基本的な計画方針は次のようなものになる．これらは，なるべく多くの被写体を様々な構図で撮影する，すなわちバラエティに富んだショットを撮影するカメラワークと考えることができる．

- 映像全体を通して様々な楽器を撮影する．
- 同時刻に各カメラで様々な楽器を撮影する．
- 印象が異なるショットを撮影する．
- カメラの設置位置を考慮する．

また，シナリオや被写体との関連性は次のようになることがわかった．

- フレーズ単位でカメラワークを計画する．
- ショットは 4 種類（指揮者，全体，メロディー楽器，組み合わせ）．
- 演奏楽器数によってショットが変わる．

次節以降，このカメラワークをどのように自動的に計画していくのかについて述べる．

4.3　提案手法

本節では，4.2.3 節の分析の結果得られた知見を利用したカメラワークの計画手法について述べる．提案手法の流れを以下に示す（図 4.5）．

(1) 最初にシナリオを一括して読み込む．

(2) 読み込んだシナリオを解析する．

(3) 編集時にユーザに必要とされる演奏楽器（被写体候補）を抽出する．

(4) (3) の被写体候補の優先度を計算する．

(5) 最も優先度の高い候補を，最も良く撮影可能なカメラに割り当てる．

(6) 何も撮影していないカメラ（空きカメラ）がなくなるまで，(4) の優先度付けステップ以降を繰り返す．

(7) 最後のフレーズまで，(2) からの処理を繰り返す．

このように提案手法では，スコアから生成したシナリオを入力として，“ある時間帯 t において”，“被写体 s を”，“カメラ c で撮影する” というカメラワークを先頭フレーズから 1 つずつ自動的に決定していく．その際，映像文法を “バラエティに富んだショット” を撮影するための技法としてとらえ，被写体の種類や構図の変化に応じて優先度を計算し，その値に基づいて限られた台数のカメラへ被写体を割り当てていく．

なお，2.3.12 節に述べたように，シナリオと実際のシーンの状況は完全には一致せず，被写体の位置やカメラワークのタイミングに微妙なズレが生じる可能性がある．これらに関しては既存の研究 [77, 78] が扱っているため本研究では対象としない．

ここで，ステップ 3 の用語について言及する．オーケストラにおける演奏者の数は，用意されるカメラの台数よりも多い．多くの楽器が同時に演奏した場合は，全ての演奏パートにカメラを割り当てることができない．そこで，シナリオで何らかの役割（演奏）を果たし，カメラが割り当てられる可能性のあるものを “被写体候補” と呼ぶ．一方，カメラワークを計画した結果，実際にカメラが割り当てられたものを “被写体” と呼ぶこととする．

以降，計画の開始から終了までの各ステップの詳細を述べる．

4.3.1　シナリオの読み込み

計算機がスコアをシナリオとして扱いシーンの演奏状況を把握するには，そのデータ形式が定義されている必要がある [94]．本研究では XML で次の 2 点を記述したものをシナリオとして扱う．なお，スコアからシナリオへの作成は事前に手動で行うものとする．

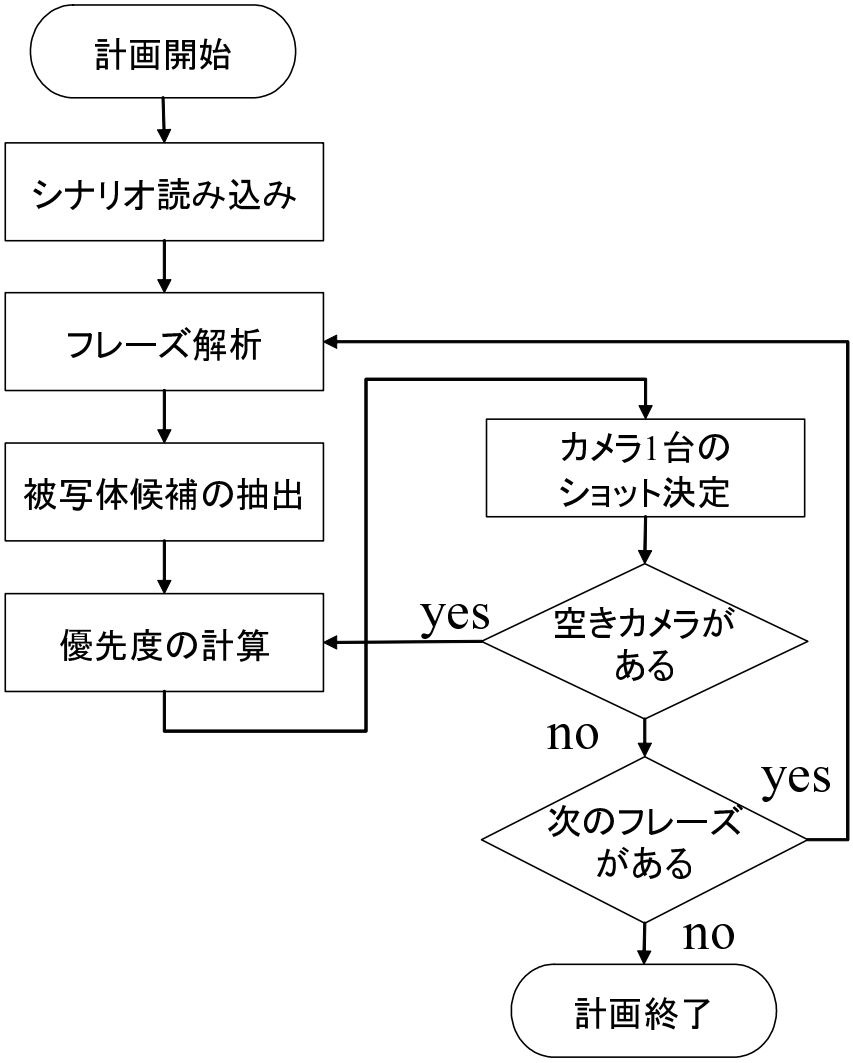


図 4.5: カメラワークの計画手法

```
<!ELEMENT stage_map (stage, orchestra)>
<!ELEMENT stage (name, w, h, d, sw, sh, sd)>
.....
<!ELEMENT fl (level, position)>
<!ELEMENT ob (level, position)>
<!ELEMENT cl (level, position)>
<!ELEMENT fg (level, position)>
.....
```

図 4.6: シナリオの DTD（舞台情報）

```
<!ELEMENT music (info, phrase*)>
<!ELEMENT info (title*, composer*, year*)>
<!ELEMENT title (#PCDATA)>
.....
<!ELEMENT phrase (no, start, end, main_part*, sub_part*)>
<!ELEMENT no (#PCDATA)>
<!ELEMENT start (#PCDATA)>
<!ELEMENT end (#PCDATA)>
<!ELEMENT main_part (#PCDATA)>
<!ELEMENT sub_part (#PCDATA)>
```

図 4.7: シナリオの DTD（フレーズ情報）

楽器の編成および位置

オーケストラでは，楽器の編成およびその配置は楽曲によって決まる．そこで舞台の大きさ，楽器の編成および配置を記述しておく．これを記述した DTD(Document Type Definition) の一部を図 4.6 に示す．舞台の大きさ等（stage），舞台上に存在する楽器名の略称（fl，ob 等），および三次元座標（position）が記述されている．

フレーズと各楽器の役割

カメラワークはフレーズ単位で計画されるため，シナリオにはフレーズに関する情報を記述しておく．DTD の一部を以下に示す．フレーズ番号（no），開始時刻（start），終了

時刻（end），メロディーを演奏している楽器のリスト（main_part），伴奏を演奏している楽器のリスト（sub_part）が記述されている．

4.3.2　フレーズの解析と被写体候補の抽出

読み込んだシナリオを解析し，ユーザの注目が集まりやすい被写体候補を抽出する．この決定には，4.2.3 節の分析結果に基づき演奏楽器の数を利用する．フレーズ $i(i = 1, 2, \cdots, n)$ における演奏率 $E(i)$ は，シナリオから抽出可能な舞台上の全楽器数 A と演奏楽器数 $I(i)$ から次のように求まる．

$$E(i) = \frac{I(i)}{A} \tag{4.1}$$

次に，フレーズ i における被写体候補 H_i が a と b である場合を $H_i = \{a, b\}$ と表すものとし，この H_i と演奏率 $E(i)$ の関係を次のように定める．

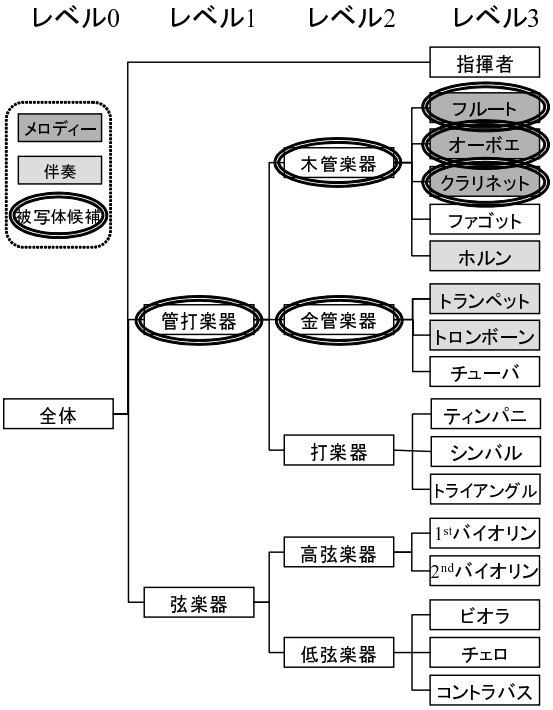
$$H_i = \begin{cases} \{M_i, C\} & (E(i) < 0.2) \\ \{M_i, G_i\} & (0.2 \leq E(i) < 0.6) \\ \{M_i, G_i, C\} & (0.6 \leq E(i)) \\ \{M_i, G_i, W\} & (0.6 \leq E(i)) \end{cases} \tag{4.2}$$

ここで M_i はフレーズ i におけるメロディー楽器，G_i は演奏楽器の組み合わせ，C は指揮者，W はオーケストラ全体の各ショットをあらわす．演奏率の境界値である 0.2，0.6 という値は 4.2.3 節のビデオ分析から決定した．

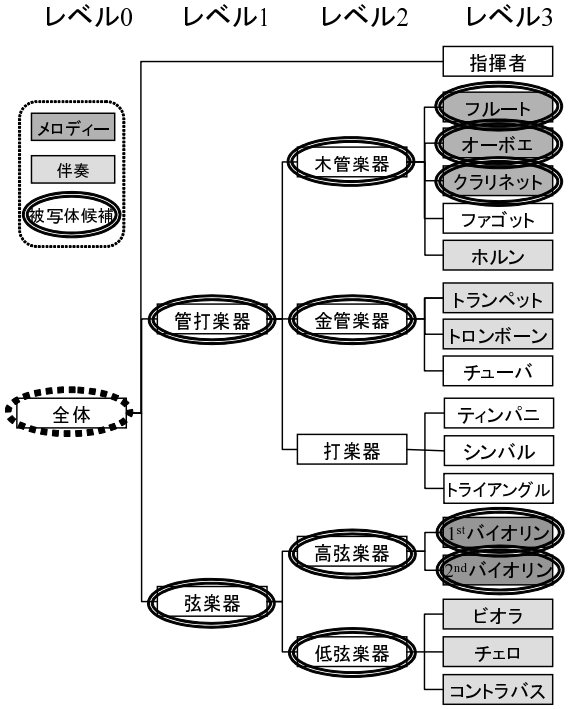
この式 (4.2) より，$E(i)$ が 2 割未満と小さい “ソロ” の場合は，メロディー楽器 M_i と指揮者 C を被写体候補とする．

$E(i)$ が 2 割以上 6 割未満の “いくつかの楽器が演奏している” 場合は，メロディー楽器と演奏楽器の組合せを被写体候補とする．この組合せを作るために，図 4.8 に示す楽器の階層関係を反映した 4 階層の構造を定義する．そして，同一階層のノードの過半数が演奏楽器である場合，その親ノードを “組合せ” として被写体候補に加えていく．

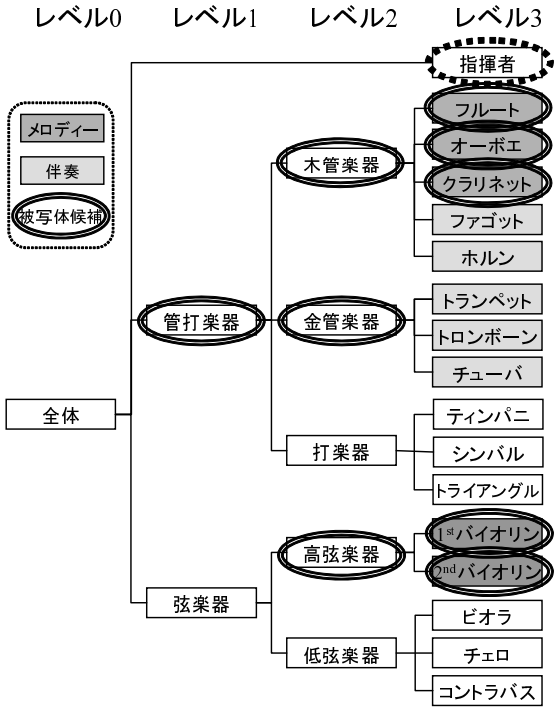
例として図 4.8(a) ではフルート・オーボエ・クラリネットがメロディーを，ホルン・トランペット・トロンボーンが伴奏を担当している（演奏率 0.38）．式 (4.2) より，まずメロディー楽器であるフルート，オーボエ，クラリネットのショット 3 つが被写体候補になる．この時，木管楽器ノードに属する子ノードの過半数がアクティブであるため，木管楽器全体を映すショットを “組合せ” として被写体候補にする．金管楽器ノードに属する子ノードの過半数もアクティブであるため，金管楽器全体を映すショットが同様に被写体候



(a) いくつかの楽器が演奏している状態
($0.2 < E(i) < 0.6$)



(b) 多くの楽器が演奏し”全体”が含まれる状態
($0.6 < E(i)$)



(c) 多くの楽器が演奏し ”指揮者” が含まれる状態
($0.6 < E(i)$)

図 4.8: 階層構造による被写体候補の決定

補になる．さらに，管打楽器の子ノードのうち 2 つがアクティブなので，管打楽器全体を映すショットも加えて計 6 つが被写体候補となる．

$E(i)$ が 6 割以上の “大合奏” にある場合は，組合せの作り方によって指揮者ショットか全体ショットを被写体候補とする．

例として図 4.8(b) はフルート・オーボエ・クラリネット・1st バイオリン・2nd バイオリンがメロディーを，ホルン・トランペット・トロンボーン・ビオラ・チェロ・コントラバスが伴奏を担当している（演奏率 0.69）．まずメロディー楽器の 5 つが被写体候補になる．さらに子ノードの状態から，木管楽器・金管楽器・高弦楽器・低弦楽器・管打楽器・弦楽器の 6 つが被写体候補となる．ここで管打楽器と弦楽器がアクティブであるため，その親ノードである全体を含めた計 12 個が被写体候補となる．

一方図 4.8(c)（演奏率 0.63）では，まずメロディー楽器の 5 つ，組合せとして 4 つが被写体候補になる．しかし全体の子ノードのうちアクティブなのは管打楽器だけであり，過半数に満たない．この場合は指揮者を被写体候補に加えることとし，計 10 個が被写体候補となる．

4.3.3　優先度の計算

式 (4.2) で決定した被写体候補の数がカメラの台数より多い場合，どの候補をカメラに割り当てるのか取捨選択する必要がある．本研究では，各被写体候補に対し優先度を計算し，この値に応じて判断を行う．現在のフレーズ i における被写体候補 x の優先度 $P(i,x)$ の計算方法を式 (4.3) に示す．

$$P(i,x) = \alpha F_p(i,x) + \beta F_f(i,x) + \gamma C(i,x) + \delta D_p(i,x) + \epsilon D_c(i,x) \tag{4.3}$$

ここで，α，β，γ，δ，ϵ は重みをあらわす定数である．以降は F_p，F_f，C，D_p，D_c の詳細について述べる．

出現頻度 F_p, F_f

候補 x がフレーズ $i(i = 1, 2, \cdots, n)$ までにカメラが割り当てられた頻度 $F_p(i,x)$ を計算し，その大きさに応じて優先度を $\alpha F_p(i,x)$ 変化させる．同様に，シナリオを先読みすることでフレーズ $i+1$ 以降に被写体候補になる頻度 $F_f(i,x)$ を計算し，優先度を βF_f 変化させる．$F_f(i,x)$ および $F_p(i,x)$ はそれぞれ次の式 (4.4)(4.5) であらわすことができる．

$$F_p(i,x) = \sum_{k=1}^{i-1} s_{k,x} \tag{4.4}$$

$$F_f(i,x) = \sum_{k=i+1}^{n} c_{k,x} \tag{4.5}$$

ただし

$$s_{k,x} = \begin{cases} 1 & (x \text{ が被写体のとき}) \\ 0 & (\text{それ以外}) \end{cases} \tag{4.6}$$

$$c_{k,x} = \begin{cases} 1 & (x \text{ が被写体候補の時}) \\ 0 & (\text{それ以外}) \end{cases} \tag{4.7}$$

であり，$F_p(0,x)$，$F_f(n,x)$ は 0 とする．

本研究が目標とするカメラワークは，全体を通して多くの被写体をまんべんなく撮影する方針で計画する．$F_p(i,x)$ が大きい候補は，過去に何度もカメラが割り当てられ，多くのショットが撮影されている．このような候補にカメラを割り当てると，同じ被写体のショットばかりになってしまうため，本研究では重み α を負の値とする．

また，$F_f(i,x)$ が大きい候補ほど，フレーズ i 以降の未来で被写体候補になる回数が多く，結果としてカメラが割り当てられる可能性が高い．本研究では $F_p(i,x)$ と同様の理由により，重み β を負の値として登場機会の少ない候補を優先的に撮影するようにする．

逆に α と β を正の値とすると，登場回数の多い被写体を優先的に撮影することができる．そのようなカメラワークは，ドラマの主役のように限られた被写体が連続して画面に登場する映像を制作するのに適している．

$F_p(i,x)$ は被写体に関連する値であり，直前フレーズまでのカメラワークが決定した後でなければ計算できないため，フレーズ毎に逐次求める必要がある．これに対して $F_f(i,x)$ は被写体候補に関連する値なので，カメラワークが決定している必要は無い．ステップ (1) でシナリオを一括して読み込んだ時点で，式 (4.2) を用いてすべて計算することができる．

前後フレーズにおける類似度 C

候補 x がフレーズ $i-1$，つまり直前フレーズにおいてカメラに割り当てられたショットと類似性がある場合に優先度を変化させる．この類似度の判定には図 4.8 の階層構造を利用する．

例として，フレーズ $i-1$ において 3 台のカメラがフルート，金管楽器，ティンパニのショットを撮影していたとする．一方フレーズ i ではトランペット，指揮者，1st バイオリン，フルートの 4 つが被写体候補にあがっているとする．この場合，フレーズ i のトラン

ペットは直前の金管楽器と図 4.8 において親子関係にある．4.2.1 節に述べたように，オーケストラでは各楽器がグループ毎にまとめて配置されているため，金管楽器の中にはトランペットが映り，ショットに類似性があるといえる．このような場合，類似度 $C(i,x) = 1.0$ とする．一方，フルートは直前のフレーズでも撮影されており，親子関係にある場合よりもさらに類似性が高い．この場合は類似度 $C(i,x) = 2.0$ とする．

同様に，シナリオを先読みし，直後のフレーズ $i+1$ における類似性も判定する．フレーズ $i+1$ では，まだどの候補をどのカメラに割り当てるか決定していないため，自分と同じ，もしくは親子関係にある候補があれば，上記と同様の類似度を設定し，優先度を変化させる．

本研究が目的とするカメラワークでは，常に注目を集める主役が存在せず，切替えの前後で異なる被写体を撮影するほうが良いとしている．よって γ の値を負に設定し，類似性の高い候補の優先度を下げていく．逆に，γ を正の値とすると，類似性の高い候補が優先的に撮影される．そのようなカメラワークは，ドラマのように主役のいる映像を制作するのに適している．

ショットサイズの差 D_p, D_c

ショットサイズの差に基づく優先度を計算する．このショットサイズの差は，図 4.8 におけるレベル値の差で表現する．例えば “全体（レベル 0）” と “金管楽器（レベル 2）” の差は 2 である．

まず，直前フレーズとの差を考える（図 4.9）．フレーズ i における被写体候補 x と，フレーズ $i-1$ におけるすべてのショットとのサイズ差 $D_p(i,x)$ は次の式で表される．

$$D_p(i,x) = \sum_{j=1}^{N_c} |L(x) - L(S(i-1,j))| \tag{4.8}$$

ここで N_c はカメラの台数，$S(i,j)$ はフレーズ i におけるカメラ j のショット，$L(S)$ はショット S のレベル値をあらわす．

この D_p は，直前フレーズにおける 3 つショットと比べてサイズの差が大きく，構図の印象が異なるほど大きな値を示す．この値に応じて優先度を δD_p 変化させる．

4.2.3 節の分析より，フレーズの前後ではショットサイズを変化させる傾向があることがわかっている．よって本研究では δ を正の値に設定する．これにより，構図が異なる演出効果の高い候補の優先度を上げることができる．逆に δ を負の値に設定すれば，ショットサイズの差が少ない候補の優先度を上げることができる．そのようなカメラワークは，スイッチング時に構図の変化が少ない安定した映像を制作するのに適している．

次に，同一フレーズとの差を考える（図 4.10）．被写体候補 x と，フレーズ i において既に決定したショットとのサイズ差 $D_c(i,x)$ は次の式で表される．

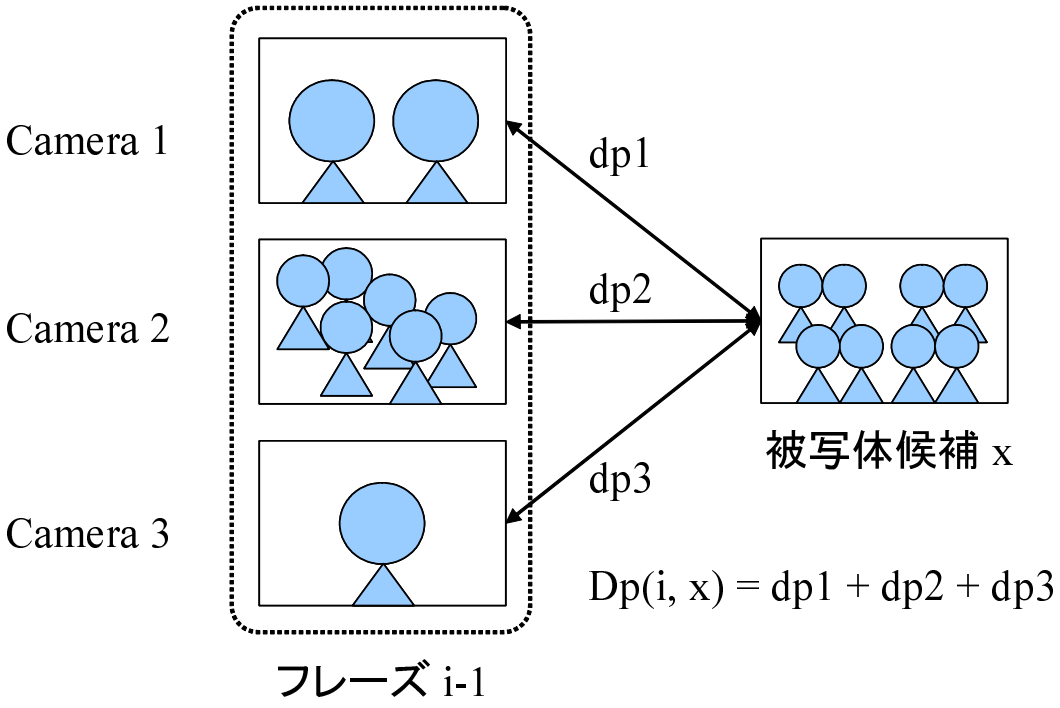


図 4.9: フレーズ間のショットサイズの差

$$D_c(i,x) = \sum_{j=1}^{N_c} |L(x) - L(S(i,j))| \tag{4.9}$$

ただし $S(i,j)$ が empty（カメラ j のショットが決定していない）の場合，$L(x) - L(S(i,j)) = 0$ として計算する．

この D_c は，現在のフレーズですでに決定したショットと比べてサイズの差が大きく，印象が異なるほど大きな値を示す．この値に応じて優先度を ϵD_c 変化させる．

本研究では，ϵ を正の値に設定することで，同一フレーズ内で様々なサイズのショットを確保するようにする．サイズに差がないと，直前のフレーズ $i-1$ からスイッチングする際に構図に変化をつけることができないからである．

4.3.4　位置関係を考慮したショット決定

フレーズ i において，すべての被写体候補の優先度計算が終わった後，その優先度に従い被写体候補をソートする．そして最も優先度の高い被写体候補を，最も良く撮影可能なカメラに割り当てる．

この “映り具合” の判断には，撮影環境におけるカメラの設置状況を記録したカメラマップを利用する．この DTD を図 4.11 に示す．カメラの三次元座標（position）と，シ

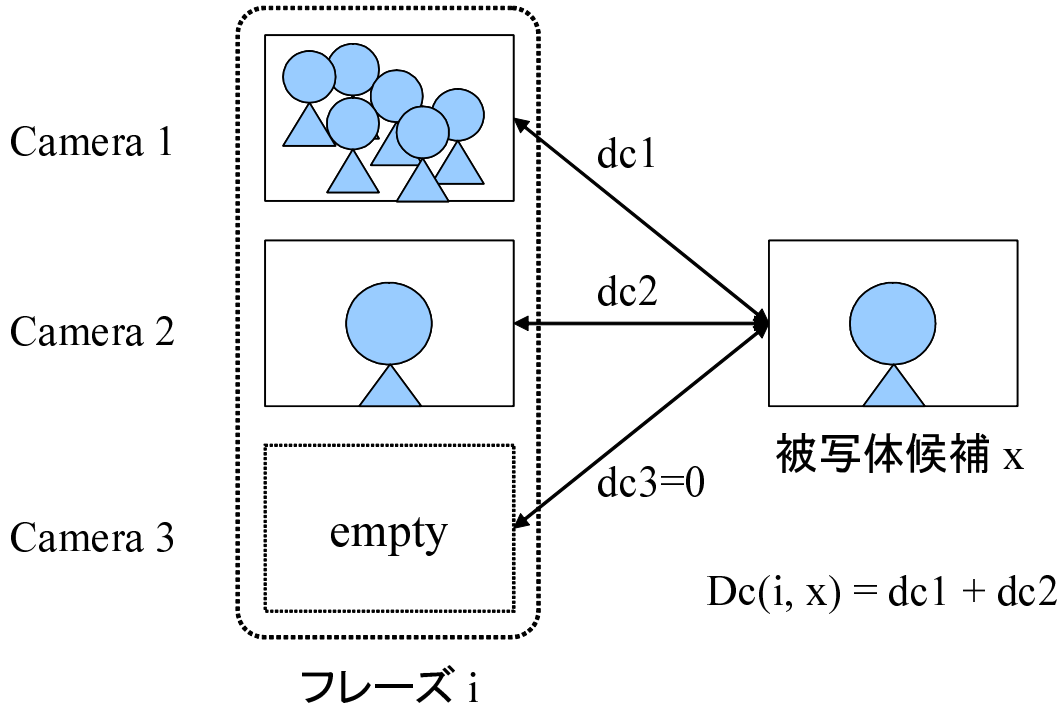


図 4.10: カメラ間のショットサイズの差

ナリオから抽出可能な演奏者の位置と向きを利用し，カメラから見た被写体の向きが正面 > 斜め > 横の順に “より良く撮影可能” と定義する．被写体とカメラの位置関係を考慮することで，撮影環境の違いを考慮したカメラワークを計画することができる．また，背後からのショットなど映像表現上不適切なものを排除できる．

カメラと被写体のペアが 1 つ決定するごとに，ステップ (4) の優先度付け以降を繰り返し，すべてのカメラに被写体を割り当てるまで処理を繰り返す．

4.4　実装

4.4.1　プロトタイプシステム

4.3 節の提案手法に基づいて，シナリオからカメラワークの計画を自動的に行うプロトタイプシステムを実装した．システム全景を図 4.12 に示す．

システムはまず，XML 形式で記述した事前知識であるシナリオと，各カメラの設置状況を記述したカメラマップを読み込む．次にカメラワーク計画部において，4.3 節の手法に従って，どのカメラでどの被写体を撮影するかを計画していく．計画されたカメラワークは，TVML 変換部において TVML スクリプトに変換される．カメラワーク実行部では，

```
<!ELEMENT camera_map (camera*)>
<!ELEMENT camera (name, position, pan, tilt, vangle)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT position (area, x, y, z)>
<!ELEMENT area (#PCDATA)>
<!ELEMENT x (#PCDATA)>
<!ELEMENT y (#PCDATA)>
<!ELEMENT z (#PCDATA)>
<!ELEMENT pan (#PCDATA)>
<!ATTLIST pan unit (degree | radian) "degree">
<!ELEMENT tilt (#PCDATA)>
<!ATTLIST tilt unit (degree | radian) "degree">
<!ELEMENT vangle (#PCDATA)>
<!ATTLIST vangle unit (degree | radian) "degree">
```

図 4.11: カメラマップの DTD

計画されたカメラワークの映像を楽曲の MIDI 再生と同期させながら，TVML による仮想空間の映像として確認することができる．

現段階では計画したカメラワークの映像を確認するため，TVML1.2[14] をシミュレータとして用いている．実際のカメラを用いて撮影すると，カメラの精度による実行時のズレの影響が大きくなり，本研究が目指す映像評価に支障が生じる可能性があるためである．ただし，実世界での利用を考慮し，カメラを任意の位置に移動させるといった仮想空間特有の機能は一切利用していない．

次に，システムの各機能の詳細について示す．シナリオは 4.3.1 節に示した DTD に従い，事前に手作業で作成しておいた．カメラマップは，4.3.4 節に示した DTD に従って自動的に生成する GUI（カメラマップ作成部）を用意した．

カメラワーク実行部と，TVML によるショットの例を図 4.13 に示す．画面左側にはスコアが表示され，現在の演奏地点と各楽器の役割（メロディー・伴奏）が表示されている．画面右側では，各カメラの配置と撮影範囲を確認することができる．

カメラマップ作成部，カメラワーク計画部，TVML 変換部は Java で作成した．カメラワーク実行部では，MIDI ファイルの再生と同期させるために，第 2 章の図 2.5 で触れた TVML プレイヤーの外部制御 API を Visual C++ から利用した．

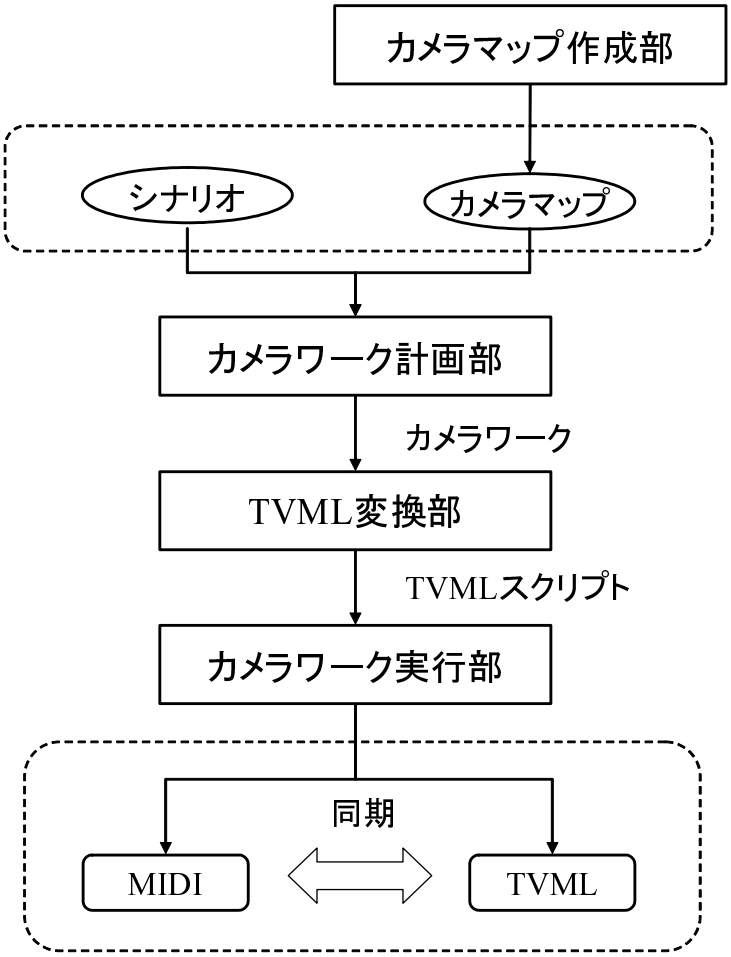


図 4.12: システム全景

4.4.2　オーケストラホール

撮影環境として想定するオーケストラホールは，TVML にあらかじめ用意されていたステージセットを利用した．座標空間は TVML の設定に従い，図 4.14 のようにステージセット中央を原点とし，横が x 軸，縦が z 軸，高さが y 軸となっている．そのパラメータを表 4.2 に示す．ホールの広さはこのステージセットの設定を利用したが，高さに関しては制限がなかったため，東京オペラシティコンサートホール [95] のスケールを参考に決定した．

4.4.3　カメラ

表 4.3 にカメラのパラメータを示す．カメラの x および z 座標はカメラマップ作成時に指定する任意の位置とする．y 座標は，舞台上のカメラの場合は山台の最上段に設置する

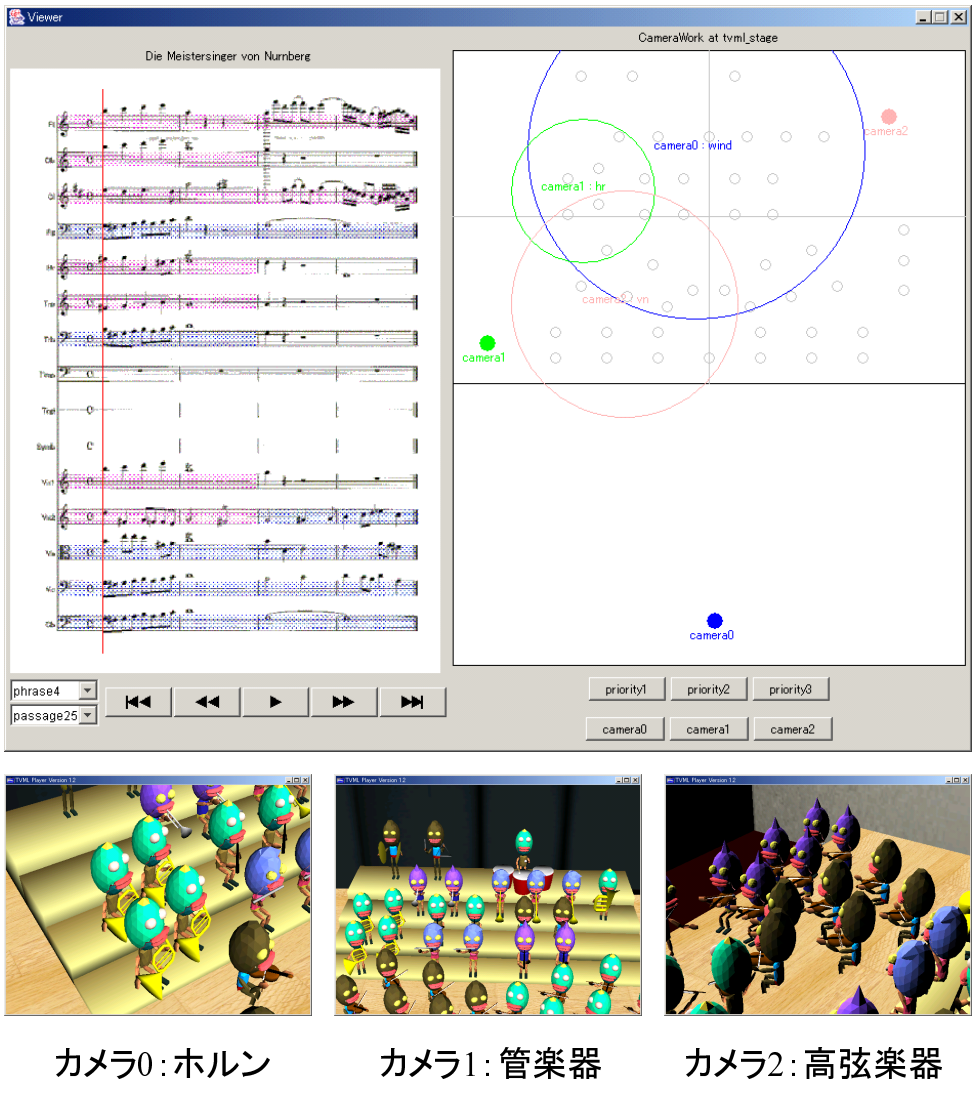


図 4.13: 実装画面とショット例

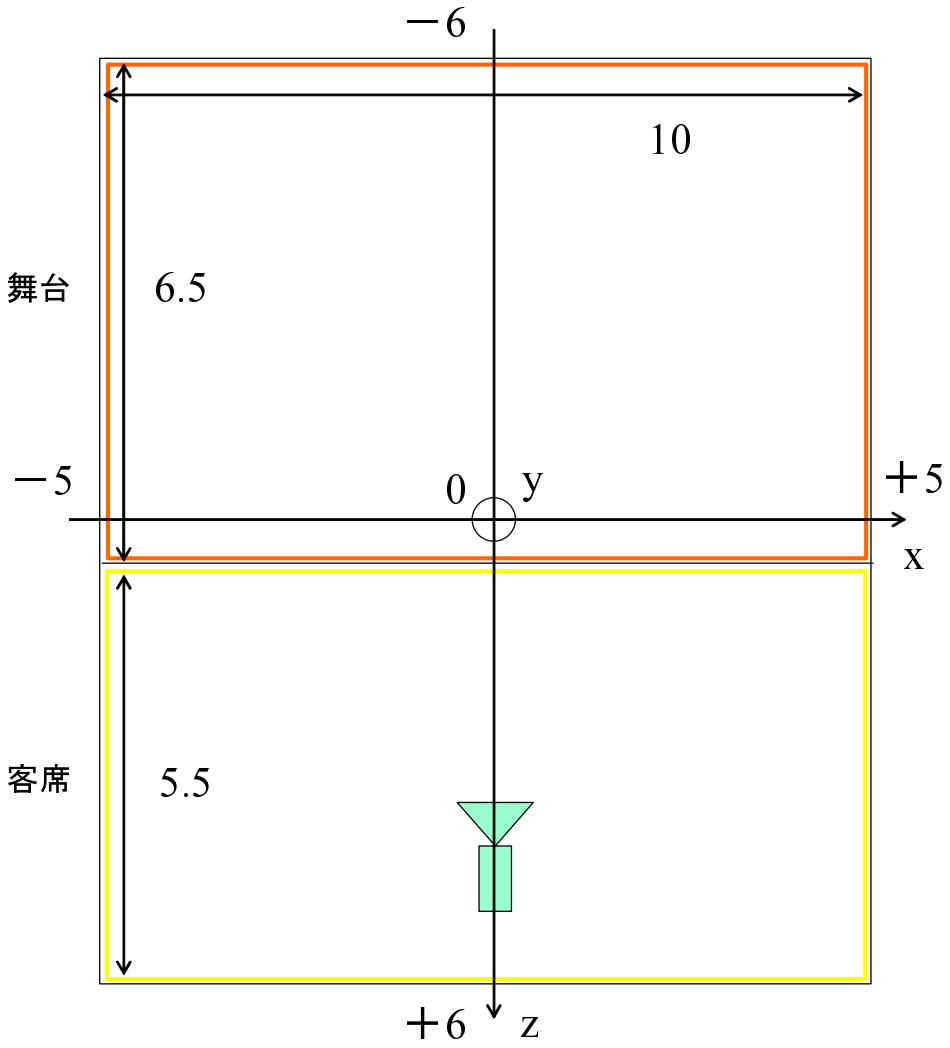


図 4.14: ホールの座標空間

表 4.2: ホールパラメータ

	舞台	客席
幅	10	10
奥行き	6.5	5.5
高さ	12	12

表 4.3: カメラパラメータ

	舞台	客席
x 座標	任意	任意
y 座標	任意	任意
z 座標	3	7
パン	左右 180 度	左右 180 度
チルト	上下 90 度	下 90 度

ことを想定して 3，客席上の場合では東京オペラシティ大ホールの 3 階席の高さを想定して 7 とした．また，パン角は左右 180 度に回転可能とした．チルト角は舞台上のカメラは上下 90 度，客席上のカメラは天井からつるすものと想定して下にのみ 90 度回転可能とした．これらの値は実際に存在するカメラとしては高機能なものであると考えられるが，TVML ではカメラは xz 平面に対して平行に，舞台後方を正面となるように初期位置が決まってしまうための措置である．

4.5　実験方法

提案手法はカメラワークの自動計画である．これを評価するため，カメラワークにしたがって撮影されたショットを編集する作業を通して，編集者が必要とするショットを撮影できているか，編集した映像にどのような特徴があるかを検討した．

4.5.1　被験者の選択傾向

マイスタージンガー前奏曲の中の連続した 10 シーン（約 2 分 30 秒）を取り出し，A，B の 2 種類の配置でカメラワークを計画した（図 4.15）．配置 A は 1 台のカメラを客席に，2 台を舞台上に設置し，配置 B では 3 台のカメラを客席に設置したものである．実験に用

いた式 4.3 のパラメータは，$\alpha = \beta = -0.1$，$\gamma = -0.5$，$\delta = \epsilon = 0.25$ とした．これらは予備的な実験に基づいて決定した．

次に，A，B 両配置に共通し，どの演奏者も無理なく撮影可能な客席正面のカメラ 0 から，単独・組合せを含めた全 22 種類の被写体を撮影した．被験者（$N = 12$．全員音楽に関する知識を有する）には，各シーンに該当するスコアの一部を提示し，状況に合致していると思うもの 3 つを 22 種類の中から選択してもらった．この際，スコアも順序を入れ替えて 10 回に分けて提示した．

このように，1 シーンから読み取れる情報のみで選択された被写体と，提案手法で計画した被写体とがどの程度一致しているのか，つまり，被験者が編集時に必要としている被写体をどの程度撮影できているのかを求めた．

4.5.2　映像編集

被験者に各シーンごとに提案手法で計画した 3 つのショットを提示し，その中から 1 つを選択して，1 本の映像作品を作ってもらう実験を行った．その際，被験者にはスコア全体を提示し，楽曲全体の構成や前後関係を考慮するよう指示した．さらに，比較対象として，4.5.1 節の実験において各フレーズで被験者の支持を集めた被写体上位 3 つをランダムにカメラに割り当てたもの（以降，比較システム）で同様の実験を行った．比較に用いた 3 つの被写体は，シーンの状況には合致しているものの，1 フレーズから読み取れる情報のみで判断しており，前後関係や全体の構成といった編集段階を意識して選択されたものではない．そこでこの二つを比較し，出来上がった映像にどのような特徴や違いがあるかを調べた．

ここで，比較システムで提示された 3 つのショットの中には被験者が必要とするショットが含まれていない場合もある．本来ならばすべてのショットを提示すべきであるが，演奏率が大きい場合，被写体候補は 20 以上にもなり，カメラが 3 台とすると 60 以上のショットの中から必要なショットを選択することになる．そのような膨大な中から選択することは困難であるため，4.5.1 節の実験によって全被験者のニーズを平均的に反映していると思われるもの 3 つを比較対象として選択した．

4.6　結果および考察

4.6.1　被験者の選択傾向との比較

4.5.1 節の結果として，プロトタイプで計画したショットと，比較システムの上位 3 つの被写体を表 4.4 に示す．プロトタイプの被写体と，比較システムの被写体が完全に一致し

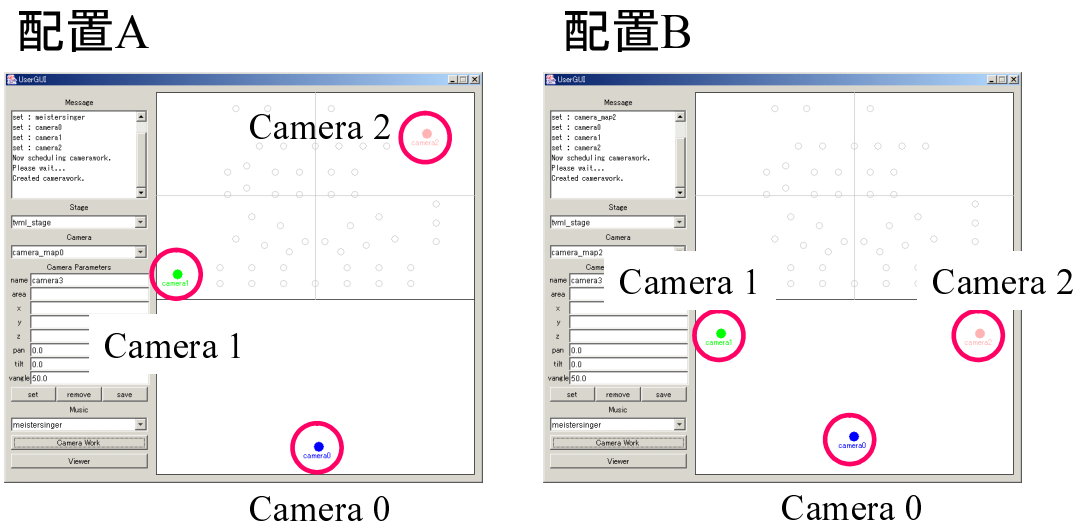


図 4.15: 実験に用いたカメラ配置

表 4.4: プロトタイプで計画した 3 ショットと比較システムの上位 3 ショット

	プロトタイプ A			プロトタイプ B			比較システム AB		
scene	camera 0	camera 1	camera 2	camera 0	camera 1	camera 2	camera 0	camera 1	camera 2
1	all	cl	vn	all	wood	vc	all	wood	wind
2	low	timp	strings	timp	vn	strings	wood	strings	timp
3	wind	trp	wood	fg	wind	low	vn	brass	low
4	fl	cl	cond	fl	fl	cond	cond	cl	fl
5	va	ob	cond	cond	va	ob	low	ob	cond
6	wood	wood	cond	wood	wood	cond	fl	cond	wood
7	strings	hr	va	va	strings	vn	strings	fg	hr
8	vn	cond	vn	vn	vn	cond	vn1	cond	vn2
9	brass	symb	wind	symb	brass	wind	wind	wood	brass
10	all	fl	low	all	hr	low	vn	all	low

略称は図 4.1 を参照．N=12

たのは，配置 A におけるシーン 4 のみであった．それ以外のシーンでも，両配置とも，比較システムの被写体と一致するショットが 1 つまたは 2 つ確保されているのがわかる．

プロトタイプで計画した被写体が，被験者のニーズとどれだけ一致しているかを見るには表 4.4 だけでは不十分である．そこで新たな評価尺度として一致率 C_i を定義する．先の 4.5.1 節の実験で，シーン i で被験者が選択した 3 つの被写体には各 1 点を与えることとし†，被験者数を N，プロトタイプが計画した被写体すべての得点の合計を N_p とすると，一致率 C_i は次の式で表される．

$$C_i = \frac{N_p}{3N} \times 100(\%) \tag{4.10}$$

†同一の被写体を重複して選択することは認めていないため，1 つの被写体の得点の最大値は N となる

各カメラ配置における一致率を表 4.5 に示す．この結果，シーン 3 が際立って低く，シーン 4，5 が高くなった．

シーン 3 は，式 (4.2) における “演奏率 6 割以上” の状態であった．舞台上では様々な楽器が演奏しているため，被験者の注目する被写体が分散したためと考えられる．シーン内に複数の被写体が存在する場合，様々な編集要求があること，演奏率が高いほど被写体候補を増やしていくという提案手法の前提が確認できたといえる．

一方で，同様に演奏率が高かったシーン 1，10 では約 50%で，シーン 3 と比較すると高い．この 2 つのシーンでは，提示したスコアから演奏の最初と最後であると被験者が判断したためか，全体を映したショットに投票が集中した．プロトタイプで計画した被写体の 1 つも全体であったため，一致率はそれほど下がらなかった．

シーン 4，5 は，演奏率が 2 割未満の状態であった．舞台上で演奏する楽器もほとんど無く，注目が集まる被写体が限定されたためだといえる．シーン 8 も演奏率が 2 割未満であったが，シーン 4 より低くなっている．これはプロトタイプで計画した被写体が 2 種類だけだった（表 4.4 参照）ことが影響している．同様のことはシーン 6 にも言える．1 つの被写体を重複して選択することを認めていないため，この 2 シーン（6, 8）は一致率を正しく評価できず，除外して考えるべきである．

シーン 7 は，配置 A と B とで一致率に差があった．シーン 7 では木管楽器と弦楽器が演奏をしている．そのため，被験者の多くがホルンなど木管楽器系の被写体を選択した．配置 A では弦楽器とともにホルンが計画されていたが，配置 B ではホルンの代わりにビオラが計画されたため値が低くなった．両配置ともシーン 6 までにホルンのショットが無く，直前・直後のシーンにおける 3 ショットが共通であること，同じスコアを利用しているため未来の優先度も同じであることを考えると，被写体とカメラの位置関係による可能性がある．

これ以外では両配置で一致率に大きな差はなく，各シーンで最も多く投票された被写体（1 位タイを含む）は，すべてプロトタイプで計画されたものと一致していた．また，一致率を評価できないシーン 6，8 を除いた平均は配置 A で 57%，配置 B で 55%であった．このことから，提案手法は撮影環境の違いに対応しつつ，そのシーンで最もニーズのある被写体のほかに，何人かの被験者が選択する被写体を確保するようにカメラワークを計画しているといえる．

4.6.2　映像編集方法の分析

4.5.2 節の結果を以下に示す．比較システムで選択したショットは，前後関係や全体の構成を意識して選ばれたわけではないため，映像編集を行う際になんらかの違いがでると考えられる．そこで，以下の 4 点について分析した．

表 4.5: 各カメラ配置における一致率

シーン	配置 A(%)	配置 B(%)
1	47	44
2	53	50
3	28	33
4	78	78
5	69	69
6	50	50
7	67	50
8	50	56
9	64	64
10	53	50

ショットサイズ

編集された映像の，スイッチング時におけるショットサイズの変化は，映像のダイナミクスをあらわす 1 つの指標である．4.2.3 節の分析で述べたように，プロが撮影したオーケストラの映像は，ショット遷移の際にサイズに変化をつける傾向がある．そこで，出来上がった映像で，図 4.8 における同一レベル値のショットを接続した回数を測定した．その結果，配置 A でプロトタイプが 2.3 回，比較システムが 3.4 回，配置 B でプロトタイプが 2.4 回，比較システムが 3.2 回となった．

提案手法では，シナリオからカメラ間距離・フレーズ間距離を計算し，ショットのサイズ差が大きくなるよう優先度を付加している．プロトタイプでは，連続したシーン 4 と 5 では演奏楽器の割合が 2 割未満のケースであったため，両配置ともに計画したショットはどれも同じ階層レベルのものであった（メロディー楽器および指揮者）．よって，実質 1.0 回はシステムによって強制されている．

これに対し比較システムでは，A，B 両配置ともに，そのような強制箇所は存在しなかったにも関わらず，同一レベルでの切替え回数が大きくなってしまった．このことから，ショットサイズによる優先度付けが映像表現の向上に一定の効果があるといえる．

カメラ切替え

ショットを接続する際，カメラの設置位置に適した被写体を撮影していれば，そのカメラの映像へ切替える回数が増えると考えられる．この切替え回数を測定したところ，配置

A でプロトタイプが平均 8.0 回，比較システムが 6.8 回．配置 B でプロトタイプが 5.6 回，比較システムが 5.6 回となった．

4.3.4 節にあるように，提案手法では被写体とカメラを結びつける際に位置関係を考慮する．配置 A では，プロトタイプが比較システムを上回り，一定の効果が見られた．一方配置 B では大きな差は見られなかった．

配置 A では，ホールの様々な位置にカメラが配置されており，各カメラからの映り具合が大きく異なる．カメラ 1 と 2 は，カメラ 0 に比べて正面方向から撮影可能な被写体は少ないが，カメラ 1 は低弦楽器全般，カメラ 2 はバイオリンと指揮者を撮影するのに適している．これらの被写体を撮影していた場合に，一時的にカメラ 1，2 へと切替える被験者が多く見られた（シーン 5，8）．このような特殊な配置では，位置による優先度が有効に機能し，撮影環境の変動に対応して適切な被写体を撮影できていることがわかる．

配置 B では，すべてのカメラが客席側にある．カメラ 1，2 ともに，正面・斜め前方から撮影できる被写体は配置 A に比べて多く，カメラ自体を切替えなくてもある程度の構図で撮影できる．配置 A に比べて切替え回数自体が少ないことからもそれが伺える．このため，位置の優先度があまり大きく働かなかったものと考えられる．将来的には，現在の正面 > 斜め > 横という単純なものではなく，例えば楽器の見え具合など位置より複雑な優先度を適用することで，B のような安定した配置でも差がでるように対応する必要があるだろう．

全体の構成

実験で用いた 10 シーンでは，打楽器の演奏機会が非常に少ない．トロンボーン，コントラバスは伴奏が多く，図 4.8 における 1 階層上の管楽器・低弦楽器ショットで撮影されることが多い．これに対しティンパニはシーン 2，シンバルはシーン 9 でしか演奏をせず，1 階層上の打楽器ショットとしても撮影されない．

プロトタイプでは，A，B 両配置ともこの 2 つのショットを確実に撮影できていた．これに対し比較システムでは，ティンパニは上位 3 つに入ったため撮影することができたものの，シンバルの順位は 4 番目であったために撮影されなかった．このことから，提案手法による過去 (F_p) および未来 (F_p) の出現頻度に対する優先度付けが有効に機能し，全体を通して様々な被写体を撮影できていることがわかる．

ショットの優先度

最後に，各シーンで選択されたショットの優先度の内訳を調べた．結果を図 4.16 および図 4.17 に示す．配置 A，B ともに，優先度の高いショットが積極的に選択されているシーンと，分散したシーンとに分かれた．

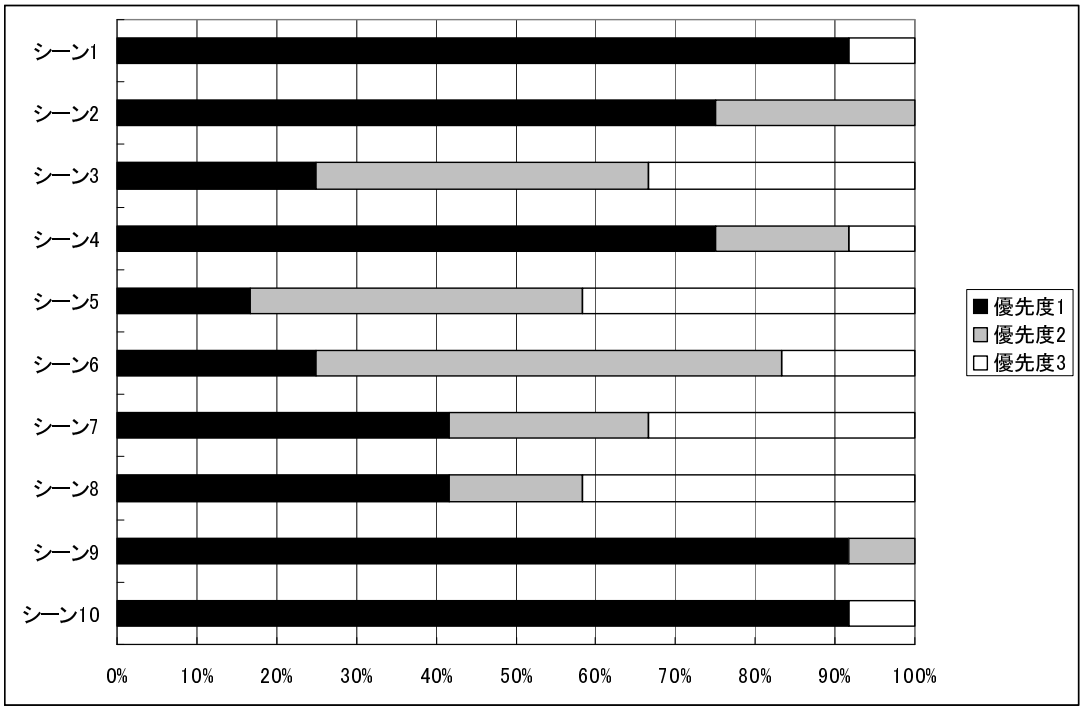


図 4.16: 優先度の内訳（配置 A）

選択が集中したシーン 1 とシーン 10 は，優先度 1 のショットは “全体” であった．これらシーンは演奏の開始と終了という特殊なシーンであったため，選択が偏ったものと考えられる．また，シーン 2 とシーン 9 はティンパニとシンバルの演奏シーンであったため，これを選択した被験者が多かった．

それ以外の演奏の中盤で選択が分散した理由としては，被験者の経験や嗜好によるところが大きい．例えば両配置のシーン 5 において，弦楽器の経験がある被験者はビオラを，木管楽器の経験がある被験者はオーボエを選択していた．各自の嗜好にあわせて必要なショットを適宜差し替える形で利用されていることがわかる．将来的には，システムを利用するユーザのプロフィールを事前に収集しておき，これを優先度付けに利用すると良いと思われる．

4.7　まとめ

本章ではストーリー型シーンの例として，シナリオ情報に基づいたカメラワークの計画手法を提案した．提案手法ではオーケストラ演奏を撮影対象とし，シナリオからシーンの状況を把握してユーザの注目する被写体の抽出方法と，優先度の高い被写体候補をカメラに割り当てていく方法を実現した．評価実験では，提案手法によって計画されたカメラ

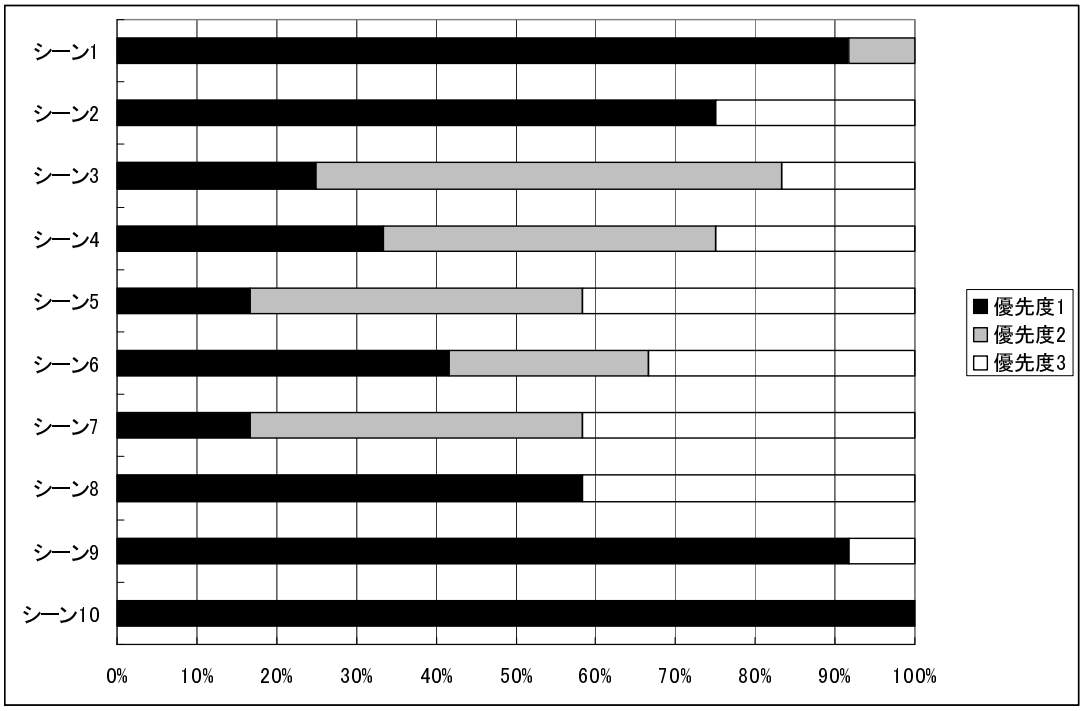


図 4.17: 優先度の内訳（配置 B）

ワークは，シナリオを意識せずに支持されたカメラワークよりも変化に富む映像を制作できることを示した．また，一般的な配置とは異なり，舞台上にカメラを設置した場合でも，それに適応したショットを提示できることがわかった．

本研究ではオーケストラを撮影対象にしたが，各要素技術は様々なストーリー型シーンに適用できると考えている．階層構造による被写体候補の作り方は，各々がグループ毎にまとめられた位置に配置されるようなシーンに適用できる．例えば結婚披露宴では，家族，親族一同，新郎側出席者などのショットを撮影することができる．一方，被写体や構図の種類に基づく優先度付けは，各パラメータの重みを変更することで，本研究が対象にした “全体を様々な構図で撮影する” 以外にも様々なカメラワークを計画することができる．出現頻度や類似度の重みを正の値にすることで，ドラマのように “限られた被写体を集中的に撮影する” カメラワークを計画することができる．また，ショットサイズの重みを負の値にすることで，“変化が少ない安定した映像を撮影する” カメラワークを計画することができる．

今後の課題としては，今回は手動で行ったスコアからシナリオを作成する部分の自動化や，実機のカメラを用いた撮影実験があげられる．実機を用いて撮影するには，2.3.12 節で述べた幾何学的・時間的ズレの補正が必要である．基本的にオーケストラは演奏者が着席したままほとんど動かないため，幾何学的ズレは大きなさほど問題にならない．しかし

時間的ズレは，楽曲の進行が指揮者の感覚で速くも遅くもなる．これにはスコアのどこを演奏しているのかを判断する音声（楽曲）認識，指揮者の動きの認識 [96]，タクトにセンサをつけるといった手法 [97] が適用できると考えられる．また，今回は比較的短いシナリオを想定していた．長いシナリオを用いた場合での優先度計算方法やパラメータの変更なども検討していく必要がある．

第5章　結論

現在，ディジタル多チャンネル時代を迎え，映像が様々なメディアを通じて供給されるようになった．このような背景から，映像制作にかかるコストを軽減させるための自動撮影に関する研究が注目を集めている．その中でも，映画やテレビのような魅力的な映像を自動で作り出せることが求められている．

本論文では，映像文法に基づいて複数台のカメラを協調動作させ，演出した映像を自動的に撮影することを目的とした．特に，撮影対象をシナリオの無いシーン，シナリオのあるシーンに分類し，それぞれにおける具体的な撮影対象を設定し，その課題を映像文法をもとに解決するアプローチをとった．

以下では，本論文の各章を振り返りながら結論をまとめていく．

まず第 1 章では，本研究の目的と概要，および本論文の構成について述べた．

第 2 章では，本研究の背景，関連研究，そして位置づけについて述べた．研究の背景としては，映像コンテンツがどのように分類，制作されるかについて述べた．また関連研究については，映像制作の各フェーズにおける要素技術と，それらを総合した会議や講義をはじめとする各種自動撮影システムの研究について述べた．そして本研究が目指す自動撮影システムの全体像と一連の研究との対応付けを行い本論へとつなげた．

第 3 章では，シナリオのないシーンを対象に，撮影対象の状況を認識しながら，演出用カメラワークをリアルタイムで生成することに焦点を置いた．その具体的な適用例として対面会議を取り上げた．この研究では映像文法を “正確で分かりやすい映像” を制作するための技法としてとらえ，参加者の発言の推移から会議の状況を判断し，概念的なイマジナリーラインを会議空間中に一意に設定できるようにした．そしてこのイマジナリーラインとカメラの三角形配置をもとに複数台のカメラを協調動作させ，中心的な人物間の対話を見やすく効果的に演出する自動撮影手法を提案した．プロトタイプで撮影した映像は，話者の検出やショットの精度に課題はあるものの，人手でスイッチングをした映像に比べて高い評価を得ることができた．

続く第 4 章では，シナリオのあるシーンを対象に，撮影対象の状況をシナリオで認識しながら，演出用のカメラワークを撮影前に自動で計画することに焦点を置いた．その具体的な適用例としてオーケストラ演奏を取り上げた．まず，映像の分析からオーケストラ演奏を撮影するためのカメラワークに関する知識を獲得した．その結果から映像文法を，編集時に必要となる “バラエティに富んだショット” を撮影するための技法としてとらえ，シナリオから抽出した被写体候補に対して，被写体の種類，類似度，構図の変化，カメラからの映り具合といった観点から優先度を計算し，限られた台数のカメラに優先度の高い被写体を割り当てる手法を提案した．提案手法で自動計画されたカメラワークは，編集時に必要となるショットをある程度確保した上で，全体の構成や前後関係を考慮しない手法で計画したカメラワークと比べて効果的な映像を編集できることを示した．

そして最後に本章では，各章を振り返りながら本研究を総括し，結論とした．

以上要するに，本論文では，映像文法に基づいて複数台のカメラを協調動作させることによって演出した映像を自動的に撮影するシステム，すなわち映像文法に基づく自動撮影システムを提案した．一連の研究は，このシステムの具体例として多方面からアプローチを行うものであり，それぞれに設定した課題を解決することができたことから，より高度な自動撮影システムの実現に向けて 1 つの方向性を提言できたと考えている．

以下では，今後の課題と新たな展望について述べる．

本論文では，計画・実行・編集という 3 つのフェーズを経た映像の “制作” までを主として扱ってきた．この制作過程で映像表現の質を高めていくことが重要なのはこれまでの議論で疑う余地は無い．しかし，今後は作った映像をどのように提供するかという “配信” の部分にも注目していく必要がある．

現在インターネットで配信される映像は，ブロードバンド用には高ビットレート，ナローバンド用には低ビットレートというように，共通の映像ソースを配信環境に適した形式で変換して提供している．つまり，双方の映像においてカメラワークは共通である．しかし，高ビットレート環境では判別容易な人物の表情は，低ビットレートでは不鮮明であり，ズームを大きめに設定したカメラワークのほうが適している場合もある．このように，今後ますます映像が配信されるメディアが増加していくことを考えると，それぞれのメディアに適したカメラワークを考慮する必要があると考えられる．

以上，考えられる課題と展望について言及した．ここに挙げた例はその一部であると思われる．本論文での成果が，将来の映像文化のさらなる発展を実現するための研究のひとつとして貢献できればと願いつつ，本論文を締めくくりたいと思う．

謝辞

本研究を進めるにあたり，岡田 謙一 教授より日々賜りましたご指導，ご鞭撻につきまして，ここに深く感謝の意を表したいと思います．

また，本論文を執筆するにあたり，貴重なご意見とご指導をくださいました，慶應義塾大学理工学部 山本 喜一 助教授，斎藤 英雄 助教授，重野 寛 助教授，東京工科大学コンピュータサイエンス学部 松下 温 教授 に厚く御礼申し上げます．松下 温 教授には，進路に関してもご相談に乗っていただきました．重ねて御礼申し上げます．さらに，本研究を進めるにあたり，的確なご助言をいただきました，東京工科大学コンピュータサイエンス学部 市村 哲 助教授に心より御礼申し上げます．

研究グループを立ち上げた時から苦楽を共にした 吉田 竜二 氏，劇的なデモ発表を共に乗り越えた 平石 絢子 氏，理論面での貢献以外にも研究を円滑に進める上で様々な配慮をしてくれた 柴 貞行 氏の協力なしに，本論文の完成はありませんでした．また，別プロジェクトでともに議論した 高久 宗史 氏，加藤 淳也 氏，住谷 哲夫 氏，津村 弘輔 氏の斬新な発想と大胆過ぎる行動力には，非常に大きなインスピレーションを与えてもらいました．強烈な個性を持った JINI 班の方々に心から感謝いたします．

そもそも学部での就職を考えていた私が，修士，博士にまで進学を決意したのは，本田 新九郎 博士をはじめとする個性的過ぎた先輩方に影響されてのことでした．また，決して優秀とは言えなかった私がここまでやってこられたのも，夜中にラーメンを食べながらとりとめも無いことを語り合った同期達や，真冬に T シャツでカレーを食べに行った 江木 啓訓 氏をはじめとする後輩達の励ましがあってのことでした．彼らと出会い，共に過ごした松下・岡田・重野研究室での日々は，私にとって生涯忘れることの出来ない大切な時間となりました．それぞれに心から感謝いたします．

最後に，私を温かい目で見守り，応援し支えてくれた母 睦子，祖母 幸子，姉 祥子，進学を快諾してくれた亡き父 雄介，可愛がっていた猫達に心から感謝いたします．

参考文献

[1] ダニエル アリホン: 映画の文法 実作品に見る撮影と編集の技法, 紀伊国屋書店 (1999).

[2] 平石絢子, 井上亮文, 重野寛, 岡田謙一, 松下温: 映画の撮影手法に基づいた会議の自動撮影, マルチメディア，分散，協調とモバイル (DICOMO2002) シンポジウム論文集, pp. 285–288 (2002).

[3] Inoue, A., Shigeno, H., Okada, K. and ka Matsushita, Y.: Introducing Grammar of the Film Language Into Automatic Shoot ing for Face-to-face Meetings, *Proceedings of IEEE/IPSJ Symposium on Applications and the In ternet (SAINT2004)*, IEEE, pp. 277–280 (2004).

[4] 井上亮文, 吉田竜二, 平石絢子, 重野寛, 岡田謙一, 松下温: 映画の映像理論に基づく対面会議シーンの自動撮影手法, 情報処理学会論文誌, Vol. 45, No. 1, pp. 212–221 (2004).

[5] 柴貞行, 井上亮文, 平石絢子, 高久宗史, 重野寛, 岡田謙一: オーケストラ撮影における楽譜を利用したカメラワークの計画, 情報処理学会研究報告, 2003-DPS-111, pp. 65–70 (2003).

[6] 柴貞行, 井上亮文, 平石絢子, 重野寛, 岡田謙一: 楽譜を利用したカメラワーク計画手法の実装と評価, マルチメディア，分散，協調とモバイル (DICOMO2003) シンポジウム論文集, Vol. 2003, No. 9, pp. 821–824 (2003).

[7] 井上亮文, 平石絢子, 柴貞行, 市村哲, 重野寛, 岡田謙一, 松下温: シナリオ情報によるオーケストラ演奏のカメラワーク生成手法, 情報処理学会論文誌, Vol. 46, No. 1, pp. 38–50 (2005).

[8] フランソワ トリュフォー: 定本 映画術―ヒッチコック・トリュフォー, 晶文社 (1990).

[9] 福井一夫: 映像メディアの動向とコンテンツ制作, 映像情報メディア学会誌, Vol. 57, No. 1, pp. 99–102 (2003).

[10] 井上誠喜: バーチャルスタジオ, 映像情報メディア学会誌, Vol. 56, No. 10, pp. 1538–1541 (2002).

[11] 住吉英樹, 有安香子, 望月祐一, 佐野雅則, 井上誠喜: データベースを中心とした番組制作支援システム, 情報処理学会研究報告, 2000-DBS-120, pp. 83–90 (2000).

[12] NHK TVML: URL: <http://www.nhk.or.jp/TVML>.

[13] TVML 道場: URL: <http://www.tvml.tv/>.

[14] 林正樹: テキスト台本からの自動番組制作〜TVML の提案, 情報処理学会研究報告, DBS-120-13, pp. 91–98 (2000).

[15] He, L., Cohen, M. F. and Salesin, D. H.: The Virtual Cinematographer: A Paradigm for Automatic Real-Time, *Proceedings of ACM SIGGRAPH'96*, New Orleans, pp. 217–224 (1996).

[16] Christianson, D., Anderson, S., He, L., Salesin, D., Weld, D. and Cohen, M.: Declarative Camera Control for Automatic Cinematography, *Proceedings of AAAI'96*, pp. 148–155 (1996).

[17] Drucker, S. M. and Zeltzer, D.: CamDroid: A System for Implementing Intelligent Camera Control, *Proceedings of Symposium on Interactive 3D Graphics*, Monterey, California, pp. 139 – 144 (1995).

[18] 西山晴彦, 大久保達真, 斉藤伸介, 松下温: 映像の知識に基づく仮想空間演出, 電子情報通信学会論文誌, Vol. J81-D-II, No. 1, pp. 146–155 (1998).

[19] 林正樹: 番組記述言語 TVML を使った情報の番組化, 情報処理学会研究報告, DBS-120-13, pp. 91–98 (2000).

[20] 牧野英二, 林正樹: TVML を使ったカメラワークシミュレータ, 映像情報メディア学会冬季大会 講演予稿集, p. 54 (1999).

[21] 道家守, 林正樹, 牧野英二: TVML を用いた番組情報からのニュース番組自動生成, 映像情報メディア学会誌, Vol. 54, No. 7, pp. 1097–1103 (2000).

[22] 加藤大一郎: 知的ロボットカメラの研究, NHK 技研 R&D, No. 52 (1998).

[23] 加藤大一郎, 津田貴生, 井上誠喜: 知的ロボットカメラ, NHK 技研 R&D, No. 64 (2000).

[24] 加藤大一郎, 石川秋男, 津田貴生, 福島宏, 下田茂, 山田光穂: カメラワーク分析と映像の主観評価実験, 映像情報メディア学会誌, Vol. 53, No. 9, pp. 1315–1324 (1999).

[25] 石川秋男, 加藤大一郎, 津田貴生, 福島宏, 下田茂, 山田光穂, 阿部一雄, 畠山祐里: 放送カメラマンのズーミング計測法の検討と静止している被写体を撮影するときのズーミング特性分析, 映像情報メディア学会誌, Vol. 53, No. 5, pp. 749–757 (1999).

[26] Kato, D., Ishikawa, A., Tsuda, T., Shimoda, S. and Fukushima, H.: Automatic control of a robot camera for broadcasting and subjective evaluation and analysis of reproduced images, *Proceedings of SPIE Human Vision and Electronic Imaging V*, Vol. 3959, pp. 468–479 (2000).

[27] 郷健太郎, 伊藤雅広, 今宮淳美: ズーム情報を利用した適応型遠隔カメラ制御法, 情報処理学会論文誌, Vol. 43, No. 2, pp. 585–604 (2002).

[28] 分散協調視覚プロジェクト: URL: <http://vision.kuee.kyoto-u.ac.jp/CDVPRJ/index-jp.html>.

[29] Nakazawa, A., Kato, H. and Inokuchi, S.: Human Tracking Using Distributed Vision System, *Proceedings of the 14th ICPR*, Vol. 1, Brisbane, Australia, pp. 593–596 (1998).

[30] 亀田能成, 石塚健太郎, 美濃導彦: 状況理解に基づく遠隔講義のための実時間映像化手法, 情報処理学会研究報告, CVIM-121-10, pp. 81–88 (2000).

[31] 浮田宗伯, 松山隆司: 移動対象の協調的追跡のための観測可能領域モデル生成・更新法, 情報処理学会論文誌, Vol. 42, No. 7, pp. 1902–1913 (2001).

[32] 浮田宗伯, 松山隆司: 能動視覚エージェント群による複数対象の実時間協調追跡, 情報処理学会論文誌, Vol. 43, No. SIG-11(CVIM-5), pp. 64–79 (2002).

[33] 日浦慎作, 村瀬健太郎, 松山隆司: ダイナミックメモリを用いた実時間対象追跡, 情報処理学会論文誌, Vol. 41, No. 11, pp. 3082–3091 (2000).

[34] 冷水明, 岡村耕二, 荒木啓二郎: 仮想雲台カメラ Union-Camera の設計と実現, マルチメディア，分散，協調とモバイル (DICOMO2001) シンポジウム論文集, pp. 241–246 (2001).

[35] Chiueh, T., Mitra, T., Neogi, A. and Yang, C.: Zodiac: A History-Based Interactive Video Authoring System, *Proceedings of ACM Multimedia*, Bristol, United Kingdom, pp. 435–444 (1998).

[36] Girgensohn, A., Boreczky, J., Chiu, P., Doherty, J., Foote, J., Golovchinsky, G., Uchihashi, S. and Wilcox, L.: A Semi-Automatic Approach to Home Video Editing, *Proceedings of ACM symposium on User interface and software technology*, San Diego, CA, pp. 81–89 (2000).

[37] 熊野雅仁, 有木康雄, 上原邦昭, 下条真司, 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, Vol. J85-D-I (2002).

[38] 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, Vol. 44, No. 3, pp. 915–924 (2003).

[39] 森山剛, 坂内正夫: ドラマ映像の心理内容に基づいた要約映像の生成, 電子情報通信学会論文誌, Vol. J84-D-II, No. 6, pp. 1122–1131 (2001).

[40] Sundaram, H. and Chang, S.: Condensing Computable Scenes Using Visual Complexity And Film, *Proceedings of IEEE International Conf. on Multimedia and Expo*, Tokyo, Japan, pp. 389–392 (2001).

[41] 西口敏司, 亀田能成, 美濃導彦, 池田克夫: 講義映像要約のための撮影ルールの構築, 電子情報通信学会 2000 年情報・システムソサイエティ大会講演論文集, PD-1-3, pp. 327–328 (2000).

[42] 三浦宏一, 浜田玲子, 井出一郎, 坂井修一, 田中英彦: 動きに基づく料理映像の自動要約, 情報処理学会論文誌, Vol. 44, No. SIG 9(CVIM 7), pp. 21–29 (2003).

[43] Hamada, R., Ide, I., Sakai, S. and Tanaka, H.: Associating Cooking Video with Related Textbook, *Proceedings of ACM Multimedia 2000 Workshop*, pp. 237–241 (2000).

[44] 廣瀬竜男, 中西吉洋, 秦淑彦, 田中克己: 被写体の写り具合に基づく多視点映像の検索と表示, データベースと Web 情報システムに関する IPSJ DBS/ACM SIGMOD Japan Chapter/JSPS-RFTF AMCP 合同シンポジウム (DBWeb2000), CVIM-121-10, pp. 81–88 (2000).

[45] Hata, T., Hirose, T. and Tanaka, K.: Skimming Multiple Perspective Video Using TempoSpatial Importance Measures, *IFIP 2.6 5th Working Conference on Visual Database Systems*, Fukuoka, Japan, pp. 219–238 (2000).

[46] 秦淑彦, 廣瀬竜男, 中西吉洋, 田中克己: カメラメタファによる多視点映像の検索, 情報処理学会論文誌：データベース, Vol. 42, No. SIG 4(TOD 9), pp. 14–26 (2001).

[47] 住吉英樹, 有安香子, 望月祐一, 佐野雅則, 井上誠喜: 階層化した情報管理構造を用いた番組制作, 映像情報メディア学会誌, Vol. 55, No. 3, pp. 397–404 (2001).

[48] 市村哲, 谷寛之, 中村亮太, 井上亮文, 松下温: MediaBlocks: マルチユーザ撮影映像共有が可能な Web 動画編集システム, 情報処理学会論文誌, Vol. 46, No. 1, pp. 15–25 (2005).

[49] Gross, R., Bett, M., Yu, H., Zhu, X., Pan, Y., Yang, J. and Waibel, A.: Towards A Multimodal Meeting Record, *Proceedings of IEEE International Conference on Multimedia and Expo (ICME2000)*, Vol. 3, pp. 1593–1596 (2000).

[50] Foote, J. and Kimber, D.: FlyCam: Practical Panoramic Video, *Proceedings of IEEE International Conference on Multimedia and Expo (ICME2000)*, Vol. 3, pp. 1419–1422 (2000).

[51] Sun, X., Foote, J., Kimber, D. and Manjunath, B. S.: Panoramic Video Capturing and Compressed Domain Virtual Camera Control, *Proceedings of the ninth ACM international conference on Multimedia*, pp. 329–338 (2001).

[52] Lee, D., Erol, B., Graham, J., Hull, J. J. and Murata, N.: Portable meeting recorder, *Proceedings of the tenth ACM international conference on Multimedia*, Juan-les-Pins, France, pp. 493–502 (2002).

[53] Rui, Y., Gupta, A. and Cadiz, J.: Viewing Meetings Captured by an Omni-Directional Camera, *Proceedings of CHI 2001*, pp. 450–457 (2001).

[54] 佐藤発樹: TV ビデオクリエーション - 演出術入門 -, オーム社 (1983).

[55] Inoue, T., Okada, K. and Matsushita, Y.: Learning from TV programs: application of TV presentation to a videoconferencing system, *Proceedings of ACM symposium on User interface and software technology*, Pittsburgh, Pennsylvania, United States, pp. 147–154 (1995).

[56] 井上智雄, 岡田謙一, 松下温: テレビ番組のカメラワークの知識に基づいた TV 会議システム, 情報処理学会論文誌, Vol. 37, No. 11, pp. 2095–2104 (1996).

[57] 大西正輝, 影林岳彦, 福永邦雄: 視聴覚情報の統合による会議映像の自動撮影, 電子情報通信学会論文誌, Vol. J85-D-II, No. 3, pp. 537–542 (2002).

[58] 大西正輝, 泉正夫, 福永邦雄: 情報発生量の分布に基づく遠隔講義撮影の自動化, 電子情報通信学会論文誌, Vol. J82-D-II, No. 2, pp. 1590–1597 (1999).

[59] 大西正輝, 泉正夫, 福永邦雄: 講義映像における板書領域のブロック分割とその応用, 電子情報通信学会論文誌, Vol. J83-D-I, No. 11, pp. 1187–1195 (2000).

[60] 大西正輝, 村上昌史, 福永邦雄: 状況理解と映像評価に基づく講義の知的自動撮影, 電子情報通信学会論文誌, Vol. J85-D-II, No. 4, pp. 594–603 (2002).

[61] Kameda, Y., Ishizuka, K. and Minoh, M.: A Live Video Imaging Method for Capturing Presentation Information In Distance Learning, *IEEE International Conference on Multimedia and Expo (ICME2000)*, Vol. 3, pp. 1237–1240 (2000).

[62] 先山卓朗, 大野直樹, 椋木雅之, 池田克夫: 遠隔講義における講義状況に応じた送信映像選択, 電子情報通信学会論文誌, Vol. J84-D-II, No. 2, pp. 248–257 (2001).

[63] 宮崎英明, 亀田能成, 美濃導彦: 複数のカメラを用いた複数ユーザに対する講義の実時間映像化手法, 電子情報通信学会論文誌, Vol. J82-D-II, No. 10, pp. 1598–1605 (1999).

[64] 井口泰典, 土居元紀, 眞鍋佳嗣, 千原國宏: スポーツ映像放送のための実時間映像解析よるマルチカメラの自動制御と自動スイッチング, 映像情報メディア学会誌, Vol. 56, No. 2, pp. 271–279 (2002).

[65] 高井勇志, 土居元紀, 千原國宏: 映像解析と VR 技術を用いたフットボール練習支援システム, 日本バーチャルリアリティ学会第 4 回大会論文集, pp. 233–234 (1999).

[66] "Carnegie Mellon Goes to the Super Bowl": `URL: <http://www.ri.cmu.edu/events/sb35/tksuperbowl.html>`.

[67] Saito, H., Baba, S. and Kanade, T.: Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room, *IEEE Transaction on Multimedia*, Vol. 5, No. 3, pp. 303–316 (2003).

[68] Kitahara, I. and Ohta, Y.: Scalable 3D Representation for 3D Video Display in a Large-scale Space, *Proceedings of IEEE Virtual Reality Conference 2003 (VR2003)*, pp. 45–52 (2003).

[69] 大田友一, 尾野徹, 秋道慎志, 藤田米春, 井上誠喜, 斎藤英雄, 北原格, 大城英裕, 藤野幸嗣, 金出武雄: 仮想化現実技術による自由視点 3 次元映像スタジアム通信, 電子情報通信学会技術研究報告, PRMU 2000-188, pp. 17–22 (2001).

[70] 石川寛享, 北原格, 大田友一: 大規模空間の他視点映像を用いた運動視差の再現可能な自由視点映像提示, 電子情報通信学会技術研究報告, PRMU 2000-190, pp. 31–38 (2001).

[71] "Virtualized Reality": URL: <http://www.ri.cmu.edu/labs/lab_62.html>.

[72] Ozeki, M., Nakamura, Y. and Ohta, Y.: Camerawork For Intelligent Video Production –Capturing Desktop Manipulations, *International Conference on Multimedia and Expo(ICME2001)*, pp. 41–44 (2001).

[73] 尾関基行, 伊藤雅嗣, 里雄二, 中村裕一, 大田友一: 複合コミュニティ空間における注目の共有〜注目誘導行動による物体への注釈付け〜, 日本バーチャルリアリティ学会論文誌, Vol. 8, No. 4, pp. 369–377 (2003).

[74] 尾関基行, 中村裕一, 大田友一: 机上作業シーンの自動撮影のためのカメラワーク, 電子情報通信学会論文誌, Vol. J86-D-II, No. 11, pp. 1606–1617 (2003).

[75] 尾関基行, 中村裕一, 大田友一: 話者の注目喚起行動による机上作業映像の自動編集—ユーザインタフェースの側面からの評価—, 情報科学レターズ, pp. 269–272 (2004).

[76] Bobick, A. and Pinhanez, C.: Controlling View-Based Algorithms Using Approximate World Models and Action Information, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 955–961 (1997).

[77] 田中彰, 東海彰吾, 松山隆司: イベント駆動型カメラワークによる動的シーンの効果的映像化, 情報処理学会研究報告, CVIM-121-10, pp. 73–80 (2000).

[78] Tokai, S. and Matsuyama, T.: Scenario-based Cooperative Camera-work Planning for Dynamic Scene Visuallization, *4th Symposium on Intelligent Information Media (IIM 1998)*, pp. 9–16 (1998).

[79] 高橋誠: 会議の進め方, 日本経済新聞社 (1987).

[80] Sauter, C., Morger, O., Muhlherr, T., Hutchison, A. and Teufel, S.: CSCW for Strategic Management in Swiss Enterprises: an Empirical Study, *Proceedings of EC-SCW'95*, pp. 117–132 (1995).

[81] Cutler, R., Rui, Y., Gupta, A., Cadiz, J., Tashev, I., wei He, L., Colburn, A., Zhang, Z., Liu, Z. and Silverberg, S.: Distributed meetings: a meeting capture and broadcasting system, *Proceedings of the tenth ACM international conference on Multimedia*, Juan-les-Pins, France, pp. 503–512 (2002).

[82] Drozd, A., Bowers, J., Benford, S., Greenhalgh, C. and Fraser, M.: Collaboratively Improvising Magic: An Approach to Managing Participation in an On-Line Drama, *Proceedings of ECSCW2001*, Bonn, Germany, pp. 159–178 (2001).

[83] Tsai, R. Y.: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, pp. 323–344 (1987).

[84] 柴貞行, 井上亮文, 高久宗史, 重野寛, 松下温: 複数カメラの協調動作システム, 情報処理学会第 64 回全国大会, Vol. 3, pp. 455–456 (2002).

[85] 高久宗史, 井上亮文, 平石絢子, 柴貞行, 重野寛, 岡田謙一, 松下温: 視差角モデルに基づいた複数カメラの協調システム - HARMONICS, 情報処理学会研究報告, 2002-IAC-3, pp. 35–40 (2002).

[86] Inoue, A., Hiraishi, J., Takaku, H., Shiba, S., Shigeno, H., Okada, K. and Matsushita, Y.: Multi-Camera Collaboration System based on Parallax Angle Model, *Proceedings of the IASTED International Conference NPDPA2002*, Tsukuba, Japan, pp. 437–442 (2002).

[87] 灰塚凡樹, 井上誠喜: カメラワーク生成に関する一考察, 電子情報通信学会技術研究報告, PRMU96-202, pp. 69–74 (1997).

[88] Pinhanez, C. and Bobick, A.: Approximate World Models: Incorporating Qualitative and Linguistic Information into Vision Systems, *Proceedings of the AAAI'96*, pp. 1116–1123 (1996).

[89] 満富俊郎: オーケストラとは何か, 新潮社 (1992).

[90] 近衛秀麿: オーケストラを聞く人へ, 音楽之友社 (1999).

[91] 鈴木織衛: オーケストラを読む本, ヤマハミュージックメディア (2000).

[92] 岩宮眞一郎: 音楽と映像のマルチモーダル・コミュニケーション, 九州大学出版会 (2000).

[93] ジェレミー ヴィンヤード: 映画技法完全レファレンス, フィルムアート社 (2002).

[94] 矢向正人: XML を用いた長唄譜のデータ表現, 情報処理学会研究報告, MUS-42-7, pp. 43–48 (2001).

[95] 東京オペラシティコンサートホール施設ガイド: URL: <http://www.operacity.jp/guide/ch.html>.

[96] Pinhanez, C. S., Mase, K. and Bobick, A.: Interval Scripts: a Design Paradigm for Story-Based Interactive Systems, *Proceedings of CHI'97*, Atlanta, GA, pp. 287–294 (1997).

[97] 大照完, 橋本周司: 仮想音楽空間, オーム社 (1996).

論文目録

主論文に関する原著論文

(1) 井上亮文, 吉田竜二, 平石絢子, 重野寛, 岡田謙一, 松下温: 映画の映像理論に基づく対面会議シーンの自動撮影手法, 情報処理学会論文誌, Vol.45, No.1, pp.212-221, Jan. 2004.

(2) 井上亮文, 平石絢子, 柴貞行, 市村哲, 重野寛, 岡田謙一, 松下温: シナリオ情報によるオーケストラ演奏のカメラワーク生成手法, 情報処理学会論文誌, Vol.46, No.1, pp.38-50, Jan. 2005.

その他の刊行論文

(1) 宇田隆哉, 砂田智, 井上亮文, 重野寛, 松下温: ソフトウエアベースの音楽配信プラットフォーム, 情報処理学会論文誌, Vol.41, No.8, pp.2237-2245, Aug. 2000.

(2) 市村哲, 谷寛之, 中村亮太, 井上亮文, 松下温: MediaBlocks: マルチユーザ撮影映像共有が可能な Web 動画編集システム, 情報処理学会論文誌, Vol.46, No.1, pp.15-25, Jan. 2005.

国際会議

(1) Uda, R., Sunada, A., Inoue, A., Shigeno, H., Matsushita, Y.: Secure Network System for Digital Music Contents with Self-Extracting Capsule, *Proceedings of IASTED International Conference PDCS*, pp.436-441, Nov. 1999.

(2) Yoshida, R., Inoue, A., Hiraishi, J., Shigeno, H., Matsushita, Y.: EXWeb: Remotely Operating Devices in the Home Network, *Proceedings of IEEE IWNA*, pp.267-274, Jan. 2002.

(3) Inoue, A., Hiraishi, J., Takaku, H., Shiba, S., Shigeno, H., Okada, K., Matsushita, Y.: An Implementation of Multi-camera Cooperation System Based on Parallax Angle Model, *Proceedings of IASTED International Conference NPDPA*, pp.437-442, Oct. 2002.

(4) Inoue, A., Shigeno, H., Okada, K., Matsushita, Y.: Introducing Grammar of the Film Language Into Automatic Shooting for Face-to-face Meetings, *Proceedings of IEEE SAINT*, pp.277-280, Jan. 2004.

(5) Sumiya, T., Inoue, A., Shiba, S., Kato, J., Shigeno, H., Okada, K.: A CSCW System for Distributed Search/Collection Tasks using Wearable Computers, *Proceedings of IEEE WMCSA*, pp.20-27, Dec. 2004.

研究会

(1) 宇田隆哉, 砂田智, 井上亮文, 重野寛, 松下温: 自己展開型カプセルを用いた安全なデジタル音楽コンテンツ配信システム, 情報処理学会マルチメディア・分散・協調とモバイル (DICOMO'99) シンポジウム, Vol.99, No.7, pp.447-452 (Jun. 1999).

(2) 井上亮文, 吉田竜二, 重野寛, 松下温: 会議における参加者の変動と端末の多様性に注目した 電子資料配布方法の検討, 情報処理学会マルチメディア・分散・協調とモバイル (DICOMO2000) シンポジウム, Vol.2000, No.7, pp.589-594 (Jun. 2000).

(3) 井上亮文, 吉田竜二, 平石絢子, 重野寛, 松下温: 分散機器間連携によるエンドユーザ環境の構築, 情報処理学会研究報告 2001-DPS-103, pp.7-12 (Jun. 2001).

(4) 吉田竜二, 井上亮文, 平石絢子, 重野寛, 松下温: Web ベースのモバイル－ホームネットワーク連携方式の提案, 情報処理学会マルチメディア・分散・協調とモバイル (DICOMO2001) シンポジウム, Vol.2001, No.7, pp.411-416 (Jun. 2001).

(5) 高久宗史, 井上亮文, 平石絢子, 柴貞行, 岡田謙一, 重野寛, 松下温: 視差角モデルに基づいた複数カメラの協調システム - HARMONICS, 情報処理学会研究報告 2002-IAC-3, pp.35-40 (Jun. 2002).

(6) 平石絢子, 井上亮文, 重野寛, 岡田謙一, 松下温: 映画の撮影手法に基づいた会議の自動撮影, マルチメディア, 分散, 協調とモバイル (DICOMO2002) シンポジウム論文集, Vol.2002, No.9, pp.285-288 (Jul. 2002).

(7) 高久宗史, 井上亮文, 柴貞行, 加藤淳也, 重野寛, 岡田謙一: ウェアラブルネットワーク環境での Uplink サービスフレームワーク, 情報処理学会研究報告 2002-DPS-110, pp.61-66 (Nov. 2002).

(8) 柴貞行, 井上亮文, 平石絢子, 高久宗史, 重野寛, 岡田謙一: オーケストラ撮影における楽譜を利用したカメラワークの計画, 情報処理学会研究報告 2003-DPS-111, pp.65-70 (Feb. 2003).

(9) 井上亮文, 高久宗史, 柴貞行, 加藤淳也, 重野寛, 岡田謙一: W4: ウェラブルサーバによる個人情報発信型アーキテクチャ, 情報処理学会研究報告 2003-DPS-112, pp.49-54 (Mar. 2003).

(10) 柴貞行, 井上亮文, 平石絢子, 重野寛, 岡田謙一: 楽譜を利用したカメラワーク計画手法の実装と評価, マルチメディア, 分散, 協調とモバイル (DICOMO2003) シンポジウム論文集, Vol.2003, No.9, pp.821-824 (Jun. 2003).

(11) 加藤淳也, 井上亮文, 柴貞行, 重野寛, 岡田謙一: 複数ウェアラブルコンピュータによる情報の生成・取得方式, 情報処理学会研究報告 2003-BCCgr-5, pp.27-32 (Jul. 2003).

(12) 柴貞行, 井上亮文, 加藤淳也, 住谷哲夫, 重野寛, 岡田謙一: ウェアラブルコンピュータを用いた身体コラボレーション支援, 情報処理学会研究報告 2003-GN-49, pp.85-90 (Oct. 2003).

(13) 住谷哲夫, 井上亮文, 柴貞行, 重野寛, 加藤淳也, 岡田謙一: ウェアラブルコンピュータを用いた分散型協調活動支援, 情報処理学会第 28 回 MBL 研究会, pp.119-124 (Mar. 2004).

(14) 加藤淳也, 井上亮文, 柴貞行, 住谷哲夫, 重野寛, 岡田謙一: ウェアラブルコンピュータと蓄積情報を用いた分散同期協調活動支援システムの実装と評価, マルチメディア, 分散, 協調とモバイル (DICOMO 2004) シンポジウム, pp.651-654 (Jul. 2004).

(15) 中村亮太, 市村哲, 井上亮文, 岡田謙一, 松下温: 複数の生体情報を用いた心理状況判別法, グループウェアとネットワークサービスワークショップ, pp.93-98 (Nov. 2004).

(16) 市村哲, 中村亮太, 井上亮文, 松下温: 同じイベントを撮影した人たちが映像を共有できる動画編集 Web システム, グループウェアとネットワークサービスワークショップ, pp.75-80 (Nov. 2004).

(17) 兵頭和樹, 田胡和哉, 伊藤雅仁, 井上亮文, 宇田隆哉, 市村哲, 星徹, 松下温: ロボット用オペレーティングシステム NOAH の構想, コンピュータシステム・シンポジウム, pp.113-120 (Nov. 2004).

口頭発表

(1) 井上亮文, 砂田智, 宇田隆哉, 重野寛, 松下温: 二次利用を考慮した著作権管理システムの提案, 情報処理学会第 58 回全国大会, Vol.3, pp.229-231 (Mar. 1999).

(2) 吉田竜二, 井上亮文, 重野寛, 松下温: Jini 仲介サービスによる対面会議支援, 情報処理学会第 60 回全国大会, Vol.3, pp.429-430 (Mar.2000).

(3) 平石絢子, 井上亮文, 吉田竜二, 重野寛, 松下温: Web ベースのモバイルーホームネットワーク連携, 情報処理学会第 62 回全国大会, 特別トラック 5, pp.139-142 (Mar. 2001).

(4) 柴貞行, 井上亮文, 平石絢子, 高久宗史, 重野寛, 松下温: 複数カメラの協調動作システム, 情報処理学会第 64 回全国大会, Vol.3, pp.455-456, (Mar. 2002).

(5) 加藤淳也, 井上亮文, 柴貞行, 重野寛, 岡田謙一: ユーザプロファイルを用いたウェアラブル機器のサーバ間連携手法, 情報処理学会第 65 回全国大会, Vol.3, pp.447-448 (Mar. 2003).

(6) 井上亮文, 神谷謙吾, 中村亮太, 市村哲, 松下温: RFID によるアクセス制御機構を持つネットワークスイッチ, 情報処理学会グループウェアとネットワークサービスワークショップ, pp.49-50 (Nov. 2004).