

論文 / 著書情報
Article / Book Information

題目(和文)	多視点ロボットカメラを用いた三次元映像表現の研究
Title(English)	
著者(和文)	池谷健佑
Author(English)	Kensuke Ikeya
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10998号, 授与年月日:2018年9月20日, 学位の種別:課程博士, 審査員:山口 雅浩,小林 隆夫,熊澤 逸夫,中本 高道,金子 寛彦
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10998号, Conferred date:2018/9/20, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

学位論文

多視点ロボットカメラを用いた三次元映像表現の研究

指導教員 山口雅浩 教授

東京工業大学 工学院 情報通信系

池谷健佑

2018年8月

目次

第 1 章 序論	1
1.1 研究目的	1
1.2 本論文の構成	2
第 2 章 多視点映像表現技術とインテグラル立体映像技術	3
2.1 はじめに	3
2.2 多視点映像表現技術	3
2.3 インテグラル立体映像技術	4
2.4 まとめ	5
第 3 章 多視点ロボットカメラを用いた多視点映像表現	6
3.1 はじめに	6
3.2 多視点ロボットカメラ	6
3.2.1 ロボットカメラの設置	7
3.2.2 カメラキャリブレーション	7
3.2.3 ロボットカメラの協調制御	7
3.3 多視点映像表現	10
3.3.1 被写体の決定	10
3.3.2 カメラキャリブレーション	10
3.3.3 仮想的カメラ制御	11
3.3.4 多視点映像表現の映像生成	12
3.4 システム	13
3.5 実験	14
3.5.1 撮影実験	14
3.5.2 番組利用	16
3.6 考察	19
3.7 まとめ	21
第 4 章 多視点カメラからのインテグラル立体映像生成手法	22
4.1 はじめに	22
4.2 提案手法	22
4.2.1 距離推定による 3 次元モデルの生成	24
4.2.1.1 DAISY による特徴量の記述	24

4.2.1.2	Belief Propagation を用いた距離推定	25
4.2.2	信頼度に基づく 3次元モデルの補正	26
4.2.2.1	エネルギー分布の尖度による信頼度の評価	27
4.2.2.2	3次元モデルの補正	28
4.2.3	3次元モデルからの要素画像郡の生成と統合	29
4.2.3.1	3次元モデルからの要素画像郡の生成	29
4.2.3.2	要素画像郡の統合	30
4.3	実験	30
4.3.1	距離推定アルゴリズムの性能評価実験	30
4.3.2	信頼度の性能評価実験	32
4.3.3	番組で使用された多視点映像を用いた実験	35
4.3.3.1	3次元モデルの生成実験	35
4.3.3.2	インテグラル立体映像コンテンツの制作	37
4.4	まとめ	39
第 5 章	多視点ロボットカメラを用いたインテグラル立体撮影手法	40
5.1	はじめに	40
5.2	提案手法	41
5.2.1	多視点ロボットカメラを用いた撮影	41
5.2.2	3次元モデルの生成	46
5.2.3	要素画像の生成	48
5.3	実験	49
5.3.1	多視点ロボットカメラの協調制御の性能検証	49
5.3.1.1	シミュレーション	50
5.3.1.2	システム開発	52
5.3.1.3	撮影実験	54
5.3.2	インテグラル立体映像の生成	56
5.4	考察	58
5.5	まとめ	61
第 6 章	結論	62
	謝辞	65
	参考文献	66
	研究業績	70

図目次

図 1.1	本論文の構成	2
図 2.1	多視点カメラの撮影領域	4
図 3.1	カメラの制御	7
図 3.2	デプス制御	9
図 3.3	デプスの決定	10
図 3.4	多視点映像表現の映像編集方法	12
図 3.5	ロボットカメラ	13
図 3.6	操作インターフェース	13
図 3.7	多視点映像表現処理装置	13
図 3.8	システム系統図	14
図 3.9	カメラ配置	15
図 3.10	バレーボールの多視点映像表現	15
図 3.11	体操の多視点映像表現	15
図 3.12	バスケットボールの多視点映像表現	16
図 3.13	柔道の多視点映像表現	16
図 3.14	会場の様子	17
図 3.15	カメラ配置	17
図 3.16	NHK 杯フィギュアスケート選手権大会 2013 の多視点映像表現	17
図 3.17	NHK 杯体操選手権 2013 の多視点映像表現	18
図 3.18	NHK 杯フィギュアスケート選手権大会 2014 の多視点映像表現	18
図 3.19	連続テレビ小説「とと姉ちゃん」の多視点映像表現	19
図 3.20	NHK 学生ロボコン 2017 の多視点映像表現	19
図 3.21	ABU ロボコン世界大会 2017 の多視点映像表現	19
図 3.22	Free D による多視点映像表現	20
図 3.23	4D Replay による多視点映像表現	20
図 3.24	多視点ロボットカメラによる多視点映像表現	21
図 4.1	提案手法の処理の流れ	23
図 4.2	DAISY 記述子	25
図 4.3	エネルギー分布の例	27
図 4.4	エラー領域と信頼度の比較	28

図 4.5	信頼度 R による 3 次元モデルの補正	28
図 4.6	立体像の欠落	29
図 4.7	3 次元モデルからのインテグラル画像の生成	30
図 4.8	視差推定結果	32
図 4.9	Middlebury の評価サイトより出力された正解領域とエラー領域	33
図 4.10	信頼度の視覚化	34
図 4.11	信頼度 R による正解領域とエラー領域の判定	34
図 4.12	カメラ配置	35
図 4.13	3 次元モデル生成実験の結果	37
図 4.14	距離画像におけるエラーが生じた領域	37
図 4.15	異なる視点位置から観測したインテグラル立体映像	38
図 4.16	相撲のインテグラル立体コンテンツ	39
図 5.1	多視点ロボットカメラの協調制御と配置	41
図 5.2	インテグラル立体ディスプレイの再現領域とカメラ配置	42
図 5.3	カメラ配置	43
図 5.4	多視点ロボットカメラの協調制御	46
図 5.5	要素画像の生成	49
図 5.6	シミュレーターのフローチャート	51
図 5.7	多視点ロボットカメラと被写体の配置	52
図 5.8	再現領域の撮影映像	52
図 5.9	多視点ロボットカメラ	53
図 5.10	操作インターフェース	53
図 5.11	システム系統図	54
図 5.12	多視点ロボットカメラの配置	55
図 5.13	インテグラル立体ディスプレイの再現領域を重畳した多視点映像	55
図 5.14	インテグラル立体ディスプレイ	56
図 5.15	3 次元モデル	57
図 5.16	要素画像	57
図 5.17	インテグラル立体映像	58
図 5.18	多視点ロボットカメラの協調制御の性能評価	59
図 5.19	多視点ロボットカメラの協調制御	60
図 5.20	三次元モデルの生成	61

表目次

表 3.1	システムを利用した番組	18
表 4.1	提案手法のパラメーター	31
表 4.2	視差推定エラーの割合	32
表 4.3	正解領域とエラー領域における信頼度の平均値	34
表 4.4	正解領域とエラー領域の判定結果の適合率	34
表 4.5	提案手法と従来の BP のパラメーター	36
表 4.6	距離画像においてエラーが生じた画素数およびその割合	37
表 5.1	インテグラル立体ディスプレイのパラメーター	57

第1章 序論

1.1 研究目的

放送技術は、より臨場感のある放送の実現にむけて、これまで進歩、発展を遂げてきた。1953年にテレビ放送が始まってから、1960年にカラーテレビ放送、1991年にハイビジョン放送、2016年にスーパーハイビジョン放送と、その映像は白黒からカラー化、ワイド化、高精細化と進歩してきた。一方、これらの進歩は2次元映像を対象としたものであり、さらなる臨場感のある放送の実現にむけては、放送に3次元映像表現技術を応用し、撮影対象の3次元空間情報を表現することが期待される。

3次元映像表現技術の一つに多視点映像表現技術がある。バレットタイムやタイムスライスとも呼ばれるこの映像表現技術は、被写体の周囲に複数のカメラ（多視点カメラ）を配置して同時に撮影し、カメラの並びに沿って映像を切り替えることで、時間を停止もしくはスローにして被写体の周囲を視点が回り込むような映像表現を実現する。視点が回り込む際の運動視差によって立体的に見え、視聴者の目を引く高い視覚効果があると共に、被写体をさまざまな視点から分かりやすく表現できるというメリットがある。

また、その他の3次元映像表現技術にインテグラル立体映像技術がある。空間像再生型の一方式であるこの立体映像技術は、高解像度ディスプレイやプロジェクターとレンズアレイを組み合わせることで、被写体の光線を忠実に再現し、視聴者の視点に応じた立体像を生成する。特殊なめがねが不要で、視聴時の疲労が少なく、自然な立体像を生成できるというメリットがある。

視聴者に新しい体験を提供し続けるために、放送は現在から未来まで絶え間なく進歩、発展することが望ましい。多視点映像表現技術の映像は、現在普及しているテレビやディスプレイに表示することが可能であるため、現在の放送への応用が可能である。一方、インテグラル立体映像技術の映像は、現在一般的には普及していないレンズアレイを取り付けたディスプレイやプロジェクターでしか表示できないため、表示デバイスや周辺技術の進歩を見据え、未来の放送への応用が相応である。これらの3次元映像表現技術を放送へ応用することで、現在、そして未来の放送においてさらなる臨場感の実現が期待できる。その一方で、放送に応用するためには、双方の技術において、スポーツの選手やドラマの俳優などさまざまな被写体を撮影できなくてはならない。ダイナミックに動く被写体や広い空間に点在する被写体、遠方の被写体といった多様な条件下の被写体を撮影可能な3次元映像撮影技術が必要である。また、ハイビジョン、4K、スーパーハイビジョンと高精細化をたどる放送に応用するためには、双方の技術において映像の画質を向上させることが重要である。

そこで、本論文では、多視点ロボットカメラを用いた3次元映像表現を提案する。多視点ロボットカメラは、複数台のロボットカメラの協調制御システムで、1人のカメラマンが1台のロボットカメラを操作して被写体を撮影すると、すべてのロボットカメラが被写体へ一斉に方向制御され、被写体の多視点映像を撮影するという新たな3次元映像撮影デバイスである。本カメラによってダイナミックに移動する被写体や広い空間に点在する被写体など、多様な被写体の多視点映像を高画質に撮影することができる。本論文では、まず、現在の放送に多視点映像表現技術を応用するため、多視点ロボットカメラの制御手法、そのシステムの設計方

法, その撮影映像を用いた多視点映像表現の制作手法を提案すると共に, 多視点ロボットカメラによる撮影から多視点映像表現の出力まで準リアルタイムに処理するシステムを開発する. そして, 未来の放送にインテグラル立体映像技術に応用するために, 多視点カメラの撮影映像から 3 次元モデルを生成し要素画像に変換するインテグラル立体映像生成手法を提案する. さらに, その撮影に多視点ロボットカメラを導入したインテグラル立体撮影手法を提案する. これらが, 放送に 3 次元映像表現技術に応用し, 撮影対象の 3 次元空間情報を表現するうえで有効な技術であることを示す.

1.2 本論文の構成

本論文の構成を図 1.1 に示す.

第 2 章では, 多視点映像表現技術およびインテグラル立体映像技術の従来技術について概説し, それらを踏まえた研究課題と課題解決にむけたアプローチについて述べる.

第 3 章では, 多視点ロボットカメラを用いた多視点映像表現技術を提案する. 多視点ロボットカメラの制御手法およびシステムの設計手法, 多視点映像表現の制作手法を提案し, それらを踏まえたシステムを開発して放送番組に利用することで有効性を示す.

第 4 章では, 多視点カメラを用いたインテグラル立体映像生成手法を提案する. 多視点カメラを用いた 3 次元モデル生成手法および距離推定の信頼度を用いた 3 次元モデル補正手法を提案し, 放送番組における多視点カメラの撮影映像からインテグラル立体映像を生成することで有効性を示す.

第 5 章では, 多視点ロボットカメラを用いたインテグラル立体撮影手法を提案する. インテグラル立体映像として再現される実空間領域の多視点映像を撮影する多視点ロボットカメラの制御手法, その撮影映像からの 3 次元モデル生成手法, 3 次元モデルから要素画像への変換手法を提案し, 撮影実験を行うことでその有効性を示す.

最後に第 6 章で本論文をまとめる.

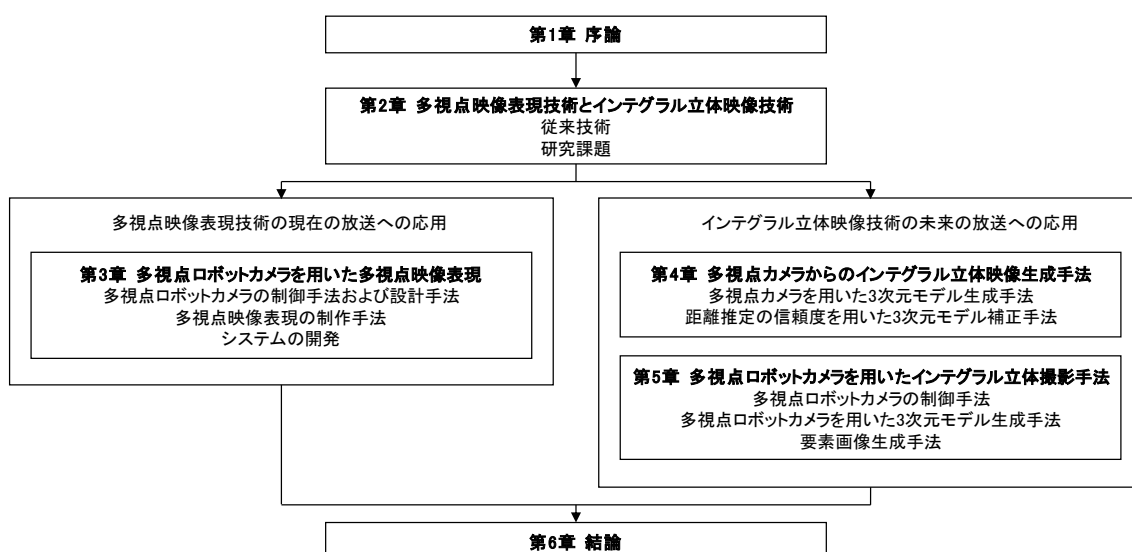


図 1.1 本論文の構成

第2章 多視点映像表現技術とインテグラル立体映像技術

2.1 はじめに

本章では, 3次元映像表現技術における多視点映像表現技術とインテグラル立体映像技術に関する従来技術について述べる. そして, それらを放送に導入するうえで, 本論文で取り組むべき課題について説明し, 課題解決にむけたアプローチについて示す.

2.2 多視点映像表現技術

多視点映像表現技術は, 1999年に映画「マトリックス」で俳優が銃弾をよけるシーンに利用され, 大きな話題となった. 複数のカメラを俳優の周囲に円周状に配置して撮影を行い, 撮影映像を切り替える手法を用いた. リプレイテクノロジー社が開発した「free D」は, スポーツ中継のリプレイで多視点映像表現を利用することで, 視聴者に分かりやすく伝えている[1]. 例えば, 野球におけるホームのクロスプレーに多視点映像表現を利用することで, セーフかアウトかの判定を分かりやすく伝えている. 複数のカメラで被写体を撮影し, 撮影映像から被写体の3次元形状を復元することで滑らかに視点が切り替える手法を用いている. 2001年に金出らが開発した「Eye Vision」は, アメリカンフットボールでフィールド上をダイナミックに移動する選手に多視点映像表現を利用することでタッチダウンの瞬間やフォーメーションを分かりやすく伝えた[2,3]. 複数の可動式ロボットカメラを使用し, それらを選手へ一斉に方向制御して撮影を行い, 撮影映像を切り替える手法を用いた. また, 多視点映像表現に関連してスポーツシーンの自由視点映像表現の研究も行われている. 自由視点映像表現では, 視聴者の操作によってスポーツシーンをインタラクティブに任意の視点から見るができる. GrauやHiltonは, 多視点映像から3次元モデルを生成し, サッカーの自由視点映像を実現した[4,5]. 北原, 斉藤らは, 計算コストを抑えるため, 多視点映像から3次元モデルを生成するのではなくビルボードを用いた自由視点映像表現システムを開発した[6,7,8]. また, Germannは, 多視点映像を用いるのではなく, 2台のカメラ映像から視点を内挿し, 自由視点映像表現を実現している[9].

「マトリックス」での撮影システムおよび「free D」は, 撮影に固定型の多視点カメラを使用しているため, ダイナミックに動く被写体や広い空間に点在する被写体の撮影が困難だという課題がある. (図 2.1(a)) そのような被写体を撮影するためには, 被写体の動く領域や点在する領域を広角で撮影し, デジタルズームで被写体を切り出すため, 相対的に被写体の解像度が低下し, 映像の画質が低下する. 同様に, 自由視点映像表現においても固定型カメラでフィールド上の全ての被写体を含めるように広角で撮影するため, 相対的に被写体の解像度が低下し, 映像の画質が低下するという課題がある. 「Eye Vision」は, 複数の可動式ロボットカメラを使用することで, その課題を解決しているが, 一方で撮影できる被写体が限定されるという課題がある. 「Eye Vision」のロボットカメラの制御手法には, 被写体は設定した2次元平面上に常に位置するという拘束条件が用いられているため, その2次元平面上の被写体しか撮影できない. (図 2.1(b)) 例えば, カメラを高い位置に設置し, 地面を2次元平面として俯瞰して撮影した場合, 地面上の被写体は撮影できるが, 地面から離れた高い位置にある被写体を撮影することはできない. また, システムを運用するための事前のカメラキャリブレーションに数日間かか

るという課題もある[10].

上記を踏まえ、放送に多視点映像表現技術を応用するため、本論文では、3次元空間をダイナミックに移動する被写体や3次元空間に点在する被写体の多視点映像を高解像度に撮影すること、また、スポーツ中継などの生放送に対応するため、撮影完了から準リアルタイムで多視点映像表現の映像を制作することを課題とする。この課題を解決するため、多視点ロボットカメラを開発する。これは、複数台のロボットカメラの協調制御システムで、1人のカメラマンが1台のロボットカメラを操作して3次元空間中の被写体を撮影すると、すべてのロボットカメラが被写体へ一斉に方向制御され、3次元空間中の被写体の多視点映像を撮影するという3次元映像撮影デバイスである。(図2.1(c))3次元空間をダイナミックに移動する被写体や3次元空間に点在する被写体の多視点映像を高解像度に撮影することが期待できる。また、多視点ロボットカメラによる撮影から多視点映像表現の出力まで準リアルタイムに処理するシステムを開発し、スポーツ中継などの生放送での利用を試みることで提案手法およびシステムの有効性を検証する。

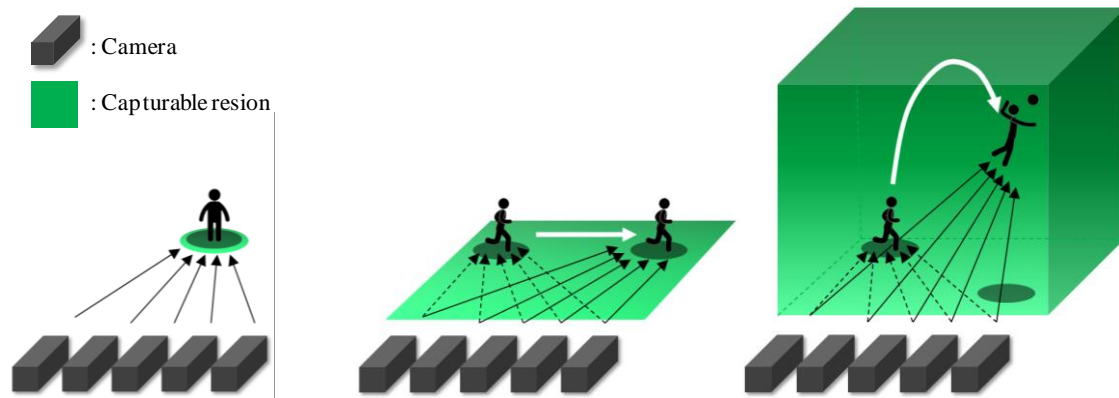


図 2.1 多視点カメラの撮影領域

(a) マトリックスおよび freeD の多視点カメラ

(b) Eye Vision

(c) 多視点ロボットカメラ

2.3 インテグラル立体映像技術

裸眼立体映像技術は高い臨場感を再現できる技術として期待されており、様々な研究が行われてきた。その中でもインテグラル立体映像技術は、像再生型の立体方式の1つで、特殊なメガネを必要とせず、水平、垂直方向に視差をもつ特徴がある[11]。3D映画[12,13]、物体追跡および抽出[14,15]、ホログラム立体映像への変換[16-18]、リフォーカス[19-23]、ユーザーインターフェース[24]などさまざまな分野に応用されてきた。インテグラル立体映像技術の表示では、レンズアレイを用いて被写体からの光線群を再現することでインテグラル立体映像を生成する。この光線群は撮影時に被写体から発せられた光線群と等しくなっている。インテグラル立体表示装置は、レンズアレイと高解像度ディスプレイもしくは高解像度プロジェクターを組み合わせた構成になっている[25-27]。レンズアレイを構成する一つのレンズに対応した画像を要素画像と呼び、要素画像内の各画素が発した光線がレンズアレイを通じ、被写体が発した光線群を再現することで立体像を生成する。インテグラル立体の撮影には、レンズアレイと高解像度カメラを組み合わせたインテグラル立体カメラを使用し、被写体が発する光線を撮影する[28-36]。実空間

の光線をそのまま取得することができ、リアルタイムでインテグラル立体表示が可能である。

一方、インテグラル立体カメラには、遠方の被写体を撮影する場合、視差を取得するために立体像の奥行き位置を調整する奥行き制御レンズが大きくなるという課題がある。例えば、スポーツ中継などの番組では、被写体の近くにカメラを置くことは難しいため、遠方からダイナミックに移動する被写体をパン、チルト、ズームして撮影する。インテグラル立体カメラで撮影する場合、巨大な奥行き制御レンズを設置したカメラでパン、チルトして被写体を撮影することになるが、番組制作でそのようなカメラの運用は現実的に困難である。

上記を踏まえ、放送にインテグラル立体映像技術を応用するため、本論文では、遠方においてダイナミックに移動する被写体や広い空間に点在する被写体のインテグラル立体映像を生成することを課題とする。この課題を解決するために、まず、多視点カメラを用いてインテグラル立体映像を生成する手法を提案する。この手法では、多視点カメラの撮影映像から3次元モデルを生成することで、仮想空間で被写体が発する光線を取得し、インテグラル立体映像を生成する。奥行き制御レンズを使用する必要がなく、多視点カメラ間のベースラインを調整することで遠方の被写体の撮影が期待できる。一方、疎に配置された多視点カメラから距離推定を用いて3次元モデルを生成するうえで、距離推定におけるエラーによって3次元モデルの品質が低下するという課題がある。そこで距離推定におけるエラーを抑制するため、多視点カメラを用いた3次元モデル生成手法および距離推定における信頼度を用いた3次元モデル補正手法を提案する。そして、撮影に2.1で述べた多視点ロボットカメラを導入したインテグラル立体映像の生成手法を提案する。この手法では、インテグラル立体映像として再現される実空間領域の多視点映像を高解像度に撮影し、3次元モデルを生成してインテグラル立体映像に変換する。遠方においてダイナミックに移動する被写体や広い空間に点在する被写体のインテグラル立体映像を高品質に生成することが期待できる。

2.4 まとめ

本章では、3次元映像表現技術における多視点映像表現技術とインテグラル立体映像技術に関する従来技術とその課題、課題解決にむけたアプローチについて示した。多視点映像表現技術では、3次元空間をダイナミックに移動する被写体や3次元空間に点在する被写体の多視点映像を高解像度に撮影すること、また、スポーツ中継などの生放送に対応するため、撮影完了から準リアルタイムで多視点映像表現の映像を制作することが課題であり、この課題を解決するため、多視点ロボットカメラおよび撮影から多視点映像表現の出力まで準リアルタイムに処理するシステムを開発する。インテグラル立体映像技術では、遠方においてダイナミックに移動する被写体や広い空間に点在する被写体のインテグラル立体映像を生成することが課題であり、この課題を解決するため、多視点カメラを用いてインテグラル立体映像を生成する手法およびその撮影に多視点ロボットカメラを導入したインテグラル立体映像の生成手法を提案する。

第3章 多視点ロボットカメラを用いた多視点映像表現

3.1 はじめに

本章の研究目的は、3次元空間をダイナミックに移動する被写体や3次元空間中に広く点在する被写体の多視点映像表現を実現することである。また、スポーツ中継のリプレイでの利用を目指し、放送現場でのシステムの事前準備が短時間で完了し、準リアルタイムで多視点映像表現を生成可能なシステムを構築することである。これらにより、より汎用的かつ実用的な多視点映像表現システムを実現する。目的を達成するためには、以下の要求条件を満たす必要がある。

- ・要求条件 1

滑らかに視点が切り替わる多視点映像表現を実現するため、3次元空間をダイナミックに移動する被写体に対しても、多視点カメラを高い精度で方向制御することが可能であること。

- ・要求条件 2

生放送のスポーツ中継において、競技中のシーンの多視点映像表現を競技直後のリプレイで放送可能な処理時間で生成できること。

- ・要求条件 3

システムの事前準備でのカメラキャリブレーションが短時間で完了すること。

この目的を達成するため、多視点ロボットカメラを開発する。本システムでは、複数台のロボットカメラを1人のカメラマンの操作によって3次元空間中の被写体に向けて一斉に方向制御し、多視点映像を撮影する。そして、撮影映像を計算機に取り込み、射影変換を用いてカメラを仮想的に被写体へ再方向制御することで、カメラマンの操作ミスやロボットカメラの機械的な制御誤差により生じた方向制御エラーを補正し、多視点映像表現を準リアルタイムで生成する。

3.2 多視点ロボットカメラ

多視点ロボットカメラは、複数台のズーム制御が可能な2軸（パン，チルト）のロボットカメラを1人のカメラマンの操作によって3次元空間中の被写体に向けて一斉に方向制御し多視点映像を撮影する「機械的方向制御」（図 3.1(a)）と、撮影映像を計算機に取り込み、射影変換を用いてカメラを仮想的に被写体へ再方向制御することで、カメラマンの操作ミスやロボットカメラの機械的な制御誤差により生じた方向制御エラーを補正する「仮想的方向制御」（図 3.1(b)）を組み合わせることで多視点映像表現を生成する。

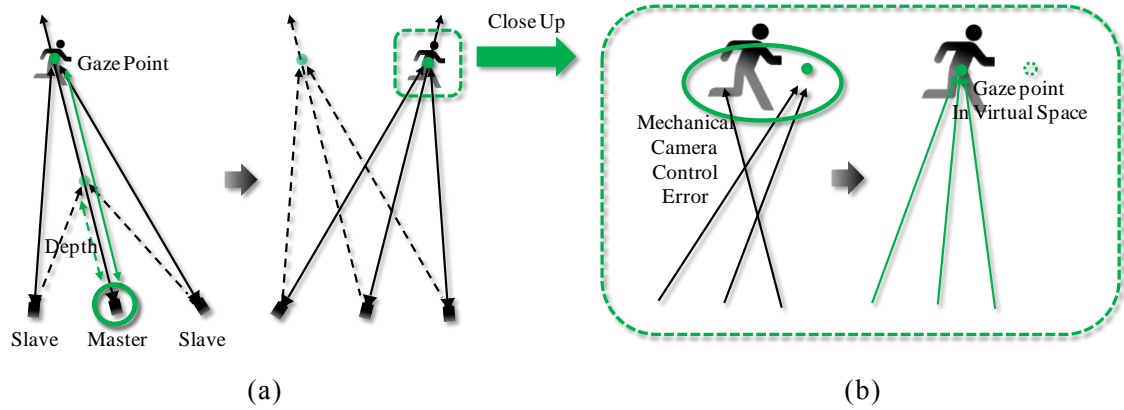


図 3.1 カメラの制御
(a) 機械的方向制御 (b) 仮想的方向制御

3.2.1 ロボットカメラの設置

被写体の移動領域や位置する領域を考慮して撮影領域を決定する．撮影領域を囲むように多視点ロボットカメラを円弧状もしくは直線状に並べて設置する．その際、両端のカメラの光軸が撮影領域の中心で、多視点映像表現の視点を回り込ませる角度で輻輳するように設置する．

3.2.2 カメラキャリブレーション

各ロボットカメラを Cam_n ($1 \leq n \leq N$) と定義し、 n はカメラ番号、 N はカメラの台数とする．すべてのロボットカメラを手動で方向制御し、撮影領域内にある適当な被写体の多視点映像を撮影する．撮影映像に弱校正カメラキャリブレーション [37] を施し、 Cam_n のカメラパラメーター（回転行列 R_{0n} 、並進行列 t_{0n} ）と撮影時のエンコーダー値（パン P_{0n} 、チルト T_{0n} ）を取得する．世界座標 x からカメラ座標 x_n への変換式を式(3.1)に示す．

$$x_n = [R_{0n} \quad t_{0n}] \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (3.1)$$

3.2.3 ロボットカメラの協制御

概要を図 3.1(a)に示す．すべてのロボットカメラが、3次元空間中の点である注視点を常に追跡するように協制御を行う．注視点は、カメラマンによるマスターカメラの操作によって、3次元空間中の任意の位置に指定することができ、注視点を被写体の位置に指定することで被写体の多視点映像を撮影できる．

まず、 Cam_n から 1 台のマスターカメラ Cam_M を選択し、それ以外をスレーブカメラとする．マスターカメラの光軸 o_M 上には、注視点 g が設定される． o_M は式(3.2)(3.3)(3.4)より求める．

$$R_M = R_{0M}^{-1} \begin{bmatrix} \cos(P_M - P_{0M}) & 0 & \sin(P_M - P_{0M}) \\ 0 & 1 & 0 \\ -\sin(P_M - P_{0M}) & 0 & \cos(P_M - P_{0M}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(T_M - T_{0M}) & -\sin(T_M - T_{0M}) \\ 0 & \sin(T_M - T_{0M}) & \cos(T_M - T_{0M}) \end{bmatrix} \quad (3.2)$$

$$\mathbf{R}_M = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (3.3) \quad \mathbf{o}_M = \begin{bmatrix} r_{13} \\ r_{23} \\ r_{33} \end{bmatrix} \quad (3.4)$$

ここで、 P_M , T_M はマスターカメラのパン，チルトのエンコーダ値を示す．また， \mathbf{g} は式(3.5)より求める．

$$\mathbf{g} = -\mathbf{R}_{0M}^{-1}T_{0M} + D \cdot \mathbf{o}_M \quad (3.5)$$

ここで，式(3.5)における D はデプス値を表す．デプスとは，マスターカメラと注視点間の距離のことをいう．デプスを連続的に増減することによって，注視点は光軸に沿って移動する．カメラマンは，マスターカメラのパン，チルト操作により光軸の方向を，デプス操作により光軸上の注視点の位置を変えることができる．ここで，デプスの操作デバイスの操作量と線形にデプスを増減させた場合，図 3.2(a)のように，注視点に追従するスレーブカメラの制御角度が，注視点がマスターカメラと近い位置にある時は大きく変化し，遠い位置にある時は小さく変化するため，直感的な操作が難しい．そこで，式(3.6)を用いてデプスを操作する．

$$D = B \cdot \sin(\theta \cdot d/d_{max}) / \sin(\theta \cdot (1 - d/d_{max})) \quad (3.6)$$

式(3.6)を用いてデプスを操作するために，まず，図 3.2(b)のように，複数台のスレーブカメラから 1 台を選択する．式(6)における B は，マスターカメラと選択したスレーブカメラ間の距離値， θ は，選択したスレーブカメラの方向制御の範囲であり，選択したスレーブカメラがマスターカメラの位置へ方向制御された時の角度とマスターカメラと同じ方向へ方向制御された時の角度の幅である． d はデプスの操作デバイスの値， d_{max} はその最大値である．式(3.6)により，デプスの操作デバイスの操作量と線形に注視点に追従するスレーブカメラの制御角度を変えることができ，直観的な操作を行うことができる．そして，すべてのスレーブカメラを注視点へ方向制御させる．カメラ座標系における各カメラから \mathbf{g} へのベクトル \mathbf{o}_n を式(3.7)より求める．

$$\mathbf{o}_n = \mathbf{R}_{0n} \frac{\mathbf{g} + \mathbf{R}_{0n}^{-1} \mathbf{t}_{0n}}{\|\mathbf{g} + \mathbf{R}_{0n}^{-1} \mathbf{t}_{0n}\|} \quad (3.7)$$

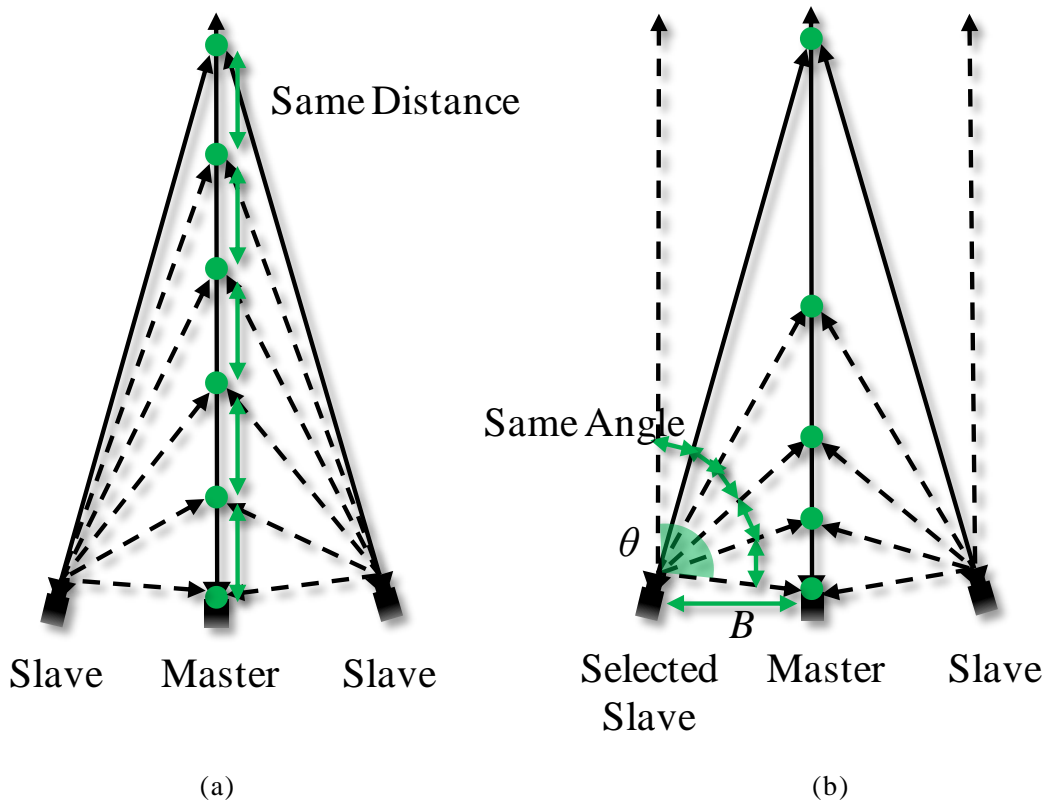


図 3.2 デプス制御

- (a) デプス操作デバイスの操作量に比例してデプスが増減
 (b) デプス操作デバイスの操作量に比例してスレーブカメラの制御角度が増減

各カメラを注視点へ方向制御させるための制御角（パン P_n ，チルト T_n ）を式(3.8)(3.9)(3.10)より求める．

$$\mathbf{o}_n = [e_{n1} \ e_{n2} \ e_{n3}]^T \quad (3.8)$$

$$P_n = \tan^{-1}(e_{n3}/e_{n1}) + P_{0n} \quad (3.9)$$

$$T_n = \sin^{-1}(e_{n2}) + T_{0n} \quad (3.10)$$

各カメラを P_n ， T_n だけ制御させることで注視点に追従させる．また，光学ズームを用いて注視点上の被写体をすべてのカメラで同じサイズで撮影するために，各カメラから注視点までの距離 D に応じて，各カメラの焦点距離 f_0 を式(3.11)で制御させる．

$$f_{0n} = f_{0M}(D_n / D_M) \quad (3.11)$$

カメラマンは，注視点を被写体位置に指定して多視点映像を撮影するため，まず，マスターカメラのパン，チルトを操作して被写体を捉える．次に，デプスを操作して，注視点を光軸に沿わせて被写体位置まで移動させる．カメラマンは，マスターカメラの他に，最低 1 台のスレーブカメラの撮影映像を見て操作し，スレーブカメラの撮影映像における被写体の位置によってデプスが合っているか判断する．例えば，図 3.3(a)のようにデプスが合っていない時は，マスターカメラの撮影映像

では画面の中心に被写体が位置しているが，スレーブカメラでは画面の中心に被写体は位置していない．一方，図 3.3(b)のようにデプスが合っている時は，マスターカメラとスレーブカメラの双方で画面の中心に被写体が位置しており，カメラマンはデプスが合っていることを判断することができる．

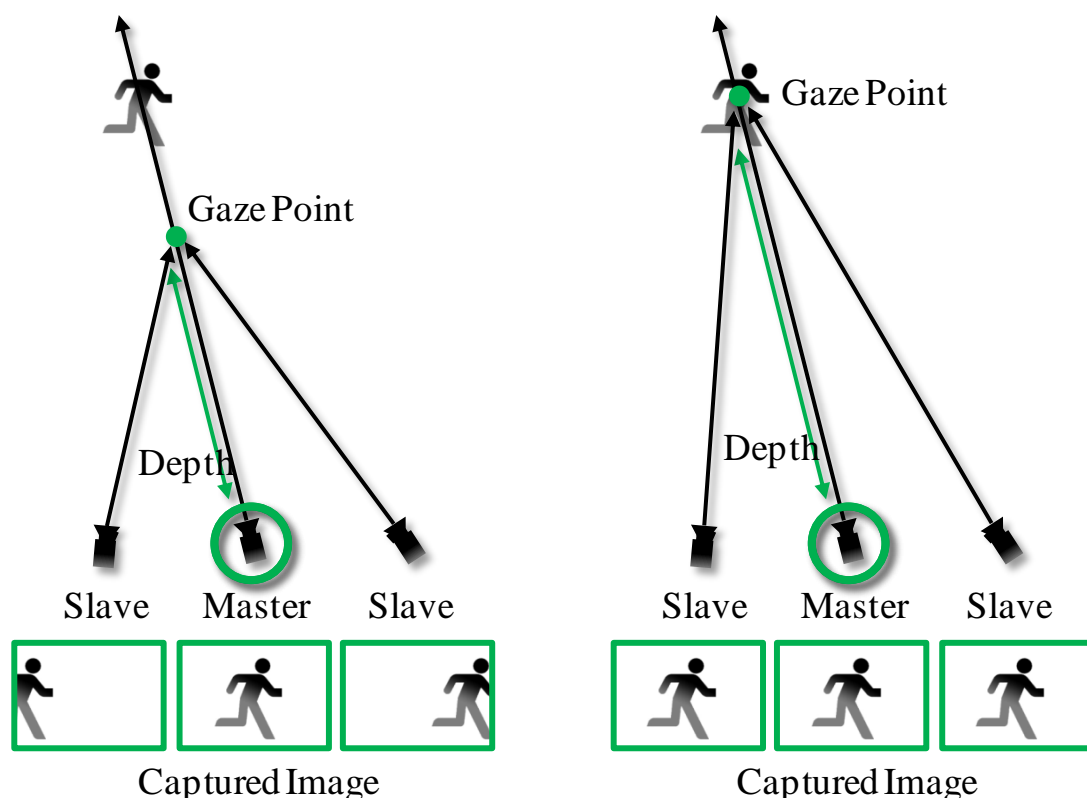


図 3.3 デプスの決定
 (a) デプスが正しくない時
 (b) デプスが正しいとき

3.3 多視点映像表現

3.3.1 被写体の決定

機械的方向制御を行い撮影した多視点映像と撮影時の焦点距離を収録する．収録完了後，ユーザーは収録した多視点映像から多視点映像表現を行うフレーム F と被写体を決定する．多視点映像のうち任意の2台のカメラ Cam_a , Cam_b の映像においてその被写体の位置の画像座標 (u_a, v_a) および (u_b, v_b) を取得する．

3.3.2 カメラキャリブレーション

多視点映像表現を行うフレーム F の多視点映像と収録した撮影時の焦点距離を用いて弱校正カメラキャリブレーション[37]を行い，カメラパラメーター（回転行列 R'_{0n} ，並進ベクトル t'_{0n} ，焦点距離 f_{0n} ）を取得する．世界座標 x から画像座標 (u_n, v_n) への変換式を式(3.12)(3.13)(3.14)に示す．

$$\omega \begin{bmatrix} u_n \\ v_n \\ 1 \end{bmatrix} = \mathbf{S}_n \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (3.12) \quad \mathbf{S}_n = \mathbf{A}'_{0n} [\mathbf{R}'_{0n} \quad \mathbf{t}'_{0n}] \quad (3.13) \quad \mathbf{A}'_{0n} = \begin{bmatrix} f_{0n} & 0 & C_x \\ 0 & f_{0n} & C_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.14)$$

ここで、 ω は画像距離、 f_{0n} は焦点距離、 (C_x, C_y) は画像中心を表す。

3.3.3 仮想的カメラ制御

3.3.1 で選択した被写体の位置に注視点を \mathbf{g}' として再指定し、射影変換を用いてカメラを \mathbf{g}' へ仮想的に再方向制御する。 \mathbf{g}' を式(3.15)(3.16)(3.17)より算出する。

$$\mathbf{g}' = \mathbf{M}^+ \mathbf{b} \quad (3.15)$$

$$\mathbf{M} = \begin{bmatrix} S_{a31}u_a - S_{a11} & S_{a32}u_a - S_{a12} & S_{a33}u_a - S_{a13} \\ S_{a31}v_a - S_{a21} & S_{a32}v_a - S_{a22} & S_{a33}v_a - S_{a23} \\ S_{b31}u_b - S_{b11} & S_{b32}u_b - S_{b12} & S_{b33}u_b - S_{b13} \\ S_{b31}v_b - S_{b21} & S_{b32}v_b - S_{b22} & S_{b33}v_b - S_{b23} \end{bmatrix} \quad (3.16) \quad \mathbf{b} = \begin{bmatrix} S_{a14} - S_{a34}u_a \\ S_{a24} - S_{a34}v_a \\ S_{b14} - S_{b34}u_b \\ S_{b24} - S_{b34}v_b \end{bmatrix} \quad (3.17)$$

ここで、 S_{nij} は \mathbf{S}_n の i 行 j 列目の要素、 \mathbf{M}^+ は \mathbf{M} の一般化逆行列を表す。カメラを仮想的に方向制御した際にカメラを地面と水平に保つため、地面との垂直軸を算出する。パン、チルトの2軸のロボットカメラのチルト軸は、地面と水平になっているため、垂直軸は任意の2台のカメラ Cam_s 、 Cam_t におけるチルト軸の外積によって算出できる。垂直軸 \mathbf{v} を式(3.18)(3.19)より算出する。

$$\mathbf{R}'_{0n} = \begin{bmatrix} r_{n1} \\ r_{n2} \\ r_{n3} \end{bmatrix} \quad (3.18) \quad \mathbf{v} = \frac{\mathbf{r}_{s1} \times \mathbf{r}_{t1}}{\|\mathbf{r}_{s1} \times \mathbf{r}_{t1}\|} \quad (3.19)$$

カメラを \mathbf{g}' へ仮想的に方向制御した時の回転行列 \mathbf{R}'_n を式(3.20)(3.21)(3.22)(3.23)より算出する。

$$\mathbf{R}'_n = \begin{bmatrix} \mathbf{e}_{nx}^T \\ \mathbf{e}_{ny}^T \\ \mathbf{e}_{nz}^T \end{bmatrix} \quad (3.20)$$

$$\mathbf{e}_{nz} = \frac{\mathbf{g}' + \mathbf{R}'_{0n}{}^{-1} \mathbf{t}'_{0n}}{\|\mathbf{g}' + \mathbf{R}'_{0n}{}^{-1} \mathbf{t}'_{0n}\|} \quad (3.21) \quad \mathbf{e}_{nx} = \frac{\mathbf{e}_{nz} \times \mathbf{v}}{\|\mathbf{e}_{nz} \times \mathbf{v}\|} \quad (3.22) \quad \mathbf{e}_{ny} = \frac{\mathbf{e}_{nz} \times \mathbf{e}_{nx}}{\|\mathbf{e}_{nz} \times \mathbf{e}_{nx}\|} \quad (3.23)$$

ここで、 \mathbf{e}_{nx} 、 \mathbf{e}_{ny} 、 \mathbf{e}_{nz} は、それぞれカメラのチルト軸、パン軸、ロール軸を表す。カメラを \mathbf{g}' へ仮想的に方向制御した際、すべてのカメラの被写体の大きさをマスターカメラ Cam_M の被写体の大きさに揃える。式(3.24)(3.25)より、各カメラの焦点距離を、カメラから注視点までの距離に応じて調整し、内部行列 \mathbf{A}'_n を算出する。

$$f_n = k \cdot f_{0M} \cdot \frac{\|\mathbf{g}' + \mathbf{R}'_{0n}{}^{-1} \mathbf{t}'_{0n}\|}{\|\mathbf{g}' + \mathbf{R}'_{0M}{}^{-1} \mathbf{t}'_{0M}\|} \quad (3.24) \quad \mathbf{A}'_n = \begin{bmatrix} f_n & 0 & C_x \\ 0 & f_n & C_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.25)$$

ここで、 k は射影変換後のズーム率である．カメラを \mathbf{g}' へ仮想的に方向制御するために、式(3.26)(3.27)より射影変換を行う． (u_n, v_n) は、射影変換前の画素の画像座標、 (u'_n, v'_n) は、射影変換後の画像座標を表す．

$$H_n = A'_n R'_n R'^{-1}_{0n} A'^{-1}_{0n} \quad (3.26)$$

$$\begin{bmatrix} u'_n \\ v'_n \\ 1 \end{bmatrix} = H_n \begin{bmatrix} u_n \\ v_n \\ 1 \end{bmatrix} \quad (3.27)$$

3.3.4 多視点映像表現の映像生成

ユーザーは専用のソフトウェアを用いて図 3.4 のように多視点映像シーケンス中の複数のフレーム（キーフレーム） F を選択し、各キーフレームにおいて、注視点と射影変換後のズーム率、どの視点から見るかという情報を入力する．選択されたキーフレームで弱校正カメラキャリブレーションを用いてカメラパラメータを求め、射影変換行列 H を生成する．キーフレームにおける被写体をどの視点から見るかという情報からカメラパスを決定し、カメラパス上のキーフレーム以外のフレームにおける射影変換行列をキーフレームの射影変換行列を用いて補間する．ここで補間はキーフレームとの距離に応じて線形補間で行われる．これによりキーフレーム間でも滑らかに視点がつながるようにカメラを仮想的に方向制御することができる．最後に、カメラパス上の画像に射影変換を施し、画像を切り替えることで多視点映像表現を行う．

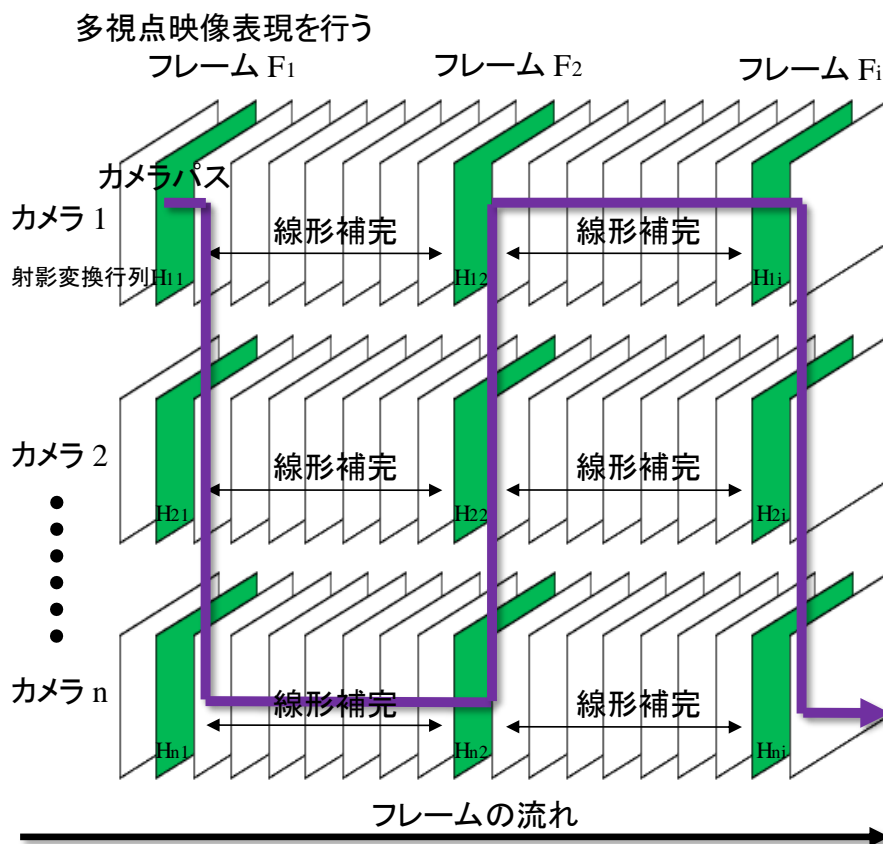


図 3.4 多視点映像表現の映像編集方法

3.4 システム

多視点ロボットカメラは、主に多視点ロボットカメラ、操作インターフェース、多視点映像表現処理装置で構成されている。多視点ロボットカメラは、小型ハイビジョンカメラを電動雲台上に設置し、ボードコンピューターを搭載した複数台のロボットカメラである。(図 3.5) 外部信号によってパン、チルト、ズーム、フォーカス、アイリスの遠隔操作が可能である。電動雲台のパン、チルト制御モーターにはサーボモーターを導入しており、高精度で方向制御を行うことができる。操作インターフェースは、マスターカメラのパン、チルト、デプスとすべてのカメラのズームを操作するもので、バーチャルリアリティ用の操作インターフェースを改修したものである。(図 3.6) フォーカスデマンドを用いてデプスを操作し、フォーカスはデプスから自動調整される。そのため、1人のカメラマンによる従来のテレビカメラとほぼ同じ操作で多視点ロボットカメラを操作できる。多視点映像表現処理装置は、収録装置から取得した撮影映像に仮想的方向制御を施し、多視点映像表現の映像を生成するワークステーションである。(図 3.7) 準リアルタイムで多視点映像表現の映像を生成するため、ユーザーによる操作(多視点映像表現を行う時間と被写体の決定)を素早く行うためのソフトウェアや、高速で射影変換を行うための GPU が搭載されている。また、このソフトウェアによって多視点映像表現における視点の切替え方向や、再生スピード、デジタルズーム率などを任意に変更可能である。



図 3.5 ロボットカメラ



図 3.6 操作インターフェース



図 3.7 多視点映像表現処理装置

図 3.8 にシステムの系統図を示す。まず、操作インターフェースは、操作時のパン、チルト、ズーム、デプスのデータをマスターカメラへ送信する。次に、受信データに基づき制御されたマスターカメラは、パン、チルト、ズーム、デプスのデータを注視点算出 PC へ送信する。注視点算出 PC は、パン、チルト、デプスのデータから算出した注視点の世界座標とズームのデータをスレーブカメラへ送信する。スレーブカメラは受信データに基づき方向制御される。多視点ロボットカメラの撮影映像と撮影時の焦点距離は、収録装置に収録される。多視点映像表現処理装置は、ユーザーインターフェース、弱校正カメラキャリブレーション部、射影変換部、多視点映像表現の映像生成部によって構成される。ユーザーインターフェースを用いて多視点映像表現を行うフレームと被写体を決定し、フレームのデータを弱校正カメラキャリブレーション部と射影変換部へ、被写体の位置の画像座標を射影変換部へ送信する。弱校正カメラキャリブレーション部は、受信したフレーム番号の多視点画像と焦点距離を収録機から取得し、弱校正カメラキャリブレーションを行い、カメラパラメータを射影変換部へ送信する。射影変換部では受信したフレーム番号の多視点画像を収録機から取得し、受信したカメラパラメータを用いて射影変換を行う。多視点映像表現の映像生成部は、射影変換画像を編集し、多視点映像表現の映像を生成する。

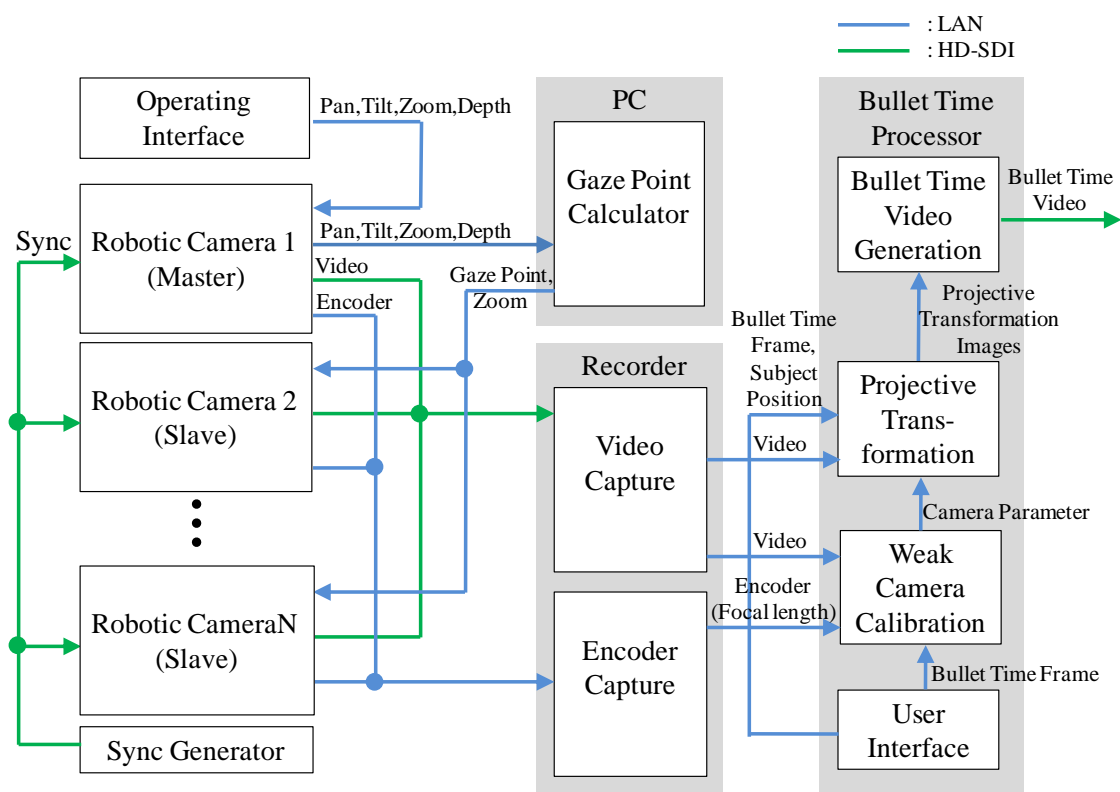


図 3.8 システム系統図

3.5 実験

3.5.1 撮影実験

本システムの性能を評価するために、バレーボールにおける 3 次元空間をダイナミックに移動する選手や 3 次元空間中に広く点在する選手を撮影し、多視点映像表現の映像を生成する実験を行った。図 3.9 のように 9 台のロボットカメラを 90°円弧上に配置して撮影を行った。カメラから被写体までの距離は約 11m、カメ

ラ間隔は 2.3m であった。生成した多視点映像表現の映像を図 3.10 に示す。また、体操、バスケットボール、柔道においても同様の実験を行い、生成した多視点映像表現の映像を図 3.11、図 3.12、図 3.13 に示す。

実験の結果、バレーボールにおける 3 次元空間をダイナミックに移動する選手や 3 次元空間中に広く点在する選手の多視点映像表現を撮影完了から約 1 分で生成できた。また、システムの事前準備でのカメラキャリブレーションを約 5 分で完了できた。体操、バスケットボール、柔道においても同様に多視点映像表現を生成することができた。



図 3.9 カメラ配置

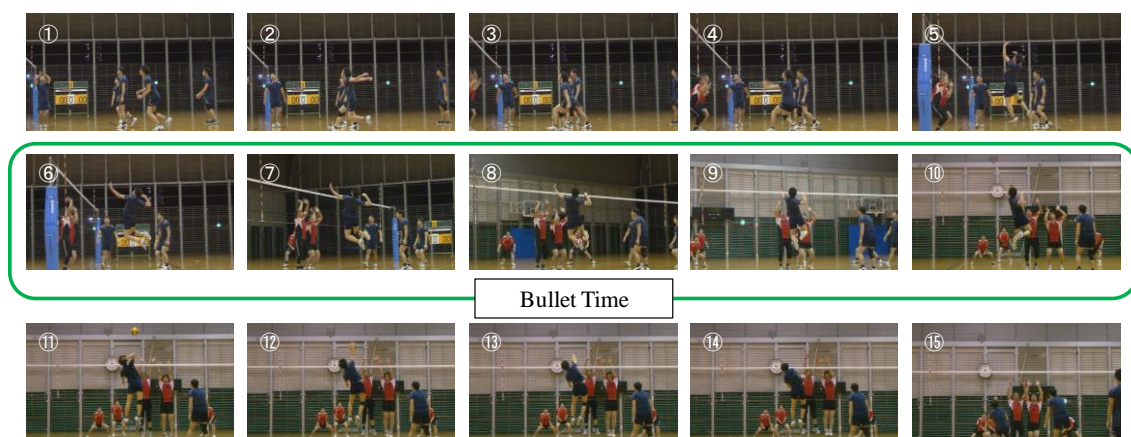


図 3.10 バレーボールの多視点映像表現（番号はフレームの流れ）

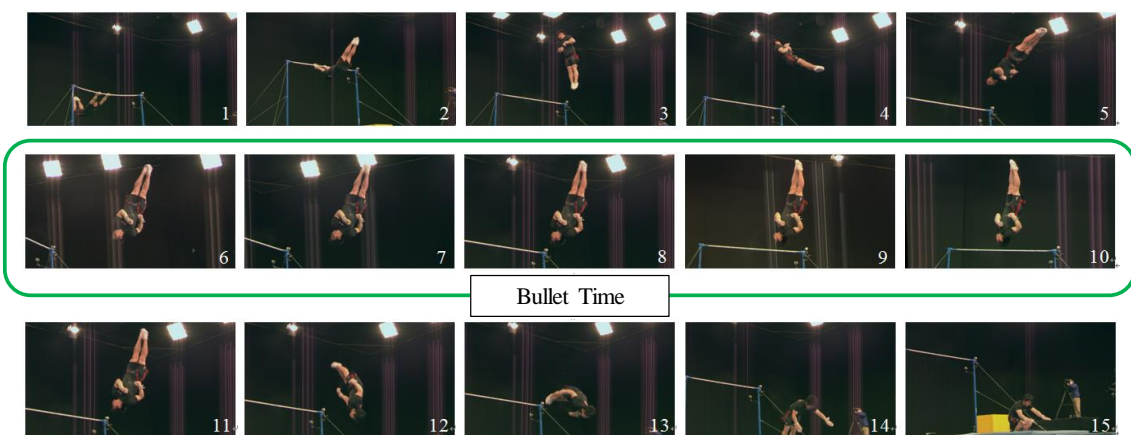


図 3.11 体操の多視点映像表現

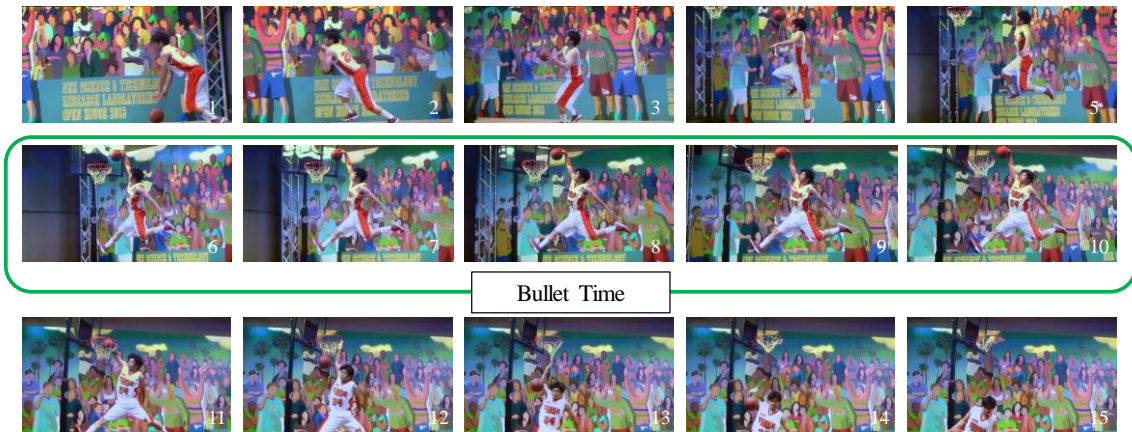


図 3.12 バスケットボールの多視点映像表現



図 3.13 柔道の多視点映像表現

3.5.2 番組利用

本システムを 2013 年 11 月に生中継された「2013 年 NHK 杯フィギュアスケート選手権大会」の競技のリプレイで番組利用した。図 3.14 の観客席に図 3.15 のように 9 台のロボットカメラを直線状に配置して撮影を行った。カメラから被写体までの距離は約 30m, カメラ間隔は約 2m であった。放送された多視点映像表現を図 3.16 に示す。

実験の結果, 広いリング上をダイナミックに移動する選手のジャンプなどの多視点映像表現を準リアルタイムで生成し, 競技直後のリプレイで放送することができた。選手のジャンプにおける蹴り出しの姿勢, 空中の姿勢, 着地の姿勢などをさまざまな視点から分かりやすく表現することができた。

また, 本システムはフィギュアスケートを含む表 3.1 に示す番組で利用された。スポーツ以外にドラマや教育番組など幅広い番組で利用された。放送された多視点映像表現を図 3.17 から図 3.21 に示す。

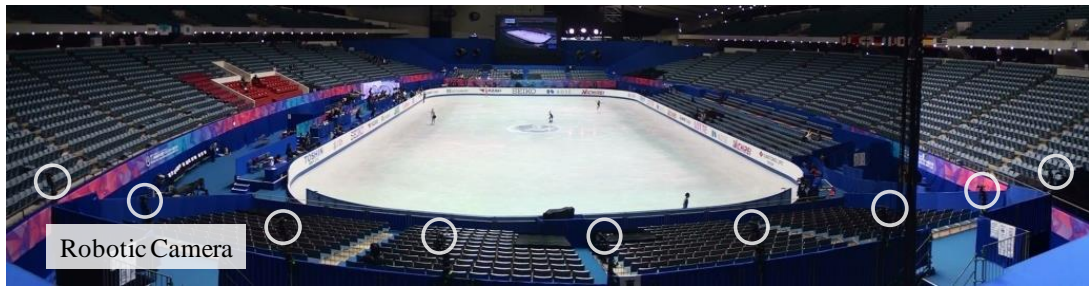


図 3.14 会場の様子

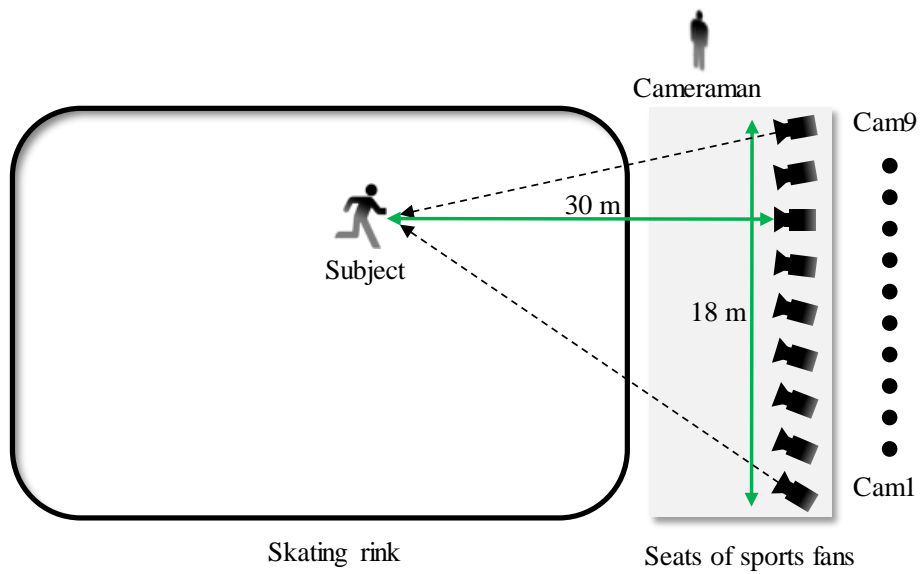


図 3.15 カメラ配置

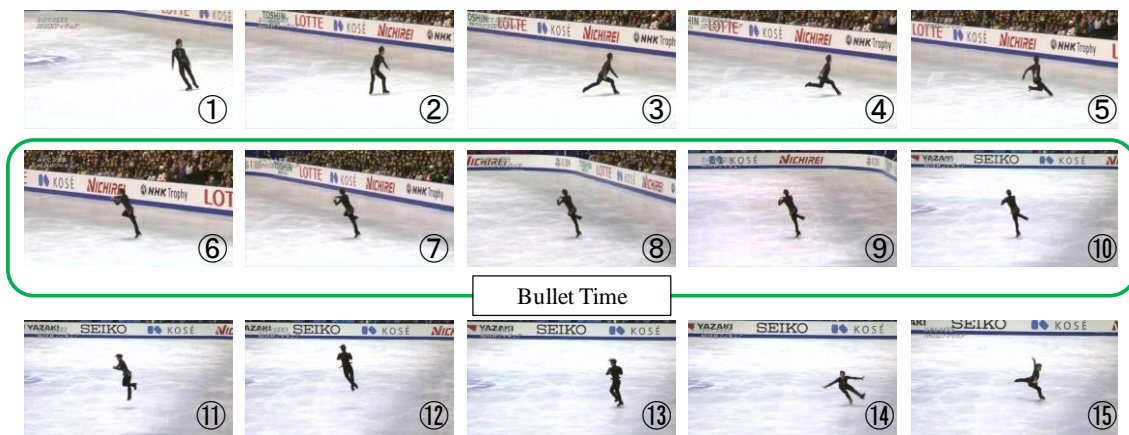


図 3.16 NHK 杯フィギュアスケート選手権大会 2013 の多視点映像表現
(番号はフレームの流れ)

表 3.1 システムを利用した番組

放送年	番組名
2013	NHK 杯体操選手権 2013
	NHK 杯フィギュアスケート選手権大会 2013
2014	NHK 杯フィギュアスケート選手権大会 2014
2016	連続テレビ小説「とと姉ちゃん」
	NHK スペシャル ミラクルボディー（柔道）
2017	NHK 学生ロボコン 2017
	ABU ロボコン世界大会 2017

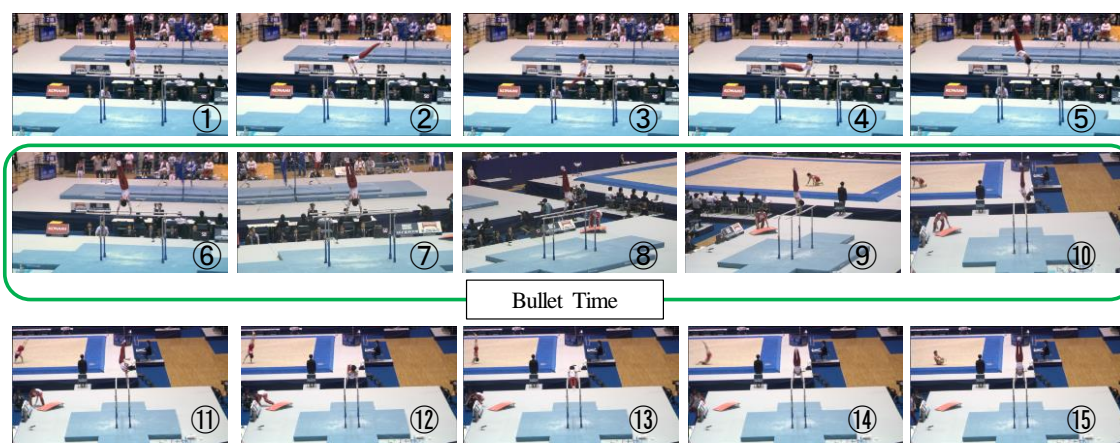


図 3.17 NHK 杯体操選手権 2013 の多視点映像表現

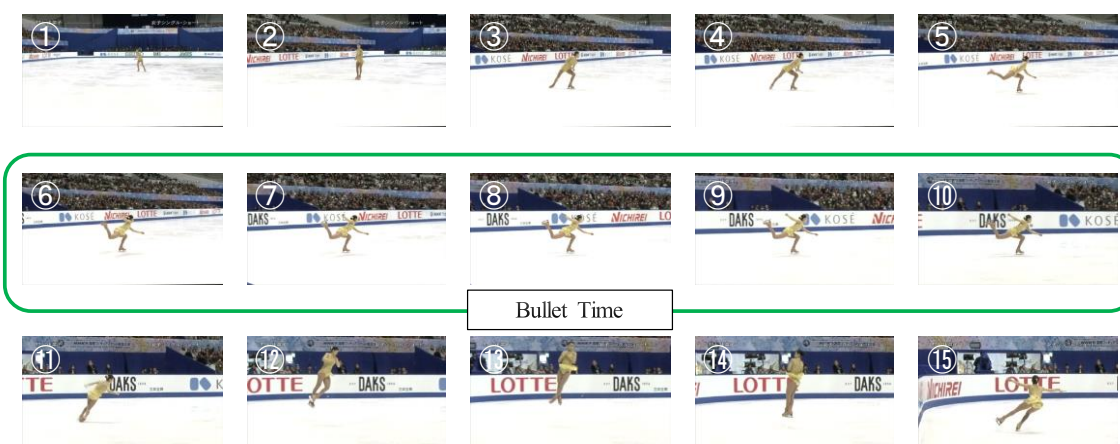


図 3.18 NHK 杯フィギュアスケート選手権大会 2014 の多視点映像表現

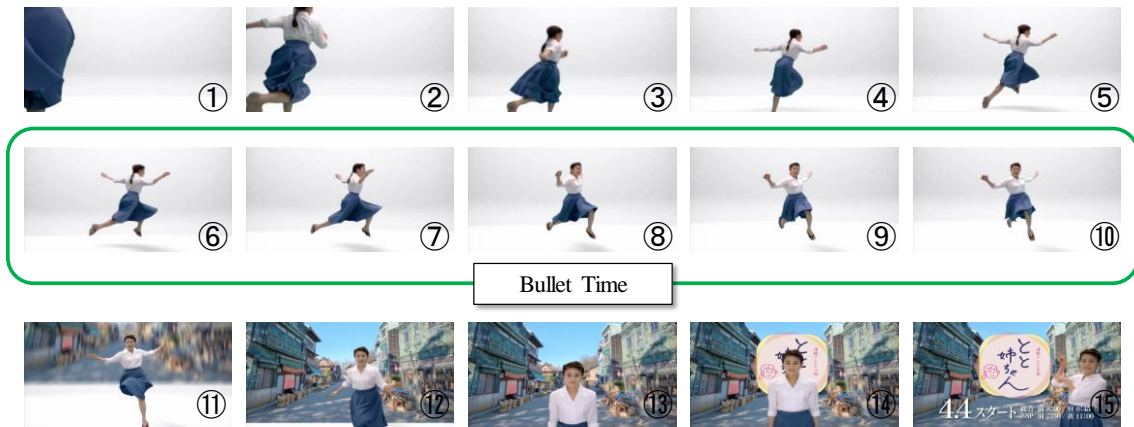


図 3.19 連続テレビ小説「とと姉ちゃん」の多視点映像表現



図 3.20 NHK 学生ロボコン 2017 の多視点映像表現



図 3.21 ABU ロボコン世界大会 2017 の多視点映像表現

3.5 考察

本システムが、3.1 で挙げた要求条件を満たすか検証した。要求条件 1 は、滑らかに視点が切り替わる多視点映像表現を実現するため、3次元空間をダイナミックに移動する被写体に対しても、多視点カメラを高い精度で方向制御することが可能であることであった。本システムは、多視点ロボットカメラの機械的方向制御および仮的方向制御により、バレーボールやフィギュアスケートなどにおける3次元空間をダイナミックに移動する選手に対して多視点カメラを高い精度で方向制御することが可能である。方向制御精度を評価するため、図 3.10 の多視点映像表現の映像における注視点位置の画像座標と画像中心位置の画像座標のユークリッド距離を計測した結果、9台のカメラの平均は1.4ピクセルであった。多視点映像表現を生成するうえで問題ない方向制御精度であり、高い精度で方向制御できると言うことができる。要求条件 2 は、生放送のスポーツ中継において、競技中のシーンの多視点映像表現を競技直後のリプレイで放送可能な処理時間であることであった。本システムは、多視点映像表現を撮影完了から約1分で生成することができ、「2013年NHK杯国際フィギュアスケート競技大会」の競技直後のリプレイで番組利用した。要求条件 3 は、システムの事前準備でのカメラキャリブレーションが短時間で完了することであった。本システムは、事前準備でのカメラキャリブレーションを約5分で完了することが可能である。これらにより、本システムが、すべての要求条件を満たすことが分かった。

本システムと他の多視点映像表現技術を比較した。現在、放送に使用されている

主な技術に Intel 社の「Free D」と 4D Replay 社の「4D Replay」がある。「Free D」は、固定型の多視点カメラを 38 台から 72 台配置し、撮影映像を用いて仮想空間に被写体の 3 次元モデルを生成後、仮想カメラで任意の視点から被写体を撮影することで多視点映像表現を行う。仮想カメラを用いるため、視点を任意の位置に設定できると共に、視点を滑らかに切り替えることができる。一方で、図 3.22 のように仮想カメラで被写体をズームアップした場合、被写体を構成する 3 次元モデルの点群が顕著に視認され、映像が破綻するため、ズームアップすることは困難である。「4D Replay」は、固定型の多視点カメラを 100 台から 120 台配置し、撮影映像に画像処理を施し、カメラの並び順に切り替えることで多視点映像表現を行う。「Free D」のように 3 次元モデルを生成するのではなく、基本的に撮影映像を直接切り替えるため、加工のない、よりリアルな映像表現となる。一方で、視点の切り替えはカメラの並んでいる位置に限定され、視点の切り替えの滑らかさは設置するカメラ台数に依存する。また、例えばサッカーなどでダイナミックに移動する被写体や広い空間に点在する被写体をズームアップする場合、図 3.23 のように広い画角で撮影し、撮影映像にデジタルズームを行う必要があるため、解像度が低下し、画質が劣化する。本システムは、撮影に多視点ロボットカメラを使用し、撮影映像に射影変換を施し、カメラの並び順に切り替えることで多視点映像表現を行う。「4D Replay」と同様に、加工のない、よりリアルな映像表現となるが、視点の切り替えはカメラの並んでいる位置に限定され、視点の切り替えの滑らかさは設置するカメラ台数に依存する。一方で、被写体をズームアップする場合、図 3.24 のように多視点ロボットカメラのパン、チルト、ズーム制御によって光学的にズームアップを行うことができ、解像度の低下および画質の劣化を抑制することができる。スポーツ中継などでは、被写体の手元や足元などの細かな所作をズームアップして視聴者に分かりやすく伝えることにニーズがある。本システムは、上述した他の多視点映像表現技術と比較し、被写体をズームして分かりやすく表現できる点において優位性があることが分かった。

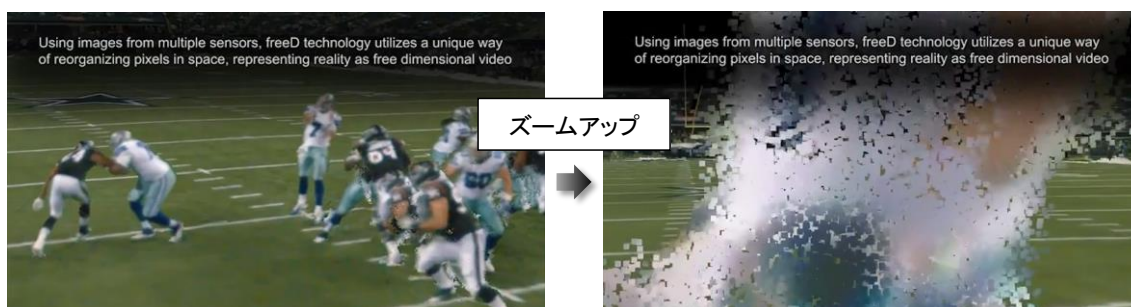


図 3.22 Free D による多視点映像表現



図 3.23 4D Replay による多視点映像表現



図 3.24 多視点ロボットカメラによる多視点映像表現

3.6 まとめ

多視点ロボットカメラを用いた多視点映像表現技術を提案した。撮影実験や番組利用を通じて本システムが、3次元空間をダイナミックに移動する被写体や3次元空間中に広く点在する被写体の多視点映像表現を実現できることを確認した。また、放送現場でのシステムの事前準備が短時間で完了し、準リアルタイムで多視点映像表現を生成可能なシステムであることが分かった。これらより、本システムが、従来の多視点映像表現システムと比較し、より汎用的かつ実用的なシステムであることを確認した。

第4章 多視点カメラからのインテグラル立体映像生成手法

4.1 はじめに

従来のインテグラル立体カメラには、遠方の被写体を撮影する場合、視差を取得するために立体像の奥行き位置を調整する奥行き制御レンズが大きくなるという課題があった。例えば、スポーツ中継などの番組では、被写体の近くにカメラを置くことは難しいため、遠方からズームして被写体を撮影する。インテグラル立体カメラで撮影する場合、巨大な奥行き制御レンズを設置したカメラで被写体を撮影することになるが、番組制作でそのようなカメラの運用は現実的に困難である。

そこで、本章における研究目的は、遠方の被写体のインテグラル立体映像を生成し、番組で扱うようなシーンのインテグラル立体コンテンツを生成することとする。この目的を達成するために、多視点カメラを用いてインテグラル立体映像を生成する手法を提案する。この手法では、多視点カメラの撮影映像から3次元モデルを生成することで、仮想空間で被写体が発する光線を取得し、インテグラル立体映像を生成する。多視点カメラ間のベースラインを調整することで遠方の被写体を撮影することができる。一方、疎に配置された多視点カメラから距離推定を用いて3次元モデルを生成するうえで、距離推定におけるエラーによって3次元モデルの品質が低下するという課題がある。そこで距離推定におけるエラーを抑制するため、多視点カメラを用いた3次元モデル生成手法および距離推定における信頼度を用いた3次元モデル補正手法を提案する。そして、本手法を用いて実際に放送された相撲のシーンのインテグラル立体コンテンツを制作する。

4.2 提案手法

提案手法では、多視点カメラの映像から Belief Propagation(BP)を用いた距離推定により複数視点の3次元モデルを生成し、BPのエネルギー分布形状の尖度を用いて距離推定結果の信頼度を評価する。複数視点の3次元モデルのうち、信頼度が高く距離推定エラーが生じていない領域のみを抽出し、統合することで距離推定エラーの影響を抑制したインテグラル立体映像を生成する。提案手法の処理の流れを図4.1に示す。以下で各処理について説明する。

(1) 多視点映像の撮影

複数台のカメラを被写体に向けて配置し、全てのカメラに同期をかけ、被写体の多視点映像を撮影する。

(2) 距離推定による3次元モデルの生成

撮影した多視点映像からBPを用いた距離推定により、各視点の距離画像を生成し、3次元モデルを生成する。隣あう3台のカメラ映像をペアとし、3台のうち中央のカメラ映像に対して3次元モデルを生成する。詳細を4.2.1で述べる。

(3) 信頼度に基づく3次元モデルの補正

(2)の処理の段階では、3次元モデルに距離推定エラーが生じている。そこで、距離推定結果の信頼度を評価し、3次元モデルのうち信頼度が高く、エラーが生じていない領域を抽出することで3次元モデルを補正する。信頼度は、BPにおけるエネルギー分布形状の尖度より算出する。詳細を4.2.2で述べる。

(4) 3次元モデルからの要素画像群の生成と統合

(3)の処理で生成した各視点の3次元モデルから要素画像群を生成し，統合することで，距離推定エラーの影響を抑制したインテグラル立体映像を生成する．各視点の要素画像群を統合する処理では，各視点の3次元モデルにおいて前景の裏側などカメラに映らなかった領域や，(3)の処理で信頼度が低く，抽出されなかった領域を異なる視点の3次元モデルの情報で補完する．詳細を4.2.3で述べる．

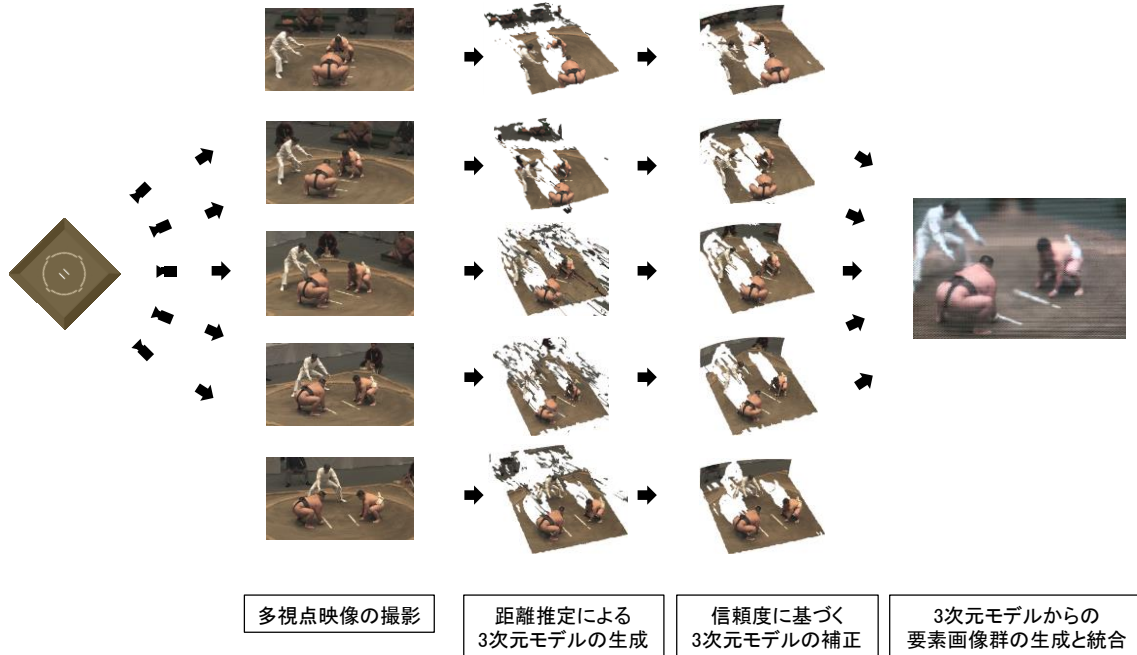


図 4.1 提案手法の処理の流れ

4.2.1 距離推定による3次元モデルの生成

撮影した多視点映像から BP を用いた距離推定により 3 次元モデルを生成する[38]. BP は、距離情報に MRF (Markov Random Field) モデルを仮定し、画像全体で大域的に最適化した距離値を推定する手法である. 大域的に最適化するために、隣接画素間でメッセージと呼ばれるエネルギーを繰り返し伝搬する. このエネルギーは、画素にどの距離値をラベルとして割り当てるかという確率を表しており、割り当てられる確率が高い距離値には低いエネルギーが、確率が低い距離値には高いエネルギーが蓄積される. 処理対象画素からメッセージを 4 近傍の画素に伝搬し、伝搬先の画素では、伝搬されたメッセージから新たなメッセージを生成し、さらに 4 近傍の画素に伝搬する. この伝搬を画素間で繰り返し行い、伝搬が完了後、エネルギーが最小となった距離値を各画素に割り当てる. このメッセージを生成するためには、距離推定の誤差エネルギーが必要であり、提案手法では DAISY を用いてこれを生成している[39]. DAISY は、画像の各画素における特徴量を記述する手法で、特徴量は、画素値の方向微分値に対してガウスクアーネルを畳み込むことで算出する. DAISY を用いることで、撮影画像内の被写体の向きに回転が生じるようなベースラインが長い多視点画像に対しても高い性能で距離推定を行うことができる. また、多視点画像間に明るさや色の差が生じた場合、対応点間の色情報に基づく距離推定手法を用いると推定精度が低下するが、DAISY は、色情報ではなく、輝度の勾配情報に基づいた手法であるため、推定精度の低下を軽減できる. 以下に、アルゴリズムについて述べる.

4.2.1.1 DAISY による特徴量の記述

DAISY によって画像の各画素における特徴量を求める. 図 4.2 に DAISY の記述子を示す. 特徴量の記述には、画素値の方向微分値を用いる. 処理対象画素から半径 r 画素の範囲に、 S 個の円領域を放射状に設定する. それぞれの円領域で、領域内の方向微分値から、 H 方向の特徴ベクトルである畳み込み方向マップを作製する. 畳み込み方向マップは式(4.1)より計算する.

$$G_o^s = G_\Sigma * \left(\frac{\partial I}{\partial o} \right)^+ \quad (4.1)$$

ここで、 G_Σ はガウスクアーネルによる畳み込み、 I は入力画像、 o は H 方向いずれかの方向、 $(a)^+$ は $\max(a, 0)$ を表し、 a は実数である. S 個の各点に対し H 方向の畳み込み方向マップを作製するので、 $S \times H$ 次元の特徴ベクトルを特徴量として得ることができる. 各パラメータの初期設定は $r=15, S=25, H=8$ である. 上記のようにして、各カメラ映像の各画素に対して特徴量を求める.

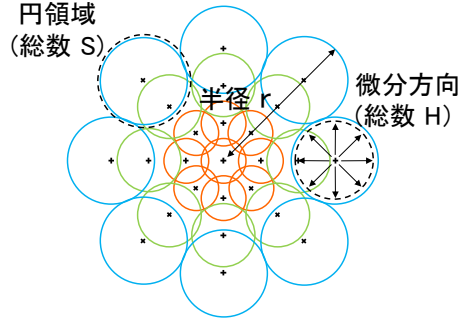


図 4.2 DAISY 記述子

4.2.1.2 Belief Propagation を用いた距離推定

DAISY より求めた特徴量から初期エネルギーを生成し，BP による距離推定を行い，距離画像を生成する．距離画像は，多視点映像のうち，隣合う 3 台のカメラ映像を用いて生成する．ここで，3 台のカメラを左，中央，右カメラとし，左と中央カメラを左ペア，中央と右カメラを右ペアとする．まず，左右のペア画像でそれぞれ BP を用いて距離推定を行い，中央カメラに対する 2 枚の距離画像を生成する．BP を用いた距離推定では，初期エネルギーとして，メッセージの生成に必要なデータ項 D およびスムーズ項 V を生成する．データ項とはステレオ画像の画素間の誤差に関するエネルギー，スムーズ項とは隣接画素間の距離情報の滑らかさに関するエネルギーである．このデータ項の生成に，DAISY 特徴量を用いる．データ項 D の生成式を式(4.2),(4.3)に，スムーズ項 V の生成式を式(4.4)に示す．

$$D_p(f_p) = \begin{cases} \lambda \Delta_{f_p} & \text{if } \Delta_{f_p} \leq T_{data} \\ \lambda T_{data} & \text{if } \Delta_{f_p} > T_{data} \end{cases} \quad (4.2)$$

$$\Delta_{f_p} = |G_o^\Sigma(p) - G_o^{\prime\Sigma}(p+d_p)| \quad (4.3)$$

$$V(f_p - f_q) = \begin{cases} |f_p - f_q| & \text{if } |f_p - f_q| \leq T_{smooth} \\ T_{smooth} & \text{if } |f_p - f_q| > T_{smooth} \end{cases} \quad (4.4)$$

ここで， p は処理対象画素， q は p の 4 近傍の画素のうちの 1 つの画素で，メッセージの伝搬先の画素である． f は距離値のラベル， d は f に対応する視差値であり，これは距離値とカメラ間のベースラインおよびカメラパラメーターより算出する． G_o^Σ は基本画像である中央カメラ画像の画素の DAISY 特徴量， $G_o^{\prime\Sigma}$ は参照画像で，左ペアであれば左カメラ画像，右ペアであれば右カメラ画像の画素の DAISY 特徴量， λ はデータ項の重み係数， T_{data} ， T_{smooth} はデータ項およびスムーズ項の制限値を表す．次に，メッセージを式(4.5)より生成する．

$$m_{p \rightarrow q}^t(f_p) = \min_{f_p} \left(V(f_p - f_q) + D_p(f_p) + \sum_{s \in N(p)/q} m_{s \rightarrow p}^{t-1}(f_p) \right) \quad (4.5)$$

ここで、 m はメッセージで初期値は 0、 t は反復回数、 $N(p)/q$ は q 以外の p の 4 近傍の画素集合、 s はその集合の要素画素を表す。式(5)は漸化式になっており、繰り返しメッセージを更新し、伝搬する。

メッセージの更新および伝搬が T 回完了した段階で、 q の f に関するエネルギー b を式(4.6)より生成する。

$$b_q(f_q) = D_q(f_q) + \sum_{p \in N(q)} m_{p \rightarrow q}^T(f_q) \quad (4.6)$$

エネルギー b が最小になる f を各画素の距離値とし、距離画像を生成する。上記のようにして、左右のペア画像で中央カメラに対する 2 枚の距離画像を生成する。

次に、この 2 枚の距離画像には、オクルージョンによる距離推定エラーが含まれているため、双方の距離情報で補間して、オクルージョンを埋め、1 枚の距離画像を生成する。まず、エネルギー b のエネルギー分布より、式(4.7)のようにオクルージョンの評価値 O を定義する。

$$O = (\max(b) - \min(b)) / \max(b) \quad (4.7)$$

ここで、 $\max(b)$ は、処理対象画素の b のエネルギー分布における最大値、 $\min(b)$ は最小値を表す。 O はエネルギー分布における最大値と最小値の落差に相当し、オクルージョンによりエラーが生じている画素ほど低い値となる特徴がある。この特徴を利用し、生成した 2 枚の距離画像の各画素で O を求め、双方の同座標の画素で O を比較し、値が大きい方の距離値を採用することで、オクルージョンを埋め、1 枚の距離画像を生成する。そして、同様に、各視点のカメラ映像に対して距離画像を生成する。距離画像から、各画素の 3 次元座標を求め、それらを頂点とし、頂点間にポリゴンを張ることで 3 次元モデルを生成する。

4.2.2 信頼度に基づく 3 次元モデルの補正

4.2.1 の処理で生成した 3 次元モデルには、3 台のカメラのうち中央カメラには映っているが、左右のカメラには映っていない領域のオクルージョンや、テクスチャの不足が原因で生じる距離推定エラーが含まれている。この 3 次元モデルをインテグラル立体表示した場合、エラーが立体的に飛び出る場合があり目立つ。そこで、距離推定結果の信頼度に基づき、3 次元モデルの補正を行う。

まず、距離推定結果の信頼度を BP におけるエネルギー b の分布形状の尖度によって算出する。尖度とは、統計学における分布の‘とがり’具合のことをいう。次に、信頼度に基づき、3 次元モデルを補正する。距離推定の範囲のうち、3 次元モデルとして復元する範囲を設定し、その範囲より奥の領域は、壁状の 3 次元モデルにする。これを、背景形状と呼ぶ。背景

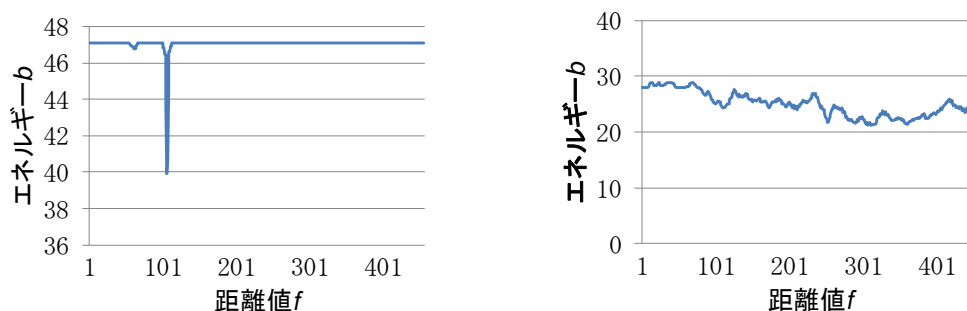
形状を設けることで、無限遠やそれに近い領域に対する距離推定の計算を省くことができる。信頼度が高い領域のみを抽出し、それ以外の信頼度が低く、エラーが生じた領域を背景形状とする。背景形状とすることで、4.2.3 の処理により、他の視点の3次元モデルの情報でこれらの領域を補間することができる。以下に、アルゴリズムについて述べる。

4.2.2.1 エネルギー分布の尖度による信頼度の評価

エネルギー b のエネルギー分布の尖度により、式(4.8)のように距離推定結果の信頼度 R を定義する。

$$R = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_f \left(\frac{b_f - \bar{b}}{s} \right)^4 - 3 \frac{(n-1)^2}{(n-2)(n-3)} \quad (4.8)$$

ここで f は距離値のラベル、 n は f の数、 b_f は f のエネルギー、 \bar{b} は b_f の平均値、 s は b_f の標準偏差を表す。 R はオクルージョンやテクスチャの不足など距離推定エラーが生じた画素ほど低い値となる特徴がある。例えば、距離が正確に求まった画素と距離推定エラーが生じた画素があった時、それぞれのエネルギー分布は図4.3のようになり、前者では、鋭くピークが立ち尖度は160.78、後者では、ピークが立つことはなく尖度は-0.96と R は後者の方が低くなる。また、距離画像における各画素で R を求め、 R の大きさを輝度値の高さで視覚的に表現すると図4.4(a)のようになり、生成した距離画像において距離推定エラーを赤い領域で示すと図4.4(b)のようになる。この赤い領域は手動で設定した。図4.4(a)の輝度が低い領域と、図4.4(b)におけるテクスチャが不足した土俵後方の床や、オクルージョンとなった右力士の手足間の領域の距離推定エラーはほぼ一致する。4.2.1の処理で生成した距離画像の各画素について R を求める。



(a) (b)

図4.3 エネルギー分布の例

- (a) 距離が正確に求まった画素 (尖度 : 160.78)
- (b) 距離推定エラーとなった画素 (尖度 : -0.96)



(a) (b)
 図4.4 エラー領域と信頼度の比較
 (a) 信頼度 R を視覚化した画像
 (b) 距離推定エラー領域（赤い領域）

4.2.2.2 3次元モデルの補正

信頼度 R に関する閾値を設定する。信頼度 R が閾値より高かった領域を抽出し、それ以外の領域を距離推定エラーと判定する。距離推定エラーと判定された領域は、背景形状にする。この時、力士や土俵に距離推定エラーと判定された領域があった時、その領域も背景形状となるため、力士や土俵に欠損が生じた状態になる。この欠損が生じた領域は、4.2.3 の3次元モデルからの要素画像群の生成と統合処理により、他の視点の3次元モデルの情報や、近傍の領域の情報で補間される。図4.5に補正を行った3次元モデルを示す。次の4.2.3の処理では、各視点のカメラ映像に映る共通の領域で、各視点の3次元モデルの情報を要素画像群上で比較する。そのため、各視点の3次元モデルのうち、任意のものを選択し、その世界座標系が基準になるように、全ての3次元モデルの座標を変換する。最終的に、ここで選択した3次元モデルの視点のインテグラル立体映像が生成される。



(a) (b)
 図4.5 信頼度 R による3次元モデルの補正
 (a) 補正なし (b) 補正あり

4.2.3 3次元モデルからの要素画像群の生成と統合

4.2.2の処理で生成した各視点の3次元モデルから要素画像群を生成し、統合することで、インテグラル立体ディスプレイに立体像を表示する。1視点の3次元モデルからインテグラル立体映像を生成すると、図4.6のように、力士の後方や、土俵上などに立体像が欠落した領域が生じる。これは、前景に隠れた領域や、距離推定エラーと判定された領域で、3次元モデル形状が復元されなかった領域である。そこで、各視点の3次元モデルから要素画像群を生成し、欠落した領域を補完、内挿するようにそれらを統合して、インテグラル立体ディスプレイに表示することで、欠落のない立体像を生成する。以下に、生成手法について述べる。



図 4.6 立体像の欠落（再撮）

4.2.3.1 3次元モデルからの要素画像の生成

図4.7のように計算機内の仮想空間に、3次元モデルを、仮想ディスプレイや仮想レンズアレイとともに配置する。そして、仮想ディスプレイ上の対象画素と近傍の仮想レンズの光学主点を結ぶ光線を追跡し、光線と3次元モデルとの交点の色情報を対象画素に割り当てることで、要素画像群を生成する。3次元モデル形状が復元されておらず、光線と3次元モデルの交点が存在しなかった場合、色情報は対象画素に割り当てられない。実装にあたっては、光線を1本ずつ追跡する代わりに、図4.7のように3次元モデルを斜投影して得られる斜投影画像を用いることで、同方向の光線を一括して取得する。この斜投影画像を要素画像の画素数だけ取得し、要素画像群の配列に再構成する。上記の処理によって、各視点の3次元モデルを変換し、複数の要素画像群を生成する。

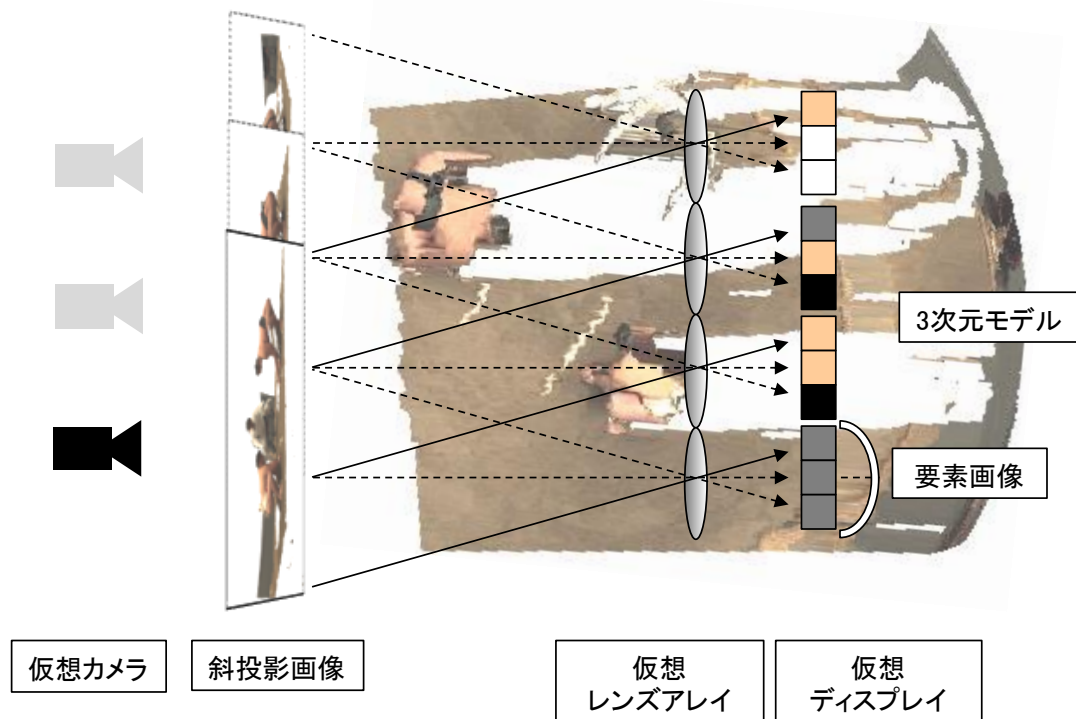


図 4.7 3次元モデルからのインテグラル画像の生成

4.2.3.2 要素画像群の統合

3.2.3.1 で生成した複数の要素画像群における同座標の画素で、色情報が割り当てられている画素数を調べる。3 つ以上の画素があった場合、それらの色情報の中間値を採用する。色が割り当てられた画素がなかった場合、近傍画素の色情報で内挿する。ここで、インテグラル立体方式の表示は、サンプリングされた有限数の光線によって立体像を生成するため、エイジアリングが生じる。これを抑制するためにフィルタリング処理を行う[41]。

4.3 実験

提案手法に関する3つの実験を行った。1つ目の実験では、提案手法で用いている距離推定アルゴリズムの基本的な性能を評価した。2つ目の実験では、距離推定結果の信頼度の性能を評価した。3つ目の実験では、実際の番組で使用された多視点映像を用いて、提案手法により3次元モデルの生成、および、インテグラル立体コンテンツの制作を行った。以下に詳細を述べる。

4.3.1 距離推定アルゴリズムの性能評価実験

ステレオデータセットを用いて、提案手法における距離推定手法の性能評価実験を行った。ステレオデータセットには Middlebury のものを使用した。Middlebury のステレオマッチング評価サイトには、このステレオデータセットに対する視差画像を入力すると、視差推定エラー率やエラーが生じた領域などを出力し、その視差推定アルゴリズムの性能評価を

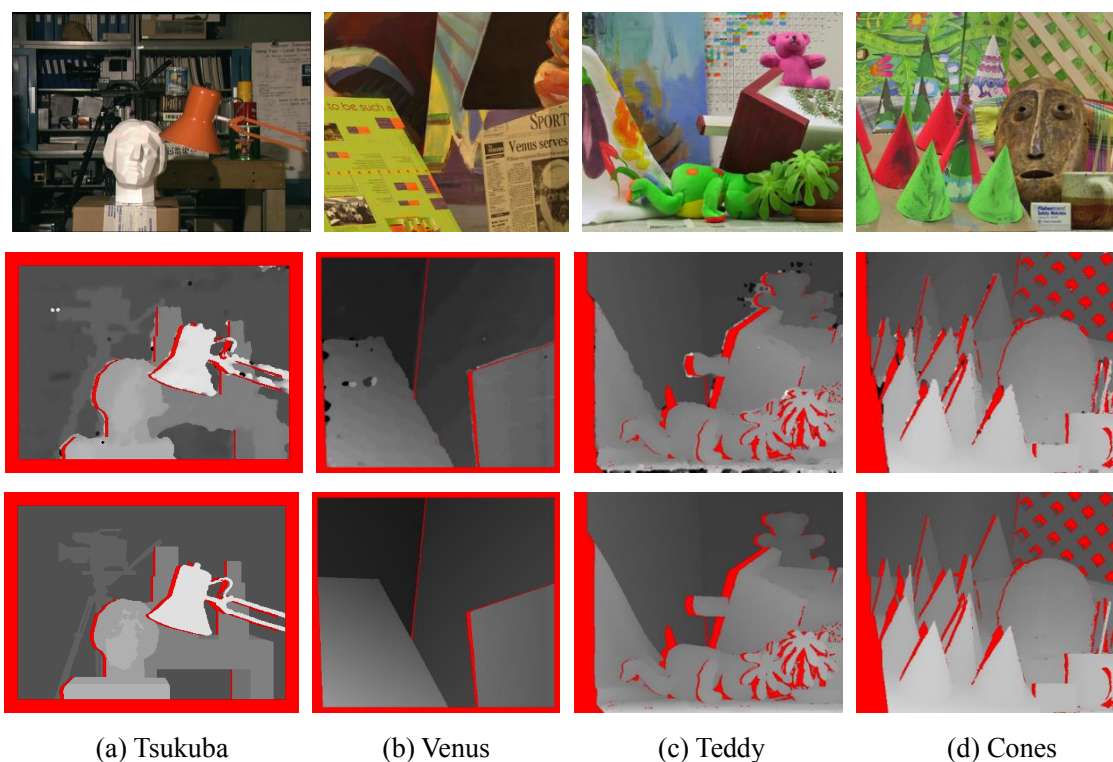
自動的に行う機能がある。本実験もこのサイトの性能評価方法および結果に基づいて行った。

提案手法では、左右ペアそれぞれでステレオマッチングにより視差を推定し、カメラパラメーターを用いて距離情報に変換している。この視差推定アルゴリズムを用いて、Middle Bury ステレオデータセットにおける Tsukuba, Venus, Teddy, Cones の画像の視差推定を行った。実験時の提案手法のパラメーターを表 4.1 に示す。提案手法では、オクルージョンは後段の処理で抑制するため、ここではオクルージョンを除いた領域に対して視差推定を行った。入力画像、生成した視差画像、Grand truth を図 4.8 に示す。図 4.8 では、上段が入力画像、中段が提案手法による視差推定結果、下段が Grand truth を示している。また、図 4.8 中の赤い領域はオクルージョンを表しており、評価サイトより出力されたものである。次に、画像全体における視差推定エラーの割合である視差推定エラー率について評価し、他のアルゴリズムと比較した。ここでは、推定した視差値が Grand truth より 0.5%以上異なっているものを視差推定エラーとした。他のアルゴリズムのうち BP に基づいたものを選択し、比較した結果を表 4.2 に示す。

実験の結果、提案手法により概ね良好な視差推定を行えることが確認できた。提案手法を用いて生成した視差画像には、BP の特徴でもある滑らかな視差のグラデーションが生じており、Venus, Teddy, Cones などの奥行きが徐々に変化する領域においても正確に視差を推定できた。一方で、視差画像における被写体の輪郭が、実際の輪郭と比較して若干歪み、視差推定にエラーが生じる場合があった。また、BP の特徴により、Teddy の右上の人形付近など、前景と重なる背景のテクスチャが不足している領域では、前景の視差情報が背景に侵食するような視差推定エラーが生じた。表 4.2 では、BP + DirectedDiff が最もエラー率が低かったが、この手法はベースラインが狭い多視点映像を対象とした 3 次元モデル生成に着目しており、本システムでは、カメラ配置に自由度を持たせることで、ベースラインが広い多視点映像が対象となるため、採用しなかった。

表 4.1 提案手法のパラメーター

Parameters of proposed method			
λ	T_{data}	T_{smooth}	T
0.4	5.0	1.7	10



(a) Tsukuba (b) Venus (c) Teddy (d) Cones

図 4.8 視差推定結果

(上段：入力画像 中段：提案手法による視差推定結果 下段：Ground truth)

表 4.2 視差推定エラーの割合 単位：% (エラー判定閾値：0.5)

	Tsukuba	Venus	Teddy	Cones
<u>BP + DirectedDiff [42]</u>	6.21	1.59	7.78	4.73
AdaptOverSegBP [43]	5.98	3.66	13.0	9.48
Our Method	10.6	4.77	14.0	8.67
AdaptingBP [44]	19.1	4.84	12.8	7.02
CSBP [45]	22.0	7.60	19.4	15.1

4.3.2 信頼度の性能評価実験

BP のエネルギー分布の尖度より評価した信頼度が、距離を正確に推定できた領域 (以下、正解領域) と、テクスチャの不足やオクルージョンなどで距離推定エラーとなった領域 (以下、エラー領域) を、どの程度正確に表しているか、2 つの実験より評価した。1 つ目の実験として、正解領域とエラー領域における信頼度を調査した。4.3.1 の実験で生成した Teddy と Cones の視差画像における正解領域とエラー領域を図 4.9 に示す。図 4.9 では、正解領域を白、エラー領域を黒で表しており、エラー領域は、距離推定エラーとオクルージョンの領域である。次に、4.3.1 の実験で距離推定を行った際の距離推定結果の信頼度を視覚化した。結果を図 4.10 に示す。図 4.10 では、信頼度が高いほど高い輝度で、低いほど低い輝度で表

しており、信頼度-5 から 30 の範囲を輝度の範囲で正規化している。そして、図 4.10 の正解領域とエラー領域における信頼度の平均値を算出した。結果を表 4.3 に示す。2 つ目の実験として、信頼度による正解領域とエラー領域の判定精度を調査した。信頼度に対し閾値を設定し、正解領域とエラー領域を判定した結果を図 4.11 に示す。図 4.11 では、正解領域であると判定した領域（以下、正解判定領域）を白、エラー領域である判定した領域（以下、エラー判定領域）を黒で示している。そして、正解領域とエラー領域に対する、図 4.9 と図 4.11 の適合率を算出した。結果を表 4.4 に示す。

実験の結果、信頼度が正解領域とエラー領域を、高い精度で正確に表していることが分かった。1 つ目の実験結果から、正解領域では信頼度は高く、エラー領域では信頼度が低いことを視覚的に、数値的に確認した。図 4.9 と図 4.10 を視覚的に比較すると、図 4.9 のエラー領域は、図 4.10 で信頼度が低いことが分かる。特に、オクルージョンや 4.3.1 の実験の Teddy のようにテクスチャが不足している領域では、その傾向が顕著にみられる。また、表 4.3 から、正解領域とエラー領域では、信頼度の平均値に顕著な差があり、信頼度が双方の領域を高い正確性をもって表現していることが分かる。2 つ目の実験結果から、信頼度によって正解領域とエラー領域を高い精度で判定できることを確認した。図 4.9 と図 4.11 を視覚的に比較した結果、正解領域と正解判定領域、エラー領域とエラー判定領域がほぼ一致していることが分かる。また、表 4.4 から、図 4.9 と図 4.11 の適合率が Teddy で 83%、Cones で 87% と高い値を示していることが分かり、これらの結果から、信頼度の判定精度の高さを視覚的、数値的に確認することができた。提案手法ではこの判定処理を、最終的に違う視点の 3 次元モデルの情報によってエラー判定領域が補間されることを前提に行う。そのため、閾値を上げることが可能であり、エラー領域をさらに高い精度で判定することができる。このように、信頼度によってエラー領域を高い精度で判定できるため、提案手法では、距離推定エラーの影響を抑制した 3 次元モデルおよびインテグラル立体映像を生成することが可能である。



図 4.9 Middlebury の評価サイトより出力された正解領域とエラー領域

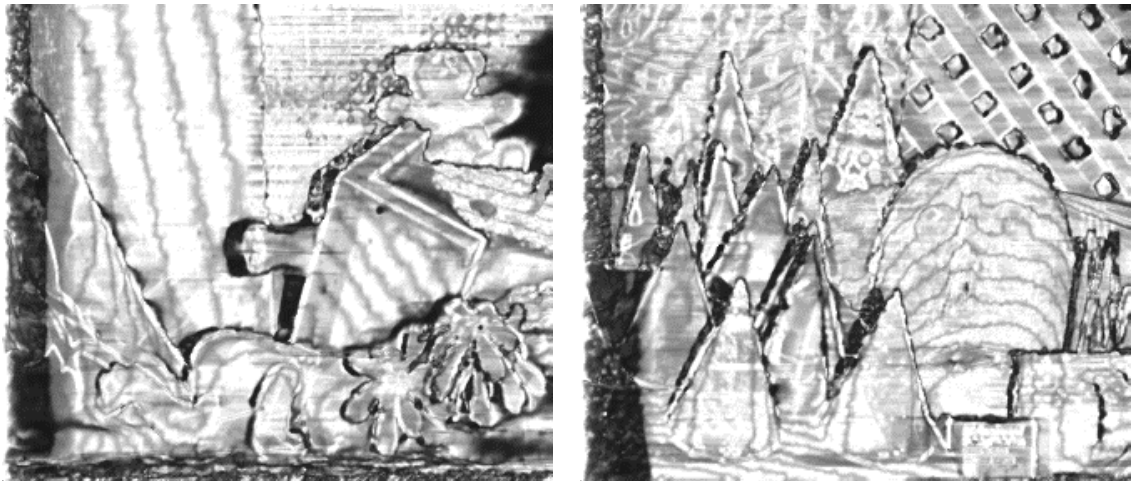


図 4.10 信頼度の視覚化

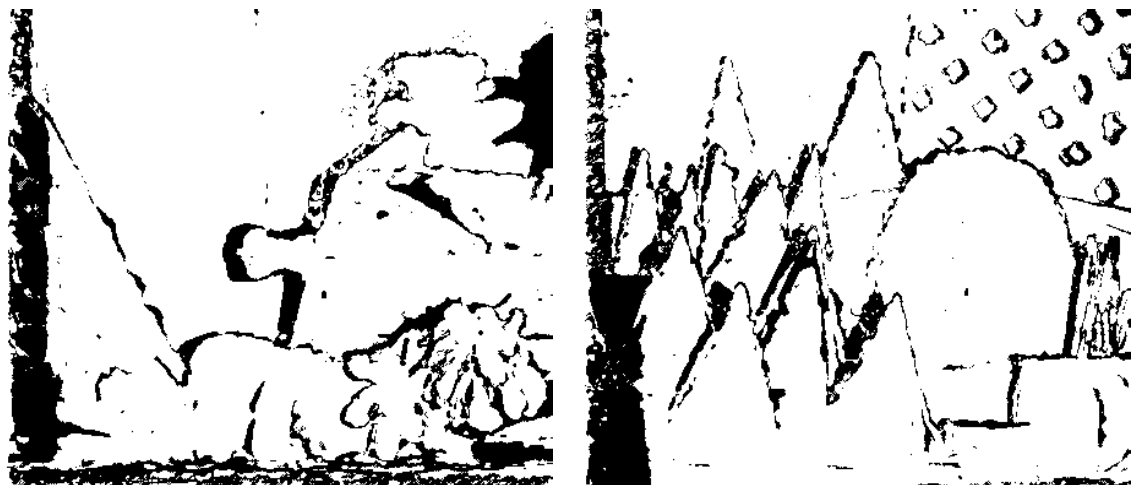


図 4.11 信頼度 R による正解領域とエラー領域の判定

表 4.3 正解領域とエラー領域における信頼度の平均値

	正解領域	エラー領域	全領域
Teddy	23.30	11.83	20.52
Cones	23.31	11.77	20.76

表 4.4 正解領域とエラー領域の判定結果の適合率

	閾値	適合率(%)
Teddy	10.5	83.9
Cones	11.4	87.3

4.3.3 番組で使用された多視点映像を用いた実験

実験では、番組で使用された多視点映像を用いて 3 次元モデルの生成、および、インテグラル立体コンテンツの制作を行った。多視点映像は、「第 55 回全日本相撲選手権大会」の映像である。図 4.12 のように土俵を囲うように 11 台のハイビジョンカメラを配置し、同期をかけて多視点映像を撮影した。カメラから被写体までの距離は約 25m、カメラ間隔は約 2m、撮影時の焦点距離は約 30mm であった。11 台のカメラで土俵の約 90 度を覆っているため、隣あうカメラの輻輳角は約 9 度となる。カメラキャリブレーションは画像内の特徴点を用いて行った[37]。以下に、詳細を述べる。

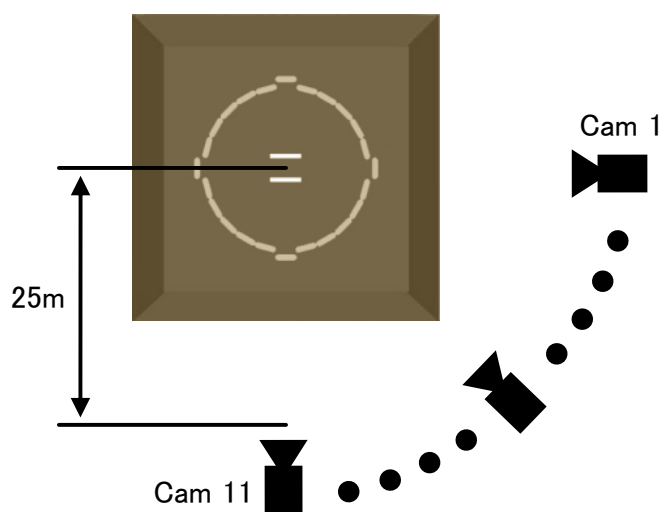


図 4.12 カメラ配置

4.3.3.1 3次元モデル生成実験

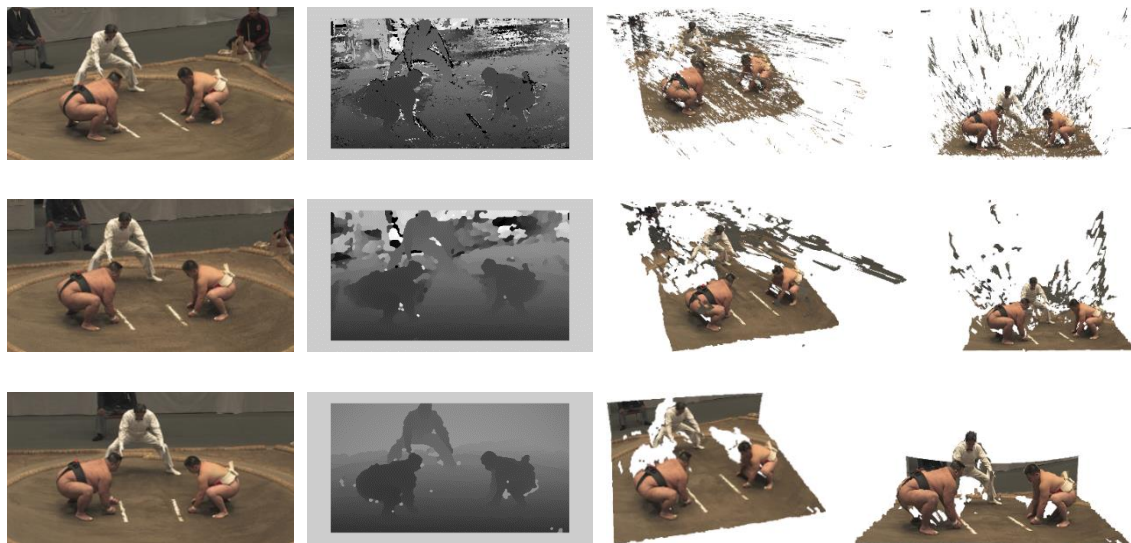
提案手法と従来手法で、3 視点のカメラ画像より 3 次元モデルを生成し、距離推定精度を評価した。距離推定は、カメラより 20m から 30m の範囲を 2cm 刻みで行った。距離推定を行う際には、撮影映像を 480×270 の解像度にダウンコンバートし、そのうち中央の 400×230 の領域を距離推定の対象とした。従来手法には、SSD(Sum of Squared Difference)と従来の BP [38]を 3 視点に拡張したものを用いた。実験時の提案手法と従来の BP のパラメーターを表 4.5 に示す。SSD のブロックサイズは 3×3 であった。また、従来手法において、左右ペアで生成した距離画像の情報で、オクルージョンを埋めた距離画像を生成する際には、双方の距離画像における同座標の画素でエネルギーの最小値を比較し、2 倍以上の差があった場合には、エネルギーの小さい方の距離値を採用し、それ以外の場合は双方の平均の距離値を採用した。入力画像には、Cam1,2,3 の撮影映像の 1 フレームを用いた。図 4.13 に入力画像、距離推定によって生成した距離画像、3 次元モデルを示す。ここで、距離画像は、距離が遠い程、高い輝度で表現しており、距離推定を行った距離の範囲を距離画像の輝度の再現範囲で正規化している。図 4.14 に従来の BP と提案手法の距離画像においてエラーが生じた領域

を赤色で表したものを示す。また、表 4.6 にエラーが生じた画素数および画像全体の画素数に対するエラーが生じた画素数の割合を示す。

実験の結果、提案手法によって高い精度で距離推定を行うことができ、従来手法と比較し、距離推定エラーを抑制した距離画像および 3 次元モデルを生成することができた。SSD では、図 4.13 の上段のように距離推定画像および 3 次元モデルで画面全体に砂状の距離推定エラーが生じた。このような 3 次元モデルからインテグラル立体映像を生成した場合、距離推定エラーが立体的に飛び出してくる状態となり目立つ。一方、図 4.13 の中段および下段のように提案手法および従来の BP では、BP の特性によって砂状のエラーは抑制され、土俵では手前から奥に滑らかに変化する距離値を高い精度で推定できている。しかし、従来の BP では、図 4.14(a)のように力士によるオクルージョンや背景などのテクスチャが不足している領域で、広範囲にわたり距離推定エラーが生じており、SSD と同様にインテグラル立体映像を生成した場合、エラーが目立つ。提案手法では、図 4.5(a)のように距離推定結果の信頼度を評価することで、図 4.14(b)のような距離推定エラーを抑制した距離画像および 3 次元モデルを生成することができた。BP では前景と重なる背景のテクスチャが不足している領域で、前景の距離情報が背景に浸食する距離推定エラーが生じやすい。従来の BP では審判と背景が重なる領域で、そのようなエラーが生じていたが、提案手法では信頼度によってエラー領域を判定することができ、エラーを抑制できた。表 4.6 で示すように、画像全体の画素数に対するエラーが生じた画素数の割合は、従来の BP が 30.7%であったのに対し、提案手法は 1.2%であった。これにより従来手法と比較し、提案手法によって距離推定エラーをより抑制できることを確認した。

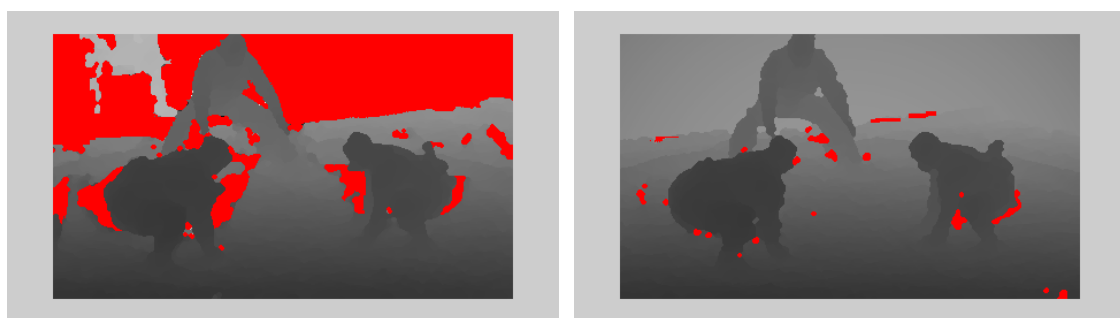
表 4.5 提案手法と従来の BP のパラメーター

	λ	T_{data}	T_{smooth}	T
提案手法	0.5	2.0	1.7	10
従来の BP	0.07	15.0	1.7	10



(a) 入力画像 (b) 距離画像 (c) 3次元モデル

図 4.13 3次元モデル生成実験の結果（上段：SSD 中段：従来のBP 下段：提案手法）



(a) (b)

図 4.14 距離画像におけるエラーが生じた領域（赤色の領域）

(a) 従来のBP (b) 提案手法

表 4.6 距離画像においてエラーが生じた画素数およびその割合

	距離画像の画素数	エラーが生じた画素数	エラーが生じた画素数の割合
従来のBP	92000	28216	30.7%
提案手法	92000	1145	1.2%

4.3.3.2 インテグラル立体コンテンツの制作

提案手法を用いて、相撲のインテグラル立体コンテンツを制作した。11台のカメラ映像から両端のカメラを除いた9視点の3次元モデル生成し、それらを統合してインテグラル立体映像を生成した。インテグラル立体ディスプレイには、所有しているものの中で一番奥行き再現範囲が広い3840×2400高解像度ディスプレイに160×118のレンズアレイを装着し

たものを使用した。奥行き再現範囲は、レンズアレイ面から手前、奥方向にそれぞれ約 20cm である。まず、このディスプレイに表示させたインテグラル立体映像を、視点位置を上下左右に移動させ、運動視差を生じさせて撮影した。結果を図 4.15 に示す。図 4.15 の中の力士の頭部、土俵の端部分の赤い線は、運動視差が理解しやすくなるよう手動で描いたものである。要素画像群を動画再生したインテグラル立体コンテンツを再撮した結果を図 4.16 に示す。

実験の結果、力士の取組を高い奥行き感をもって立体的に表現することができ、臨場感のあるインテグラル立体コンテンツを制作することができた。図 4.15 の力士の頭部と土俵の端部分などの位置関係から、視点を変えることで大きく運動視差が再現されていることが分かる。図 4.16 のインテグラル立体コンテンツでは、ぶつかりあう力士達の奥行き方向の位置関係や、手前から奥に滑らかに続く土俵の奥行きも再現することができた。提案手法により距離推定エラーの影響を抑制できたため、エラーが立体的に飛び出ることにはなかった。また、1 視点の 3 次元モデルを要素画像群に変換した際に生じていた欠落領域も、視点の異なる要素画像群で補完することができた。スポーツ番組で扱うシーンを被写体としたインテグラル立体コンテンツは、これまで例がなく、今回の実験によって初めて実現することができた。また、ディスプレイの解像度は 160×118 (レンズアレイ数と同等) であったが、実際に立体像を見た時、それ以上の解像度があるように感じると共に、力士の細かな筋肉の動きも視認でき、インテグラル立体テレビの潜在的な発展性を確認することができた。



図 4.15 異なる視点位置から観測したインテグラル立体映像
(左：左視点 中上：上視点 中下：下視点 右：右視点)



図 4.16 相撲のインテグラル立体コンテンツ (再撮)

4.4 まとめ

多視点カメラを用いてインテグラル立体映像を生成する手法を提案した。提案手法では、多視点カメラの撮影映像から 3 次元モデルを生成することで、仮想空間で被写体が発する光線を取得し、インテグラル立体映像を生成した。また、距離推定を用いて 3 次元モデルを生成するうえでの距離推定エラーを抑制するため、多視点カメラを用いた 3 次元モデル生成手法および距離推定における信頼度を用いた 3 次元モデル補正手法を提案した。撮影実験により、提案手法によって 3 次元モデルを生成するうえでの距離推定エラーを抑制し、遠方の被写体のインテグラル立体映像を生成できることを確認した。また、実際に放送で扱った相撲のシーンのインテグラル立体コンテンツを制作することができた。これらにより、本手法の有効性を確認した。

第5章 多視点ロボットカメラを用いたインテグラル立体撮影手法

5.1 はじめに

前章の多視点カメラを用いたインテグラル立体映像生成手法によって遠方の被写体の撮影が可能になった。一方で、多視点カメラの姿勢および画角が固定されているため、ダイナミックに動く被写体や広い空間に点在する被写体の撮影が困難だという課題がある。第3章では、多視点映像表現技術における同様の課題を解決するため、多視点ロボットカメラを撮影に用いたが、これをインテグラル立体の撮影に適応した場合、多視点ロボットカメラの協調制御および配置がインテグラル立体の撮影に適していないため、生成されるインテグラル立体映像の品質が低下する。高品質なインテグラル立体映像を生成するためには、インテグラル立体ディスプレイで再現される領域の多視点映像を高解像度に撮影すること、また、視域内の水平および垂直方向の光線を取得する必要がある。多視点映像表現技術における多視点ロボットカメラの協調制御は、カメラを被写体上の空間中の1点（注視点）へ方向制御し、被写体までの距離が異なるすべてのカメラで被写体と同じ大きさで撮影されるよう撮影画角を制御したものであり、インテグラル立体ディスプレイの再現領域を撮影することは困難であった。撮影画角に再現領域が収まらない場合、図 5.1(a)の撮影領域内の赤い部分のようにインテグラル立体映像に欠損が生じ、再現領域を収めるため撮影画角を広くすると図 5.1(b)のように相対的に再現領域の解像度が低下し、インテグラル立体映像の品質が低下する。また、図 5.1(a) (b)のようにカメラを水平方向に一次元状に配置していたため、視域内の赤い部分のような垂直方向の光線を取得することが困難であり、インテグラル立体映像に欠損が生じることがあった。

この課題を解決するために、多視点ロボットカメラを用いたインテグラル立体撮影手法を提案する。提案手法では、図 5.1(c)のように多視点ロボットカメラの協調制御によりインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で高解像度に撮影する。また、カメラを二次元状に配置することで垂直方向の光線を取得し、二次元状に配置したカメラの多視点映像を相補的に活用して3次元モデルを生成する。そして、カメラマンによるパン・チルト・ズームなどのカメラワークを再現した要素画像およびインテグラル立体映像を生成する。実験では、提案手法を用いて空手を被写体としたインテグラル立体映像を生成し、その有効性を確認する。

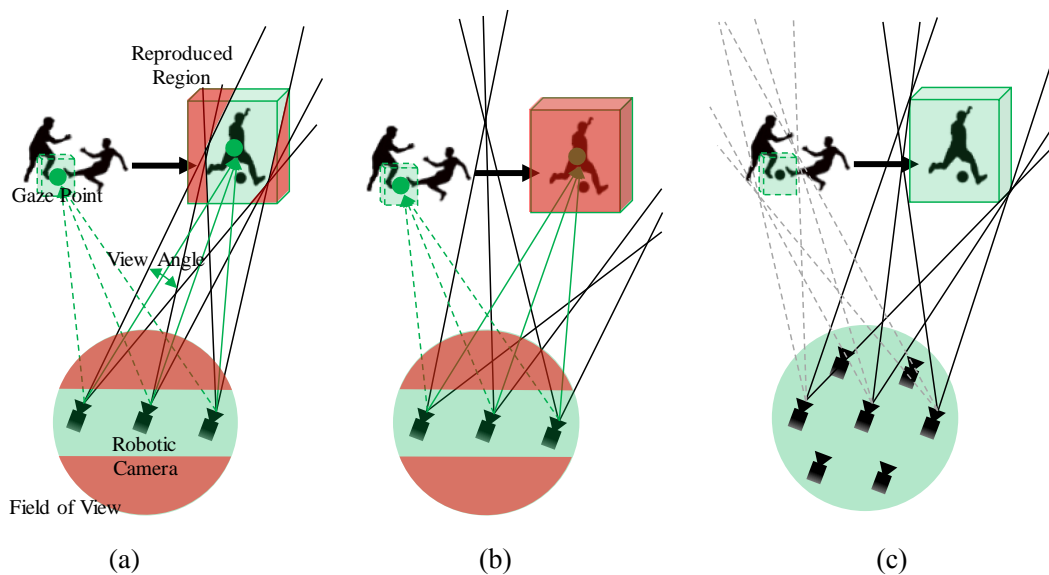


図 5.1 多視点ロボットカメラの協調制御と配置

(a) 従来手法（撮影画角が狭い場合） (b) 従来手法（撮影画角が広い場合） (c) 提案手法

5.2 提案手法

提案手法では、多視点ロボットカメラを用いて被写体を撮影し、多視点映像から 3 次元モデルを生成後、要素画像に変換することでインテグラル立体映像を生成する。各処理について述べる。

5.2.1 多視点ロボットカメラによる撮影

従来手法の多視点ロボットカメラの協調制御は、カメラを被写体上の空間中の 1 点へ方向制御し、すべてのカメラの撮影画角を一定に制御したものであり、インテグラル立体ディスプレイの再現領域を撮影することは困難であった。また、カメラを水平方向に一次元状に配置していたため、垂直方向の光線を取得することが困難であった。提案手法では、多視点ロボットカメラの新たな協調制御手法によりインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で高解像度に撮影する。また、カメラを二次元状に配置することで垂直方向の光線を取得する。

実空間にインテグラル立体ディスプレイの再現領域を規定する方法について述べる。提案手法では、カメラで被写体を撮影した際の被写体の位置および撮影画角に応じて再現領域の位置とサイズが変わる。被写体を表示するインテグラル立体ディスプレイの画面サイズを W 、奥行再現範囲を D 、被写体を立体像にした時のインテグラル立体ディスプレイからの飛び出し量を Δ 、カメラからの被写体までの距離を d 、撮影画角を θ とした時、実空間にインテグラル立体ディスプレイの再現領域を規定した時のサイズの縮尺比 k は式(5.1)によって求まり、再現領域は図 5.2 のように規定される。図 5.2 は再現領域を上から見たものであり、実空間における再現領域の形状は四角錐台となる。

$$k = d \cdot \tan(\theta/2) / \left\{ \frac{W}{2} - \Delta \tan(\theta/2) \right\} \quad (5.1)$$

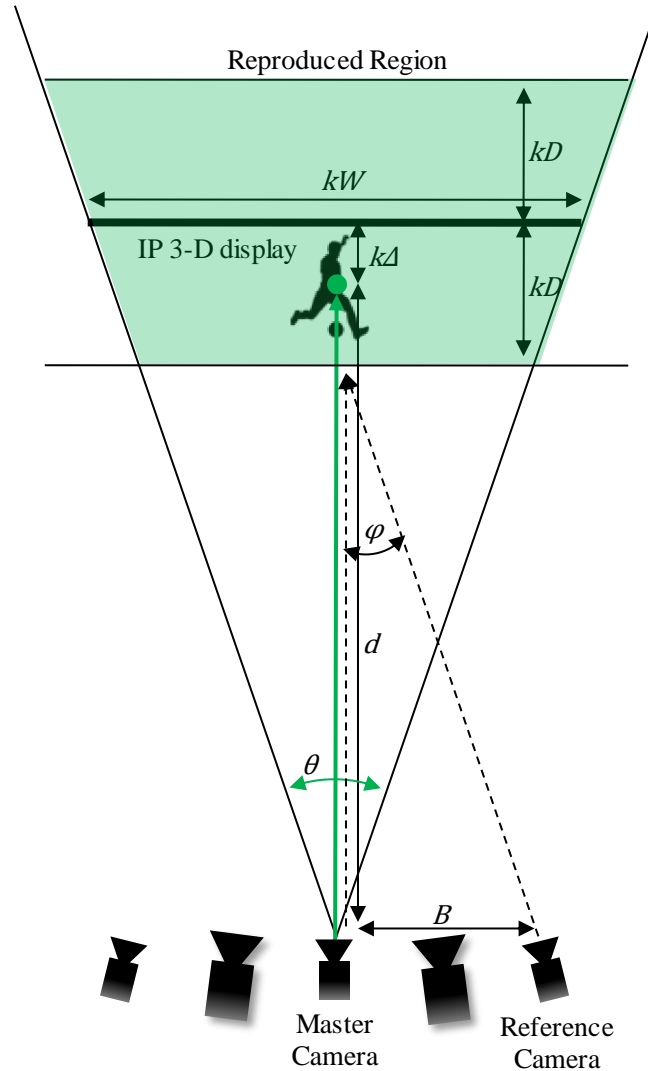


図 5.2 インテグラル立体ディスプレイの再現領域とカメラ配置

処理について述べる。まず，図 5.3 のように多視点ロボットカメラを正六角形状に 6 台とその中心に 1 台配置する。中心の 1 台をマスターカメラ，正六角形状の 6 台をリファレンスカメラと呼ぶ。この形状に配置することで垂直方向の光線を取得することができる。また，本手法では，5.2.2. で述べるように，多視点映像からステレオマッチングを用いて三次元モデルを生成する。そのため，ベースラインをステレオマッチングで許容される 2 台のカメラの輻輳角（許容視差角）に応じた一定の距離以下にしなければならない。この形状では，隣接するカメラ間のベースラインを等しくその一定の距離に設定することができるため，少ないカメラ台数でより広い視域の光線を取得できる。ベースラインを決めるために，まず，

撮影時のおおよその被写体位置と撮影画角の見当をつける．その時のカメラからの被写体までの距離を d_0 ，撮影画角を θ_0 として式(5.1)に代入し，算出した縮尺比を k_0 とする．図 5.2 のように，カメラの輻輳角が最大となるのは，再現領域内のカメラから最も近い点であるため，その点でカメラの輻輳角が許容視差角以下になるようにベースラインを決める．許容視差角を Φ とした時，ベースライン B を式(5.2)で求める．

$$B = \tan \Phi \{d_0 - k_0(D - \Delta)\} \quad (5.2)$$

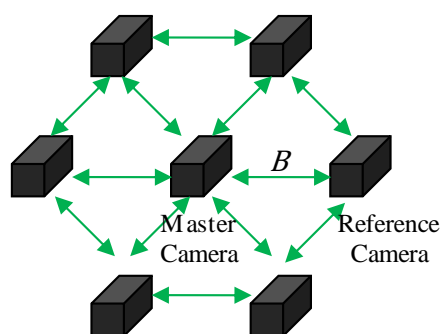


図 5.3 カメラ配置

次に，カメラキャリブレーションを行う[37]．ロボットカメラのカメラ番号を n ，回転行列を \mathbf{R}_0 ，並進行列を \mathbf{t}_0 とした時，世界座標 \mathbf{X} からカメラ座標 \mathbf{x} への変換には式(5.3)を用いる．

$$\mathbf{x}_n = [\mathbf{R}_{0n} \quad \mathbf{t}_{0n}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \quad (5.3)$$

図 5.4 に多視点ロボットカメラの協調制御を示す．カメラマンは，マスターカメラのパン・チルト・ズーム・デプスを操作して被写体を撮影する．ここで，デプスとはカメラからの距離値のことをいい，パン・チルト・デプスにより空間中の1点（注視点）を指定することができる．カメラマンは，この注視点が被写体の位置にくるようロボットカメラを操作する．マスターカメラのカメラ番号を M ，カメラキャリブレーション時のパン・チルトの値を P_0, T_0 ，カメラマンがロボットカメラを操作した時のパン・チルトの値を P, T とした時，マスターカメラの光軸 o_M は式(5.4), (5.5), (5.6)で表される．

$$\mathbf{R}_M = \mathbf{R}_{0M}^{-1} \begin{bmatrix} \cos(P_M - P_{0M}) & 0 & \sin(P_M - P_{0M}) \\ 0 & 1 & 0 \\ -\sin(P_M - P_{0M}) & 0 & \cos(P_M - P_{0M}) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(T_M - T_{0M}) & -\sin(T_M - T_{0M}) \\ 0 & \sin(T_M - T_{0M}) & \cos(T_M - T_{0M}) \end{bmatrix} \quad (5.4)$$

$$\mathbf{R}_M = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (5.5)$$

$$\mathbf{o}_M = \begin{bmatrix} r_{13} \\ r_{23} \\ r_{33} \end{bmatrix} \quad (5.6)$$

デプスを d とした時、注視点 \mathbf{g} は式(5.7)で表される。

$$\mathbf{g} = -\mathbf{R}_{0M}^{-1}\mathbf{t}_{0M} + d \cdot \mathbf{o}_M \quad (5.7)$$

カメラマンが、デプスを調整する際に、注視点が被写体の位置にきたかどうか確認するために、例えば、3.2.3 で述べたような、ロボットカメラをもう 1 台設置し、注視点に常に方向制御させ、画面内の被写体の位置によって注視点が被写体の位置にきたか確認する方法がある。

次に、実空間にインテグラル立体ディスプレイの再現領域を規定する。撮影画角を θ_M 、カメラの水平垂直アスペクト比を $w : h$ とした時、マスターカメラの撮影領域を構成する 4 本の直線 l_k ($1 \leq k \leq 4$) を式(5.8)で求める。

$$\mathbf{l}_k = \mathbf{R}_{0M}^{-1} \begin{bmatrix} \cos(p_k) & 0 & \sin(p_k) \\ 0 & 1 & 0 \\ -\sin(p_k) & 0 & \cos(p_k) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(t_k) & -\sin(t_k) \\ 0 & \sin(t_k) & \cos(t_k) \end{bmatrix} \mathbf{R}_{0M} \mathbf{o}_M \quad (5.8)$$

$$p_k = \left\{ \frac{\theta}{2}, -\frac{\theta}{2}, -\frac{\theta}{2}, \frac{\theta}{2} \right\}$$

$$t_k = \left\{ \frac{\theta h}{2w}, \frac{\theta h}{2w}, -\frac{\theta h}{2w}, -\frac{\theta h}{2w} \right\}$$

インテグラル立体ディスプレイの奥行再現範囲を D 、被写体を立体像にした時のインテグラル立体ディスプレイからの飛び出し量を Δ 、実世界とインテグラル立体ディスプレイの再現する空間像の縮尺比 k とした時、マスターカメラの光軸と再現領域のマスターカメラから向かって手前の交点と奥の交点 \mathbf{g}' を式(5.9)で求める。

$$\mathbf{g}' = \begin{cases} -\mathbf{R}_{0M}^{-1}\mathbf{t}_{0M} + (d - kD + k\Delta) \cdot \mathbf{o}_M, & \text{near} \\ -\mathbf{R}_{0M}^{-1}\mathbf{t}_{0M} + (d + kD + k\Delta) \cdot \mathbf{o}_M, & \text{far} \end{cases} \quad (5.9)$$

実空間に再現領域を構成する 8 つの頂点を規定する。8 つの頂点 \mathbf{v}_i ($1 \leq i \leq 8$) を式(5.12)より算出する。

$$\mathbf{v}_i = -\mathbf{R}_{0M}^{-1}\mathbf{t}_{0M} + l_k \frac{o_M(\mathbf{g}' + \mathbf{R}_{0M}^{-1}\mathbf{t}_{0M})}{o_M l_k} \quad (5.12)$$

次に、リファレンスカメラがインテグラル立体ディスプレイの再現領域を領域が収まる最小の画角で撮影するようにパン・チルト・ズームを制御する。リファレンスカメラのカメラ番号を n とした時、リファレンスカメラから \mathbf{v}_i に向かうカメラ座標系におけるベクトル \mathbf{o}'_{ni} を式(5.13)で求める。

$$\mathbf{o}'_{ni} = \mathbf{R}_{0n} \frac{\mathbf{v}_i + \mathbf{R}_{0n}^{-1}\mathbf{t}_{0n}}{\|\mathbf{v}_i + \mathbf{R}_{0n}^{-1}\mathbf{t}_{0n}\|} \quad (5.13)$$

\mathbf{v}_i に方向制御した場合のパン・チルトの値 P_{ni} , T_{ni} は式(5.14), (5.15), (5.16)となる。

$$\mathbf{o}'_{ni} = [e_{ni1} \quad e_{ni2} \quad e_{ni3}]^T \quad (5.14)$$

$$P_{ni} = \tan^{-1}(e_{ni1}/e_{ni3}) + P_{0n} \quad (5.15) \quad T_{ni} = \sin^{-1}(e_{ni2}) + T_{0n} \quad (5.16)$$

リファレンスカメラで再現領域を領域が収まる最小の画角で撮影するためのパン・チルトの値 P_n , T_n を式(5.17), (5.18)で求め、パン・チルト制御する。

$$P_n = \frac{\max_i P_{ni} + \min_i P_{ni}}{2} \quad (5.17) \quad T_n = \frac{\max_i T_{ni} + \min_i T_{ni}}{2} \quad (5.18)$$

次に、カメラの水平垂直アスペクト比を考慮し、領域が収まる最小の画角 θ_n を式(5.19)で求め、画角が θ_n になるようにズーム制御する。

$$\theta_n = \max \left[\max_i P_{ni} - \min_i P_{ni}, \frac{w}{h} \left(\max_i T_{ni} - \min_i T_{ni} \right) \right] \quad (5.19)$$

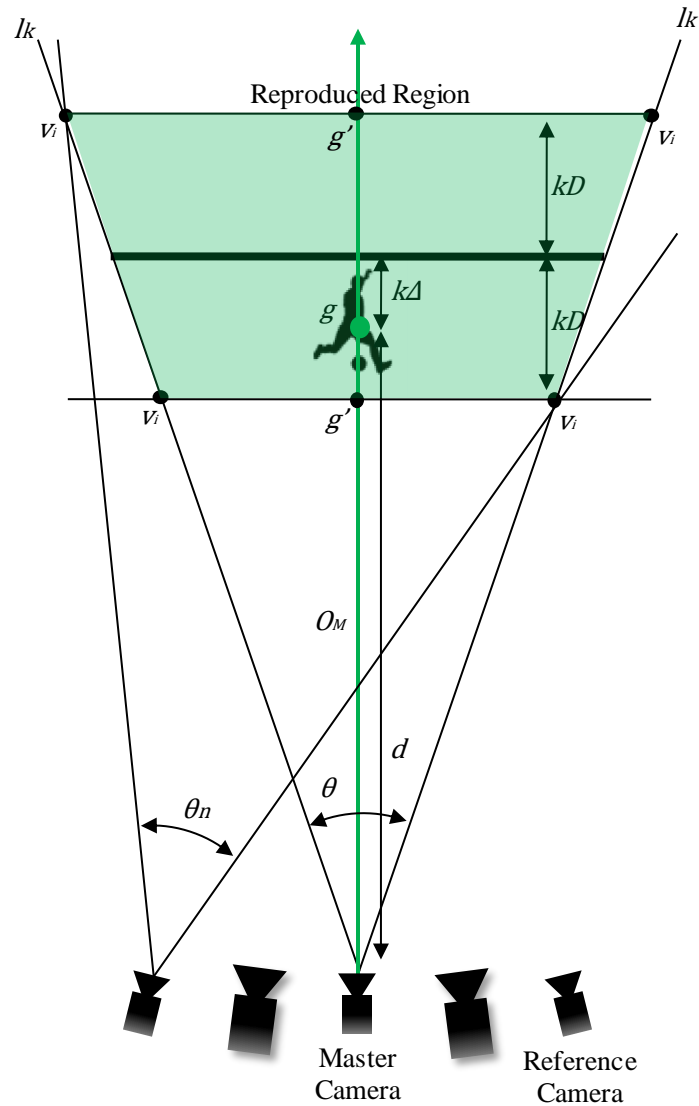


図 5.4 多視点ロボットカメラの協調制御

5.2.2 3次元モデルの生成

前章の多視点カメラを用いたインテグラル立体映像生成手法における3次元モデルの生成では、水平方向に一次元状に配置されたカメラのうち2台のみのカメラの情報を用いてステレオマッチングを行っていたため奥行推定精度に課題があった。また、ステレオマッチングに Belief Propagation [38] を用いており、PCのメモリを大量に必要としたため、高解像度の多視点画像を扱うことが困難であった。提案手法では、二次元状に配置された7台のカメラの情報を相補的に活用してマルチベースラインステレオとコストボリュームフィルタリングを行い奥行推定精度の改善を試みている。また、よりメモリの消費量を抑えた手法にすることで高解像度の多視点画像を扱うことを可能にしている。処理について述べる。

はじめに、7台の各視点のカメラ映像を用いてカメラキャリブレーションを行う[37]。次

に、各視点のカメラ映像において奥行推定を行う。奥行推定には、マルチベースラインステレオとコストボリュームフィルタリング[46]を用いる。まず、奥行推定の対象のカメラとその他のカメラで6つのカメラペアをつくる。奥行推定の対象のカメラにおける各画素に各奥行値のラベルを割り当てた時のコスト D を式(5.20)より求める。

$$D_{m,p,l} = \min \left[-\frac{\sum_{(i,j) \in R(p)} (I(i,j) - \bar{I})(I'_m(i+d(l),j) - \bar{I}'_m)}{2\sqrt{\sum_{(i,j) \in R(p)} (I(i,j) - \bar{I})^2 \times \sum_{(i,j) \in R(p)} (I'_m(i+d(l),j) - \bar{I}'_m)^2}} + 0.5, T_D \right] \quad (5.20)$$

ここで、 m はカメラペア番号、 p は処理対象の画素、 l は奥行値のラベル、 $R(p)$ は p を中心としたブロック内の画素集合、 i, j はそのブロック内の画素のインデックス、 I は奥行推定の対象となったカメラの画素値、 I' はそのペアとなったカメラの画素値、 \bar{I} および \bar{I}' はそれらのブロック内の画素値の平均値、 $d(l)$ は奥行値のラベルに対応した視差値、 T_D はコストの閾値を表す。式(5.20)は、正規化相互相関 (ZNCC) をベースにしており、多視点画像間に輝度差が生じた場合にも安定してコストを求めることができる。また、 T_D を設けることで、奥行推定における距離値がカメラから近く、カメラペアにおけるカメラ間の輻輳角が許容視差角を超え、有効なコストを得られない場合にはその情報を排除し、一方、距離値がカメラから遠く、輻輳角が許容視差角以内になり、有益なコストを得られた場合にはその情報を採用している。これにより、奥行推定の対象のカメラと隣接していない離れたカメラの情報も有効に活用することができる。次に、奥行推定の対象のカメラにおける各ペアのコストを統合したコスト C を式(5.21)より求める。

$$C_{p,l} = \frac{1}{6} \sum_m D_{m,p,l} \quad (5.21)$$

奥行値が l の時のコストを全画素分まとめることでコストマップを生成し、フィルタリングを行う。フィルタリング後のコスト C' を式(5.22)より求める。

$$C'_{p,l} = \sum_q W_{p,q}(l) C_{q,l} \quad (5.22)$$

ここで、 W はフィルタ、 q はフィルタ内の画素を表す。フィルタにはガイドドフィルタ【47】を使用し、ガイダンス画像には I を用いた。処理対象の画素 p にフィルタリング後のコストが最小となった奥行値のラベル l を割り当てる。画素 p における奥行値 f を式(5.23)より求める。

$$f_p = \underset{l}{\operatorname{argmin}} C'_{p,l} \quad (5.23)$$

この手法を用いて7台すべてのカメラ映像において奥行推定を行う。次に、3次元モデリングを行う。まず、奥行推定結果の信頼性が高い画素を抽出する。ここでは、式(5.23)によって奥行値 l を割り当てた時のコスト C が T_c 以下の画素を抽出する。そして、すべてのカメラの抽出した画素を世界座標系へ投影し3次元点群モデルを生成する。これにより、1視点のカメラの3次元点群モデルに生じるオクルージョンおよび信頼性が低く欠けた領域が補完される。

5.2.3 要素画像の生成

カメラマンはシーンや被写体をカメラマンの意図をもって表現するためにパン・チルト・ズームなどのカメラワークを用いて撮影する。提案手法では、仮想空間に撮影時のカメラ、被写体、ディスプレイの位置関係を再現して要素画像を生成することで、カメラマンによるカメラワークを再現したインテグラル立体映像を生成する。ここで、提案手法におけるカメラワークとは、図 5.2 のように撮影画角に比例して奥行再現範囲が変わるものとする。例えば、ある被写体にズームインした時、視聴者からは、画角内の被写体のサイズが上がると共にその奥行感も上がって見える。

処理について述べる。まず、仮想空間に撮影時のカメラ、被写体、インテグラル立体ディスプレイの位置関係を再現する。3次元モデルを実空間と同じスケールに変換し、式(5.3)より世界座標系の3次元モデルをマスターカメラのカメラ座標系に変換する。図 5.5 のように、ディスプレイのサイズを kW とし、マスターカメラから距離 $d+kf$ のところにディスプレイの中心がマスターカメラの光軸を通り、レンズ面が光軸と垂直の姿勢になるように配置する。

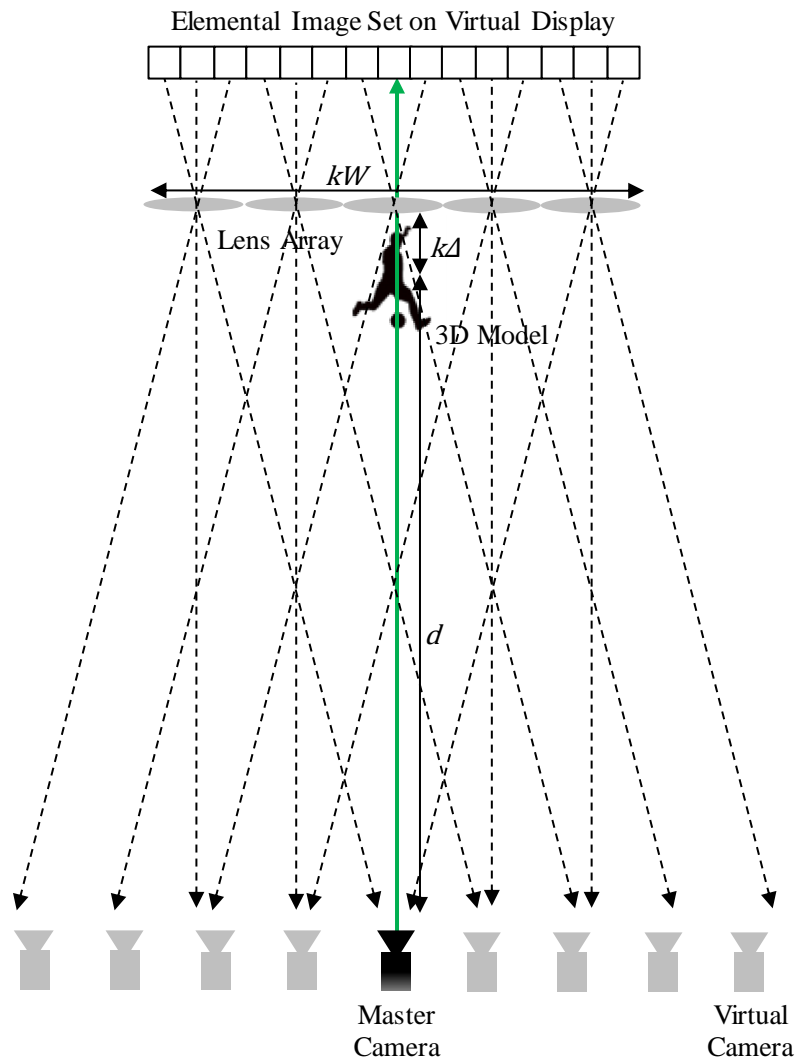


図 5.5 要素画像の生成

5.3 実験

実験では、多視点ロボットカメラの協調制御によってインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で撮影できるか検証した。また、提案手法によって撮影した多視点映像からインテグラル立体映像を生成できるか検証した。

5.3.1 多視点ロボットカメラの協調制御の検証

まず、シミュレーションを用いた実験により提案手法における多視点ロボットカメラの協調制御の再現性を検証した。次に、提案手法を実装したシステムを用いた撮影実験により実用性を検証した。

5.3.1.1 シミュレーション

シミュレーターを開発し、仮想空間内で多視点ロボットカメラを協調制御させてインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で撮影できるか検証した。シミュレーターでは、仮想空間に多視点ロボットカメラと被写体の3次元モデルを配置し、提案手法を用いてロボットカメラを協調制御させることで、被写体およびインテグラル立体ディスプレイの再現領域がどのように撮影されるか確認することができる。シミュレーターのフローチャートを図 5.6 に示す。

ニューススタジオのキャスターを撮影対象とし、図5.7のように多視点ロボットカメラと被写体の3次元モデルを配置した。パラメーターを $\{W, D, \Delta, \phi, d, \theta\} = \{243\text{mm}, 100\text{mm}, 50\text{mm}, 20^\circ, 5000\text{mm}, 59^\circ\}$ とした。マスターカメラおよびリファレンスカメラによる再現領域の撮影映像を図5.8に示す。図5.8では、中心の映像がマスターカメラ、その周囲の映像がリファレンスカメラの映像を表示している。マスターカメラの映像では、再現領域と撮影画角が完全に一致しているため、青い線で描画された領域は見えない。一方で、リファレンスカメラの映像では、再現領域の横幅もしくは縦幅が、撮影画角の横幅と縦幅に一致していることが分かる。実験の結果、すべてのカメラでインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で撮影することができ、提案手法の再現性を検証できた。

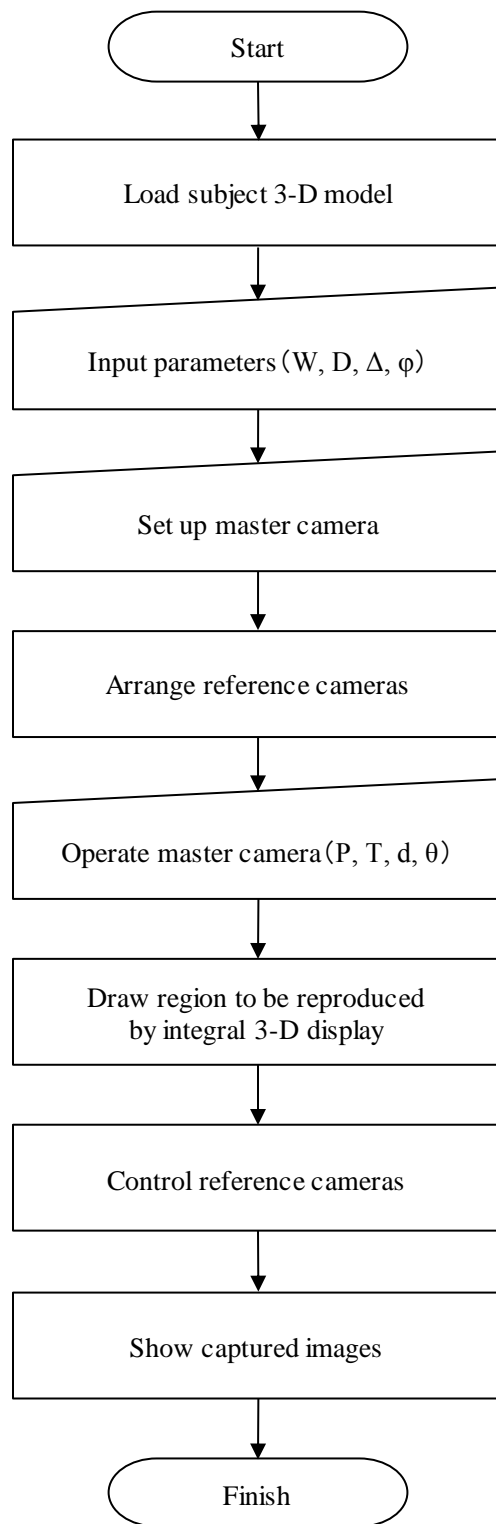


図 5.6 シミュレーターのフローチャート



図 5.7 多視点ロボットカメラと被写体の配置

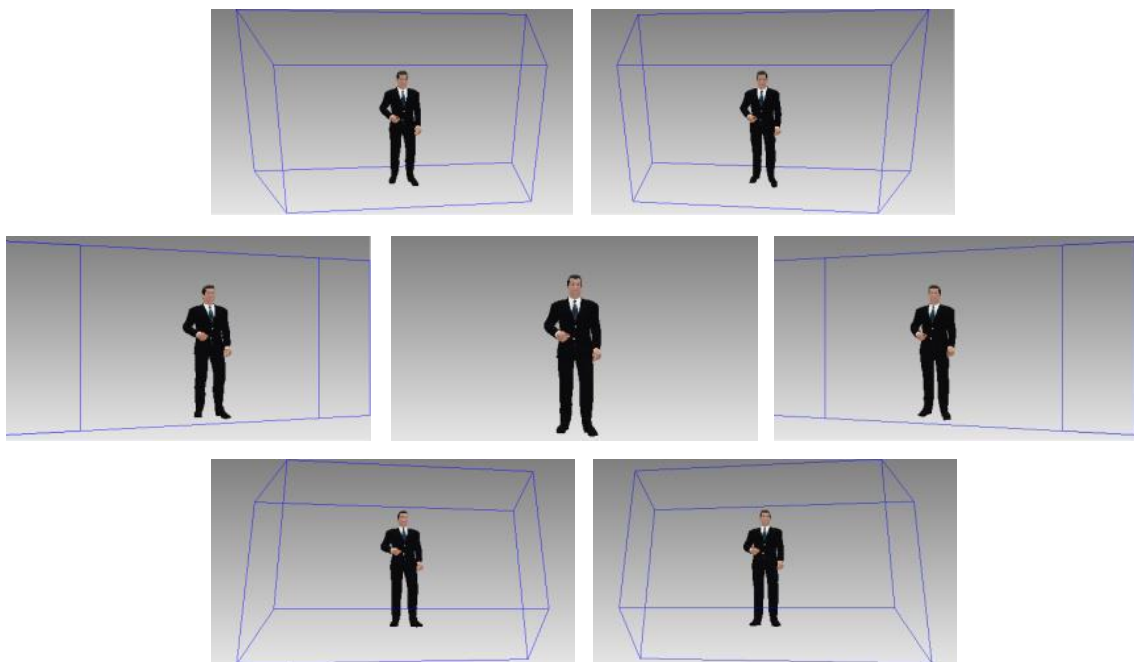


図 5.8 再現領域の撮影映像（青枠が再現領域）

5.3.1.2 システム開発

提案手法を実装したシステムを開発した。システムは、主に小型 HD カメラとそれを載せた電動雲台、マイクロコンピューターからなるロボットカメラ、カメラマンがマスターカメラを操作するための操作インターフェース、多視点映像収録装置、3次元モデルおよび要素画像を生成する多視点映像処理装置から構成される。図 5.9 に多視点ロボットカメラ、図 5.10 に操作インターフェース、図 5.11 にシステムの系統図を示す。



図 5.9 多視点ロボットカメラ



図 5.10 操作インターフェース

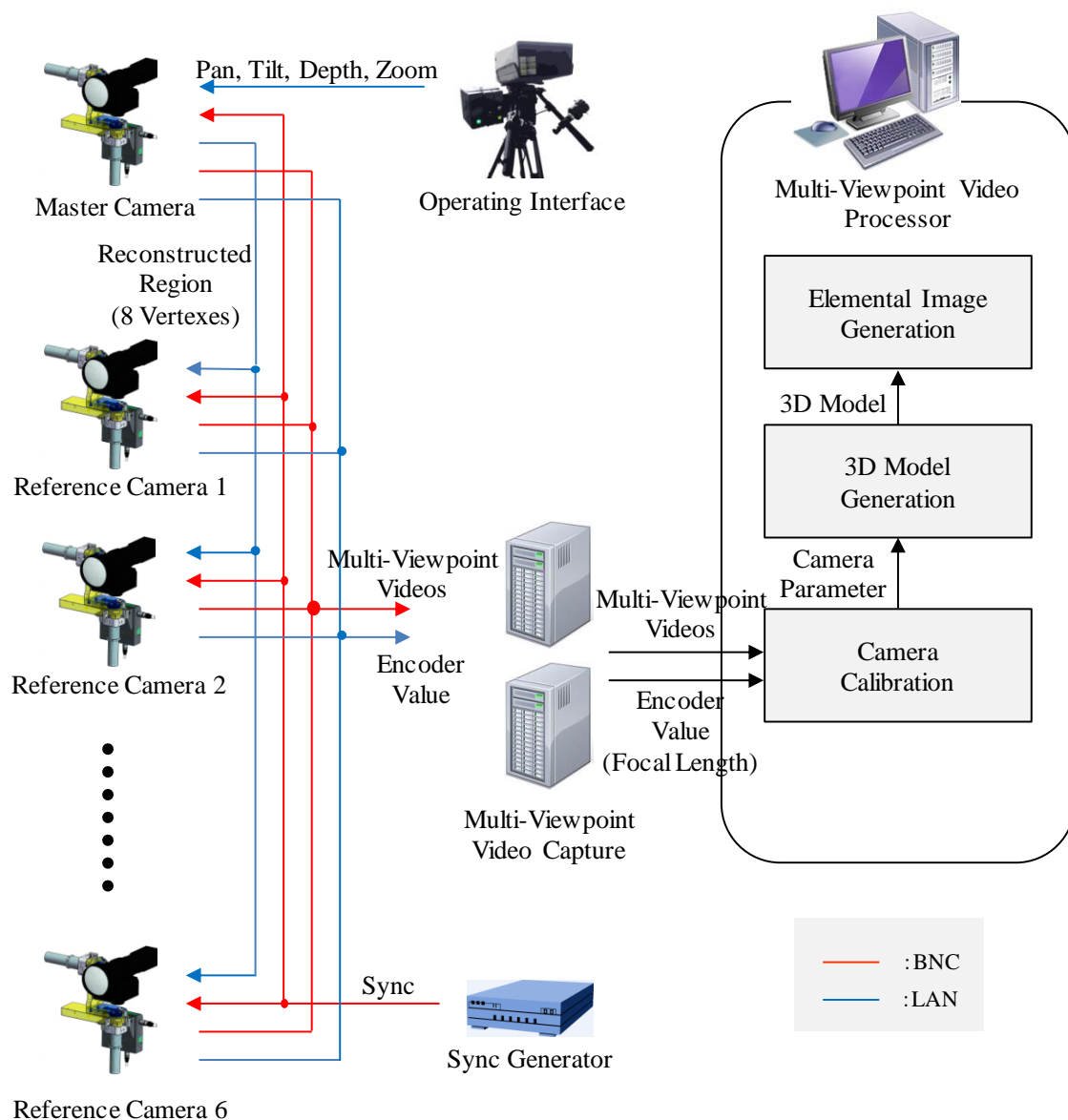


図 5.11 システム系統図

5.3.1.3 撮影実験

提案手法を実装したシステムで被写体を撮影し、撮影映像にインテグラル立体ディスプレイの再現領域を可視化して重畳することで、機械的制御誤差などを含んだ再現領域の撮影精度を検証した。空手選手を被写体とし、図 5.12 のように多視点ロボットカメラを配置して多視点映像を撮影した。パラメーターを $\{W, D, \Delta, \varphi, d, \theta\} = \{293\text{mm}, 45\text{mm}, 20\text{mm}, 12^\circ, 5000\text{mm}, 47^\circ\}$ とした。次に、撮影映像にインテグラル立体ディスプレイの再現領域を可視化して重畳するために、まず、弱校正のカメラキャリブレーションを行い、撮影時のカメラパラメーターを取得した。撮影時のマスターカメラの画角と注視点位置、インテグラル立体ディスプレイのサイズと奥行き再現範囲、取得したカメラパラメーターを用いて再現領域

の世界座標を算出し、多視点映像に投影した。撮影した多視点映像にインテグラル立体ディスプレイの再現領域を可視化して重畳したものを図 5.13 に示す。5.3.1.1 での実験結果とは異なり、再現領域と撮影画角が完全には一致しなかった。そこで、機械的制御誤差などを含んだ再現領域の撮影精度を求めるため、各視点の映像における撮影画角と再現領域の誤差を計測した。計測の結果、誤差は平均 44.2pixel であった。従来手法では、撮影時にインテグラル立体ディスプレイの再現領域は考慮されず、再現領域が画角に納まらないか、画角よりも小さく撮影されるためその誤差は大きくなる。提案手法を用いることによってその誤差を抑制できることを確認し、提案手法の実用性を検証できた。

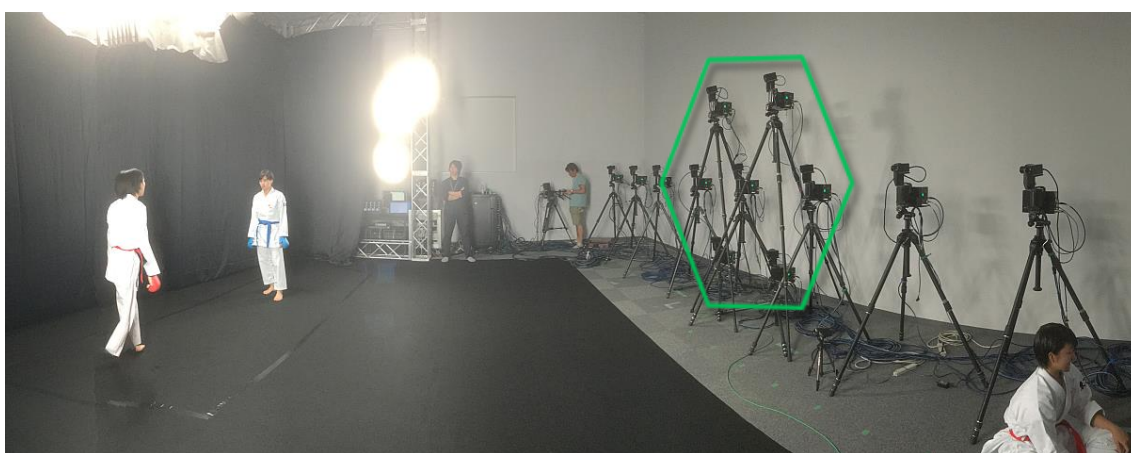


図 5.12 多視点ロボットカメラの配置（緑枠内のカメラを使用）



図 5.13 インテグラル立体ディスプレイの再現領域を重畳した多視点映像
(緑枠が再現領域)

5.3.2 インテグラル立体映像の生成

5.3.1.3 で撮影した多視点映像から提案手法における 3 次元モデルおよび要素画像の生成によってインテグラル立体映像を生成した。3 次元モデルの生成における奥行推定では、パラメーターを $\{R, T_d, T_c\} = \{9 \times 9, 0.2, 0.9\}$ とし、再現領域内を 10mm ごとにサンプリングして推定を行った。インテグラル立体映像の表示には図 5.14 のようなインテグラル立体ディスプレイを用いた。インテグラル立体ディスプレイを表 5.1 に示す。生成した 3 次元モデルを図 5.15 に、3 次元モデルを変換した要素画像を図 5.16 に、要素画像から生成したインテグラル立体映像を再撮したものを図 5.17 に示す。実験の結果、提案手法を用いることで 3 次元モデル、要素画像およびインテグラル立体映像を生成することができた。要素画像の生成によって実空間における撮影時のカメラ、被写体、ディスプレイの位置関係がインテグラル立体映像で再現されているか検証した。図 5.12 のシーンでは、実空間においてカメラから被写体までの距離が 5000mm、インテグラル立体ディスプレイまでの距離が 5314mm、撮影画角が 47° であった。インテグラル立体ディスプレイの中心を原点とした時の被写体のサイズは 1239mm、被写体の部位で最もディスプレイから飛び出ている被写体の膝の位置は、 $x: -400\text{mm}$, $y: -350\text{mm}$, $z: 139\text{mm}$ であった。表 5.1 のインテグラル立体ディスプレイに表示した時、インテグラル立体ディスプレイの中心を原点とした時の被写体のサイズは 79mm、被写体の膝の位置は $x: -27\text{mm}$, $y: -25\text{mm}$, $z: 7\text{mm}$ であった。実世界と再生された立体像の計算上の縮尺比は 15.73 : 1 であり、計測した被写体の位置およびスケールの比もほぼ同じであることから、実空間における撮影時のカメラ、被写体、ディスプレイの位置関係がインテグラル立体映像で再現されていることが分かった。

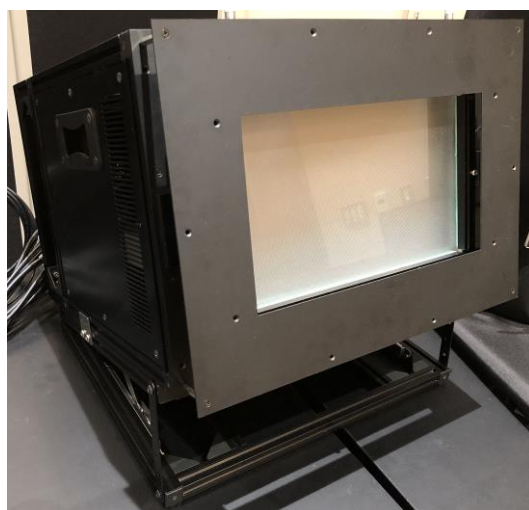


図 5.14 インテグラル立体ディスプレイ

表 5.1 インテグラル立体ディスプレイのパラメーター

レンズアレイ数	294×191
レンズピッチ	1.003mm
レンズの焦点距離	1.74mm
ディスプレイの解像度	7680×4320
画素ピッチ	0.03825
ディスプレイのサイズ	11inch (水平 293mm)

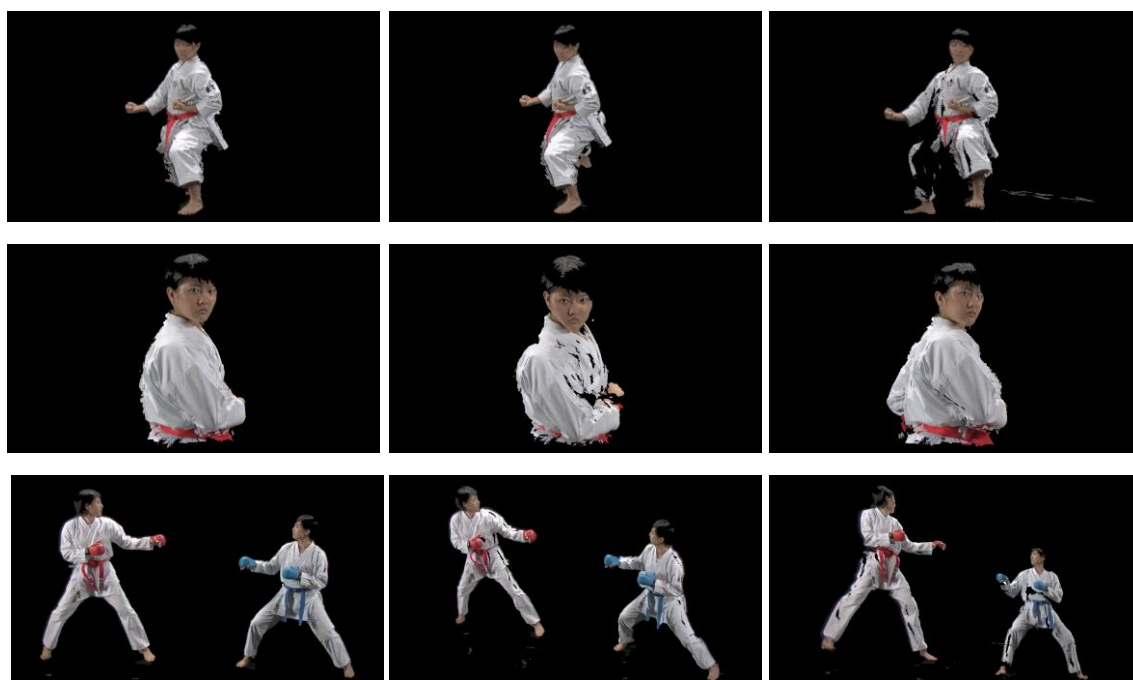


図 5.15 3次元モデル (左：視点正面 中：視点斜め上 右：視点斜め下)



図 5.16 要素画像

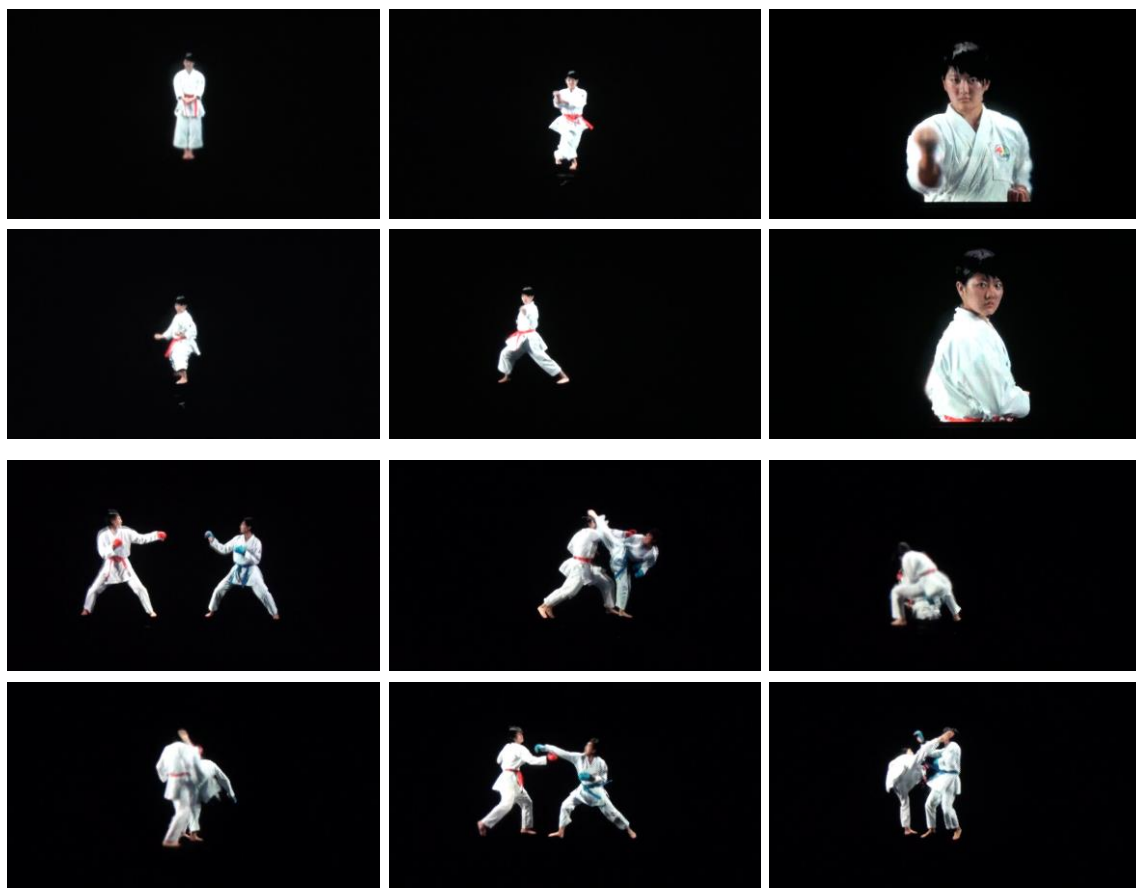


図 5.17 インテグラル立体映像（再撮）

5.4 考察

提案手法の顕著な特徴である多視点ロボットカメラの協調制御について性能を検証した。ダイナミックに動く空手選手のバーストショットのインテグラル立体映像を生成することを考える。図 5.18(a)は、カメラの姿勢や画角が固定された一般的に市販されている多視点カメラで撮影した画像、図 5.18(b)は、その画像から生成した 3 次元モデル、図 5.18(c)は、多視点ロボットカメラの協調制御を用いて撮影した画像、図 5.18(d)は、その画像から生成した 3 次元モデルである。双方の多視点画像は HD カメラ 7 台を正六角形状に配置して撮影し、3 次元モデルは 5.2.2 で述べた手法を用いて生成した。一般的な多視点カメラで空手選手のようなダイナミックに動く被写体のバーストショットを撮影する場合、図 5.18(a)のように動く被写体が画角に納まるよう広い画角で撮影しなくてはならない。そのため、相対的にバーストショットの領域の解像度は低下し、図 5.18(b)のように 3 次元点群が不足し、被写体の表面を形成できない。点群のサイズを大きくしたり、点群間でポリゴンを張ったりすることで表面を形成することはできるが、3 次元モデルおよびインテグラル立体映像の解像度や鮮鋭度が低下する。一方で、多視点ロボットカメラの協調制御を用いることで、図 5.18(c)のようにバーストショットの高解像度な多視点画像を撮影することができる。また、図 5.18(d)

のようにその撮影映像から生成した3次元モデルでは、3次元点群により被写体の表面を形成することができ、解像度や鮮鋭度の低下を抑制できる。図5.18(b)の3次元モデルの点群数が215,277点であったのに対し、図5.18(d)の3次元モデルの点群数は、その約10倍の2,166,035点であった。多視点ロボットカメラの協調制御により高品質な3次元モデルおよびインテグラル立体映像を生成でき、その有効性を確認した。

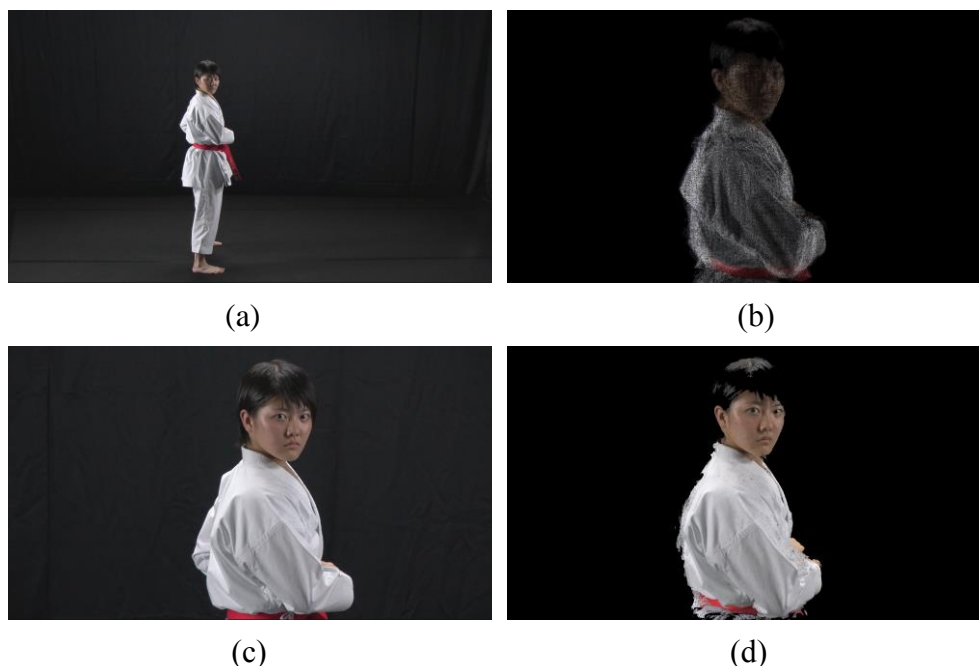


図 5.18 多視点ロボットカメラの協調制御の性能評価
 (a) 固定型の多視点カメラで撮影した多視点映像
 (b) 固定型多視点カメラを用いて生成した3次元モデル
 (c) 多視点ロボットカメラで撮影した多視点映像
 (d) 多視点ロボットカメラを用いて生成した3次元モデル

多視点ロボットカメラの協調制御に関して従来手法と提案手法で性能を評価した。図5.19のシーンにおいて従来手法を用いて協調制御した場合、リファレンスカメラのパン角度は -7.83° 、チルト角度は -9.40° 、撮影画角は 47.08° と制御され、図5.19(a)のような撮影映像となり、提案手法を用いて協調制御した場合、リファレンスカメラのパン角度は -5.34° 、チルト角度は -9.81° 、撮影画角は 54.91° と制御され、図5.19(b)のような撮影映像となった。従来手法と提案手法では、パン角度では -2.49° 、チルト角度では 0.41° 、撮影画角では -7.83° の誤差があり、その結果、従来手法では、図5.19(a)のようにリファレンスカメラで被写体が画角に収まらず、生成したインテグラル立体映像に欠損が生じてしまうことがあったが、提案手法では、図5.19(b)のようにリファレンスカメラでインテグラル立体ディスプレイの再現領域を踏まえて撮影しているため、被写体を画角に収めることができ、インテグラル立体映像の欠損を抑制することができた。

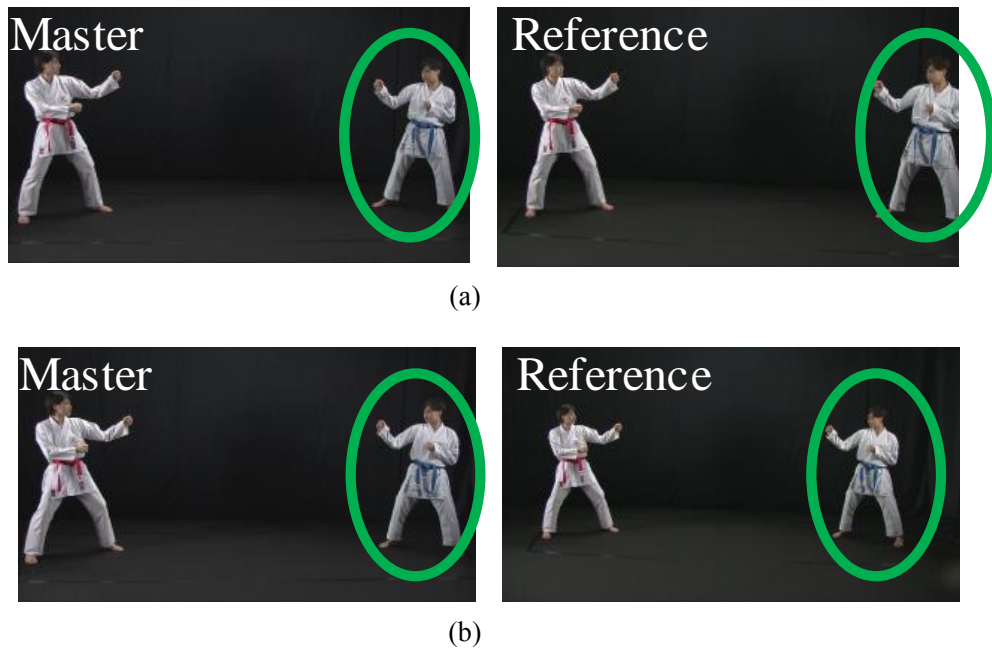
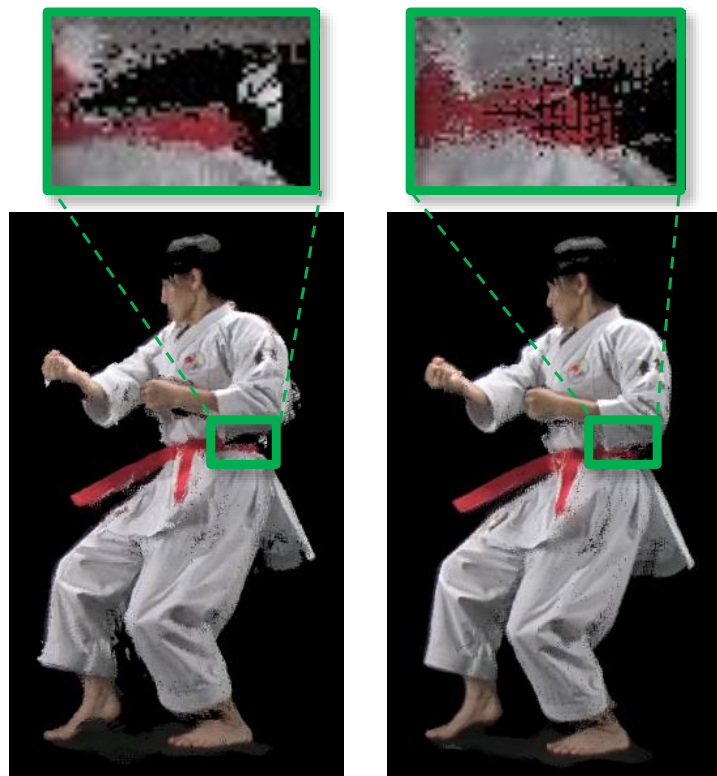


図 5.19 多視点ロボットカメラの協調制御
(a) 従来手法 (b) 提案手法

また、多視点ロボットカメラの配置において、従来手法では水平方向に一次元状に配置しており、垂直方向の光線を取得できないため、図 5.20 のように 3 次元モデルに欠損が生じたが、提案手法では二次元状に配置することで、垂直方向の光線も取得できるため、欠損を抑制することができた。



(a) (b)
 図 5.20 三次元モデルの生成
 (a) 従来手法 (b) 提案手法

5.5 まとめ

多視点ロボットカメラを用いたインテグラル立体撮影手法を提案した。提案手法では、多視点ロボットカメラの協調制御によりインテグラル立体ディスプレイの再現領域の多視点映像を領域が収まる最小の画角で高解像度に撮影した。また、カメラを二次元状に配置することで垂直方向の光線を取得し、二次元状に配置したカメラの多視点映像を相補的に活用して3次元モデルを生成した。そして、カメラマンによるパン・チルト・ズームなどのカメラワークを再現した要素画像およびインテグラル立体映像を生成した。実験では、シミュレーションと撮影実験により、提案手法における多視点ロボットカメラの協調制御の再現性と実用性を確認した。また、提案手法を用いて空手を被写体とした多視点映像を撮影し、3次元モデルおよび要素画像を生成することでインテグラル立体映像を生成し、その有効性を確認した。

第6章 結論

本論文では、多視点ロボットカメラを用いた三次元映像表現として多視点映像表現技術とインテグラル立体映像技術を提案した。

第2章では、多視点映像表現技術およびインテグラル立体映像技術の従来技術について概説し、それらを踏まえた研究課題と課題解決にむけたアプローチについて述べた。

第3章では、多視点ロボットカメラを用いた多視点映像表現技術として多視点ロボットカメラの制御手法およびシステムの設計手法、多視点映像表現の制作手法を提案し、それらを踏まえたシステムを開発した。撮影実験や番組利用を通じて本システムが、3次元空間をダイナミックに移動する被写体や3次元空間中に広く点在する被写体の多視点映像表現を実現できることを確認した。また、放送現場でのシステムの事前準備が短時間で完了し、準リアルタイムで多視点映像表現を生成可能なシステムであることが分かった。これらより、本システムが、従来の多視点映像表現システムと比較し、より汎用的かつ実用的なシステムであることを確認した。

第4章では、多視点カメラを用いたインテグラル立体映像生成手法を提案した。撮影実験により、提案手法によって遠方の被写体のインテグラル立体映像を生成できることを確認した。また、実際に放送で扱った相撲のシーンのインテグラル立体コンテンツを制作することができ、本手法の有効性を確認した。

第5章では、多視点ロボットカメラを用いたインテグラル立体撮影手法としてインテグラル立体映像として再現される実空間領域の多視点映像を撮影する多視点ロボットカメラの制御手法、その撮影映像からの3次元モデル生成手法、3次元モデルから要素画像への変換手法を提案した。実験では、シミュレーションと撮影実験により、提案手法における多視点ロボットカメラの協調制御の再現性と実用性を確認した。また、提案手法を用いて空手を被写体とした多視点映像を撮影し、3次元モデルおよび要素画像を生成することでインテグラル立体映像を生成し、その有効性を確認した。

本論文では、臨場感のある放送の実現にむけて、放送に3次元映像表現技術を応用し、撮影対象の3次元空間情報を表現することを目標とした。そして、放送への応用の過程で、視聴者に新しい体験を継続して提供し続けるために、多視点映像表現技術に関しては現在の放送、インテグラル立体映像技術に関しては未来の放送への応用を目指した。一方で、放送に応用するためには、双方の技術において、スポーツの選手やドラマの俳優などさまざまな被写体の撮影を可能にすることが課題であった。ダイナミックに動く被写体や広い空間に点在する被写体、遠方の被写体といった多様な条件下の被写体を撮影可能な3次元映像撮影技術が必要であった。また、ハイビジョン、4K、スーパーハイビジョンと高精細化をたどる放送に応用するためには、双方の技術において映像の画質を向上させることが課題であった。これらの課題を解決するため、上述のような多視点ロボットカメラを用いた多視点映

像表現技術およびインテグラル立体映像技術を提案した。提案した多視点映像表現技術は、実用化され、現在のスポーツやドラマといったさまざまな放送で利用された。また、提案したインテグラル立体映像技術は、次世代の立体テレビとして研究開発が進められているインテグラル立体テレビの撮像システムとして利用された。提案した多視点ロボットカメラを用いた三次元映像表現は、現在の放送および未来の放送に応用され、これらにより、視聴者に新しい体験を提供することができた。放送に3次元映像表現技術を応用することができ、さらなる臨場感のある放送の実現に寄与することができた。

一方、多視点ロボットカメラを用いた多視点映像表現技術とインテグラル立体映像技術には課題が残されている。多視点ロボットカメラを用いた多視点映像表現技術では、多視点映像表現のさらなる品質の向上が課題である。提案した多視点映像表現技術では、撮影映像から多視点映像表現の映像を生成する処理の過程でデジタルズームをかける必要があるため、出力映像の解像度は撮影したカメラの解像度より低下するという課題がある。そのため、例えば、現在の放送のハイビジョンに多視点映像表現映像を映す場合には、4Kやスーパーハイビジョンなどの高解像度カメラを導入した多視点ロボットカメラで撮影することで解像度の低下を抑制することが望ましい。その場合には、情報量、処理量が大幅に増加するため、システムにおける収録部の大容量化および処理部の高速化が必須となる。また、提案した多視点映像表現技術では、注視点から離れた領域ほど視点切り替えの際に、視点が離散的に切り替わるという課題がある。提案したインテグラル立体映像技術における3次元モデル生成の奥行き推定手法を応用することでカメラ間の視点を内挿し、視点が滑らかに切り替わる多視点映像表現を行うことができるが、一方で、奥行き推定におけるさらなる高精度化および高速化が必須となる。また、システムのさらなる運用性の向上も課題である。番組制作現場では、システムの設定時間は限られている。例えば、カメラ間のケーブル布線作業を削減するためのカメラ間通信のワイヤレス化や、可搬性の向上のためのシステムの小型化や一体化などさまざまな番組制作現場で簡易に利用できるように汎用的なシステムを開発していく必要がある。多視点ロボットカメラを用いたインテグラル立体撮像技術では、インテグラル立体表示装置の高解像度化、広視域化に伴う3次元モデルの高品質化および広視域化が課題である。未来の放送における番組制作現場での運用を見据え、より少ないカメラ台数で360°など広視域の3次元モデルを高品質に生成することが望ましい。3次元モデルの生成精度のさらなる向上とオクルージョン領域やカメラを設置できず撮影できなかった領域の補間処理が必須となる。また、インテグラル立体表示装置の奥行き再現範囲は原理的に有限で、範囲を越えた位置にある被写体の立体像にはぼやけが生じるという課題がある。その課題を解決するために、現在、3次元モデルの奥行きを圧縮することで被写体をその再現範囲内に収め、立体像のぼやけを抑制するという研究が行われている[48]。最終的に奥行きが圧縮されることを踏まえ、3次元モデルの生成時に奥行き圧縮量に応じて奥行き分解能を調整することで3次元モデルの品質向上や処理時間の削減を実現できる可能性がある。

本論文における提案技術は、異なる目的への発展、応用の可能性がある。本論文では、3次元映像表現技術のうち多視点映像表現技術とインテグラル立体映像技術を扱ったが、多視点ロボットカメラはその他の3次元映像表現技術の撮影デバイスとして応用可能である。例えば、コンピュータ上で現実に似た仮想世界を作り出し、あたかもそこにいるかのような感覚を体験できる技術である仮想現実(Virtual Reality)や、現実世界にデジタル情報を付与し、CGなどで作った仮想現実を現実世界に反映する技術である拡張現実(Augmented Reality)では、基本的に2次元映像ではなく3次元映像を扱う。多視点ロボットカメラをそれらの技術の撮像デバイスとして応用することで、ダイナミックに動く被写体や広い空間に点在する被写体、遠方の被写体などの高画質な実写の3次元映像コンテンツを扱うことが可能になり、表現の幅を広げることができる可能性がある。また、立体映像技術においては、インテグラル立体方式の他にも、ホログラフィー方式やレンチキュラー方式などさまざまな技術があり、多視点ロボットカメラは、それらの撮像デバイスとしても応用することができる。3次元映像表現以外にも、3Dプリンターで扱う3DCGデータの入力デバイスとして応用することで、上述のようなさまざまな被写体の3Dオブジェクトを造形できる可能性がある。

本研究における課題が解決されると共に、さまざまな形で発展、応用されることで、本研究が放送やメディア、そして社会や文化のさらなる発展に寄与することを期待する。

謝辞

本論文は2017年9月から1年間、東京工業大学工学院情報通信系の社会人博士課程およびNHK放送技術研究所に在籍し執筆したものです。

本研究を行うにあたり、多大なるご指導ご鞭撻を賜った本学 山口雅浩教授に心から感謝いたします。論文審査等において貴重なご指導とご助言を頂いた本学 小林隆夫教授、熊澤逸夫教授、中本高道教授、金子寛彦教授に深く感謝いたします。

本研究を進める機会を頂き、最後までご理解、ご支援頂いたNHK放送技術研究所の黒田徹前所長、三谷公二所長、今井亨副所長に深く感謝いたします。本研究に関して多大なるご指導ご鞭撻を頂いた本研究所立体映像研究部の岩館祐一前部長、菊池宏前部長、本研究所空間表現メディア研究部の三科智之部長、本研究所研究企画部の洗井淳副部長に感謝申し上げます。本研究にあたり、有益なご討論を頂いた本研究所空間表現メディア研究部の久富健介上級研究員、片山美和主任研究員、加納正規研究員、空間表現メディア研究部3次元映像システムグループの皆様にご感謝申し上げます。

また、システムの番組利用の際にお世話になった番組制作現場の関係各所の皆様、システムの開発に尽力頂いたメーカーの皆様、システムの運用に携わって頂いた番組制作会社の皆様にご感謝申し上げます。

そして、本研究において有益な議論をさせて頂いた山口研究室の中村友哉助教、山口研究室の皆様、事務手続きなどでお世話になりました研究室秘書の志賀久美子様に感謝申し上げます。

最後に、いつもすぐそばで温かく見守ってくれる妻の史と娘の碧に心より感謝します。

参考文献

- [1] free D, <https://www.intel.co.jp/content/www/jp/ja/sports/technology/intel-freed-360-replay-technology.html>
- [2] T. Kanade, P. Rander, and P. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34–47, 1997.
- [3] R.Collins and T.Kanade. Multi-Camera Tracking and Visualization for Surveillance and Sports. *Fourth International Workshop on Cooperative Distributed Vision Proceedings*, page 27-55, March 2001
- [4] O. Grau, M. Prior-Jones, and G. Thomas. 3d modelling and rendering of studio and sport scenes for tv applications. in *Proc. of WIAMIS*.
- [5] A. Hilton, J. Guillemaut, J. Kilner, O. Grau, and G. Thomas. 3D-TV production from conventional cameras for sports broadcast. *IEEE Trans. Broadcasting*, 57(2):462–476, Jun 2011
- [6] Y. Ohta, I. Kitahara, Y. Kameda, H. Ishikawa and T. Koyama. Live 3D Video in Soccer Stadium. *International Journal of Computer Vision (IJCV)*, 75(1):173-187, 2007
- [7] K. Kimura and H. Saito. Player viewpoint video synthesis using multiple cameras. in *IEEEur. Conf. Vis. Media Prod.*, page 112–121, 2005.
- [8] N. Inamoto and H. Saito. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *IEEE Trans. Multimedia*, 9(6):1155–1166, 2007.
- [9] M. Germann, A. Hornung, R. Keiser, R. Siegler, S. Wurmlin, and M. Gross. Articulated billboards for video-based rendering. *Comput. Graphics Forum*, 29(2):585–594, 2010
- [10] 伊佐, 小川, 見持, 川内, 鈴木, 「世界初!プロ野球中継における Eye Vision の活用」, 放送技術, vol.54, no.11, pp.1406-1415(2001)
- [11]F. Okano, J. Arai, H. Hoshino, and I. Yuyama, “Three-dimensional video system based on integral photography”, *Opt. Eng.* 38, pp. 1072-1077, (1999)
- [12]X. Xiao, B. Javidi, M. Martinez-Corral, A. Stern, "Advances in three-dimensional integral imaging: Sensing display and applications [invited]", *Appl. Opt.*, vol. 52, no. 4, pp. 546-560, (2013)
- [13]B. Javidi, J. Sola-Pikabea, M. Martinez-Corral, "Breakthroughs in photonics 2014: Recent advances in 3-D integral imaging sensing and display", *IEEE Photon. J.*, vol. 7, no. 3, Jun. (2015)
- [14]M. Cho, B. Javidi, "Three-dimensional tracking of occluded objects using integral imaging", *Opt. Lett.*, vol. 33, no. 23, pp. 2737-2739 (2008)

- [15] Y. Zhao, X. Xiao, M. Cho, B. Javidi, "Tracking of multiple objects in unknown background using Bayesian estimation in 3D space", *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 28, no. 9, pp. 1935-1940 (2011)
- [16] R. V. Pole, "3-D Imagery and holograms of objects illuminated in white light", *Appl. Phys. Lett.*, vol. 10, no. 1, pp. 20-22 (1967)
- [17] T. Mishina, M. Okui, F. Okano, "Calculation of holograms from elemental images captured by integral photography", *Appl. Opt.*, vol. 45, no. 17, pp. 4026-4036 (2006)
- [18] K. Wakunami, M. Yamaguchi, B. Javidi, "High-resolution three-dimensional holographic display using dense ray sampling from integral imaging", *Opt. Lett.*, vol. 37, no. 24, pp. 5103-5105 (2012)
- [19] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, P. Hanrahan, "Light field photography with a hand-held plenoptic camera", Stanford Univ., Stanford, CA, USA, Tech. Rep. CSTR 2 (2005)
- [20] [online] Available: <https://www.lytro.com>.
- [21] M. Levoy, R. Ng, A. Adams, M. Footer, M. Horowitz, "Light field microscopy", *Proc. ACM SIGGRAPH*, pp. 924-934 (2006)
- [22] A. Lumsdaine, T. Georgiev, "The focused plenoptic camera", *Proc. IEEE Int. Conf. Comput. Photogr.*, pp. 1-8, Apr. (2009)
- [23] C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, M. Pesch, "Refocusing distance of a standard plenoptic camera", *Opt. Exp.*, vol. 24, no. 9, pp. 21521-21540 (2016)
- [24] V. J. Traver, P. Latorre-Carmona, E. Salvador-Balaguer, F. Pla, B. Javidi, "Human gesture recognition using three-dimensional integral imaging", *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 31, no. 10, pp. 2312-2320 (2014)
- [25] J. Arai et al., "Integral imaging system with 33 mega-pixel imaging devices using the pixel-offset method", SPIE, Bellingham, 8167 0X, 2011, pp. 81670X.1-81670X.9 (2011)
- [26] N. Okaichi et al., "Integral 3D display using multiple LCD panels and multi-image combining optical system", *Optics Express*, vol. 25, no. 3, pp. 2805-2817 (2017)
- [27] H. Watanabe et al., "Wide viewing angle projection-type integral 3D display system with multiple UHD projectors", IS&T International Symposium on Electronic Imaging 2017, SD&A-358, p. 67-73 (2017)
- [28] F. Okano, H. Hoshino, J. Arai, I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography", *Appl. Opt.*, vol. 36, no. 7, pp. 1598-1603 (1997)
- [29] J. Arai, F. Okano, H. Hoshino, I. Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images", *Appl. Opt.*, vol. 37, no. 11, pp. 2034-2045 (1998)

- [30] J.-S. Jang, B. Javidi, "Three-dimensional synthetic aperture integral imaging", *Opt. Lett.*, vol. 27, no. 13, pp. 1144-1146 (2002)
- [31] J.-S. Jang, B. Javidi, "Formation of orthoscopic three-dimensional real images in direct pickup one-step integral imaging", *Opt. Eng.*, vol. 42, no. 7, pp. 1869-1870, Jul. (2003)
- [32] J.-S. Jang, B. Javidi, "Three-dimensional projection integral imaging using micro-convex-mirror arrays", *Opt. Exp.*, vol. 12, no. 6, pp. 1077-1083(2004)
- [33] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, G. Saavedra, "Integral imaging with improved depth of field by use of amplitude-modulated microlens arrays", *Appl. Opt.*, vol. 43, no. 31, pp. 5806-5813(2004)
- [34] J. Arai et al., "Compact integral three-dimensional imaging device", *Proc. SPIE*, vol. 9495, pp. 94950I, May (2015)
- [35] M. Miura, N. Okaichi, J. Arai, T. Mishina, "Integral three-dimensional capture system with enhanced viewing angle by using camera array", *Proc. SPIE*, vol. 9391, pp. 939106, Mar. (2015)
- [36] J. Arai et al., "Progress Overview of Capturing Method for Integral 3-D Imaging Displays", *Proceedings of the IEEE* vol. 105, no. 5, pp. 837-849 (2017)
- [37] Noah Snavely, Steven M. Seitz, Richard Szeliski : "Photo Tourism: Exploring image collections in 3D", *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, 25, 3, (2006)
- [38] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient Belief Propagation for Early Vision", *CVPR*, 1, pp.261-268(2004)
- [39] Engin Tola, Vincent Lepetit and Pascal Fua, "DAISY: An Efficient Dence Descriptor Applied to Wide-Baseline Stereo", *PAMI*, 32, 5, pp. 815-830(2010)
- [40] P. Rander, "A Multi-Camera Method for 3D Digitization of Dynamic Real-World Events", *CMU-RI-TR-98-12*(1998)
- [41] M. Katayama and Y. Iwate, "A method for converting three-dimensional models into auto-stereoscopic images based on integral photography", *Proc. SPIE*, 6805, pp.68050 Z.1-68050 Z.8(2008)
- [42] A. Banno, K. Ikeuchi, "Disparity Map Refinement and 3D Surface Smoothing via Directed Anisotropic Diffusion", *3DIM* (2009)
- [43] Y. Taguchi, B. Wilburn, and L. Zitnick, "Stereo reconstruction with mixed pixels using adaptive over-segmentation", *CVPR*, pp. 1-8(2008)
- [44] A. Klaus, M. Sormann and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure", *ICPR*, pp. 15-18(2006)
- [45] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching", *CVPR*, pp. 1458-1465(2010)
- [46] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, M. Gelautz, "Fast Cost-Volume Filtering for Visual Correspondence and Beyond", *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

35, 2, p. 504 - 511 (2013)

[47] K. He, J. Sun, X. Tang, "Guided image filtering", in Proceed. of the 11th European Conference on Computer Vision, vol. 1, pp. 1-14 (2010)

[48] Y. Sawahata and T. morita, "Estimating depth range required for 3-D displays to show depth-compressed scenes without inducing sense of unnaturalness", IEEE Transaction on Broadcasting, vol. 64, no. 2, pp. 488-497 (2018)

研究業績

論文

1. K.Ikeya, J.Arai, T.Mishina, and M.Yamaguchi, “Capturing method for integral three-dimensional imaging using multiviewpoint robotic cameras”, Journal of Electronic Imaging, vol. 27, no. 2, 023022(2018).
2. K. Ikeya and Y. Iwadate, “Multi-viewpoint robotic cameras and their applications,” ITE Trans. Media Technol. Appl. 4(4), 349–362 (2016).
3. 池谷, 久富, 片山, 岩館: “多視点映像からのインテグラル立体映像生成手法 — スポーツシーンの3次元モデル生成と立体映像変換 —”, 映情学誌, 67, 7, pp.J229-J240(2013)

国際会議

1. K.Ikeya, K.Hisatomi, M.Katayama, T.Mishina, and Y.Iwadate,: “Bullet Time Using Multi-Viewpoint Robotic Camera System”, 11th European Conference on Visual Media Production (CVMP 2014), ACM open access, available from <http://delivery.acm.org/10.1145/2670000/2668932/a1-ikeya.pdf?ip> (2014)
2. K.Ikeya, K.Hisatomi, M.Katayama, and Y.Iwadate, “Depth estimation from three cameras using belief propagation: 3D modelling of sumo wrestling,” in Conf. for Visual Media Production (CVMP '11), pp. 118–125 (2011).

研究会・全国大会など

1. 池谷, 「多視点ロボットカメラを用いた「ぐるっとビジョン」」, 映像情報メディア学会スポーツ情報処理時限研究会 (SIP) (2018)
2. 池谷, 洗井, 三科, 山口, 「多視点ロボットカメラを用いたインテグラル立体撮影技術」, 映像情報メディア学会技術報告, vol.41, no.42, 3DIT2017-42, IST2017-72, IDY2017-57, p.15-18(2017)
3. 池谷, 三科, 「多視点ロボットカメラを用いたインテグラル立体撮影技術」, 2017年映像情報メディア学会年次大会講演予稿集,34D-2(2017)
4. 池谷, 三科, 「三次元復元のための多視点ロボットカメラの開発」,2017年電子情報通信学会総合大会講演論文集, 情報・システム 2,D-11-3, p.3(2017)
5. 池谷, 「多視点ロボットカメラの研究」電波技術協会報 FORN, no.313, 2016, p.18-21(2016)
6. 池谷, 「多視点ロボットカメラによる「ぐるっとビジョン」を用いた新しい映像表現技術の開発」,HBF [放送文化基金報],no.85, p.22-23(2016)
7. 池谷, 「多視点ロボットカメラのスポーツへの応用」,2016年映像情報メディア学会年次

- 大会講演予稿集,S4-5, (2016)
8. 池谷, 「多視点ロボットカメラのスポーツへの応用」,映像情報メディア学会誌,vol.70, no.5, p.719-721(2016)
 9. 池谷, 洗井, 「多視点カメラを用いたインテグラル立体における立体再現領域の撮影手法」,2016年電子情報通信学会総合大会講演論文集, 情報・システム 2,D-11-25, p.25(2016)
 10. 池谷, 「多視点ロボットカメラのスポーツ番組利用」,画像ラボ,vol.27, no.1, p.8-13(2016)
 11. 池谷, 加納, 久富, 片山, 三科, 岩館, 「多視点ロボットカメラを用いた Best-Viewpoint Vision」,2015年映像情報メディア学会年次大会講演予稿集,24C-1(2015)
 12. 池谷, 「多視点ロボットカメラを用いたスポーツ番組制作」,映像情報メディア学会誌, vol.69, no.4, p.288-291(2015)
 13. 池谷, 久富, 片山, 三科, 岩館「複数フレームで視点切り替え可能なぐるっとビジョン」,2014年映像情報メディア学会冬季大会講演予稿集,9-1(2014)
 14. 池谷, 久富, 片山, 岩館, 「多視点カメラを用いたインテグラル立体コンテンツ制作手法」,画像ラボ,vol.25, no6, p.36-46(2014)
 15. 池谷, 久富, 片山, 三科, 岩館, 「多視点ロボットカメラを用いたインテグラル立体コンテンツの試作」,2013年映像情報メディア学会冬季大会講演予稿集,5-9(2013)
 16. 池谷, 久富, 片山, 三科, 岩館, 「多視点ロボットカメラを用いた映像制作」,放送技術,vol.66, no.11, p.102-105(2013)
 17. 池谷, 久富, 片山, 岩館, 「多視点ロボットカメラを用いたぐるっとビジョンシステム」, 2013年映像情報メディア学会年次大会講演予稿集,6-2(2013)
 18. 池谷, 久富, 片山, 岩館「Multi-Viewpoint Robotic Camera System」,第16回画像の認識・理解シンポジウム論文集 [MIRU 2013],情報処理学会, OS4-5(2013)
 19. 池谷, 久富, 片山, 岩館, 「弱校正多視点ロボットカメラにおける方向制御手法」,2012年映像情報メディア学会冬季大会講演予稿集,6-4(2012)
 20. 池谷, 久富, 片山, 岩館, 「多視点ロボットカメラを用いた映像表現手法」,2012年映像情報メディア学会年次大会講演予稿集,21-1(2012)
 21. 池谷, 久富, 片山, 岩館, 「被写体の動きをパンフオー可能な多視点ロボットカメラ」,2011年映像情報メディア学会冬季大会講演予稿集,9-3, (2011)
 22. 池谷, 久富, 片山, 岩館, 「テレビ的インテグラル式立体コンテンツの制作に関する一提案 —信頼度伝搬法を用いた距離推定—」,映像情報メディア学会技術報告,vol.35, no.42, 3DIT2011-80, IDY2011-50, IST2011-59, p.5-8(2011)
 23. 池谷, 久富, 片山, 岩館, 「信頼度伝搬法による多視点画像からの距離推定手法 —相撲コンテンツの3次元モデリング—,Depth Estimation from Multi-view Images using Belief Propagation Method —3D Modeling of Sumo Contents—」,2011年映像情報メディア学会年次大会講演予稿集,4-6(2011)
 24. 池谷, 久富, 片山, 岩館, 「階層型信頼度伝搬法を用いた視差推定の性能評価」,第14回

画像の認識・理解シンポジウム論文集 [MIRU 2011] ,情報処理学会, IS4-30, p.1486-1493(2011)

25. 池谷, 久富, 片山, 岩館, 「多視点ロボットカメラの方向制御手法に関する検討」,2010年映像情報メディア学会冬季大会講演予稿集,8-5(2010)
26. 池谷, 久富, 片山, 岩館, 「階層型信頼度伝搬法による視差推定実験」,2010年映像情報メディア学会年次大会講演予稿集,12-3(2010)
27. 池谷, 久富, 片山, 岩館, 「階層型信頼度伝搬法による視差推定に関する一検討」,映像情報メディア学会技術報告,vol.34, no.12, 3DIT2010-40, HI2010-88, p.55-58(2010)

機関紙論文

1. 池谷, 「多視点カメラを用いたインテグラル立体映像の生成手法」, HHK 技研 R&D, No146, p.49-55(2014)

受賞

1. 公益財団法人放送文化基金 第42回放送文化基金賞(2016)
「多視点ロボットカメラによる「ぐるっとビジョン」を用いた新しい映像表現技術の開発」
2. 公益財団法人通信文化協会 第61回前島密賞(個人)(2016)
「多視点ロボットカメラの開発」
3. NHK 会長特賞(個人)(2014)
「多視点ロボットカメラの開発」
4. 映像情報メディア学会 鈴木記念奨励賞(2011)
5. NHK 本部・関東甲信越技術報告会 最優秀賞(2008)
6. 映像情報メディア学会 学生優秀発表賞(2006)
7. 電気通信大学目黒会 目黒会賞(2006)

特許

1. 特開 2011-180675, 特 05200042, 視差推定装置およびそのプログラム
2. 特開 2012-114593, 特 05509044, 多視点ロボットカメラ, 多視点ロボットカメラ制御装置及びプログラム
3. 特開 2013-077131, 特 05764027, 距離指標情報推定装置及びそのプログラム
4. 特開 2013-077132, 特 05764028, 距離指標情報推定装置及びそのプログラム
5. 特開 2013-076621, 距離指標情報推定装置及びそのプログラム
6. 特開 2014-027528, 射影変換映像生成装置及びそのプログラム, 並びに, 多視点映像表現装置
7. 特開 2014-103474, ロボットカメラ制御装置, そのプログラム及び多視点ロボットカメ

ラ

8. 特開 2014-103490, ロボットカメラ制御装置, そのプログラム及び多視点ロボットカメラ
9. 特開 2014-103491, ロボットカメラ制御装置, そのプログラム及び多視点ロボットカメラ
10. 特開 2015-126401, ロボットカメラ制御装置, そのプログラム及び多視点ロボットカメラ
11. 特開 2015-126402, ロボットカメラ制御装置, そのプログラム及び多視点ロボットカメラ
12. 特開 2016-046746, 多視点映像表現装置及びそのプログラム

その他の国際会議

1. M.Kano, H.Okubo, M.Takahashi, K.Ikeya, K.Hisatomi, T.Mishina, “Accurate Calibration of Multiple Pan-Tilt Cameras for Live Broadcasts”, 5th International Conference on 3D Vision (3DV 2017), p.594-602(2017)
2. K.Hisatomi, M.Kano, K.Ikeya, M.Katayama, T.Mishina, Y.Iwadate, K.Aizawa, “Depth Estimation Based on an Infrared Projector and an Infrared Color Stereo Camera by Using Cross-based Dynamic Programming with Cost Volume Filter”, IEEE Transactions on Circuits and Systems for Video Technology, vol.27, no.10, p.2086-2097(2017)
3. M.Takahashi, K.Ikeya, M.Kano, H.Okubo, T.Mishina, ”Robust Volleyball Tracking System using Multi-View Cameras, Proceedings of the 23rd International Conference on Pattern Recognition (ICPR 2016), WePT2.9, p.2741-2746(2016)
4. K.Hisatomi, M.Kano, K.Ikeya, M.Katayama, T.Mishina, K.Aizawa, “Depth Estimation Based on an Infrared Projector and an Infrared Color Stereo Camera by Using Cross-based Dynamic Programming with Cost Volume Filter”, 3rd International Conference on 3D Vision (3DV 2015), p.580-588(2015)
5. K.Hisatomi, K.Ikeya, M.Katayama, Y.Iwadate, K.Aizawa, ” Depth Estimation by Cost Volume with Spatial-temporal Cross-based Local Multipoint Filter using Projecting Infrared Patterns”, 10th European Conference on Visual Media Production (CVMP 2013),10(2013)
6. K.Hisatomi, K.Ikeya, M.Katayama, Y.Iwadate, K.Aizawa, ”Depth Estimation based on Stereo Camera Pairs of Color and Infrared using Cross-based Local Multipoint Filter”, 5th International Conference on 3D Systems and Applications 2013 (3DSA 2013), S1-3(2013)

その他の研究会・全国大会など

1. 大久保, 高橋, 加納, 池谷, 三科, 「3次元被写体追跡スポーツグラフィックスシステム」,画像ラボ,vol.28, no.10, p.30-36(2017)

2. 高橋, 盛岡, 池谷, 横澤, 今井, 三ツ峰, 三科, 「進化するスポーツ中継向け放送技術～新映像表現と新しい考え方の回折放送「音声ガイド」～」, 放送研究と調査, vol.67, no.10, p.94-103(2017)
3. 関口, 池谷, 黒住, 西村, 「MPEG-DASH を用いた多視点映像ストリーミング方式の検討」, 2017 年映像情報メディア学会年次大会講演予稿集, 22D-1, (2017)
4. 加納, 大久, 池谷, 高橋, 久富, 三科, 「センサを用いた複数のパン・チルト・ズームカメラの校正手法」, 第 20 回画像の認識・理解シンポジウム論文集[MIRU 2017], 情報処理学会, PS3-3, (2017)
5. 大久保, 高橋, 加納, 池谷, 三科, 「被写体追跡可能なスポーツグラフィックスシステムの試作—複数の可動カメラを利用したボールの三次元リアルタイム追跡—」, 映像情報メディア学会技術報告, vol.41, no.26, ME2017-85, p.9-12(2017)
6. 高橋, 池谷, 加納, 大久保, 三科, 「多視点カメラを利用したバレーボール追跡処理の頑健化」, Robust Volleyball Tracking Method using Multi-View Cameras, 映像情報メディア学会技術報告, "vol.40, no.41, 3DIT2016-35, IDY2016-44, IST2016-69, p.5-8(2016)
7. 池谷, 高橋, 大久保, 加納, 三科, 「多視点ロボットカメラを用いたスポーツシーンの空間情報解析と映像表現～」, 2016 年映像情報メディア学会年次大会講演予稿集, 32B-1(2016)
8. 大久保, 高橋, 加納, 池谷, 三科, 「スポーツグラフィックスのためのリアルタイム三次元ボール追跡システムの試作」, 2016 年映像情報メディア学会年次大会講演予稿集, 32B-2(2016)
9. 加納, 大久保, 池谷, 高橋, 久富, 三科, 「様々なカメラ配置に対応可能な可動式多視点カメラの校正手法」, 2016 年映像情報メディア学会年次大会講演予稿集, 21C-3(2016)
10. 高橋, 池谷, 加納, 大久保, 三科, 「多視点カメラを利用した相補的オブジェクト追跡手法」, 2016 年電子情報通信学会総合大会講演論文集, 情報・システム 2, D-12-78, p.147(2016)
11. 大久保, 加納, 高橋, 池谷, 三科, 「雲台カメラによる多視点映像と CG を合成するための三次元位置算出に関する検討」, 2016 年電子情報通信学会総合大会講演論文集, 基礎・境界, H-3-11, p.304(2016)
12. 高橋, 池谷, 加納, 大久保, 山内, 三科, 「オブジェクト追跡技術を利用したスポーツ映像解析」, ビジョン技術の実利用ワークショップ ViEW2015 [精密工学会画像応用技術専門委員会主催], OS3-01, p.215-220(2015)
13. 加納, 大久保, 池谷, 久富, 三科, 「現場設置を考慮した可動式多視点カメラの校正手法」, 2015 年映像情報メディア学会年次大会講演予稿集, 24A-2(2015)
14. 加納, 大久保, 池谷, 久富, 三科, 「角度センサを用いた多視点カメラの校正手法と映像表現への適用」, 映像情報メディア学会技術報告, vol.39, no.7, HI2015-31, ME2015-31, AIT2015-31, MMS2015-33, CE2015-33, p.273-278(2015)

15. 加納, 大久保, 池谷, 久富, 三科, 「角度センサを用いた多視点カメラの校正手法」,5-10(2014)
16. 久富, 加納, 池谷, 片山, 三科, 相澤, 「赤外線ドットパターンと赤外線カラーステレオカメラを用いた奥行き推定手法」,映像情報メディア学会技術報告,vol.38, no.51, ME2014-109, p.17-20(2014)
17. 岩館, 片山, 久富, 池谷, 相澤, 苗村, 広田, 趙志鎬, 久保, 山崎, 及川, 小池, 宇都木, 「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」(課題 143) 課題イ 三次元映像通信・放送のための中核的要素技術～インテグラル立体コンテンツ生成のための要素技術の開発～, NICT 産学連携フォーラム講演予稿集, p.64-69(2012)
18. 久富, 池谷, 片山, 岩館, 「複数距離画像からのインテグラル立体映像生成手法」,映像情報メディア学会技術報告,vol.36, no.24, 3DIT2012-34, IDY2012-19, p.7-10(2012)
19. 岩館, 相澤, 苗村, 広田, 山崎, 及川, 小池, 片山, 久富, 池谷, 趙志鎬, 久保, 宇都木, 「インテグラル立体コンテンツ生成のための要素技術の開発」,映像情報メディア学会技術報告,"vol.35, no.42, 3DIT2011-79, IDY2011-49, IST2011-58, p.1-4(2012)
20. 岩館, 片山, 久富, 池谷, 岡野, 「超臨場感立体コンテンツ生成技術の基盤構築～リッチなコンテンツ制作を目指して～」,NICT 超臨場感コミュニケーションシンポジウム [情報通信研究機構主催] (2009)
21. 池谷, 富山, 岩館, 「多視点映像における移動物体抽出とその CG 表現に関する検討」, 電子情報通信学会技術研究報告 IE 画像工学,vol.105, no.611, ITS2005-109, IE2005-316, p.165-170(2006)
22. 池谷, 富山, 岩館, 「テレビ方式,多視点映像における CG 合成による移動物体表現法」,2005 年映像情報メディア学会冬季大会講演予稿集,8-8(2005)