

博士論文

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と その映像制作への応用



東京大学
THE UNIVERSITY OF TOKYO

指導教官 相澤 清晴 教授

東京大学大学院

情報理工学系研究科 電子情報学専攻

48-127420 久富 健介

2015年 12月 11日 提出

概要

本研究は、1台の赤外線プロジェクターと2台の赤外線カラーカメラで構成されたカメラアレイを用いて、映像制作のための被写体の奥行きを推定する手法について検討したものである。プロジェクターは赤外線のドットパターンを照射し、赤外線カラーカメラは、ドットパターンが投影された赤外線画像とドットパターンがなく被写体の表面テクスチャーのみが写ったカラー画像を同時に撮影する。撮影された赤外線ステレオ画像とカラーステレオ画像から Cost volume を生成し、Cost volume filter を用いて、被写体の制約を緩和しながら、安定してデプスマップを推定できる手法を提案する。

本研究では、はじめに赤外線ステレオ画像とカラーステレオ画像の基本的な特性を調査し、これらのステレオ画像から単一フレームのデプスマップを推定する基本処理を検討した。基本処理では、Cross-based Local Multipoint Filter (CLMF) による Cost volume filter を施した Cost volume を用いて、各画素について最小 Cost となる視差を選択することでデプスマップを推定する手法を提案する。また、赤外線ステレオ画像とカラーステレオ画像のそれぞれから生成した2つの Cost volume に対して、それぞれ Cost volume filter を施して算出した2つの Cost volume を、適応型カーネルの大きさにより選択して1つに統合した Cost volume から、デプスマップを推定する手法を提案する。これにより、カメラ間の距離が狭い赤外線カラーステレオカメラから、高い精度のデプスマップを推定することができた。本研究では、この基本処理をベースに、3つの拡張を行うことでデプスマップの推定精度の改善をはかった。

第1に、奥行き解像度を改善するための拡張を行う。映像制作への応用にあたっては奥行き解像度を高くする必要があるが、そのためにはカメラの解像度が一定であるとき、カメラ間を広げる必要である。しかし、カメラ間が広がると射影変換の影響で、カメラに正対していない領域を中心に、左右の画像上で形状が大きく変形する。Cost volume filter は、近傍画素の同視差の Cost を共有することにより、推定精度の改善を図っているが、これは被写体表面がカメラに正対していることを前提としているため、射影変換に対する形状の変化に対応していない。そこで、Cost volume filter を実行したのち、最小 Cost を選択する代わりに、2次元のグラフカットを導入することで、形状変化にも頑強なデプスマップ推定手法を提案する。これにより、奥行き解像度の高いデプスマップを安定して推定することが可能となった。

第2に、複数フレームのデプスマップで構成された時系列デプスマップ上での推定精度を改善するための拡張を行う。Cost volumeと CLMF を時間方向に拡張したのち、Cost volume filter を時間方向に拡張した時空間フィルターを提案し、フレーム間の連続性を考慮したデプスマップ推定手法を提案する。これにより、フレーム毎に独立にデプスマップを推定していた際に、発生していたフリッカーを抑制した時系列デプスマップを推定することができた。

第3に、推定処理の効率化のための拡張を行う。2次元グラフカットを用いることにより、通常の3次元グラフカットを用いたデプスマップ推定処理よりも効率的な奥行き推定を実現したが、グラフカット自体が計算時間と使用メモリ量が大きいアルゴリズムである。そのため、映像制作の応用するにあたっては、多くのフレームを処理して時系列デプスマップを推定する必要があることから、効率的にデプスマップを推定する必要がある。そこで、2次元グラフカットの代わりに、2回の Dynamic Programming(DP)を導入し、それを Cost volume filter 時の2次元平滑フィルターの代わりに用いることで、Cost volume filter 内に最適化アルゴリズムを組み込んで、効果的にデプスマップを推定する手法を提案する。これにより、カメラに正対していない被写体の領域についても、高精度なデプスマップを効率的に推定することが可能となった。

最後に、本研究の最終目標は映像制作への応用であることから、提案した手法を用いて推定した複数の時系列デプスマップを統合して、時系列3次元モデルを生成し、それらを、2次元映像および3次元映像の制作に応用する方法について述べ、提案手法が映像制作への応用の可能性を示す。

目次

第 1 章	序論	11
1.1	研究背景	11
1.2	研究目的	13
1.3	本論文の構成	15
第 2 章	従来のデプスマップ推定手法	17
2.1	受動型カメラを用いたデプスマップ推定手法	18
2.1.1	単眼視	18
2.1.2	ステレオ視	18
2.1.3	多眼視	18
2.2	能動型カメラを用いたデプスマップ推定手法	19
2.2.1	構造化照明法	19
2.2.2	TOF 法	20
2.2.3	照度差ステレオ法	21
2.3	映像制作への応用にあたっての要求条件	21
第 3 章	撮影システム	26
3.1	従来のカラーステレオカメラ	26
3.2	赤外線カラー一体型カメラアレー	26
3.2.1	赤外線プロジェクター	26
3.2.2	赤外線カラーカメラ	27
3.2.3	カメラアレー	29
3.3	赤外線カラー分離型カメラアレー	30
3.3.1	赤外線カメラ	31
3.3.2	カラーカメラ	31
3.3.3	カメラアレー	31
3.4	データセット	32

第 4 章	赤外線カラーステレオ画像からの奥行き推定	35
4.1	はじめに	35
4.2	カラーステレオ画像からのデプスマップ推定	35
4.3	赤外線ステレオ画像からのデプスマップ推定の基本的特性	37
4.3.1	赤外線ドットパターン照射の効果	38
4.3.2	ノイズの影響	39
4.3.3	ブロックサイズ	39
4.3.4	Cost	40
4.4	赤外線カラーステレオからのデプスマップ推定の基本処理	43
4.4.1	Cost volume filter	43
4.4.2	Guided Filter	44
4.4.3	Cross-based Local Multipoint Filter (CLMF)	45
4.4.4	CLMF を用いた Cost Volume Filter	47
4.4.5	赤外線とカラーの Cost volume の統合によるデプスマップ推定	47
4.4.6	赤外線カラー分離型カメラアレーにおける Cost Volume の生成方法	48
4.5	実験結果	49
4.5.1	カラーステレオ画像のみからのデプスマップ推定	49
4.5.2	Cost Volume Filter の効果	50
4.5.3	赤外線とカラーの Cost volume の統合	51
4.6	まとめ	53
第 5 章	2次元グラフカットを用いた奥行き推定	54
5.1	はじめに	54
5.2	グラフカットを用いた従来のデプスマップ推定	55
5.3	2次元グラフカットを用いた奥行き推定	56
5.4	実験結果	57
5.5	まとめ	63
第 6 章	st-Cost Volume Filter による奥行き推定	66
6.1	はじめに	66
6.2	st-Cost Volume	67
6.3	st-CLMF	68
6.4	st-Cost Volume Filter を用いた奥行き推定	69
6.5	実験結果	69
6.6	まとめ	73

第 7 章	DP Cost volume filter による奥行き推定	74
7.1	はじめに	74
7.2	Dynamic Programming	77
7.3	セグメンテーション	78
7.3.1	水平セグメンテーション	78
7.3.2	垂直セグメンテーション	80
7.4	Cross-based DP	80
7.4.1	水平方向の DP	80
7.4.2	垂直方向の DP	82
7.5	DP Cost volume filter による奥行き推定	82
7.6	実験結果	83
7.6.1	赤外線カラーステレオ画像からのデプスマップ推定結果	83
7.6.2	カラーステレオ画像からのデプスマップ推定結果	86
7.7	まとめ	88
第 8 章	映像制作への応用	90
8.1	はじめに	90
8.2	従来手法	91
8.3	複数デプスマップからの 3 次元モデル生成	92
8.3.1	表記	92
8.3.2	Visibility ベースのデプスマップ統合手法	93
8.3.3	欠落領域の補完処理	94
8.4	映像制作への応用	95
8.4.1	2 次元映像制作	95
8.4.2	3 次元映像制作	97
8.5	実験結果	99
8.6	まとめ	102
第 9 章	結論	104
参考文献		108
発表文献		113

目次

1.1	3次元モデルの応用分野	12
1.2	映像制作への応用	12
1.3	本研究の最終目標と目的	14
1.4	本研究の撮影システム	14
1.5	本研究の時系列デプスマップ推定手法	15
2.1	能動型カメラを用いた奥行き推定手法	20
2.2	市販されている能動型カメラ	20
2.3	カメラ間の距離と奥行き解像度の関係	24
2.4	カメラ間隔と撮影画像への射影変換の影響の関係	24
2.5	ブロック処理による Cost の算出	25
3.1	赤外線カラー一体型カメラアレー	27
3.2	撮影画像	27
3.3	赤外線画像内のノイズの様子	28
3.4	赤外線カラーカメラキャリブレーション用の撮影画像	29
3.5	フィッティング前後での赤外線画像とカラー画像の一致の様子	30
3.6	赤外線カラー分離型カメラアレー	31
3.7	赤外線カラー分離型カメラアレーの撮影画像	32
3.8	各データセットのカメラ配置	33
3.9	撮影したシーン	33
4.1	基本的なカラーステレオ画像からの視差画像推定処理	35
4.2	時間 t 、視差 d のときの Cost map $M_t(d)$ と Cost volume V_t	37
4.3	赤外線ステレオ画像からの視差画像推定処理	38
4.4	赤外線ドットパターン照射の効果	38
4.5	赤外線画像上のノイズの影響	39

4.6	異なるサイズのブロックを用いてブロックマッチングにより推定した視差画像 (上段: カラーステレオ画像、下段: 赤外線ステレオ画像)	40
4.7	赤外線ステレオカメラから、異なるブロックサイズを用いてブロックマッチングにより推定したデプスマップ	41
4.8	Cost の定義の違いによるデプスマップ推定精度の違い	42
4.9	“toss” データセットの赤外線ステレオ画像から、Cost 定義とブロックサイズを変えてデプスマップを推定したときの推定エラー率と RMSE	42
4.10	“mannequin75” から、Cost 定義とブロックサイズを変えて、ブロックマッチングでデプスマップを推定したときの推定エラー率と RMSE	43
4.11	赤外線ステレオ画像とカラーステレオ画像からデプスマップを推定する基本処理	44
4.12	Guided filter で用いられるサイズ固定の矩形ブロック	45
4.13	Cross-based 適応型カーネル Ω_p	45
4.14	積分画像を用いた適応型カーネルの処理過程	46
4.15	カラー画像の赤外線画像へのマッピング	49
4.16	カラーステレオ画像のみから Cost volume filter を用いて推定した結果	50
4.17	Cost volume filter の効果 (“mannequin75”)	50
4.18	赤外線 Cost とカラー Cost の統合の効果 (“toss”)	51
4.19	重み付け定数 λ の違いによる推定デプスマップの違い (“toss”)	51
4.20	カーネルサイズによる赤外線 Cost とカラー Cost の選択による奥行き推定	52
5.1	赤外線カラーステレオ画像からのデプスマップ推定手法	55
5.2	グラフカット	55
5.3	一般的なグラフカットを用いたデプスマップ推定法	56
5.4	スキャンライン毎にグラフカットを実施する際の 2 次元グラフ構造	57
5.5	グラフカットを用いず、Cost volume filter のみで推定したデプスマップ	58
5.6	2 次元グラフカットを用いた場合と用いなかった場合の推定結果	59
5.7	2 次元グラフカットを用いたときと用いなかったときの精度の定量的な比較	59
5.8	2 次元グラフカットと 3 次元グラフカットを用いて推定したデプスマップの比較	60
5.9	2 次元グラフカットと 3 次元グラフカットを用いた奥行き推定手法の処理時間と使用メモリ量の比較	60
5.10	推定したデプスマップの比較	61
5.11	グラフカットを用いたときと用いなかったときの推定率の比較	62
5.12	OpenCV のグラフカット手法を用いて推定した視差画像	62

5.13	ベースラインが長いときと、短いときの 3 次元モデルの比較	63
5.14	従来のカメラと高解像度カメラから推定したデプスマップ	64
5.15	従来のカメラと高解像度カメラから推定した 3 次元モデル	64
6.1	赤外線カラーステレオ画像からのデプスマップ推定手法	66
6.2	Cost map $M_t(d)$ と Cost volume V_t の時間拡張	67
6.3	3 次元適応型カーネル Ω_p の水平断面	68
6.4	CLMF と st-CLMF の積分画像を用いた処理過程	69
6.5	CVF(上段) と st-CVF(下段) によるデプスマップの比較	70
6.6	“mannequin460” から CVF と st-CVF で推定したデプスマップの推定率	70
6.7	“toss” と “man” から CVF と st-CVF で推定したデプスマップの前後差分画像の平均輝度値	70
6.8	CVF(上段) と st-CVF(下段) で推定したデプスマップの差分画像の比較	71
6.9	図 3.9(c) の点 A、B、C について、“man” データセットから、CVF と st-CVF で推定した奥行き値	71
6.10	現フレームと次フレーム間の差分の絶対値を 17 フレーム分累積したヒートマップ	72
7.1	2 次元グラフカットと DP Cost volume filter によるデプスマップ推定処理の比較	75
7.2	Cost volume filter と DP Cost volume filter の処理の比較	75
7.3	DP Cost volume filter を用いたデプスマップ推定手法の処理過程	76
7.4	DP の原理	77
7.5	Cross-based セグメンテーション	79
7.6	水平セグメンテーション後の各セグメント	79
7.7	垂直セグメンテーション後の各セグメント	79
7.8	Cross-based DP	81
7.9	水平方向の DP	81
7.10	異なる手法で推定したデプスマップの比較	84
7.11	入力画像の種類の違いによるデプスマップの推定結果の比較	85
7.12	“mannequin75”(data 1) と “mannequin460”(data 2) の処理時間	85
7.13	Middlebury ステレオデータセットから推定した視差画像	87
8.1	本章での各カメラと光軸上の点の表記	93
8.2	複数デプスマップを 3 次元空間に重ね合わせて統合した初期 3 次元モデル Q	94
8.3	欠落領域の補完処理	95

8.4	伝統舞踊アーカイブシステム	96
8.5	3次元モデルの2次元映像制作への応用例	96
8.6	3次元映像制作への応用	97
8.7	レンズアレーを用いた撮影方法	98
8.8	3次元モデルを用いた撮影方法	99
8.9	“comedy” データセット撮影時のカメラ配置	99
8.10	“comedy” データセットを撮影した最も外側のカメラの撮影画像 (上段は赤外線画像、下段はカラー画像)	100
8.11	デプスマップを統合した結果 (上段: 右上からの映像、下段: 上からの映像)	100
8.12	頂点のカラー値とカラー画像上の投影点のカラー値との RMSE	101
8.13	総画素数に対する投影点のパーセント	101
8.14	3次元モデルの映像制作への応用	102
8.15	インテグラル式立体像の再撮影した画像	103
9.1	赤外線ステレオ画像とカラーステレオ画像からのデプスマップ推定	105

表目次

2.1	デプスマップ推定手法の分類	17
2.2	映像制作への応用にあたっての要求条件と従来のデプスマップ推定手法 . . .	22
4.1	各 Cost の定義	36
4.2	適応型カーネルのサイズと Cost の信頼度の関係	48
8.1	各変数の表記	93
9.1	各手法の性能得失比較表	106

第 1 章

序論

本章では、はじめに本研究の背景を述べる。つぎに本研究の最終的な目標が、実空間にある被写体の時系列 3 次元モデルを映像制作へ応用することであることを述べ、その実現のために必要となる本研究の目的について述べる。最後に、本論文の構成について述べる。

1.1 研究背景

コンピューターやグラフィックボードの急速な進展に伴い、3 次元形状モデルをコンピューター内で 3 次元的にシミュレーションするコンピューターグラフィックス (CG: Computer Graphics) が広く普及している。この CG で用いる 3 次元形状モデルは、コンピューターを用いて手作業で制作することがほとんどで、この作業には膨大な作業量を必要とする。また、実空間の被写体から取得したデータからモデルを生成していないため、実物のような臨場感が少ない課題がある。

一方で、実空間の被写体から取得したデータから、直接 3 次元形状のモデルを復元する研究が盛んである [31][16]。本研究では、実空間の被写体から取得したデータから復元した、形状と表面色、メッシュ構造で構成されるモデルデータを 3 次元モデルと呼ぶ。この 3 次元モデルは、文化遺産やロボット工学、映像制作など、広い分野に応用可能である。

文化遺産の分野では、図 1.1 (a) のように有形および無形文化財の保存に用いられているほか、無形文化財については教育面での利用も期待されている [52][12][15]。ロボット工学の分野では、(b) のように、周辺環境を 3 次元モデルとしてシステムに構築することで、進路や動作の決定に用いられている [28]。映像制作の分野では、(c) のように、3 次元モデルを用いて、現在広く普及している映画やテレビのような 2 次元映像や、近年普及しはじめている 3 次元映像の制作に応用する試みが行われている [13][50]。

我々は、これまで映像制作への応用について研究をしてきた。2 次元映像の制作への応用においては、3 次元モデルを用いると被写体の複製が容易に行え、編集時にはカメラアングル自

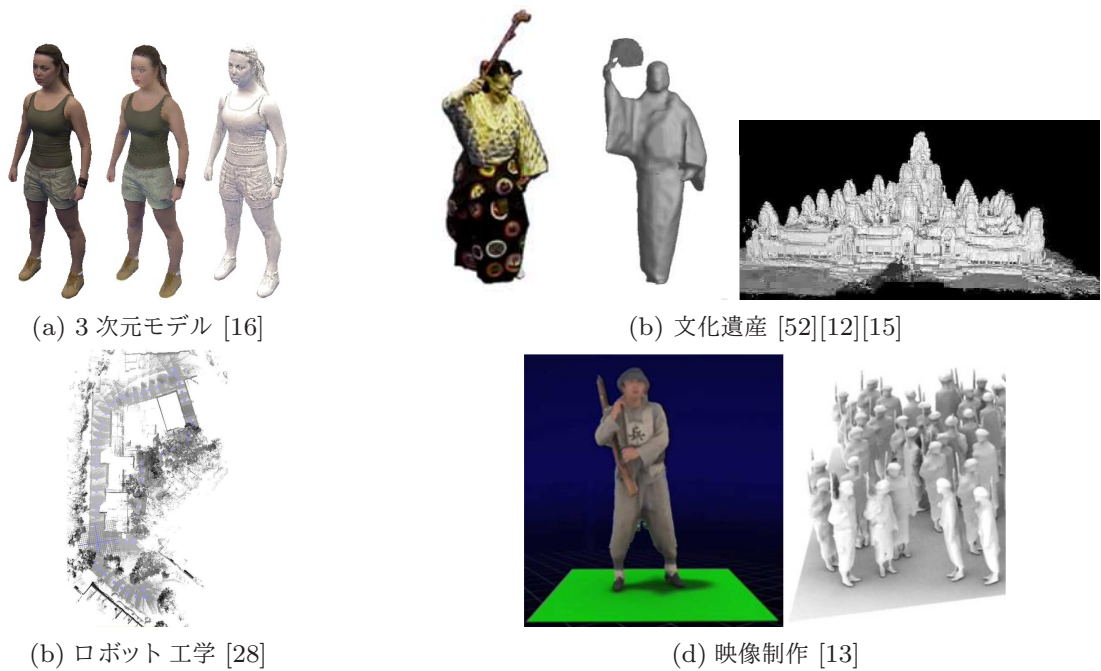


図 1.1 3次元モデルの応用分野



図 1.2 映像制作への応用

由に変更できる。そこで、役者を複数のカメラで同時に撮影した多視点映像から 3次元モデルを生成し、デジタルエキストラとして映像制作へ応用した [53]。図 1.2 (a) は、一人の役者に 1 分間演技してもらい、その様子を撮影した多視点映像から 3次元モデルとして生成したデジタルエキストラを複製して制作した群衆シーンである。この場合、多くのエキストラに演技してもらわなければならない必要がなく、3次元モデルを複製するだけで群衆シーンを効率的に制作することができる。このデジタルエキストラをデータベース化すれば、様々なシーンに再利用することもできる。この 3次元モデルは、データ形式が CG と同じであることから、現状の制作ワークフローと親和性が高い特徴もあわせもつ。

3次元映像制作においても、2眼式では右目用の光線と左目用の光線、フルパララックス式では、空間全体の光線を 3次元モデルから光線追跡法を用いて計算により算出できる。図 1.2 (b) は、3次元モデルから生成したインテグラル式の裸眼 3次元映像である。通常このような

3次元映像は、マイクロレンズが並んだレンズアレー越しにカメラで撮影して制作するが、被写体が遠方にある場合に、立体感が失われる課題がある。3次元モデルから生成する場合、計算機内の仮想空間に被写体の3次元モデルとレンズアレー、カメラを設置して、撮影をシミュレーションして3次元映像を制作することができる。この場合、現実には困難な大型のレンズアレーとカメラを仮想空間内に設置して3次元映像を生成できるため、遠方の被写体も立体感を持って再生することができる [50]。

この3次元モデルを生成する研究は、これまで長期にわたって研究されているが、そのほとんどは2台以上のカメラで構成される多視点カメラを用いて撮影した多視点映像(画像)から生成する。詳細は2章で述べる。多視点カメラの構築は複数のカメラが必要となることから価格的に困難であったが、近年のカラーカメラの小型化・高性能化により、構築が容易になってきている。

また多視点映像から生成する手法のほかにも、可視光や赤外線的光線を照射し、その反射光を観測することにより被写体の3次元モデルを取得できるセンサーも、小型化、低価格化しており、静止物だけでなく動物体についても、取得できる環境が整いはじめている [46]。近年では、各画素の奥行き情報が2次元配列として格納されたデプスマップを高フレームレートで取得できるゲーム用の機器も市販されるようになり、動く被写体についても容易に3次元モデルを取得することができるようになってきている [22]。

取得したデプスマップや3次元モデルの処理は膨大なデータとなるが、コンピューターやグラフィックボードの急速な性能向上により、それらを扱える環境も整ってきている [17]。このような背景から、実世界の3次元モデルを取得し、広い分野に応用する研究は、近年非常に盛んになっている。しかし、3次元モデルをCGと同じ要領で、2次元映像や3次元映像を制作するケースは多くない。これは3次元モデルを取得する際に被写体への制約が多く、安定して取得できないことが理由の一つとして挙げられる。

1.2 研究目的

本研究では、3次元モデルを2次元および3次元映像制作へ応用することを最終的な目標とし、動く被写体(動物体)の時系列3次元モデルを安定して高精度・高解像度に求めることが目的である。そのため、被写体への制約は極力軽減する必要がある。また、映像制作にあたっては、撮影のやり直しは困難であることから、安定して確実に3次元モデルを取得する必要がある。さらに、これまでの2次元映像制作と同様に、解像度も平面方向(XY方向)および奥行き方向(Z方向)に高く、高精度であることが重要となる。時間方向の解像度についても、現在の映像と同様に毎秒30フレーム以上の高いフレームレートが要求される。一方で、生放送でない限り、リアルタイムで処理する必要はないため、本研究ではリアルタイム性は追求せず、その他の要求条件を優先する。しかしながら、このような高解像度の3次元モデルを生成

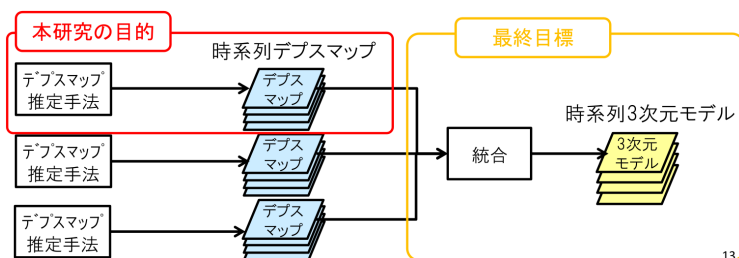


図 1.3 本研究の最終目標と目的

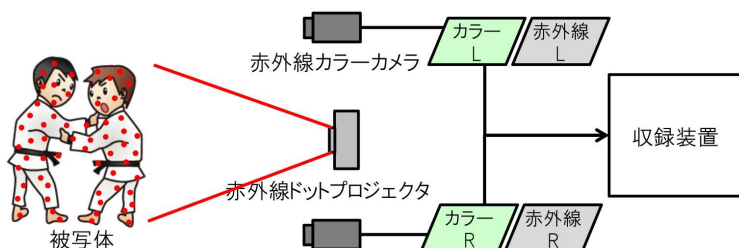


図 1.4 本研究の撮影システム

には、膨大な処理を要するため、リアルタイム性を考慮しないとしても、処理時間を考慮したアルゴリズムの開発が必要となる。

3次元モデルの生成方法には、多視点画像から一度に3次元モデルを推定する多眼的アプローチと、2つの位置から撮影したステレオ画像から、1つの撮影点からのデプスマップを推定し、複数のデプスマップを統合して3次元モデルを生成するステレオマッチング的アプローチに分けることができる。

多眼的アプローチは、一般的に被写体の対象領域をボクセルという小空間に分け、それぞれのボクセルが被写体の表面である可能性を評価して、一度に3次元モデルを推定する手法であるが、一般的にカメラを等間隔に置く必要がある。一方、ステレオマッチング的アプローチは、2台のカメラセットを離散的におけばよく、カメラの設置条件が多眼的アプローチに比べて少ない。

そこで、本研究では図 1.3 のように、映像制作への応用を考慮してステレオマッチング的アプローチにより時系列デプスマップを推定することを目的とし、複数撮影点からの複数の時系列デプスマップを統合することにより、最終的な目標である時系列3次元モデルを生成する。

しかし、ステレオマッチングは左右画像の各画素に対応を見つけ出し、三角測量によって奥行き値を求めるため、被写体表面にテクスチャがない場合に手がかりがないために、左右画像の各画素の対応付けができない原理的な課題がある。そこで、本研究では図 1.4 のように、赤外線のプロジェクターを用いて被写体表面にドットパターンを照射し、その投射の様子を、赤外線画像とカラー画像を同時に取得することができる2台の赤外線カラーカメラを用いて撮影する。このようにして撮影した赤外線画像には、赤外線プロジェクターから照射されたドッ

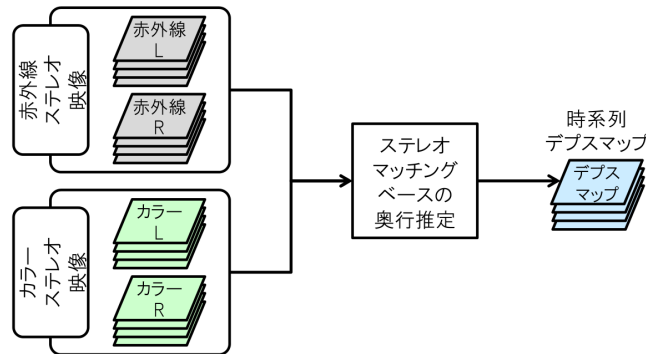


図 1.5 本研究の時系列デプスマップ推定手法

トパターンが観測され、赤外線ステレオ画像上のドットパターンを手がかりに奥行きを推定することができる。一方で、カラー画像には赤外線のドットパターンが観測されないことから、被写体の表面テクスチャーを取得することができる。映像制作への応用を考慮すると、被写体の 3 次元形状を取得することと同時に、表面テクスチャーを取得することが重要であることから、表面テクスチャーに影響を与えることなく、対応付けの手がかりを与えることは重要となる。

図 1.4 は、本研究で用いる撮影システムの概要を示したものである。この撮影システムを用いると、1 対の赤外線ステレオ画像と 1 対のカラーステレオ画像を同時に取得することができる。これらのステレオ画像から、図 1.5 のようにステレオマッチング的アプローチにより、時空間方向に高解像度で高精度な時系列デプスマップを安定して推定することが、本研究の目的である。

1.3 本論文の構成

本論文は、9 章で構成される。第 1 章では、本研究の背景と目的について述べた。第 2 章では、実空間の被写体のデプスマップを取得する従来手法を、受動的手法と能動的手法に分けて述べる。また、3 次元モデルを映像制作に応用するにあたっての要求条件を述べ、従来手法と提案手法に照らし合わせ、本研究のアプローチの必要性を示す。第 3 章では、本研究で撮影に用いたカメラアレーについて述べる。第 4 章では、第 3 章で述べたカメラアレーを用いて撮影した、赤外線ステレオ画像とカラーステレオ画像からデプスマップを推定するにあたっての基本的な特性を調査するとともに、Cross-based Local Multipoint Filter (CLMF) による Cost Volume Filter を用いて、単一フレームのデプスマップを推定する基本処理を提案する。第 5 章では、奥行き解像度を高くするための拡張を行う。Cost Volume Filter を施した Cost Volume から、画素毎に Cost が最小となる視差を選択する代わりに、最適化アルゴリズムの一つであるグラフカットを用いることにより、被写体のうち、表面がカメラに正対してい

ない領域の推定精度の改善をはかる。また、従来の 3 次元のグラフカットを用いる代わりに、2 次元のグラフカットをスキャンライン毎に実施することで、処理時間とメモリ使用量の削減をはかる。第 6 章では、複数フレームで構成された時系列デプスマップを推定のための拡張を行う。Cost Volume Filter を時間方向に拡張することによって、時間方向の連続性を考慮し、時系列デプスマップにおけるフリッカーを抑制する奥行き推定手法を提案する。第 7 章では、推定処理の効率化のために、Cost volume filter 内に最適化アルゴリズムを組み込む。Cost volume filter では 2 次元平滑化フィルターを用いるが、カラー画像を用いてセグメンテーションを行った各小領域に対して、1 次元最適アルゴリズムである Dynamic Programming(DP) を 2 回施すことにより、2 次元グラフカットによる推定精度を保持しながら、処理時間とメモリ使用量を削減する手法を提案する。第 8 章では、本研究で提案手法で推定したデプスマップを統合して時系列 3 次元モデルを生成し、2 次元映像制作と 3 次元映像制作それぞれへの応用について述べ、本研究の有効性について述べる。最後に、第 9 章で、本論文のまとめと今後の展望を述べる。

第 2 章

従来のデプスマップ推定手法

実空間の被写体のデプスマップを取得する研究は、長期に渡って研究されており、多くの手法が提案されている [31][16]。表 2.1 は、各手法を分類したものである。デプスマップの推定方法には、カメラから可視光や赤外線を照射する能動型と、照射せずに観測したデータから推定する受動型と大きく 2 つに分けることができる。本章では、受動型と能動型に分けて、それぞれの代表的な手法について述べる。また、3 次元モデルを映像制作に応用するにあたってのデプスマップ推定手法への要求条件を挙げ、各手法と照らし合わせて課題を述べる。

表 2.1 デプスマップ推定手法の分類

区分	波長	小区分	手法	課題	
受動型	可視光	単眼視	合焦法	× 動画	
			ステレオ視	対応点探索	× 任意テクスチャー対応
				グラフカット	
		確率伝搬法			
		多眼視	視体積交差法	× 凹部の復元ができない	
			Voxel Coloring	× 表面の連続性が考慮されない	
			Structure from Motion	× 高解像度	
グラフカット	× 任意テクスチャー対応				
能動型	可視光 / 赤外線	構造化照明法	スポット 光法	× 解像度 × 複数同時取得 (干渉)	
			スリット 光法		
			コード化パターン法	× 安全性 (赤外線光量), × 動画対応 × 複数同時取得 (干渉)	
			干渉法	× 安全性 (赤外線光量), × 解像度 × 複数同時取得 (干渉)	
		TOF 方式	× 解像度		
		照度差ステレオ法	× 安全性 (赤外線光量), × 動画対応 × 複数同時取得		

2.1 受動型カメラを用いたデプスマップ推定手法

受動型カメラを用いたデプスマップ推定方法は、カメラの構成により単眼視、ステレオ視、多視点に分けることができる。

2.1.1 単眼視

単眼視は、1台のカメラで撮影した映像から、デプスマップを推定する手法で、代表例として合焦法が挙げられる。合焦法は、レンズ移動法とも呼ばれ、カメラのピントを変化させた時の合焦情報から、奥行きを推定する手法である。

2.1.2 ステレオ視

ステレオ視は、2台のカメラで撮影したステレオ画像から、左右の画像中の各画素について対応関係を算出し、三角測量の原理からデプスマップを推定する手法である。この手法は長期に渡って非常に多くのアプローチが提案されている [31]。左画像の対象画素を中心としたブロックと、右画像の同位置から水平方向に視差分並進移動した位置を中心としたブロックとの類似度を算出し、類似度が最大となる視差から三角測量により奥行きを推定する。近年では、グラフカット法 [3] や確率伝搬法 [5] といった最適化アルゴリズムを用いて、高精度に奥行きを推定できるようになってきた [44]。

2.1.3 多眼視

多眼視では、被写体領域をボクセルと呼ばれる小さな 3次元領域に分けるのが一般的である。そして、各ボクセルについて 3台以上のカメラで撮影した多視点画像に投影して、被写体の表面、もしくは内部/外部である可能性を算出して、3次元モデルを一度に推定する手法である。多眼視には、視体積交差法や Voxel Coloring 法、グラフカット法などがある。視体積交差法は、多視点画像を前景と後景に分けたシルエット画像を用いて 3次元モデルを取得する。各ボクセルをすべてのシルエット画像に投影し、投影点がすべて前景に含まれる場合には、そのボクセルは残し、そうでないボクセルは削除することにより、3次元モデルを生成する。調整すべきパラメーターが少ない上に、安定してモデルを生成できるが、原理的に凹部の復元ができない課題がある [13][53][54]。Voxel Coloring 法は、各ボクセルをすべての画像に投影し、投影点の輝度値の一致度が一定以下の場合に、当該ボクセルを削除することにより、3次元モデルを生成する手法である [32][30]。シルエット画像を求めずに、3次元モデルを推定できるが、キャリブレーション誤差等により、被写体表面にも関わらず、投影点の輝度値が一

致せず、過剰に削りすぎてしまう課題がある。グラフカット法は、はじめに視体積交差法で形状の概形を求めたあと、残ったボクセルを当該ボクセルから見えるカメラ画像に投影し、それらの類似度で定義した Cost を算出する。そして、ボクセルをノード、ボクセル間の線をエッジとしてグラフを構成し、視体積交差法で求めた 3 次元モデルの表面に Source、3 次元モデルの芯に Sink を設置して、最適化 2 分割問題を解くグラフカットを用いて、被写体の内部のボクセルと外部のボクセルとを分離することにより、3 次元モデルを生成する [12][43][42][51]。

2.2 能動型カメラを用いたデプスマップ推定手法

能動型カメラは、カメラから可視もしくは赤外線的光を照射し、その投影の様子を撮影することにより、デプスマップを推定する手法である。照射する光の変調方法によって、2 種類に分類される。空間方向に変調した光を照射してデプスマップを推定する方法を構造化照明法 [46] と呼び、時間方向に変調した光を照射してデプスマップを推定する方法を TOF(Time-of-Flight) 方式 [7] と呼ぶ。

2.2.1 構造化照明法

構造化照明法は照射するパターンにより、スポット光法、スリット光法、コード化パターン法、干渉法に分類することができる。最もシンプルな手法は、スポット光法である。スポット光法は、プロジェクターからスポット光を被写体に照射し、プロジェクターを設置した位置と異なる位置に 1 台のカメラを設置して撮影する。プロジェクターとカメラの位置と方向が既知の場合、撮影されたスポット光の位置を用いて三角測量の原理からスポット光が投影された点の 3 次元座標を取得することができる。このスポット光を順次走査することにより、被写体全体の形状を取得することができる。

スポット光法の発展形として、スリット光法が挙げられる。スリット光法は光切断法とも呼ばれ、スポット光の代わりに、プロジェクターとカメラを結ぶ直線に垂直なスリット光を照射し、プロジェクターの位置と異なる位置に設置したカメラで投射の様子を撮影して、奥行き情報を取得する方法である。プロジェクターから照射された光線は、被写体までの奥行きにより、プロジェクターとカメラを結ぶ直線方向に平行移動することから、スリット光をその直線と垂直方向に照射すると、1 次元の照射で一括して、その垂直方向のすべての画素について奥行きを取得することができる。そのため、2 次元走査が必要なスポット光法に比べて、1 次元走査でデプスマップを取得することができるため、より効率的に奥行き情報を取得することができる。

コード化パターン法は、スリット光法をさらに発展させた手法で、図 2.1(a) のように、輝度値を決まったルールで変化させた縦縞パターンを照射することにより、スリット光を走査する

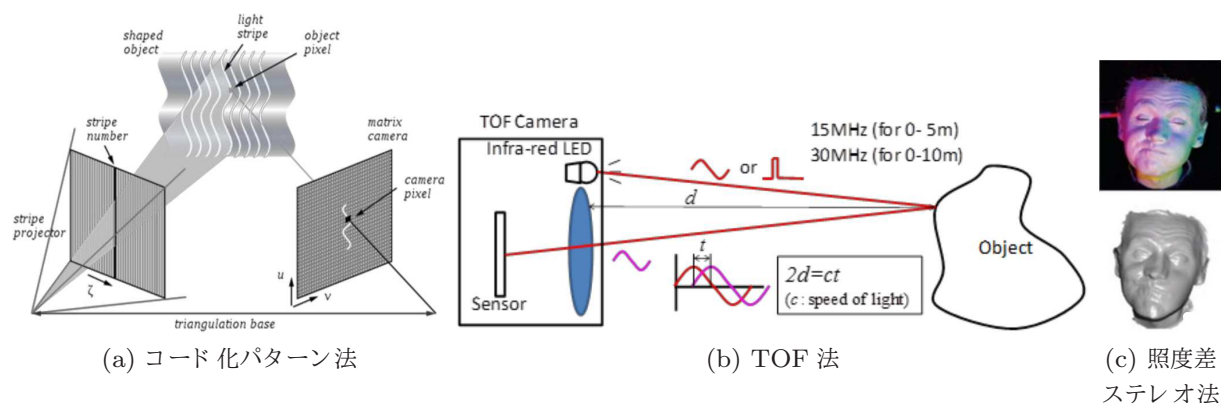


図 2.1 能動型カメラを用いた奥行き推定手法

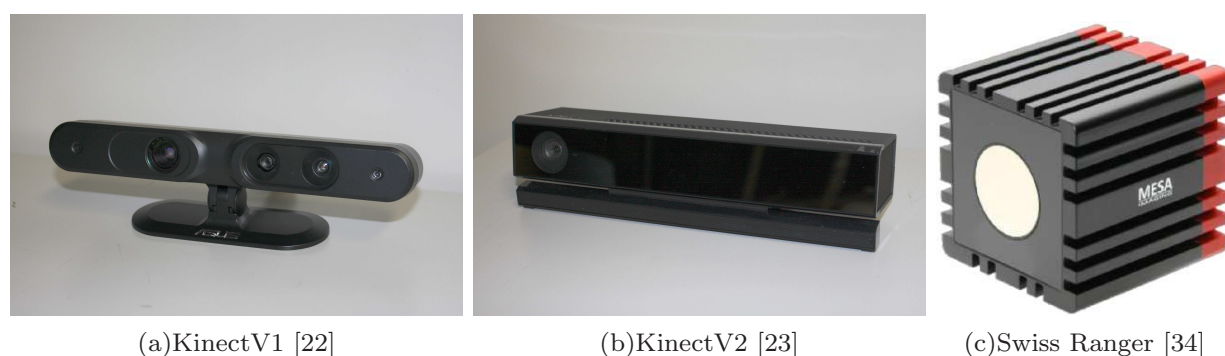


図 2.2 市販されている能動型カメラ

ことなしに、撮影画像中の各画素がどの程度平行移動したかを、既知のパターンと照らしあわせて求める手法である [46]。スリット光を 1 次元走査する必要がないため、より効率的に奥行きを取得できるほか、動物体のデプスマップをフレーム毎に推定した時系列デプスマップを取得することができる。縦縞の代わりに、ドットをランダムに散りばめたランダムドットパターンを照射する手法もある。この手法を取り入れた、図 2.2 (a) のようなゲーム機用の Kinect V1 センサーは安価なうえ、ビデオレートでデプスマップを取得できることから広く普及している [22]。

干渉法は、2 つの異なる位置から、繰り返しパターンを照射し、その干渉の様子をカメラで撮影することにより、奥行き情報を取得する手法である。奥行き解像度の非常に高いデプスマップを取得することができるが、奥行き推定範囲が狭い課題がある。

2.2.2 TOF 法

TOF 法は、図 2.1(b) にあるように、レーダーの原理と同様に、パルス波やサイン波といった時間方向に変調した光を照射して、その反射波の遅延量から奥行き距離を計測する手法である。光線が発射されてから、反射してカメラに戻ってくるまでの光線の飛行時間を、遅延量か

ら算出することができるため、その時間に光速をかけて被写体までの往復の距離を求めることができる。この手法を取り入れた、図 2.2 (a) のような Swiss Ranger [34] や、図 2.2 (b) のような KinectV2 [23] といった製品も販売されている。これらの製品は、近赤外線光を用いるため、被写体のテクスチャーに影響を与えずにビデオレートでデプスマップを取得することが可能である。この TOF 法を用いたデプスマップを取得するカメラを、レンジセンサーやレンジカメラとも呼ぶ。

2.2.3 照度差ステレオ法

照度差ステレオ法は、3 つ以上の区別可能な光源で被写体を照らし、各光源の光の反射を観測して、3 次元形状を取得する手法である [29]。この手法は、被写体の表面が一様に光を反射するランバート反射であることを前提に、反射光から表面の法線ベクトルを推定することで形状を高精度に取得することができる。通常、RGB など波長の異なる光を同時に照射するか、同様の光源をタイミングを変えて照射することで、光源を区別する。照明を当てることから、テクスチャーの同時取得ができないほか、複数回照射する必要から、高フレームでの取得ができない課題がある。

2.3 映像制作への応用にあたっての要求条件

2.1 節や 2.2 節で述べたように、受動型カメラや能動型カメラを用いてデプスマップを推定することができる。しかし、それらを映像制作に応用するためには、以下の要求条件を満たす必要があり、従来の手法でこれらの条件を満たすものはないため、映像制作に応用することは困難である。

【映像制作への応用にあたっての要求条件】

1. 高解像度：取得したデプスマップは、空間解像度が高いこと。
2. 高精度：取得したデプスマップは、精度が高いこと。
3. 形状・テクスチャーの同時取得：形状と同時にテクスチャーを取得が可能なこと。
4. 複数デプスマップの同時取得：複数のデプスマップを同時に取得できること。
5. 動画対応：動く被写体に対応していること。
6. 任意被写体 (テクスチャー)：テクスチャーに依存せず安定して取得できること。
7. 任意被写体 (形状)：被写体の形状に制約が少ないこと。
8. 任意被写体 (素材)：被写体の素材に制約が少ないこと。

1 は、映像制作では高い画質が求められるため、デプスマップも水平垂直方向、そして奥行き方向にも高い解像度を有する必要がある。同様の理由から、2 の精度についても高い精度が

表 2.2 映像制作への応用にあたっての要求条件と従来のデプスマップ推定手法

手法	解像度	精度	テクスチャ同時取得	複数デプス同時取得	動画	任意テクスチャ	任意形状	任意素材	課題
合焦法	○	×	○	○	×	×	○	×	精度が悪く、動画に対応できない
対応点探索	○	△	○	○	○	×	○	×	テクスチャがないと、推定精度が低下する カメラ間距離を広げないと奥行き解像度が低い
グラフカット	○	△	○	○	○	×	○	×	
確率伝搬法	○	△	○	○	○	×	○	×	
SfM	○	○	○	○	○	×	○	×	
視体積交差法	○	×	○	○	○	○	×	×	凹部を復元できない
Voxel Coloring	○	×	○	○	○	×	○	×	精度が悪く、テクスチャがないと精度が低下する
スポット光法	○	○	○	×	×	○	○	×	複数デプスマップを同時に取得できない
スリット光法	○	○	○	×	×	○	○	×	
コード化パターン法	○	○	×	×	○	○	○	×	
干渉法	○	○	×	×	○	○	○	×	
TOF 方式	×	○	○	△	○	○	○	×	解像度が低く、同時に複数取得すると精度が低下する
照度差ステレオ法	○	○	×	×	○	○	○	×	テクスチャーを同時に取得できず、複数同時に取得できない
提案手法	○	△	○	○	○	○	△	×	カメラに正対していない領域の精度低下に対応が必要

求められる。特に映像制作においては、3次元モデルにテクスチャーマッピングすることが前提となっているため、精度が低いと制作した映像に大きな悪影響を与える。3は、前述の通り、映像制作では、テクスチャーマッピングが前提となっているため、デプスマップと同時にマッピングする被写体のテクスチャーを取得する必要がある。4は、デプスマップが1視点からの奥行き情報しか有していないことから、視点をずらすと3次元情報のない欠落領域が多く発生してしまう。そのため、複数のデプスマップを統合して3次元モデルを生成するため、複数のデプスマップを同時に取得する必要がある。5は、映像制作の多くの被写体は動物体であることから、ビデオフレームの時系列デプスマップを推定するため、映像制作の現場では、撮影後に撮り直しをすることが困難であること、また、デプスマップ推定のために被写体の条件を狭めることは困難であることから、無地の被写体は避けるなどのテクスチャーの制約を緩和するために必要な条件である。7についても、デプスマップ推定のための被写体の形状に関する制約は極力少ないことが望ましいから設定された条件である。8も、被写体の制約を緩和するための条件で、反射や半透明などの素材の条件が極力少ないことが望ましい。

これらの要求条件と従来のデプスマップ推定手法を照らしあわせたのが、表 2.2 である。合焦法は精度が悪く、レンズを移動しながら奥行きを推定するため、動画に対応できない課題が

ある。また、対応点探索やグラフカット、確率伝搬法、Structure from Motion など、カラー画像を用いた奥行き推定手法は、総じて被写体の表面にテクスチャーが十分存在しないとデプスマップの推定精度が低下してしまう課題がある。視体積交差法については、テクスチャーが十分なくても、シルエット画像が正確に取得することができれば、3次元モデルを取得することができるが、被写体の凹部を復元できない課題がある。

能動型カメラを用いたデプスマップ推定方法は、照射するパターンを既知として奥行きを推定するため、複数のパターンを照射して、複数のデプスマップを同時に取得しようとする、照射したパターン同士が互いに干渉してしまうために精度が低下してしてしまう課題がある。TOF法は、3台程度までの同時運用までは、複数観測データの累積により、干渉によるエラーをある程度排除することができ、複数デプスマップの同時取得も可能であるが、それ以上の台数で同時運用すると、互いに干渉してしまい精度が低下してしまう。また、画素構造が複雑なことから、高解像度化が困難な点も課題である。照度差ステレオ法は、3つ以上の区別可能な光源からの照明が必要となることから、赤外線的光源を照射する際には、異なる波長の赤外線を3つ以上照射する必要がある。しかし、赤外線は目の瞳孔が反応しないため、目の安全性を考えると複数の光源を全面に常時照射することは困難である。可視光を照射する場合、表面テクスチャーに波長の異なる可視光を複数照射する必要があり、色のついた光線を照射することから、形状とともにテクスチャーを同時取得することが困難となる。また、複数のデプスマップを同時に取得する場合には、光源からのオクルージョンを考慮すると、さらに多くの光源が必要なことから、複数デプスマップの同時取得は困難である。

したがって、映像制作への応用を考慮すると、従来のデプスマップ推定手法はいずれも課題があり、新たなデプスマップ推定手法を検討する必要がある。本研究では、これらの状況を鑑みて、第1章で述べたような、能動型と受動型の両方を取り入れたハイブリッドのデプスマップ推定手法を提案するものである。

提案手法をこの要求条件に照らし合わせると、表 2.2 の最終行のようになる。第1章で述べた撮影システムでデプスマップを推定する場合には、奥行き推定精度を高くするために、カメラ間距離を広げる必要がある。これは、図 2.3 のように、カメラ間隔が広がると1視差に対応する奥行きが狭まり、奥行き解像度が高くなるためである。一方で、カメラ間隔が広がると射影変換の影響で、撮影した画像が左右のカメラで大きく変形する課題がある。図 2.4 は、赤の位置に左カメラを設置したときに、カメラ間隔が狭い青の位置に右カメラを設置したときと、カメラ間隔が広い緑の位置に右カメラを設置したとき (a) の撮影画像のエッジを重畳した画像 (b) である。カメラ間隔が狭いとき、それぞれのエッジである赤線と青線はほぼ重なり、射影変換の影響が少なく、左右画像の各画素の対応付けが容易にできる。しかし、カメラ間隔が広がると射影変換の影響を大きく受けて、赤線と緑線は大きく異なり、対応付けが困難になる。特に、カメラに正対していない床や壁といった領域は、射影変換の影響は大きく、2次元画像上での形状変化にロバストな対応付けの手法が必要となる。本研究では、この形状変化に

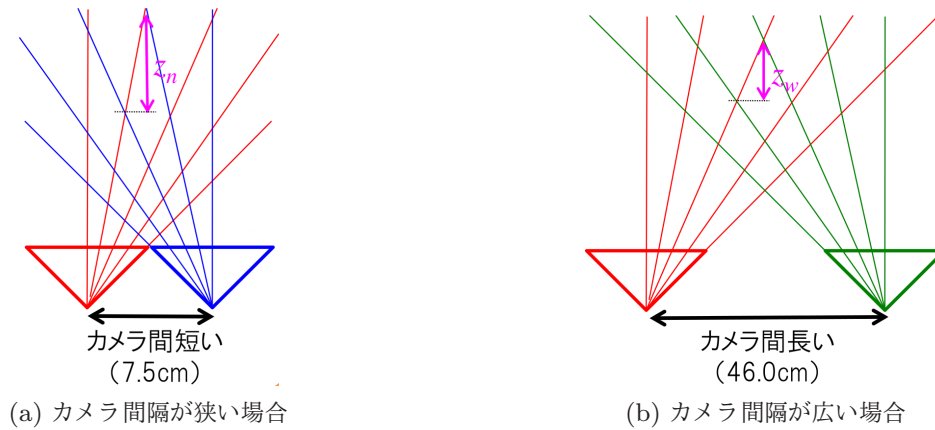


図 2.3 カメラ間の距離と奥行き解像度の関係

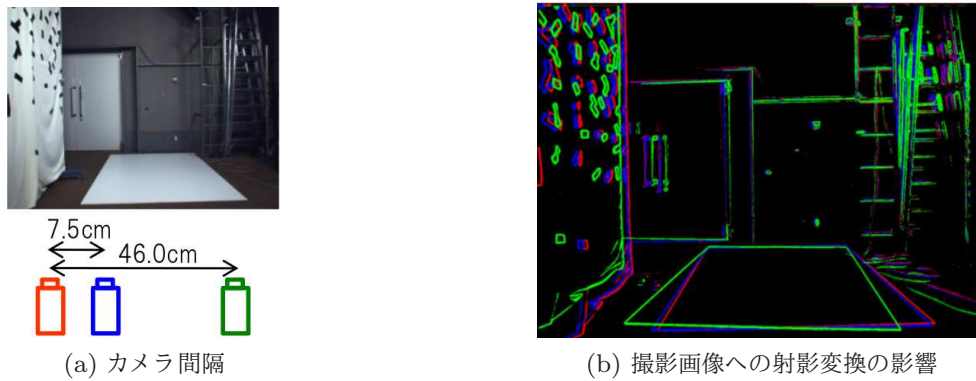


図 2.4 カメラ間隔と撮影画像への射影変換の影響の関係

ロバストな対応付け手法についても、検討を行った。

また、このカメラに正対していない領域は、奥行き推定時に用いる、左右画像の各画素の類似度で定義した Cost を算出する際にも、正しく Cost を算出できないために、デプスマップの精度が低下してしまう課題がある。図 2.5 は、Cost を算出する際に一般的に用いられるブロック処理について示したものである。ブロック処理を用いて Cost を算出する場合、(a) のように Cost を求めたい左画像上の対象画素 p の周辺にブロック L を配置し、右画像の同一にある画素 q の周辺にブロック R を配置して、ブロック L とブロック R との類似を算出する。画素 p に対応する画素 q は、奥行きに応じて右画像の水平方向に視差 d だけ平行移動するため、視差 d を変化させた画素 q の各 Cost を算出し、最も類似度の高い視差 d を対象画素 p の視差として選択し、三角測量の原理を用いて奥行き推定を推定する。

この処理を 3 次元的に示すと、(b) のような断面図になる。3 次元空間中のブロックは、(b) のようにカメラ方向に正対する平面となる。そのため、(b) の左側のように、被写体表面がカメラに正対してる場合、ブロックと表面とは重なって精度の高い Cost を算出することができる。一方で、右側のように、被写体表面がカメラに正対していない場合には、ブロックで Cost

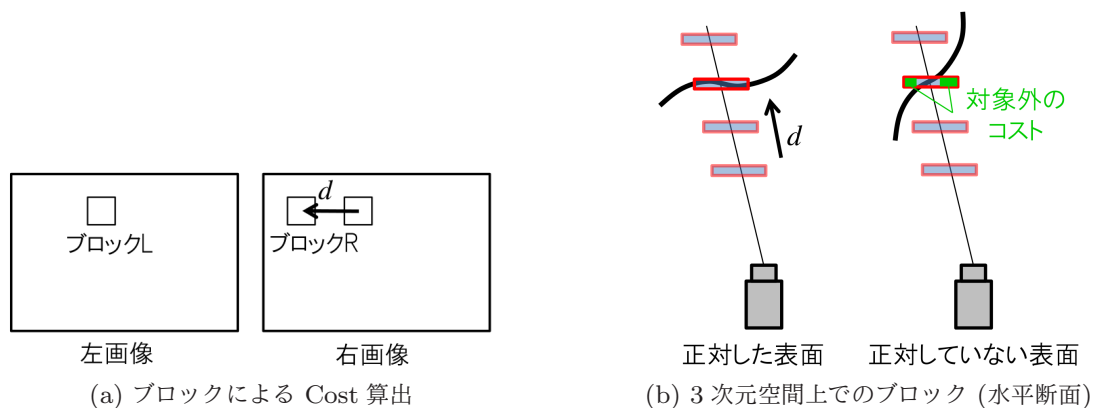


図 2.5 ブロック処理による Cost の算出

を算出する場合には、Cost には表面が含まれない誤った Cost が含まれてしまうため、Cost の精度が低下して奥行き推定の精度も低下してしまう。本研究では、カメラに正対していない領域の奥行き精度の改善も検討した。

これらの課題を検討した上で、本研究で提案する手法は、映像制作への応用に対するほとんどの要求条件を満たすことを目指すものである。透明な被写体や、鏡面反射する被写体については対応できないが、これについては現存するほとんどの手法が対応できないため、デプスマップ推定全般の今後の課題となる。

第 3 章

撮影システム

本章では、はじめに従来のデプスマップ推定手法で一般的に用いられるカラーステレオカメラについて述べる。つぎに、本研究で用いる赤外線カラー一体型カメラアレーと赤外線カラー分離型カメラアレーの 2 種類のカメラアレーについて述べる。

3.1 従来のカラーステレオカメラ

従来の受動型カメラを用いたデプスマップ推定手法では、2 台のカラーカメラを平行に配置したカラーステレオカメラを用いる。2 台のカメラは同期信号等を用いて同じタイミングで撮影することで、動物体であっても、各フレームでは剛体であると見なすことができる。このとき、左画像の各画素に対応する、右画像上の画素を見つけ、その視差からカメラパラメーターを用いて三角測量の原理からデプスマップを推定する。

3.2 赤外線カラー一体型カメラアレー

赤外線カラー一体型カメラアレーの外観を、図 3.1(a) に示す。このカメラアレーは、1 台の赤外線プロジェクターと 2 台の赤外線カラーカメラで構成される。以下、各構成機器の詳細について述べる。

3.2.1 赤外線プロジェクター

この赤外線プロジェクターは、ランダムに配置された点 (ドット) 群のパターン (ドットパターン) を照射する。時間方向には変化のない固定パターンである。プロジェクター光源は、波長は約 800nm の赤外線である。赤外線光線に対して、人の目の瞳孔は反応しないことから、照射する光量は安全が確保される量に限定される。そのため、照射範囲は被写体表面の反射率に依存するが、反射率が高い場合でも約 10m である。

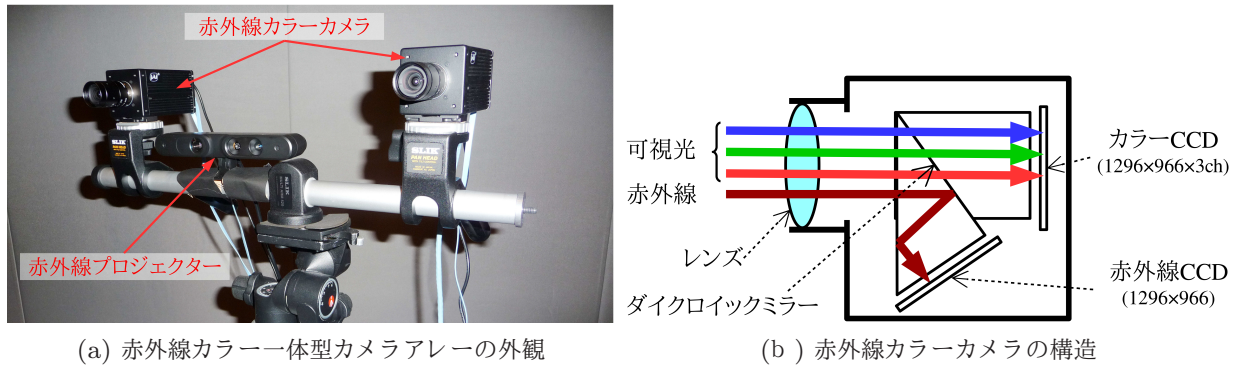


図 3.1 赤外線カラー一体型カメラアレー

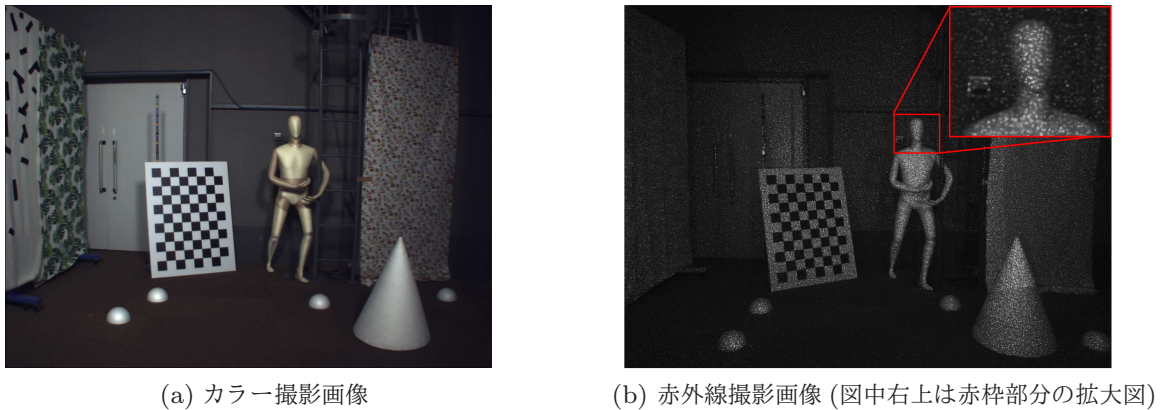


図 3.2 撮影画像

3.2.2 赤外線カラーカメラ

赤外線カラーカメラは、赤外線画像とカラー画像を同時に撮影できるカメラである。赤外線カラーカメラの構造を、図 3.1(b) に示す。レンズからプリズムに入射してきた光線のうち、赤外線光線はダイクロイックミラーで反射し、赤外線 CCD で撮像される。赤外線以外の可視光線を含む残りの光線は、ダイクロイックミラーを透過してカラー CCD で撮像される。赤外線 CCD とカラー CCD の解像度は、ともに 1296×966 画素である。ただし、カラー CCD はベイヤー配列である。

赤外線カラーカメラで撮影した画像の一例を、図 3.2 に示す。(a) カラー撮影画像には、赤外線で照射されたドットパターンは写っていない。一方で、(b) 赤外線撮影画像には、撮影画像全域に一樣にドットが写っている。これより、テクスチャーのない領域も含めて、赤外線画像上で一樣に左右画像の各画素を対応づける手がかりを付与することができるため、安定してデプスマップを推定することができる。一方で、カラー画像にはドットパターンが観測されないことから、映像制作への応用にあたって要求されるテクスチャーの同時取得を満たすことができる。この赤外線画像を見ると、ドットが一樣に付与されているため、従来のステレオマッチ

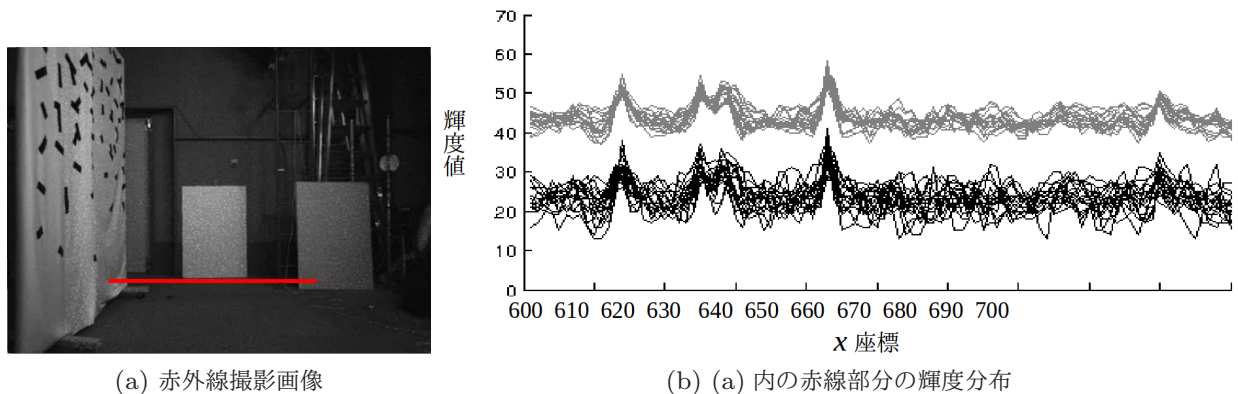


図 3.3 赤外線画像内のノイズの様子

ング手法を用いて容易にデプスマップを推定できるように予想されるが、この赤外線画像をデプスマップ推定に用いるにあたっては、いくつかの課題が存在する。

第一に、このドットパターンのある赤外線ステレオ画像をテクスチャーとして、ステレオマッチングに用いるには粗い点が挙げられる。そのため、ステレオマッチングを実施する際には、左右のブロックの類似度を算出するにあたって、サイズの大きいブロックを用いる必要がある。大きいブロックを用いると、デプスマップ上の被写体の境界付近において、膨張現象が発生してしまう。また、サイズの大きなブロックは、2.3節で述べた射影変換の影響が大きくなり、デプスマップ推定精度が低下してしまう課題もある。さらに、同節で述べたように、カメラに正対していない領域において、ブロック内に表面でない領域の Cost が多く含まれることになるため、Cost を算出する精度が低下してしまう課題も発生する。

第二に、撮影した赤外線画像の SN(Signal to Ratio) 比が低い点が挙げられる。図 3.3 は、(a) の赤外線画像中の赤線の輝度分布を 10 フレーム分多重したものを示したものである。黒線は入力画像、グレー線は各画素において 3 フレーム幅の平均化フィルターをかけてノイズが軽減されている画像の輝度分布を示している。なお、グレー線のノイズ軽減後の分布は、黒線と重ならないように輝度値 20、上にシフトして表示している。ドットが照射されているため、いずれの線にも一定のピークを確認することができるが、黒線のノイズ軽減前はフレーム毎に輝度変化が大きくなっている。これらはノイズを思われ、反射率の低い領域においては、ドットのピークがこれらノイズに埋もれてしまうため、精度が低下してしまう。一方、グレー線のノイズ軽減後の分布は狭くなっており、ノイズによってドットが埋もれている箇所も軽減していることが確認できる。

前述の通り、赤外線の光線は人の目に見えなく、瞳孔が反応しないため、目の保護の観点から、ドットパターンの出力光量は制限される。そのため、低光量の入射光を高増幅して撮像する必要のあることから、赤外線画像の SN 比は低くなってしまふ。赤外線ドットパターンを用いて奥行きを推定するにあたっては、この点についても課題となる。

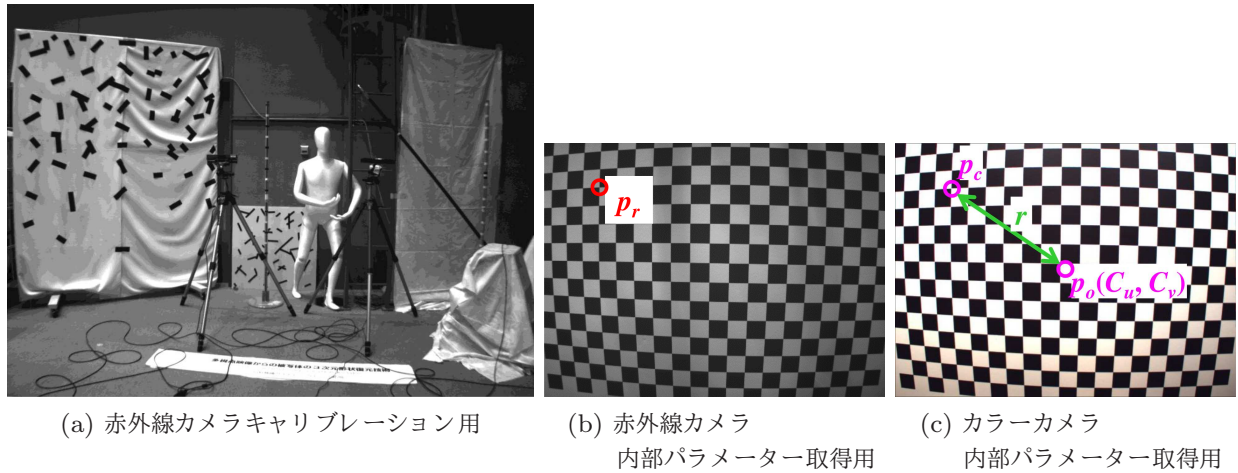


図 3.4 赤外線カラーカメラキャリブレーション用の撮影画像

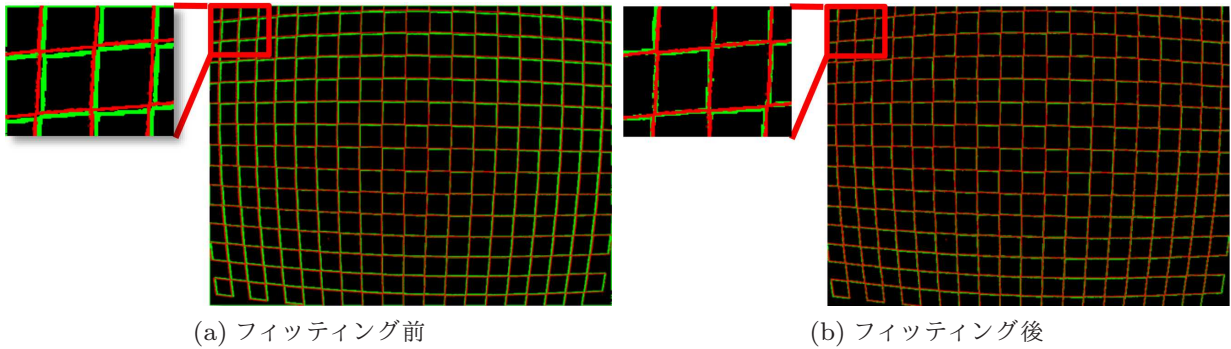
第三に、赤外線画像の階調が低い点が挙げられる。赤外線ドットパターンが照射された赤外線画像は、グレー画像として撮影されるが、ドットが照射されている画素とそうでない画素との輝度差が大きいことから、2値画像に近い階調の低い画像となる。そのため、局所的な領域に注目すると、一般的なカラー画像と比べて近傍画素との相関が低くなってしまう。これにより、デプスマップの推定精度は、キャリブレーション誤差や Cost の定義等により、大きく左右する課題もある。キャリブレーション精度が低いと、エピポラ上にドットが並ばなくなることも発生する。これらの3つの課題により、ドットパターンのある赤外線ステレオ画像を一般的なステレオマッチングの手法を用いても、同様に奥行きを推定することはできない。

3.2.3 カメラアレー

本研究では、3.3.1 節で述べた赤外線ドットプロジェクター 1 台と、2 台の赤外線カラーカメラを、図 3.1 のように配置して、赤外線カラー一体型カメラアレーを構築して被写体を撮影する。2 台の赤外線カラーカメラは、バーの上に水平に設置する。赤外線プロジェクターは、ドットパターンが照射されない領域 (パターンのオクルージョン領域) が、左右の赤外線ステレオカメラ上で極力少なくなるように、2 台の赤外線カラーカメラの中間点に設置する。

2 台の赤外線カラーカメラは、赤外線画像のキャリブレーションとカラー画像のキャリブレーションに分けて実施し、カメラパラメーターを取得する。赤外線画像のキャリブレーションは、図 3.4(a) のような画像を用いて、弱校正カメラキャリブレーション手法 [33] により実施する。この弱校正カメラキャリブレーションは、キャリブレーションパターンを用いずに、撮影画像内の自然特徴点を用いてキャリブレーションを行うため、撮影範囲全体で精度の高いカメラパラメーターを取得することができる。

一方、カラー画像については、ダイクロイックミラーを用いて、赤外線画像と同時に取得す



(a) フィッティング前

(b) フィッティング後

図 3.5 フィッティング前後での赤外線画像とカラー画像の一致の様子

るため、赤外線画像と光軸はほぼ一致しているが完全には一致していない。また、可視光線との波長の違いから屈折率が異なり、画像の外側での収差が発生するため、赤外線画像のカメラパラメータをそのまま用いることはできない。そこで、赤外線画像のカメラパラメータを共通に用いるために、カラー画像を赤外線画像にフィッティングするためのパラメータを取得する。フィッティングパラメータを求めるために、図 3.4(b) および (c) のような、チェッカーボードを全面に撮影した画像を用いて強校正カメラキャリブレーションにより、フィッティングパラメータを求める。収差は、画像中心 p_o からの距離に応じて、係数 κ_1 、 κ_2 、 κ_3 でモデル化し、水平方向および垂直方向の光軸のずれ量を C_u 、 C_v でモデル化した。このとき、これらの係数は以下の式 3.1 で求めることができる。

$$\begin{cases} (\kappa_1, \kappa_2, \kappa_3, C_u, C_v) = \arg \min \sum_i (p_r(i) - p'_c(i))^2 \\ p'_c(i) = p_c(i)(1 + \kappa_1 r(i)^2 + \kappa_2 r(i)^4 + \kappa_3 r(i)^6) \end{cases} \quad (3.1)$$

ただし、 p_r および p_c は、それぞれ赤外線画像とカラー画像上のチェッカーボードのコーナーの座標である。また、 i はコーナー番号を示す。これらのフィッティングパラメータを用いて、カラー画像を赤外線画像へ 2 次元的にフィッティングする。図 3.5 は、撮影範囲全体でチェッカーボードを撮影した赤外線画像とカラー画像のエッジを重畳した画像を、フィッティング前後で比較したものである。赤線と緑線は、それぞれ赤外線画像とカラー画像のエッジである。フィッティング前は、特に画像の周辺付近で、赤外線画像とカラー画像の画素が一致していないが、フィッティング後には一致していることが確認できる。このようにして、カラー画像は赤外線にフィッティングした後、赤外線カメラパラメータを用いて、赤外線画像と同様に扱うことができる。

3.3 赤外線カラー分離型カメラアレー

赤外線カラー分離型カメラアレーの外観を、図 3.6 に示す。このカメラアレーは、1 台の赤外線プロジェクターと 2 台の赤外線カメラ、2 台のカラーカメラで構成される。



図 3.6 赤外線カラー分離型カメラアレー

3.3.1 赤外線カメラ

赤外線カメラには、HD*¹解像度 (1920 × 1080 画素) の CMOS 撮像素子が内蔵されており、HD-SDI*²で映像を毎秒 30 フレーム (プログレッシブ) もしくは毎秒 60 フレーム (インターレース) で出力することができる。3 値同期信号を入力することで、複数台のカメラを同期して撮影することができる。レンズの前には、波長 810~890nm の光線のみを透過するガラス製のバンドパスフィルターが付けられている。

3.3.2 カラーカメラ

カラーカメラには、赤、青、緑の 3 枚の CCD が内蔵されている。各 CCD の解像度は、HD 解像度で、HD-SDI で映像を毎秒 30 フレーム (プログレッシブ) もしくは毎秒 60 フレーム (インターレース) で出力することができる。カラーカメラも 3 値同期信号を入力することで、複数台のカメラを同期して撮影することができる。

3.3.3 カメラアレー

2 台の赤外線カメラは、図 3.6 のように下段のバーの上に水平に配置され、その間に 1 台の赤外線プロジェクターが配置されている。2 台のカラーカメラは、上段のバーの上に水平に配置され、左右の赤外線カメラとカラーカメラは、極力近い位置に設置されている。カメラアレーを構成する合計 4 台のカメラには同期信号が入力され、すべてのカメラが同期して撮影することができる。図 3.7 は、赤外線カラー分離型カメラアレーで撮影した赤外線画像とカラー

*¹ High Definition の略。ITU-R BT.709 で規定されている規格。

*² HD 画質の非圧縮デジタル映像を伝送できる規格。SMPTE 292M で規定されている。

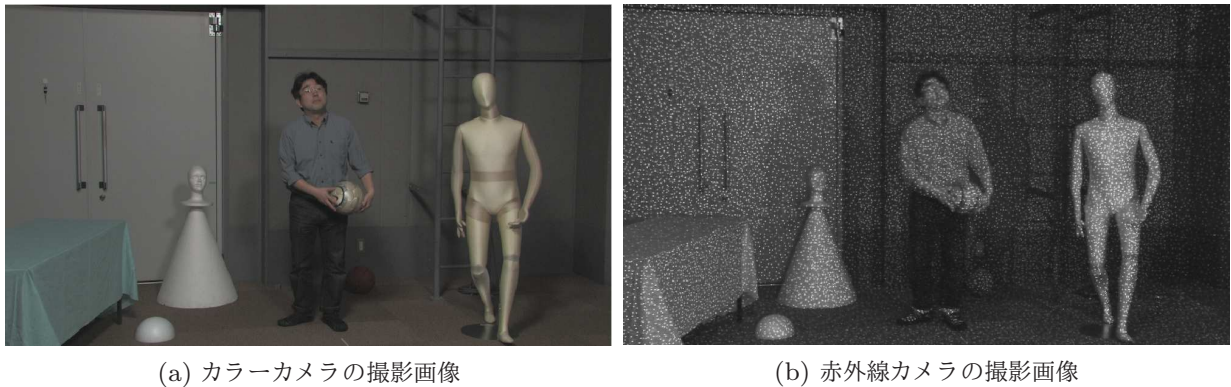


図 3.7 赤外線カラー分離型カメラアレーの撮影画像

画像の例である。赤外線カラー一体型カメラアレーと同様に、カラー画像上ではドットパターンは写っていない、被写体表面のテクスチャーを撮影できている。一方で、赤外線画像では、ドットがシーン全体に観測でき、テクスチャーのない領域にもドットパターンが照射されていることが確認できる。

各カメラのカメラパラメーターは、キャリブレーションパターンを用いる強カメラキャリブレーション [35]、もしくは撮影画像中の自然特徴点を用いる弱カメラキャリブレーション [33] を用いて取得した。弱カメラキャリブレーションの際には、カラー画像と赤外線画像の自然特徴点が大きく変わらないように、反射率の低い黒色の特徴点を配置して撮影した画像を用いた。このようにして取得したカメラパラメーターを用いて、カラー画像を赤外線画像にマッピングしてデプスマップの推定に用いた。

3.4 データセット

本章で述べた、赤外線カラー一体型カメラアレーと赤外線カラー分離型カメラアレーを用いて、4つの赤外線カラーステレオデータセット生成した。図 3.8 は、各データセットのカメラアレーの配置を上から見た断面図である。紫色は赤外線カラーカメラ、赤色と青色はそれぞれ赤外線カメラとカラーカメラを示す。図 3.9 は、各データセットで撮影したシーン示す。なお、赤外線カラー一体型カメラアレーを構成する赤外線カラーカメラには、JAI 製 AD-130GE [18] を用いた。赤外線カラー分離型カメラアレーを構成する赤外線カメラとカラーカメラには、それぞれキャノン製の XF105[47] と XH G1[6] を用いた。また、赤外線プロジェクターには Asus 製 Xtion[41] を用いた。以下に、各データセットについて述べる。

“mannequin75”

このデータセットは、赤外線カラー一体型カメラアレーを図 3.8(a) のように配置して撮影した。カメラ間の距離 (ベースライン長) は、7.5 cm と狭い。撮影した被写体を図 3.9(a) に示

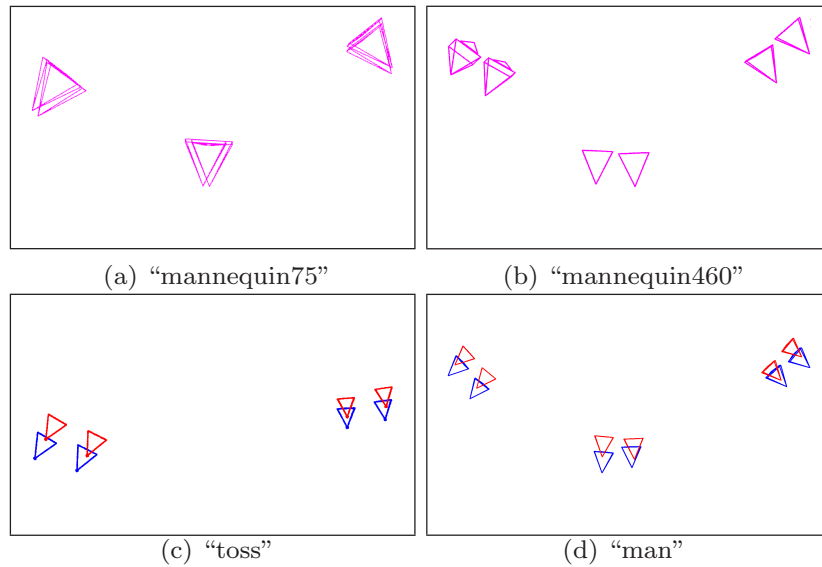


図 3.8 各データセットのカメラ配置

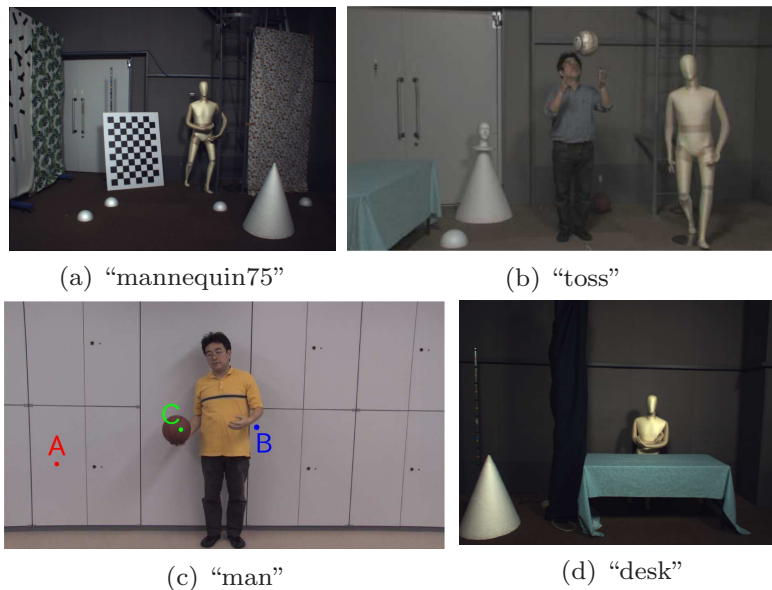


図 3.9 撮影したシーン

す。カメラアレーから被写体までの距離は約 5 m である。撮影にあたっては 3 式のカメラアレーを用いたが、実験にはこのうち中央のカメラアレーを用いた。画像内には白いドアやグレーの壁など、テクスチャーの少ない領域が含まれている。

“mannequin460”

このカメラデータは、“mannequin75”と同様に、図 3.9(a)のような被写体を赤外線カラー一体型カメラアレーを用いて撮影したものである。撮影位置もほぼ同じであるが、図 3.8(b)のように、ベースラインは 46 cm と “mannequin75” よりも広い。このデータセットについて

も、撮影にあたって 3 式のカメラアレーを用いたが、実験にはこのうち中央のカメラアレーを用いた。

“toss”

このカメラデータは、図 3.9(b) のような被写体を、赤外線カラー分離型カメラアレーで撮影したものである。カメラアレーから被写体までの距離は、約 4m である。撮影にあたっては 2 式のカメラアレーを用いたが、実験にはこのうち右側のカメラアレーを用いた。画像内には白ドアやグレーの壁など、テクスチャーの少ない領域が含まれている。

“man”

このカメラデータは、図 3.9(c) のような被写体を、赤外線カラー分離型カメラアレーで撮影したものである。カメラ間のベースラインは約 46cm と広く、カメラアレーから被写体までの距離は、約 4m である。撮影にあたっては 3 式のカメラアレーを用いたが、実験にはこのうち中央のカメラアレーを用いた。画像内には白い壁など、テクスチャーの少ない領域が含まれている。

“desk”

このカメラデータは、図 3.9(d) のような被写体を、赤外線カラー一体型カメラアレーで撮影したものである。カメラ間のベースラインは約 46cm と広く、カメラアレーから被写体までの距離は、約 4m である。画像内にはグレーの壁など、テクスチャーの少ない領域が含まれている。

第4章

赤外線カラーステレオ画像からの奥行き推定

4.1 はじめに

本章では、はじめに基本的なカラーステレオ画像からのデプスマップ推定処理について述べる。次に、第3章で述べたカメラアレーを用いて撮影した、赤外線ステレオ画像とカラーステレオ画像からのデプスマップ推定の基本的な特性を調査する。最後に、Cross-based Local Multipoint Filter (CLMF) による Cost Volume Filter を用いて、単一フレームのデプスマップを推定する基本処理 [19] [20] を提案する。

4.2 カラーステレオ画像からのデプスマップ推定

図4.1に、基本的なカラーステレオ画像からデプスマップを推定する処理を示す。入力されたカラーステレオ画像は、はじめに画像平行化処理が行われる。次に、平行化されたステレオ画像から Cost volume が生成される。

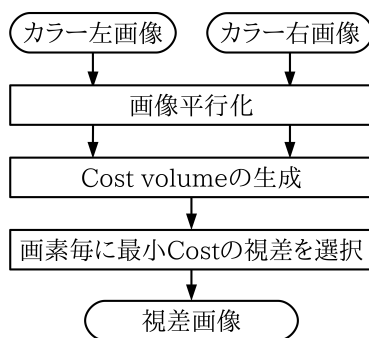


図 4.1 基本的なカラーステレオ画像からの視差画像推定処理

表 4.1 各 Cost の定義

AD	$C_t(u, v, d) = I_L(u, v, t) - I_R(u - d, v, t) $
SD	$C_t(u, v, d) = (I_L(u, v, t) - I_R(u - d, v, t))^2$
SAD	$C_t(u, v, d) = \sum_{m=-\tau}^{\tau} \sum_{n=-\tau}^{\tau} I_L(u + m, v + n, t) - I_R(u - d + m, v + n, t) $
SSD	$C_t(u, v, d) = \sum_{m=-\tau}^{\tau} \sum_{n=-\tau}^{\tau} \{I_L(u + m, v + n, t) - I_R(u - d + m, v + n, t)\}^2$
NCC	$C_t(u, v, d) = \sum_{m,n} \{I_L(u + m, v + n, t) - \bar{I}_L(u, v, t)\} \{I_R(u + m - d, v + n, t) - \bar{I}_R(u - d, v, t)\}$
ZNCC	$C_t(u, v, d) = \frac{\sum_{m,n} \{I_L(u + m, v + n, t) - \bar{I}_L(u, v, t)\} \{I_R(u + m - d, v + n, t) - \bar{I}_R(u - d, v, t)\}}{\sqrt{\sum_{m,n} \{I_L(u + m, v + n, t) - \bar{I}_L(u, v, t)\}^2 \sum_{m,n} \{I_R(u + m - d, v + n, t) - \bar{I}_R(u - d, v, t)\}^2}}$ <p style="text-align: center;">ただし、$\bar{I}(u, v, t) = \frac{1}{ \Psi } \sum_{m,n} I(u + m, v + n, t)$</p>

Cost volume の基本的な生成方法について述べる前に、はじめに Cost Volume を構成する Cost について述べる。2.3 節で述べた通り、ステレオマッチングは、左画像上の対象画素に対応する、右画像上の画素を求めることにより、その視差からカメラパラメータを用いて三角測量の原理により奥行きを推定する。画像平行化されたステレオ画像における右画像上の対応点は、左画像上の対象画素から水平方向にのみ移動する。そこで、左画像上の対象画素の座標から視差 d 水平移動した右画像上の画素との類似度を順次求め、最も類似度の高い右画像上の画素を見つけることで、左右画像の画素の対応関係を求めることができる。このときの類似度を Cost と呼ぶ。

Cost の定義は、さまざま提案されており、主な Cost には Absolute Difference (AD)、Squared Difference (SD)、Sum of Absolute Difference (SAD)、Sum of Squared Difference (SSD)、Normalized Cross-Correlation (NCC)、Zero-mean Normalized Cross-Correlation (ZNCC) がある。各 Cost は、表 4.1 のように定義される。ここで、 I_L と I_R は、それぞれ左画像、右画像を示し、 u, v, d は、それぞれ X 座標、 Y 座標、視差を示す。また、 τ と \bar{I} は、それぞれ半ブロックサイズと第 2 式に示すブロック内の平均輝度値を表す。 $|\Psi|$ は、ブロック内の画素数である。NCC と ZNCC 以外のは類似度が高いほど低い値を出力する一方、NCC と ZNCC は類似度が高いほど、高い値を出力する。そこで、本研究では NCC については $[0, 1]$ の範囲で出力するため、1 から引いた値 C'_t を Cost として用い、類似度が高いほど Cost が小さくなるように定義した。すなわち、

$$C'_t(d) = 1 - C_t(u, v, d) \quad (4.1)$$

また、ZNCC についても $[-1, 1]$ の範囲で出力するため、1 を足し、2 で割って正規化した値を

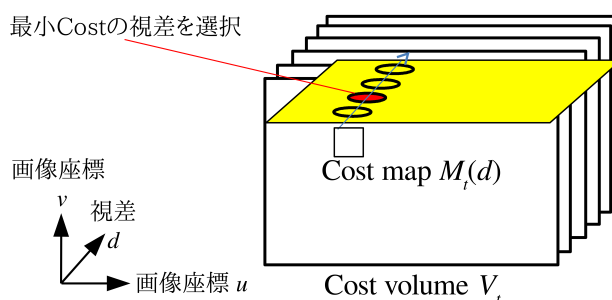


図 4.2 時間 t 、視差 d のときの Cost map $M_t(d)$ と Cost volume V_t

1 から引いた値 C'_t を Cost とした。すなわち

$$C'_t(d) = 1 - \{C_t(u, v, d) + 1\}/2 \quad (4.2)$$

これらの定義を用いて、図 4.2 のように、時間 t 、視差 d のときの Cost をすべての画素について求めたものを、Cost map $M_t(d)$ と呼ぶ。すなわち、Cost map $M_t(d)$ は以下のように定義される。

$$M_t(d) = \{C_t(u, v, d) \mid u \in [0, w], v \in [0, h]\} \quad (4.3)$$

ここで、 w と h は、それぞれ水平方向と垂直方向の画像サイズである。また時間 t における、すべての視差 $d = [d_{min}, d_{max}]$ について、Cost map を求めたものを Cost volume V_t と呼ぶ。すなわち、Cost volume V_t は以下のように定義される。

$$V_t = \{M_t(d) \mid d \in [d_{min}, d_{max}]\} \quad (4.4)$$

このようにして、生成された Cost volume の各画素について、Cost が最も小さい視差を選択すると視差画像を推定することができる。この視差画像には、左画像と右画像の両方から見えなく、原理的に推定が不可能な領域や、推定時のエラーが含まれるため、左カメラからの視差画像と右カメラからの視差画像を推定し、両方の推定結果が一致した結果のみを最終的に出力する。最終的な視差値は、左視差画像上の視差 D_L とそれに対応する画素の右視差画像上の視差値 D_R の差が閾値 T 以下のとき、正しく推定できたとして出力される。すなわち、

$$|D_L(u, v) - D_R(u, v - D_L(u, v))| < T. \quad (4.5)$$

この左右視差の一致チェックののち、出力された視差からカメラパラメーターを用いて、三角測量の原理により奥行き値に変換することで、デプスマップを推定することができる。

4.3 赤外線ステレオ画像からのデプスマップ推定の基本的特性

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デプスマップ推定手法は、過去に本格的に研究された事例がない。そこで、本節では赤外線ステレオ画像からのデプスマップ

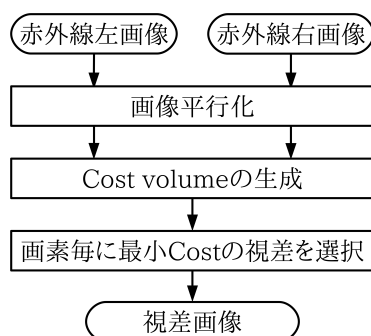


図 4.3 赤外線ステレオ画像からの視差画像推定処理

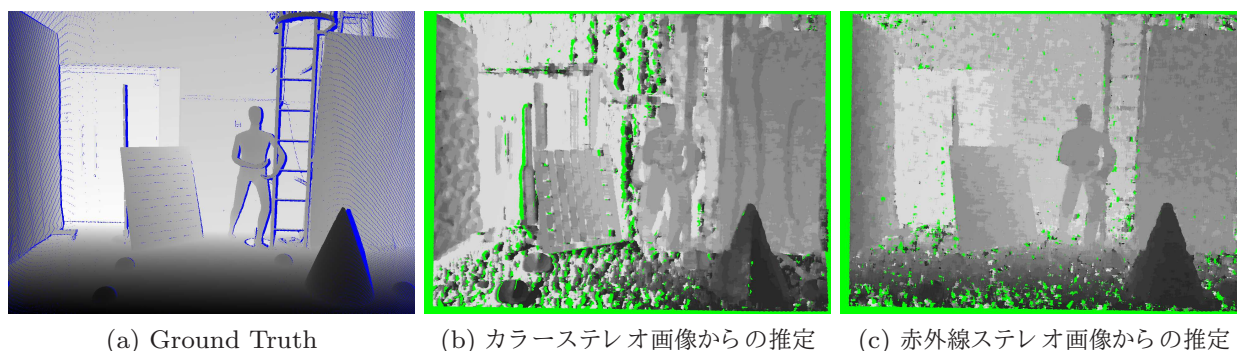


図 4.4 赤外線ドットパターン照射の効果

プ推定の基本的特性について調査したので述べる。はじめに、カラーステレオ画像からのデプスマップ推定結果と、図 4.3 のように、赤外線ステレオ画像を入力してデプスマップ推定した結果と比較し、赤外線ドットパターンの照射の効果を確認した。次に、ノイズの影響についてもノイズ除去処理の有無を比較することで調査した。最後に、赤外線ステレオ画像における Cost の定義についても、Cost の種類やサイズを変えて、その特性について調査した。

4.3.1 赤外線ドットパターン照射の効果

図 4.4 は、“mannequin75” データセットの、(b)ドットパターンのないカラーステレオ画像のみから、ブロックマッチングを用いて推定したデプスマップと、(c)ドットパターンのある赤外線ステレオ画像のみから推定したデプスマップを比較したものである。(a) は、レンジセンサーで取得した Ground Truth である。(a) の青色の画素は、レンジスキャナーとステレオカメラの左側のカメラとの撮影位置の違いによるオクルージョンのために 3 次元情報が取得できなかった領域である。緑色の画素は、式 4.5 で定義した左右視差チェックで削除された画素である。

(b) を見ると、テクスチャーが十分ある領域においては、奥行きが正しく推定されている。しかし、中央付近の灰色の壁や、左側の白いドア、また手前の床など、テクスチャーが十分に

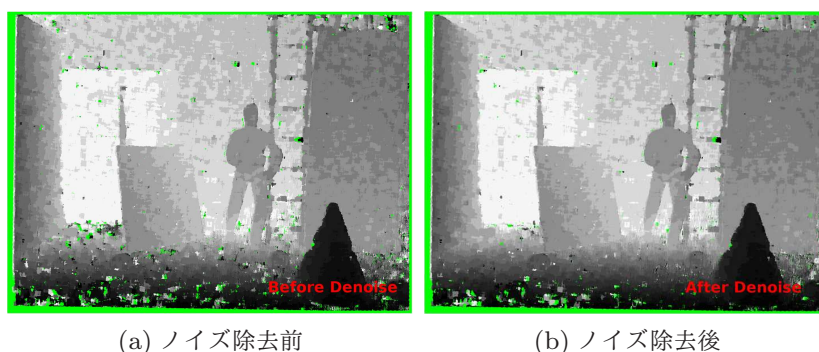


図 4.5 赤外線画像上のノイズの影響

ない領域では、奥行き推定に失敗している。一方、(c)を見ると、テクスチャーの少ない領域を含め、赤外線ドットパターンの照射により、奥行きはシーン全体で安定して求まっている。このことから、赤外線ドットパターンの照射の効果が大きいことが確認できた。

4.3.2 ノイズの影響

図 4.5 は、赤外線画像上のノイズの影響を確認するため、“mannequin75” データセットの赤外線ステレオ画像を、(a) そのまま入力画像として、ブロックマッチングで推定したデプスマップと、(b) 各画素について前後3 フレームの時間方向の平均フィルタをかけた赤外線ステレオ画像を入力画像から推定したデプスマップを比較したものである。(a) では、床部分を中心に、推定できなかった緑色の画素が点在しているが、(b) ではそれらの画素が減少した。このことから、赤外線画像上のノイズが、奥行き推定結果に悪影響を与えていることがわかった。

4.3.3 ブロックサイズ

ブロックサイズによる奥行き推定精度の違いを確認するため、カラーステレオ画像と赤外線ステレオ画像について、異なるサイズのブロックを用いてブロックマッチングにより視差画像を推定した。図 4.6 の上段は、カラーステレオ画像に対して、(a) 7×7 、(b) 13×13 、(c) 19×19 のブロックを用いて推定した視差画像を比較している。ブロックサイズが大きくなるにつれ、テクスチャーのある領域については、より安定的に推定できた。一方で、背景の壁やドアなど、テクスチャーがない領域については、ブロックサイズを大きくしても、奥行きを推定できなかった。このことから、カラーステレオ画像のみから、テクスチャーのない領域の奥行きを推定することは困難であることが確認できた。

下段は、赤外線ステレオ画像から、(d) 7×7 、(e) 13×13 、(f) 19×19 のブロックを用いて、ブロックマッチングにより視差画像を推定したものである。赤外線ドットパターンを照射された赤外線ステレオ画像を用いることで、より大きいサイズのブロックを用いると、シーン全体

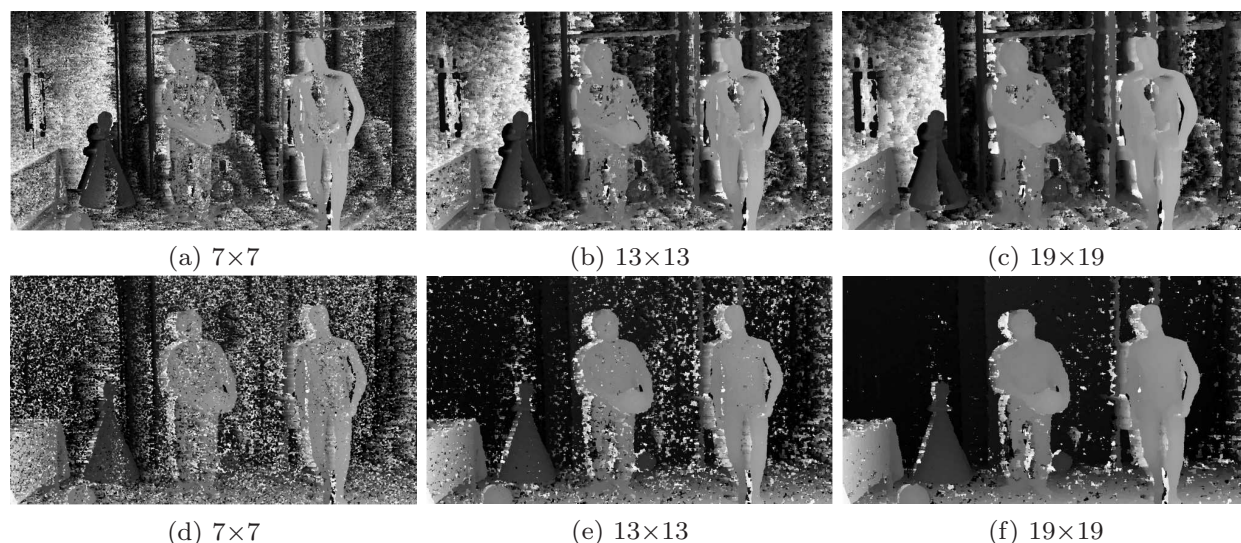


図 4.6 異なるサイズのブロックを用いてブロックマッチングにより推定した視差画像 (上段: カラーステレオ画像、下段: 赤外線ステレオ画像)

を安定的に視差推定できることがわかった。しかし、照射されるドットパターンは、テクスチャーとして用いるには粗いことから、 19×19 という大きなブロックを用いる必要があることがわかる。そのため、推定した視差画像を見ると、被写体の境界付近で前景が実際よりも大きくなる膨張現象が発生してしまった。

図 4.7 は、"mannequin75" データセットの赤外線ステレオ画像から、(a) 5×5 、(b) 9×9 、(c) 13×13 、(d) 19×19 のブロックでデプスマップを推定した結果である。緑色の画素は、左右の視差画像の推定結果をチェックして、推定エラーと判断された画素である。また、赤色の画素は、4.4(a)の Ground Truth からの誤差が閾値以上の画素を示す。この場合も、大きいブロックの方がより推定結果がよく、少なくとも 17×17 という大きいブロックが必要となることがわかった。

4.3.4 Cost

赤外線ドットパターンを照射した赤外線ステレオからのデプスマップを推定する際の、Cost の定義による精度の違いを確認するため、主な Cost を用いてブロックマッチングによるデプスマップ推定を行った。図 4.8 は、"mannequin75" データセットの赤外線ステレオ画像から、(a) SAD、(b) SSD、(c) NCC、(d) ZNCC で定義した Cost を用いて推定したデプスマップである。(a) SAD や (d) ZNCC を用いて推定したデプスマップは、(b) SSD や (c) NCC を用いて推定したものよりも、推定エラーの画素数が少ないことを確認できた。特に、ZNCC は他の Cost 定義よりも、最も推定エラーが少ないことがわかった。ZNCC は、ブロック内の平均画素値とブロック内の各画素との差分ベクトルの内積で定義されていることから、画像の交流

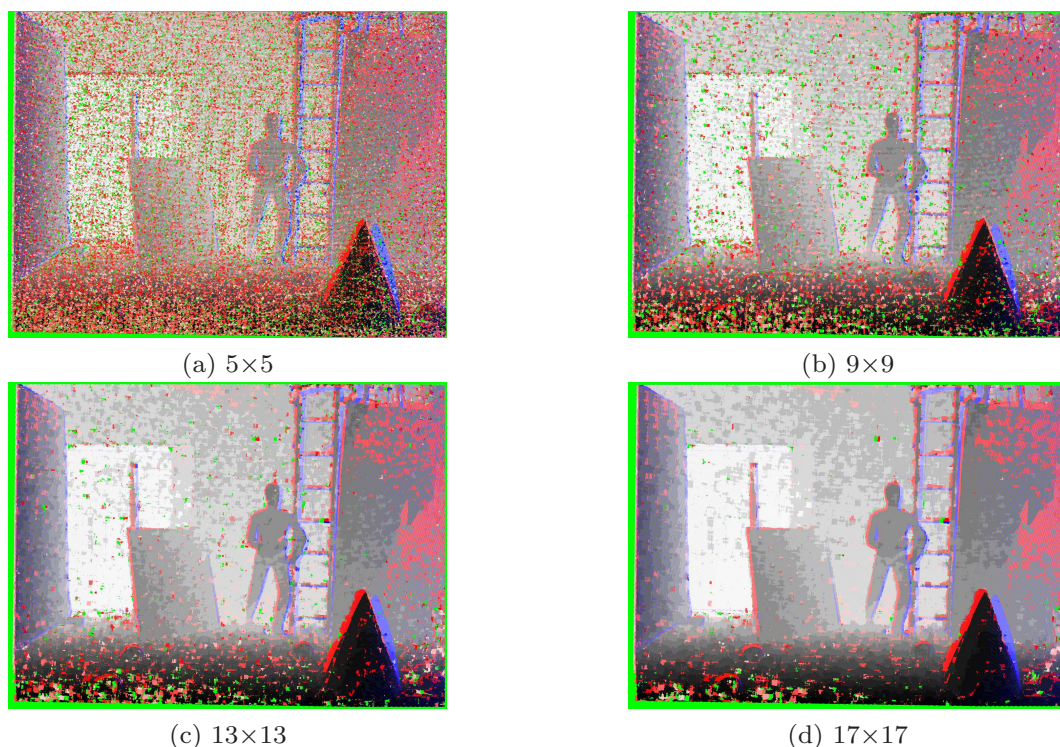


図 4.7 赤外線ステレオカメラから、異なるブロックサイズを用いてブロックマッチングにより推定したデプスマップ

成分を計測することとなる。赤外線ドットパターンの照射は、画像に交流成分を付与することに類似していることから、ZNCC は赤外線ドットパターンを照射した赤外線ステレオ画像からデプスマップを推定する際に、最も適しているのではないかと考えられる。

図 4.9 は、”toss” データセットの赤外線ステレオ画像から、各 Cost 定義について、ブロックサイズを変えてブロックマッチングを用いてデプスマップを推定したとき、左右の視差画像のチェックにより推定エラーと判定された画素のパーセント (推定エラー率) を示したものである。この定量的な実験結果から見ても、ZNCC が小さいブロックサイズで低い推定エラー率を実現できることがわかった。また、ブロックサイズが 15×15 から 17×17 で推定エラー率が最小になっていることがわかった。(b) は、左右の視差画像チェックでエラーと判定されなかった画素について、推定視差にしたがって左右のカラー画像に投影した 2 つの投影点の画素値の RMSE (Root Mean Square Error) を示したものである。RMSE は、式 4.6 で定義される。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(I_L(u, v) - I_R(u, v - D_L(u, v)) \right)^2} \quad (4.6)$$

これを見ると、SAD が NCC や ZNCC に比べて、小さいブロックで低い RMSE を実現できている。しかし、ブロックサイズが 13×13 よりも大きくなると、どの Cost の定義でも差がなくなることがわかった。(a) 推定エラー率では、ブロックサイズを 17×17 以上であることが

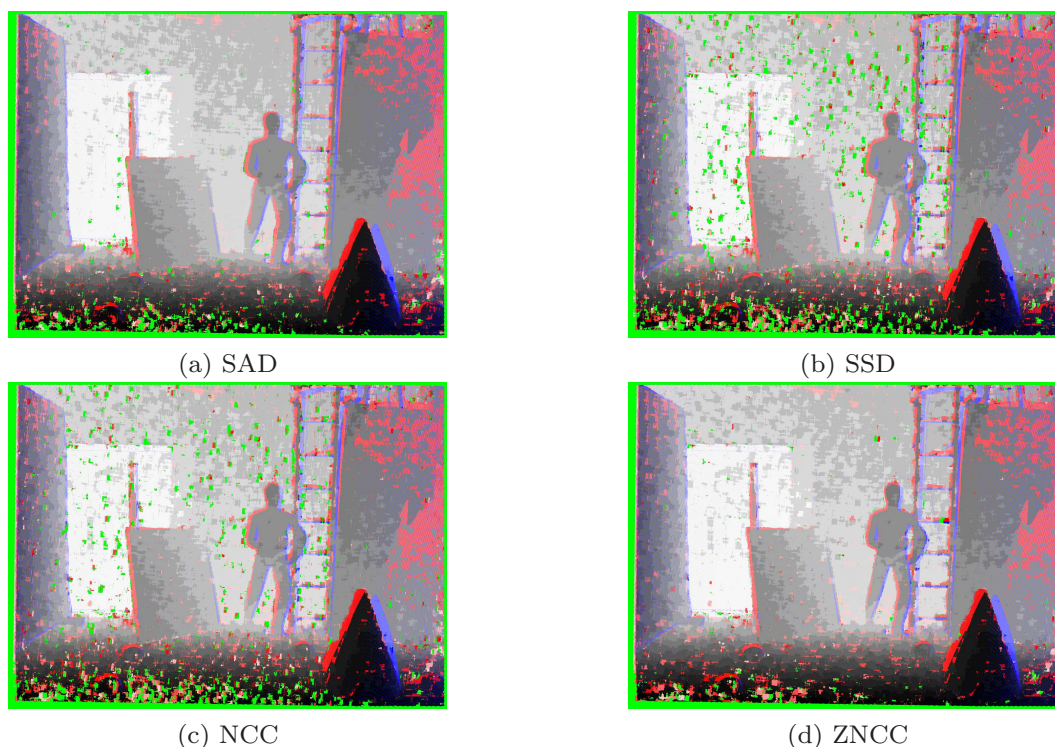


図 4.8 Cost の定義の違いによるデプスマップ推定精度の違い

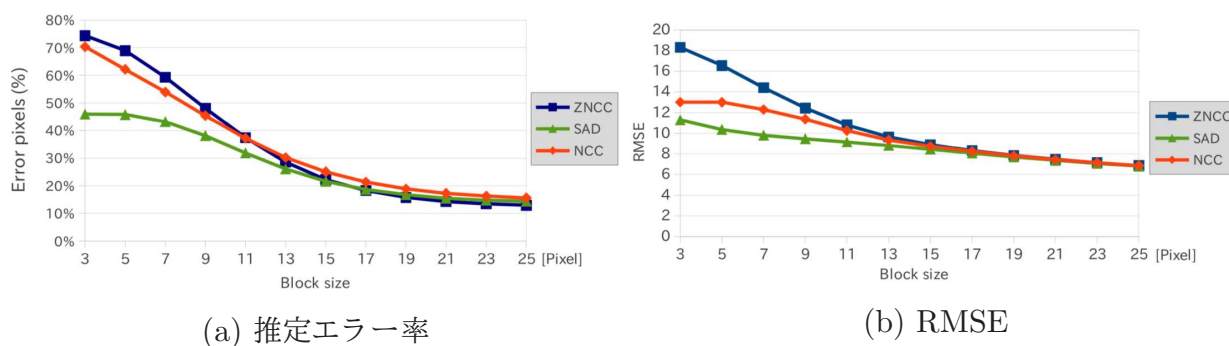


図 4.9 “toss” データセットの赤外線ステレオ画像から、Cost 定義とブロックサイズを変えてデプスマップを推定したときの推定エラー率と RMSE

望ましいことから、ZNCCを用いると推定率が最も高くなることがわかった。

図 4.10 は、“mannequin75” データセットの赤外線ステレオ画像から、各 Cost 定義について、ブロックサイズを変えてブロックマッチングによりデプスマップを推定したときの推定エラー率と RMSE を示したものである。この結果からも、SAD や SSD が、NCC や ZNCC に比べて小さいブロックで、低い RMSE を実現できている。しかし、“toss” データセットのときと同様に、推定エラー率からブロックサイズは 17×17 以上であることから、Cost として ZNCC を用いることが適していることがわかる。

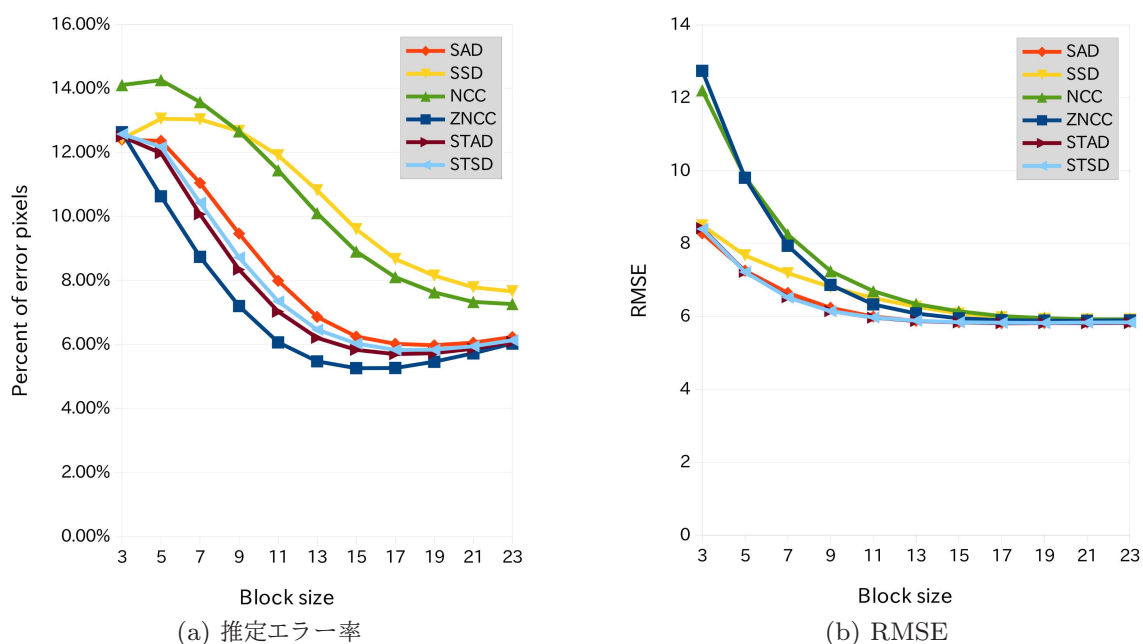


図 4.10 “mannequin75” から、Cost 定義とブロックサイズを変えて、ブロックマッチングでデプスマップを推定したときの推定エラー率と RMSE

4.4 赤外線カラーステレオからのデプスマップ推定の基本処理

4.3 節で調査した基本特性から、赤外線ドットパターンが照射された赤外線ステレオ画像から、デプスマップを推定するにはサイズの大きいブロックを用いる必要があることがわかった。一方で、ブロックサイズを大きくすると、デプスマップ上で膨張現象が発生することがわかった。そこで、赤外線ステレオ画像からデプスマップを推定するにあたって、膨張現象を抑制する効果のある Cost volume filter を導入した。図 4.11 に、赤外線ステレオ画像とカラーステレオ画像からの視差画像を推定する基本処理を示す。本節では、Cost volume filter とその中で用いられる CLMF について述べたあと、赤外線ステレオ画像とカラーステレオ画像から Cost volume filter を用いて視差画像を推定する基本処理について述べる。

4.4.1 Cost volume filter

Cost volume filter は、各 Cost map $M_t(d)$ を 2 次元平滑化フィルタを用いて平滑化するフィルタ処理である。Cost volume filter を用いたデプスマップ推定手法は、一般的に Cost volume に対して、Cost volume filter で処理したのち、各画素について Cost が最小となる視差 d を当該画素の視差値として選択して推定する [14]。この処理は、Cost $C_t(u, v, d)$ を近隣画素の同一視差 d 内で共有することができるため、特にテクスチャーが少ない領域において、

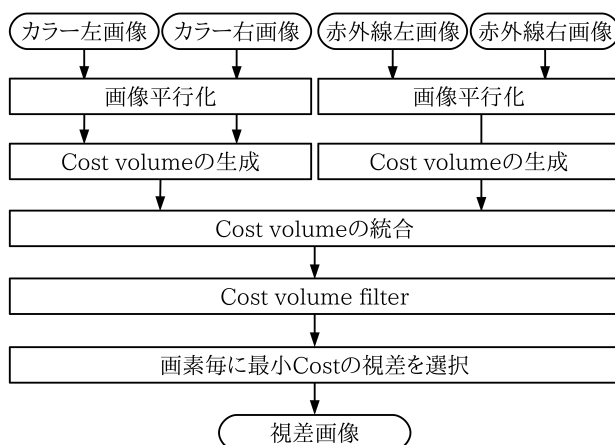


図 4.11 赤外線ステレオ画像とカラーステレオ画像からデプスマップを推定する基本処理

Cost の信頼性を高めることができる。近年では、Guided filter [8] に代表されるような、エッジ保存型の平滑化フィルターを用いることで、デプスマップ上の被写体境界付近で発生する膨張現象を抑制する手法が提案されている。

本手法では、Guided filter の代わりに、同じエッジ保存型の平滑化フィルターである CLMF [25] を用いる。CLMF は、Guided filter に適応型カーネルを適用することにより、平滑化の高精度化をはかったものである。CLMF は、Guided filter の発展形であることから、次節以降で Guided filter と CLMF について述べたあと、CLMF の Cost volume filter への適応について述べる。

4.4.2 Guided Filter

Guided filter は、平滑化する対象画像とガイド画像の 2 つの画像を入力する必要がある。2 つの画像は同一でも構わない。ステレオマッチングにおいては、平滑化する対象画像は、Cost map $M_t(d)$ であり、ガイド画像はカラー画像となる。Guided filter は、平滑後画像 S 上の画素 p の画素値は、ガイド画像の画素 p の画素値の線形変換であると仮定する。すなわち、

$$S_p^k = a_k G_p + b_k, \quad p \in \Omega_k. \quad (4.7)$$

ただし、 a_k と b_k は線形変換の 2 つの係数で、図 4.12 のように画素 k を中心とし、 p を含むサイズ固定の矩形ブロック Ω_k 内で一定である。このとき、 a_k と b_k は線形回帰を用いて、式 4.8 より求めることができる。

$$a_k = \frac{\frac{1}{|\Omega|} \sum_{p \in \Omega_k} G_p M_{p,t}(d) - \mu_k \bar{M}_{k,t}(d)}{\sigma_k^2 + \epsilon}, \quad b_k = \bar{M}_{k,t} - a_k \mu_k. \quad (4.8)$$

ここで、 μ と σ^2 は、それぞれガイド画像 G 上のブロック Ω 内の平均値と分散である。また、 ϵ と $|\Omega|$ は、拘束項とブロック Ω 内の画素数を示す。 $M_{p,t}(d)$ と $\bar{M}_{k,t}(d)$ は、Cost map $M_t(d)$

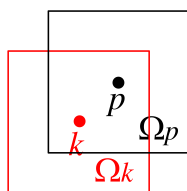


図 4.12 Guided filter で用いられるサイズ固定の矩形ブロック

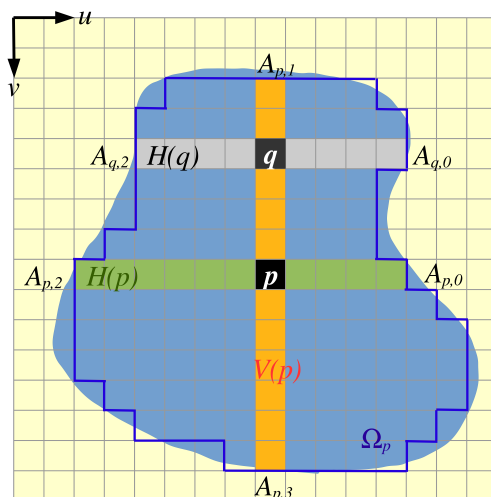


図 4.13 Cross-based 適応型カーネル Ω_p

上での画素 p の Cost とブロック Ω_k 内の Cost の平均値である。平滑化後の出力画像 S は、画素毎に式 4.7 を用いて a_k と b_k を算出して取得することができる。画素 p を含むブロック Ω_k は複数存在することから、最終的な S_p は、すべてのブロック Ω_k の出力値の平均値で算出する。すなわち、

$$S_p = \frac{1}{|\Omega|} \sum_{k \in \Omega_p} (a_k G_p + b_k) = \bar{a}_p G_p + \bar{b}_p. \quad (4.9)$$

ここで、 \bar{a}_p と \bar{b}_p は係数の平均値で、式 4.10 で定義される。

$$\bar{a}_p = \frac{1}{|\Omega|} \sum_{k \in \Omega_p} a_k, \quad \bar{b}_p = \frac{1}{|\Omega|} \sum_{k \in \Omega_p} b_k. \quad (4.10)$$

4.4.3 Cross-based Local Multipoint Filter (CLMF)

Guided filter は、サイズ固定の矩形ブロックを用いて平滑化処理を行うが、CLMF は矩形ブロックの代わりに、図 4.13 のように類似色領域にセグメンテーションした Cross-based 適応型カーネルを用いることで、高精度化をはかったものである。この適応型カーネルは、以下の処理で取得することができる。

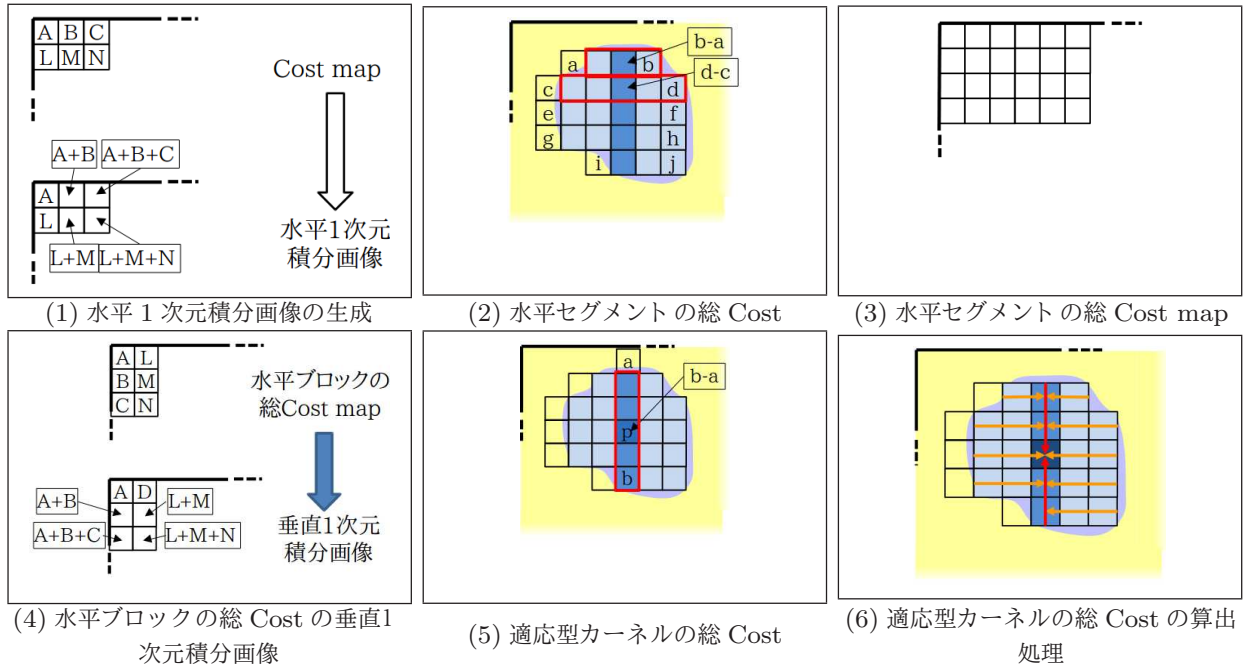


図 4.14 積分画像を用いた適応型カーネルの処理過程

1. 中央の対象画素 p から、上下方向にアームを伸ばす。
2. 現在の画素値と次の画素値との差が閾値以上の点でアームを停止して、アーム長 $A_{p,1}$ と $A_{p,3}$ を取得する。
3. 水平アーム $V(p)$ 上のすべての画素 q から、アームを左右方向に伸ばす。
4. 垂直アームと同様の条件でアームを停止させ、各画素 q におけるアーム長 $A_{q,0}$ と $A_{q,2}$ を取得する。

図 4.13 は、この処理によって取得した適応型カーネルを示す。紺色の線で囲まれた Ω_p が、青色の対象物に対応した適応型カーネルである。このとき、画素 $p = (u_p, v_p)$ を基点とした適応型カーネル Ω_p は、式 4.11 で表すことができる。

$$\Omega_p = \bigcup_{q \in V(p)} H(q)$$

$$H(p) = \{(u, v) | u \in [u_p - A_{p,2}, u_p + A_{p,0}], v = v_p\}$$

$$V(p) = \{(u, v) | u = u_p, v \in [v_p - A_{p,1}, v_p + A_{p,3}]\}.$$
(4.11)

この処理は 2 回の 1 次元積分画像 [38] [45] を用いることができるため、高速に処理をすることができる。

図 4.14 は、積分画像を用いた適応型カーネルの処理を図示したものである。はじめに、(1) のような Cost map があつたとき、各画素について左端の画素からの累積 Cost を算出して格納し、水平方向の 1 次元方向の積分画像を生成する。次に、(2) のようにアームの右端の画素

の水平 1 次元積分画像上の数値 b から、左端の画素の水平 1 次元積分画像上の数値 a を引き、この水平セグメントにおける合計 Cost を当該画素に格納する。この処理をすべての画素について行くと、(3) のような水平セグメントにおける総 Cost map を取得することができる。さらに、(4) のように、各画素について上端の画素からの累積 Cost を算出して格納し、垂直方向の 1 次元方向の積分画像を生成する。最後に、(5) のように、アームの下端の画素の垂直 1 次元積分画像上の数値 b から、上端の画素の垂直 1 次元積分画像上の数値 a を引くと、この適応型カーネルの総 Cost が求まる。この処理をまとめると、(6) のように、カーネル内の Cost をはじめに、“背骨” に集めたあと、カーネル中心に集めることで効果的にカーネル内の総 Cost を算出できる。

4.4.4 CLMF を用いた Cost Volume Filter

CLMF は Guided filter と同様に、平滑化対象画像とガイド画像の 2 種類の画像を入力する。Cost volume filter に適応するにあたっては、各 Cost map $M_t(d)$ を平滑化対象画像として、またカラー画像をガイド画像を入力して処理を行う。この処理をすべての視差 d について実行することにより、Cost volume filter 処理を行うことができる。

視差画像を推定するにあたっては、この Cost volume filter 処理後の Cost volume を用いて、各画素について Cost が最小となる視差 d を選択して推定する。本手法では、平行化された赤外線ステレオ画像とカラーステレオ画像からそれぞれ Cost volume を生成したのち、Cost volume filter 処理し、それらを 1 つに統合した Cost volume を用いてデプスマップを推定する基本処理を提案する。

4.4.5 赤外線とカラーの Cost volume の統合によるデプスマップ推定

Cost volume filter を用いてデプスマップを推定するにあたって、赤外線ステレオ画像のみから Cost volume を生成してデプスマップを推定すると、テクスチャーの細かい領域においては、適応型カーネルのサイズが小さくなり、カーネル内に入るドットの数が減るために、Cost の信頼度が低下してしまう。一方で、カーネルサイズが小さいことはテクスチャーが十分あることを意味するため、カラーステレオ画像から生成した Cost の信頼度は高まる。逆に、カーネルサイズが大きいときは、テクスチャーが十分でないことを意味するため、カラー Cost の信頼度は低下するが、赤外線 Cost の信頼度は高くなる。これらの関係を示したのが、表 4.2 である。

そこで、提案手法では赤外線ステレオ画像とカラーステレオ画像のそれぞれについて、st-Cost volume \widehat{V}_t^r と \widehat{V}_t^c を生成し、st-Cost volume filter 処理して、Cost volume $\widehat{V}_t^{r'}$ と $\widehat{V}_t^{c'}$ を生成する。そして、カーネルサイズに応じて、赤外線 Cost とカラー Cost を選択して、

表 4.2 適応型カーネルのサイズと Cost の信頼度の関係

カーネルサイズ	大きい	小さい
赤外線 Cost	✓	✗
カラー Cost	✗	✓

1 つの Cost volume V_t^i に統合することで、デプスマップの推定精度を改善はかる。 $p = (u, v)$ におけるカーネルサイズ $|\Omega_p|$ が、閾値 b^2 以上のとき、赤外線 Cost $\widehat{C}_t^{r'} \in \widehat{V}_t^{r'}$ を Cost C_t^i として、Cost volume V_t^i に格納する。逆に、カーネルサイズ $|\Omega_p|$ が閾値 b^2 より小さいとき、カラー Cost $\widehat{C}_t^{c'} \in \widehat{V}_t^{c'}$ を Cost C_t^i として、Cost volume V_t^i に格納する。すなわち、統合 Cost volume V_t^i は式 4.12 で定義される。

$$\begin{cases} C_t^i(p, d) = \widehat{C}_t^{r'}(p, d) & (|\Omega_p| > b^2) \\ C_t^i(p, d) = \widehat{C}_t^{c'}(p, d) & (|\Omega_p| \leq b^2) \end{cases} \quad (4.12)$$

このときの b^2 は、ブロック内にドットパターンのドットが平均して最低 1 つの入るブロックサイズ $b \times b$ で定義する。このようにして統合した Cost volume V_t^i の各画素について、Cost が最小となる視差 d を選択することで視差画像を推定する。

4.4.6 赤外線カラー分離型カメラアレーにおける Cost Volume の生成方法

3.3 節で述べた赤外線カラー分離型カメラアレーを用いる場合、赤外線カメラとカラーカメラの光軸は一致していない。そこで、カメラパラメーターを用いて、カラー画像を赤外線画像にマッピングして Cost volume filter のガイド画像を生成する。図 4.15 のように、画像平行化された赤外線画像上の画像座標 (u, v) の同次座標系の座標 \mathbf{p} は、視差 d のとき式 (4.13) を用いて算出することができる。

$$\mathbf{p} = \frac{z}{f}(u - C_u, v - C_v, f)^T, z = \frac{Bf}{d} \quad (4.13)$$

ここで、 B 、 (C_u, C_v) 、 f は、それぞれ赤外線カメラ間距離、赤外線画像座標系のカメラ中心の座標、画像平行化後の赤外線カメラの焦点距離を画素数で示したものである。このとき、 \mathbf{p} は式 4.14 を用いて、世界画像系の座標 \mathbf{P} にマッピングすることができる。

$$\mathbf{P} = \mathbf{R}_r^{-1} \mathbf{F}_r^{-1} \mathbf{p} - \mathbf{R}_r^{-1} \mathbf{T}_r. \quad (4.14)$$

ただし、 \mathbf{F} 、 \mathbf{R} 、 \mathbf{T} 、 r は、それぞれカメラの内部パラメーター行列、回転行列、平行移動ベクトル、画像平行化後の赤外線画像を示すインデックスである。このとき、世界画像系の点 \mathbf{P} は、

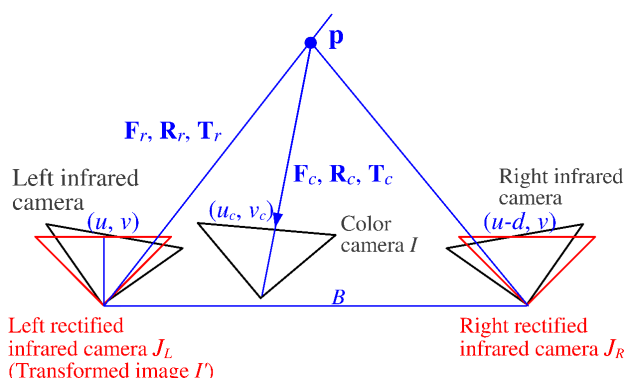


図 4.15 カラー画像の赤外線画像へのマッピング

式 4.15 を用いて、カラー画像上のカラー画像座標系の座標 (u_c, v_c) に投影することができる。

$$z_c(u_c, v_c, 1)^T = \mathbf{F}_c \mathbf{R}_c \mathbf{P} + \mathbf{F}_c \mathbf{T}_c \quad (4.15)$$

ここで、 c はカラー画像を示すインデックスで、 z_c はカラー画像座標系での奥行き値である。このようにしてマッピングされた変換後のカラー画像 I' は、座標 $I^c(u_c, v_c)$ の画像値を座標 $I'(u, v)$ に代入することにより、取得することができる。式 4.13 を見るとわかるように、変換後カラー画像 I' は視差 d に依存する。そのため、ガイド画像を生成するにあたっては、各視差 d における変換後カラー画像を求め、変換後カラー画像群 $I'_c(u, v, d)$ を用いる。

4.5 実験結果

4.5.1 カラーステレオ画像のみからのデプスマップ推定

はじめに、ドットパターンが投影された赤外線ステレオ画像を使用する必要性を再確認するために、カラーステレオ画像のみから基本処理を用いてデプスマップを推定した。図 4.16 は、カラーステレオ画像のみから、Cost volume filter を用いる基本処理で推定したデプスマップと 3 次元モデルである。(a) のような、テクスチャーが十分ある被写体を撮影したカラーステレオ画像から、CLMF による Cost volume filter で推定した視差画像が (b) である。赤色の画素は、左右の視差チェックにおいて、エラーと判定された画素である。(b) を見ると、土俵上が滑らかなグラデーションになっている。土俵は平面であることから、視差は正しく推定していることが確認できる。(c) の 3 次元モデルを見ても、土俵は正しく平面に復元できているほか、力士や行司も高精度に推定されている。

一方で、(e) は同じ手法とパラメーターを用いて、(d) のようなテクスチャーのない領域を含むシーンを推定したときの視差画像である。ドアや壁といったテクスチャーのない領域の推定精度が著しく低下していることが確認できる。(f) は、(e) から生成した 3 次元モデルであ

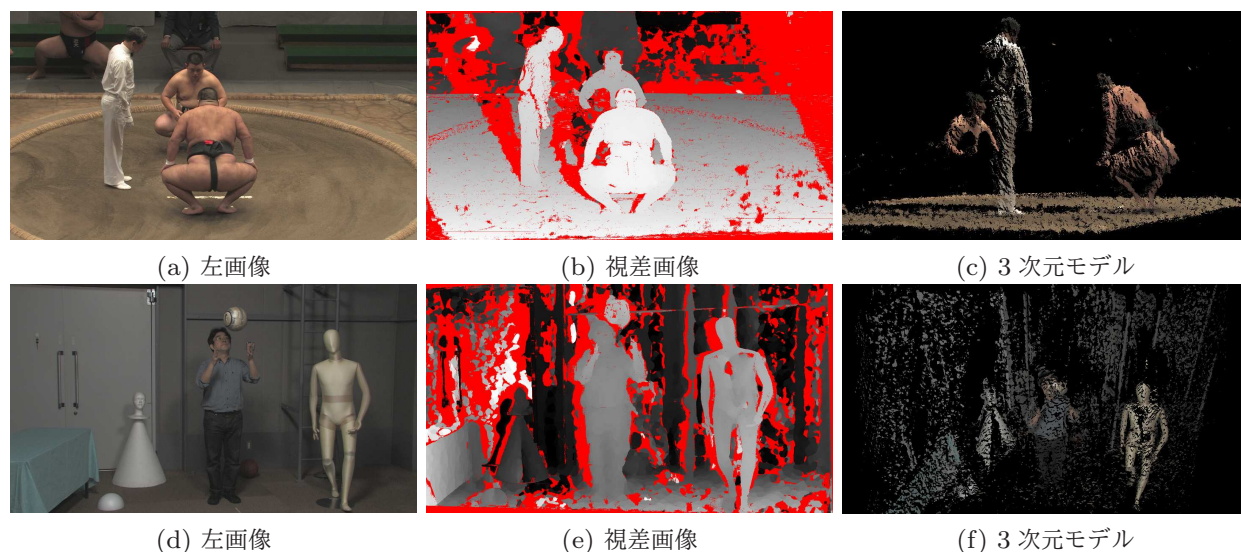


図 4.16 カラーステレオ画像のみから Cost volume filter を用いて推定した結果

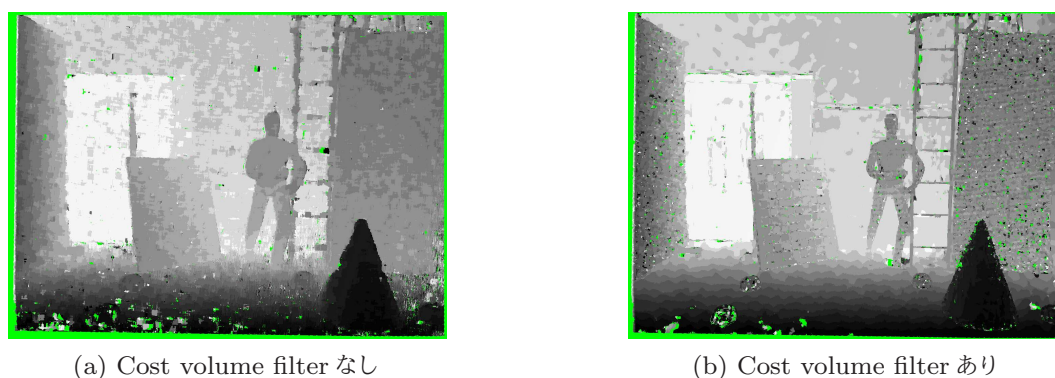


図 4.17 Cost volume filter の効果 (“mannequin75”)

る。Cost volume filter を用いてデプスマップを推定しても、カラーステレオ画像のみから推定すると多数のエラーによる浮遊物が存在し、映像制作に用いるに十分な精度を確保できないことがわかった。このことから、テクスチャーのない領域は Cost volume filter を用いても、カラーステレオ画像のみから推定することは困難であることが確認された。

4.5.2 Cost Volume Filter の効果

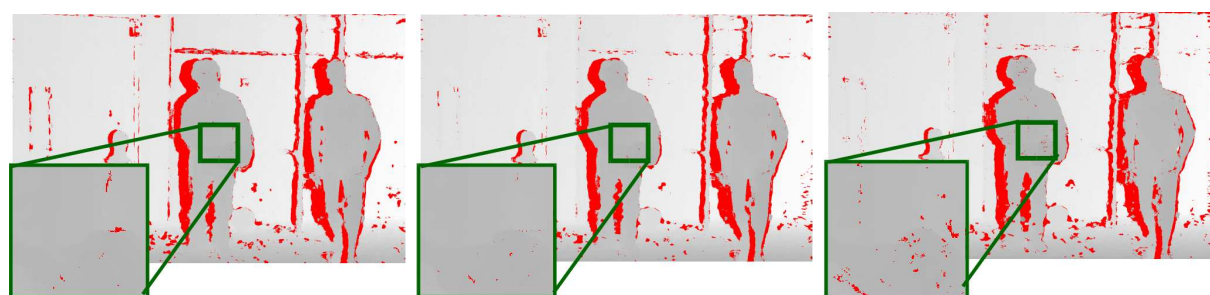
次に、Cost volume filter の効果を確認するために、赤外線ステレオ画像のみから Cost volume filter を用いて推定したデプスマップと、用いずに推定したデプスマップを比較した。図 4.17 は、“mannequin75” データセットの赤外線ステレオ画像のみから、(a) Cost volume filter を用いなかったときと、(b) 用いたときのデプスマップを比較したものである。(b) を見ると、(a) に比べて被写体境界付近の膨張現象が抑制されており、デプスマップの推定精度が改善されていることが確認できた。また、被写体表面がカメラに正対していない床などの領域



(a) 赤外線 Cost のみからの推定

(b) 赤外線 Cost とカラー Cost からの推定

図 4.18 赤外線 Cost とカラー Cost の統合の効果 (“toss”)

(a) $\lambda=0.2$ (b) $\lambda=0.5$ (c) $\lambda=0.8$ 図 4.19 重み付け定数 λ の違いによる推定デプスマップの違い (“toss”)

についても、適応型カーネルが導入されることで不適切な Cost が排除され、精度が改善している。一方で、左側側面の壁や右側の壁など、テクスチャーの多い領域については、スポット状の推定エラーが発生した。

4.5.3 赤外線とカラーの Cost volume の統合

4.5.2 節では Cost volume filter を用いても、赤外線ステレオ画像のみからデプスマップを推定するとスポット状の推定エラーが発生した。そこで、基本処理では赤外線ステレオ画像とカラーステレオ画像の両方から Cost volume を生成したのちに、Cost volume filter 処理した Cost volume をカーネルサイズで選択して、1つの Cost volume に統合する手法を提案した。ここでは、Cost volume の統合を行う基本処理の実験を行う前に、赤外線 Cost とカラー Cost の両方を導入する効果を確認するために、2つの Cost を重み付け定数を用いて Cost 統合した Cost volume から Cost volume filter を用いてデプスマップを推定した結果を示す。図 4.18 は、“toss” データセットから ZNCC で算出した Cost を用いて実験した結果を示したものである。(a) は赤外線 Cost のみから推定した視差画像である。赤色の画素は、左右の視差画像チェックで、推定エラーと判定された画素である。(b) は、赤外線 Cost とカラー Cost を重み付け定数で統合した Cost から推定した視差画像である。(a) よりも、推定に失敗した画素が



(a) 赤外線 Cost のみ (b) カーネルサイズが閾値以下の領域 (c) 赤外線 Cost とカラー Cost

図 4.20 カーネルサイズによる赤外線 Cost とカラー Cost の選択による奥行き推定

少ないことが確認できる。

また、図 4.19 は、重み付け定数 λ の違いによる推定デプスマップの違いを比較したものである。異なる重み付け λ で、推定したデプスマップに大きな違いは発生しなかったが、左下の拡大図の結果や画面中央の水平方向に伸びた推定エラーの様子から、赤外線 Cost とカラー Cost を半分づつ足しあわせた $\lambda=0.5$ の結果が最も精度高くデプスマップを推定できた。

基本処理では、この 2 つの Cost の統合方法を高度化し、Cost volume filter 処理した赤外線 Cost volume とカラー Cost volume をカーネルサイズにより、各画素について赤外線 Cost もしくはカラー Cost のいずれか 1 つを選択して、1 つの Cost volume に統合する。図 4.20 は、"mannequin75" データセットから、基本処理を用いてデプスマップを推定したデプスマップとカラー Cost volume を用いずに推定したデプスマップを比較したものである。(a) は赤外線ステレオ画像から求めた赤外線 Cost のみを用いて推定したデプスマップである。全体的に高精度にデプスマップを推定できているが、左側面の壁や右側の背景にある壁など、テクスチャーの多い領域において、白色や黒色のスポット状の奥行き推定エラーが発生している。これらの領域は、テクスチャーが細かいために、適応型カーネルのサイズが小さくなってしまったために、カーネル内に十分にドットが含まれないために推定に失敗してしまったと考えられる。(b) のオレンジ色の画素は、適応型カーネルのサイズが式 4.12 の閾値 b^2 以下となる点を示している。実験では、ブロック内にドットパターンのドットが平均 1 個以上入るように、閾値 b^2 は 49 画素とした。(a) において、白色や黒色の推定に失敗した点は、(b) のオレンジ色の画素と一致することから、ドットパターンがブロック内に十分入らないことが推定エラーを発生していることが推測される。(c) は、(b) のオレンジ色の領域ではカラー Cost を用い、それ以外では赤外線 Cost を用いて Cost volume を生成して、推定したデプスマップを示している。(a) において、スポット状に発生していた推定エラーの点が減少しており、テクスチャーが細かい領域についても高精度に奥行きを推定できている。

4.6 まとめ

テクスチャーの少ない領域のある被写体を含め、安定的にデプスマップを推定できるように赤外線ドットパターンを照射し、それを 2 台の赤外線カラーカメラで撮影して取得した赤外線ステレオ画像とカラーステレオ画像から、デプスマップを推定する基本処理を提案した。基本処理を提案するにあたって、はじめに赤外線ステレオ画像からの奥行き推定処理の基本特性を調査した。その結果、赤外線ドットパターンの照射により、シーン全体にドットが投影され、テクスチャーの少ない領域を含む被写体全体のデプスマップを安定して推定することができることを確認した。また、赤外線ドットパターンは目を保護するため、出力光量が制約されているため、赤外線ステレオ画像にはノイズが多く含まれ、それらが奥行き推定に悪影響を及ぼすことを確認した。ブロックサイズについては、異なるサイズでブロックマッチングを行い、ドットパターンはテクスチャーとしては粗く、大きなブロックサイズを必要とすることがわかった。しかし、大きなブロックを用いたブロックマッチングでは被写体境界付近の膨張現象が発生し、膨張現象を抑制する推定処理が必要となることがわかった。Cost については、主要な Cost を用いてデプスマップを推定し、赤外線ドットパターンの照射された赤外線ステレオ画像のマッチングには、ZNCC が適していることが確認された。そこで、Cost を ZNCC を用いて定義し、エッジ保存型平滑フィルターの一つである CLMF による Cost volume filter を用いたデプスマップ推定手法を基本処理として提案した。このとき、CLMF の適応型フィルターが小さいとき、赤外線 Cost の信頼度が低下することから、Cost volume を生成する際に、カーネルサイズにより赤外線 Cost とカラー Cost を選択して統合することとした。この基本処理により、膨張現象を抑制した高い精度のデプスマップを推定できた。

第 5 章

2 次元グラフカットを用いた奥行き推定

5.1 はじめに

第 4 章では、赤外線ドットパターンを照射した赤外線ステレオ画像とカラーステレオ画像から生成した Cost volume に対して、Cost volume filter 処理を行ったのち、各画素について Cost が最小となる視差を選択することにより、視差画像を推定する基本処理を提案した。これにより、赤外線ステレオ画像とカラーステレオ画像から、テクスチャーのない領域も含めて安定的に高精度なデプスマップを推定できるようになった。しかし、Cost volume filter は、同じ視差 d の Cost map に対して 2 次元エッジ保存型平滑化フィルターで平滑化処理を行うため、潜在的に被写体の表面はカメラに正対していることを前提としている。そのため、表面がカメラに正対していない領域では推定精度が低下する。特に、カメラ間のベースラインが長い場合には、カメラに正対していない領域は射影変換による左右画像の変形が大きいため、推定精度は著しく低下する。

本研究は、映像制作への応用することを考慮した時系列デプスマップを推定することにある。そのため、高解像度の 3 次元モデルを生成するために、水平垂直方向の解像度が高いことが望まれるほか、奥行き方向の解像度も高いデプスマップを推定する必要がある。奥行き方向の解像度を高くするためには、カメラ間のベースラインを長くする必要があるが、ベースラインを長くすると射影変換の影響を大きく受ける。そこで、本章では図 5.1 のように、基本処理の視差選択処理において、Cost が最小となる視差を Winner-takes-all 戦略により選択する代わりに、最適化アルゴリズムの一つであるグラフカットを導入することにより、表面がカメラに正対していなく、射影変換の影響を大きく受けても頑強にデプスマップを推定できる手法を提案する [49][20]。

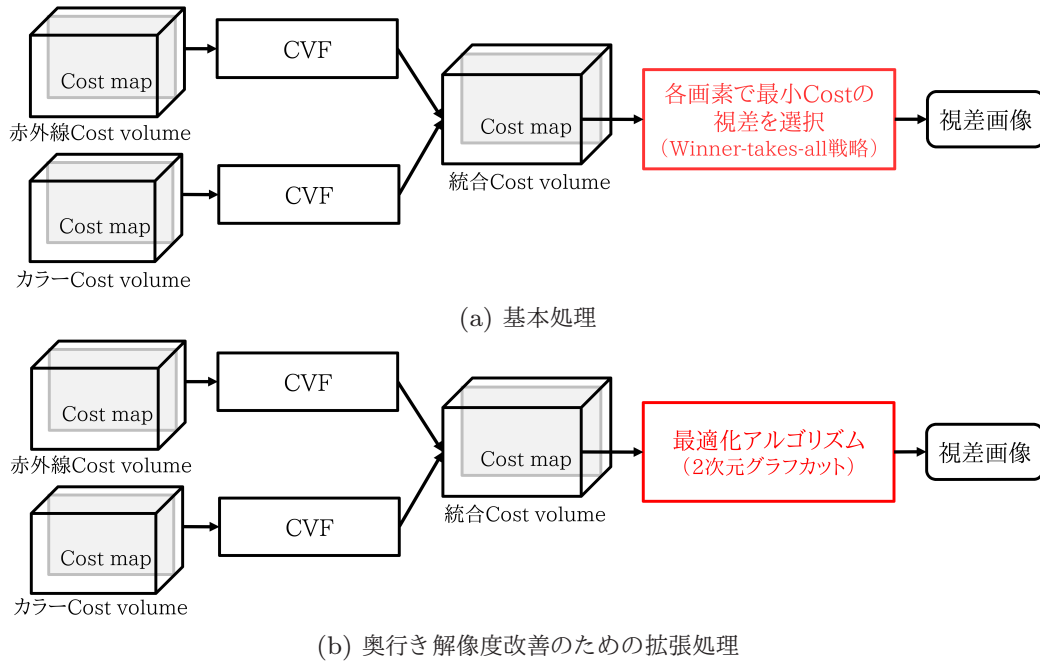


図 5.1 赤外線カラーステレオ画像からのデプスマップ推定手法

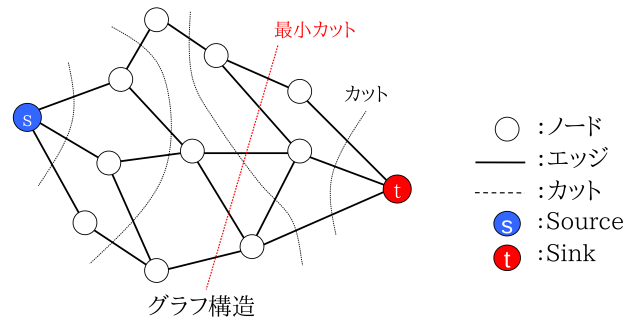


図 5.2 グラフカット

5.2 グラフカットを用いた従来のデプスマップ推定

グラフカットは、図 5.2 のように、ノード、エッジ、特別なノードである Source と Sink で構成されるグラフ構造を用いることにより、2 分割問題をエネルギー問題として扱えるアルゴリズムである。エッジはノード間を結び、各エッジにはエネルギーが与えられている。グラフカットは、このグラフ構造を Source から Sink に流れるネットワークと考え、エッジエネルギーを伝送容量として扱う。このとき、Source から Sink への最大伝送容量 (最大フロー) を考え、伝送量が飽和しているエッジを求める。最大フロー/最小カットの定理を適用すると、飽和しているエッジを結ぶことで、ノードを 2 分割する際にカットするエッジエネルギーの総和が最小となるようなカット (最小カット) を求められるため、対象を最適に 2 分割すること

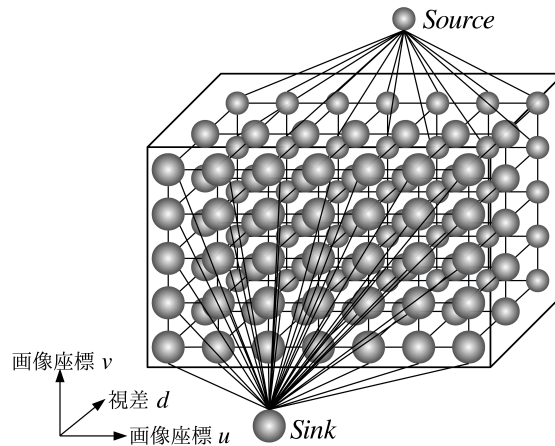


図 5.3 一般的なグラフカットを用いたデプスマップ推定法

ができるアルゴリズムである。

このグラフカットを用いて、カラーステレオ画像からデプスマップを推定する手法が提案されている [3]。一般的なグラフカットを用いたデプスマップ推定法は、図 5.3 のように Cost volume の各 Cost の存在する位置にノードを設定し、Cost を各エッジにエネルギーとして付与する。そして、Source を各画素の視差 d_{min} のノードと接続し、Sink を各画素の視差 d_{max} のノードと接続した 3次元グラフ構造上で、最大フロー/最小カットアルゴリズムを実施する。その結果、各画素について最小カットで Source 側と Sink 側に別れる視差 d を当該画素の視差 d として選択し、デプスマップを求める。

この手法は、Cost volume filter に比べて、同一視差内の近隣画素の Cost を考慮するだけでなく、近隣画素の異なる視差の Cost も考慮されるため、表面がカメラに正対していない領域も含めて高精度にデプスマップ推定することができる。しかし、アルゴリズムが複雑であるため、多くのメモリを必要とするほか、膨大な計算時間を必要とするため、映像制作へ応用するには、多くのフレームを処理して時系列デプスマップを推定する必要があるため不向きである。

5.3 2次元グラフカットを用いた奥行き推定

画像平行化処理を行ったステレオ画像は、左右画像における射影変換による変形は水平方向にのみ発生する。そこで、本章ではこの点に注目し、グラフカットをスキャンライン毎に実施して、2次元グラフ構造上で最大フロー/最小カットアルゴリズムを実施することで、使用メモリの大幅削減と計算時間の短縮を実現する手法を提案する。

図 5.4 は、スキャンライン毎にグラフカットを実施する際のグラフ構造を示したものである。ノードは、対象のスキャンライン上にある各画素のすべての視差 d について配置される。

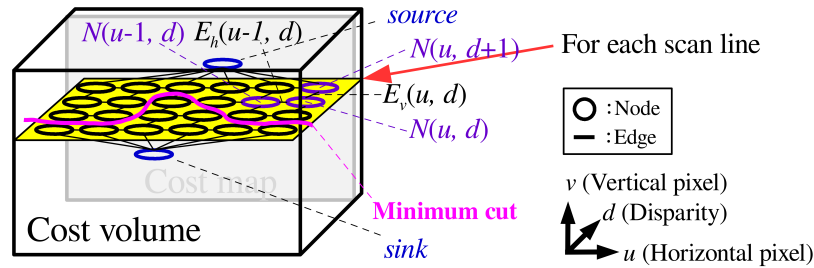


図 5.4 スキャンライン毎にグラフカットを実施する際の 2次元グラフ構造

ノード $N(u, v, d)$ と $N(u, v, d + 1)$ 間の視差方向のエッジには、Cost $C_t^i(u, v, d)$ で定義されるエネルギー $E_d(u, d)$ が付与される。ノード $N(u - 1, v, d)$ と $N(u, v, d)$ 間の水平方向のエッジには、ペナルティ一定数 ρ で定義されるエネルギー $E_h(u - 1, d)$ が付与される。各スキャンラインについて、このような 2次元グラフ構造を構築した上で、最大フロー/最小カットを実施すると、ノードを最適に 2分割することができる。各画素について、Source 側と Sink 側に分離される境目にあるノードの視差 d を抽出すると、当該スキャンラインの視差画像を取得できる。この処理をすべてのスキャンラインについて行うことで、視差画像を取得することができる。

通常、スキャンライン毎で独立にグラフカットを実施する場合、3次元グラフ構造上でグラフカットを実施する場合に比べて、垂直方向の連続性は考慮されないため、推定したデプスマップ上で、墨引き現象 (ストリーキング) が発生する。しかし、本手法では、スキャンライン毎にグラフカットを実施する前に、Cost volume filter を実施し、その段階で垂直方向も含めて近傍の連続性が考慮されているため、ストリーキングの発生を抑制することができる。また、射影変換による左右画像の形状変化については、水平方向にのみ発生するため、これらについては水平方向のみのグラフカットで対応できる。

このようにして取得した視差画像は、第 4 章で提案した基本手法と同様に、左右視差画像の推定チェックののち、カメラパラメーターを用いて、三角測量の原理からデプスマップを取得することができる。

5.4 実験結果

2次元グラフカットを用いてデプスマップを推定する提案手法の有効性を示すため、実世界を撮影した 3.4 節のデータセットを用いて実験を行った。はじめに、グラフカットを用いずに、Cost volume filter のみでデプスマップを推定した結果を図 5.5 に示す。(a) はベースラインが短い "mannequin75" データセットから推定したデプスマップである。緑色の画素は、左右視差チェックにおいて、推定エラーと判定された画素である。ベースラインが短いとき、左右の画像で射影変換の影響による形状の変化は少ないため、左右視差チェックでエラーと判定

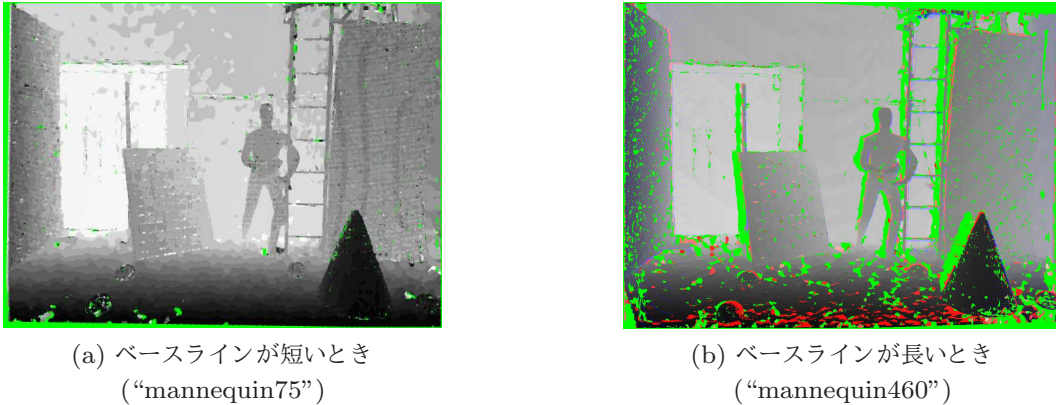


図 5.5 グラフカットを用いず、Cost volume filter のみで推定したデプスマップ

された画素が少なく、高精度なデプスマップを推定できている。しかし、デプスマップの階調を見ると粗く、これを 3次元モデルに変換すると離散的な奥行き値をもったものになってしまう。一方、(b) はベースラインが長い”mannequin460” データセットから推定したデプスマップである。赤色の画素は、図 5.6(d) の Ground Truth から閾値以上の誤差があった画素を示す。ベースラインが長いとき、推定したデプスマップの階調は高く、奥行き方向に解像度が高いデプスマップを推定できているが、推定エラーの画素が多くなってしまっている。

図 5.6 は、ベースラインが長い”mannequin460” データセットから、2次元グラフカットを用いて推定したデプスマップを、Xtion の推定結果や Ground Truth と比較したものである。(a) は、2次元グラフカットを用いずに Cost volume filter のみを用いる基本処理によるデプスマップ推定結果である。表面がカメラに正対していない領域で、多くの推定エラーが発生している。一方、(b) は、基本処理が画素毎に最小 Cost を視差として選択する代わりに、2次元グラフカットを用いて推定したデプスマップである。推定エラーの画素が大幅に低減しているのが確認できる。(c) は、赤外線ドットプロジェクターとして用いている Xtion に実装されているデプスマップ推定手法で取得した取得したデプスマップである。デプスマップの階調が低く、奥行き解像度が低くなっているほか、被写体の境界付近の精度が低下しているため、被写体中央付近のポールなどの推定に失敗している。(d) は、レンジスキャナーを用いて取得したデプスマップの Ground truth である。なお、青色の画素はレンジスキャナーと赤外線カメラの左側のカメラの位置が異なるために発生したオクルージョン領域で、奥行き情報が取得できなかった画素である。各デプスマップを (d) と比較すると、2次元グラフカットを用いた (b) が、最も高い精度で推定できていることが確認できた。

図 5.7 は、図 5.5 を定量的に解析した結果である。(a) は、表面がカメラに正対していない平面領域である床付近を指定し、その領域内の点群から最小二乗法を用いて近似平面を算出して、近似平面から各点群までの距離を平均した平均誤差を示したものである。2次元グラフカットの導入により、平均誤差が軽減されており、正対していない領域の推定精度の改善を確

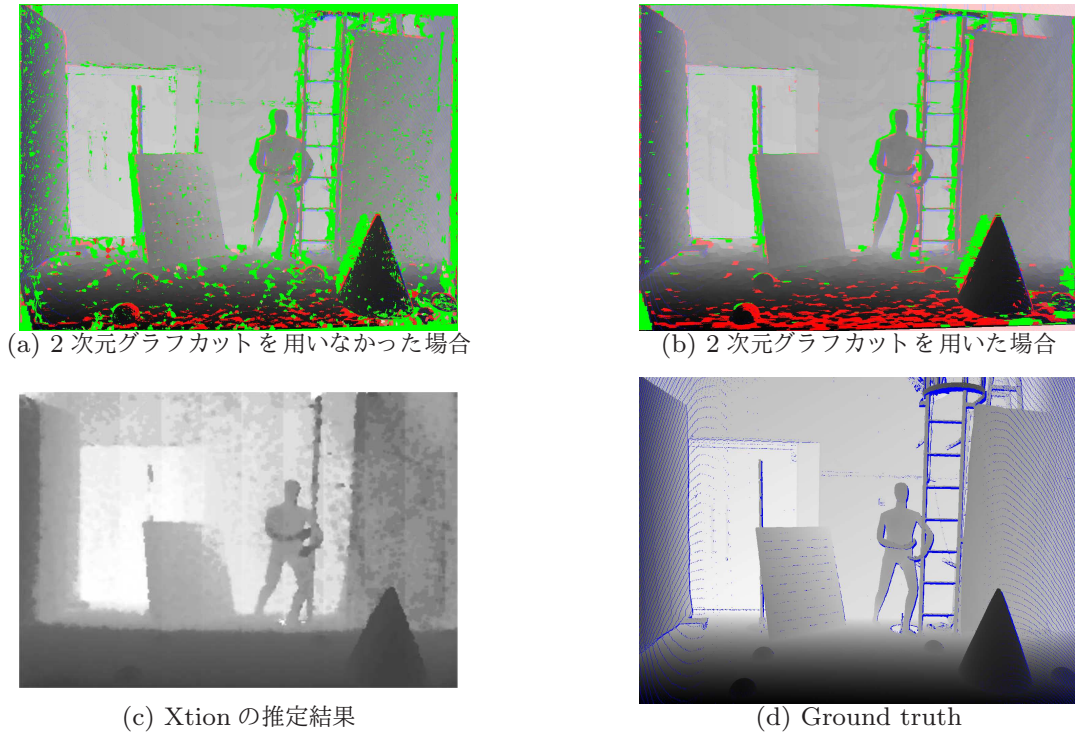


図 5.6 2次元グラフカットを用いた場合と用いなかった場合の推定結果

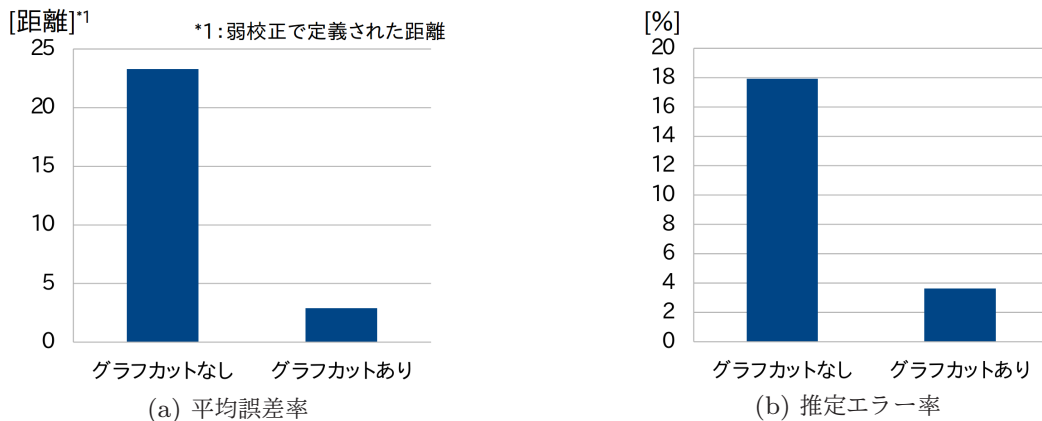
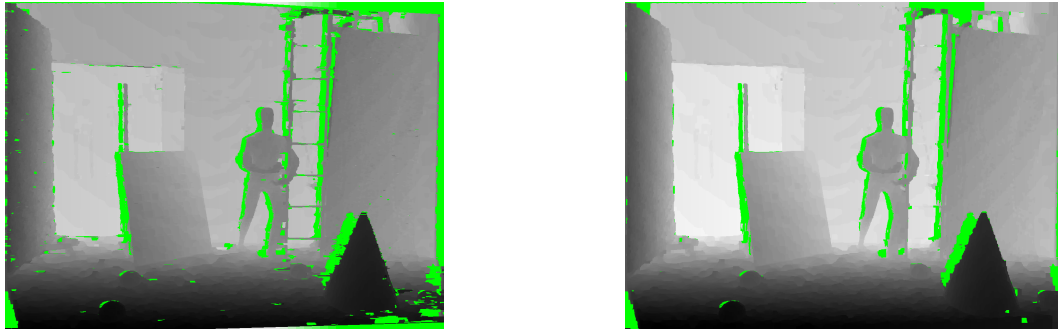


図 5.7 2次元グラフカットを用いたときと用いなかったときの精度の定量的な比較

認することができた。また、(b) は、左右視差チェックで推定エラーと判定された画素の、総画素数に対するパーセントを示した推定エラー率を示している。これについても、2次元グラフカットの導入により、大幅に軽減できることを確認できた。

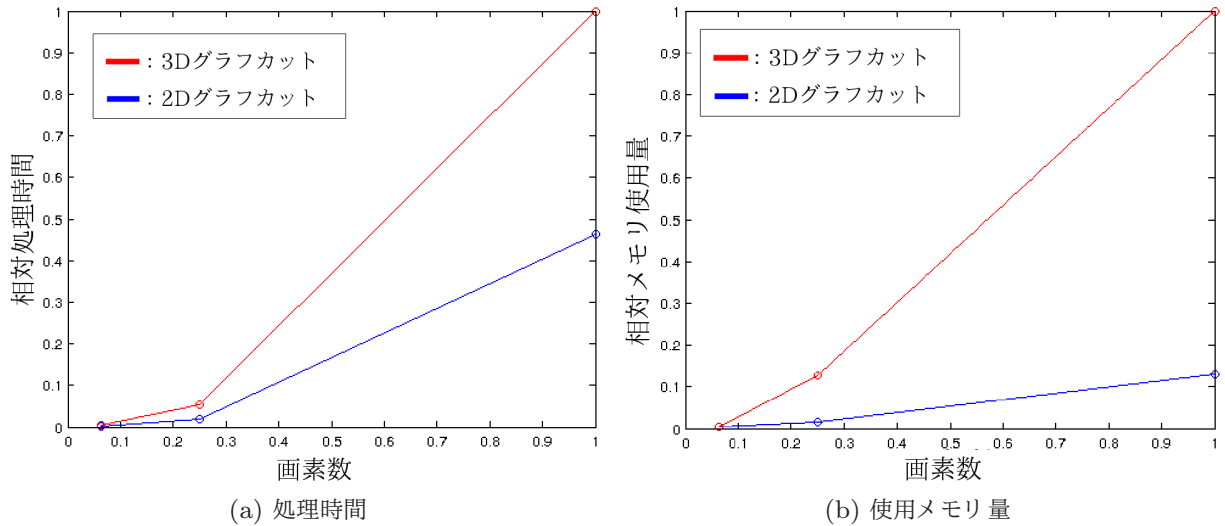
前述のとおり従来の手法では、一般的にグラフカットはすべての画素の全視差分について、ノードを設定した3次元グラフカットを用いて奥行きを推定する。図 5.9 は、提案する2次元グラフカットを用いて推定したデプスマップと一般的な3次元グラフカットを用いて推定したデプスマップを比較したものである。3次元グラフカットを用いて推定したデプスマップの



(a) 2次元グラフカット

(b) 3次元グラフカット

図 5.8 2次元グラフカットと3次元グラフカットを用いて推定したデプスマップの比較



(a) 処理時間

(b) 使用メモリ量

図 5.9 2次元グラフカットと3次元グラフカットを用いた奥行き推定手法の処理時間と使用メモリ量の比較

方が推定エラー画素が少ないものの、ほとんど精度の差のないデプスマップを取得することができた。2次元グラフカットを用いた提案手法では、スキャンライン毎で独立に処理しているが、Cost volume filterにより垂直方向の連続性が考慮され、ストリーキングの抑制されたデプスマップを推定することができた。

一方で、2次元グラフカットは、3次元グラフカットに対して、処理時間と使用メモリ量で大きな改善を図れた。図 5.9 は、2次元グラフカットと3次元グラフカットでデプスマップを推定した際の処理時間と使用メモリ量を比較したものである。処理時間は、デプスマップの水平・垂直画素数と視差数に依存するため、(a)では”mannequin460”データセットの原画像の画素数とそれを処理する時間を1としたときの、相対画素数と相対処理時間との関係を示したものである。この図より、2次元グラフカットは3次元グラフカットに比べて処理時間が約半分となっていることを確認できた。また、使用メモリ量もデプスマップの水平・垂直画素数と視差数に依存するため、(b)では”mannequin460”データセットの原画像の画素数とそれを処

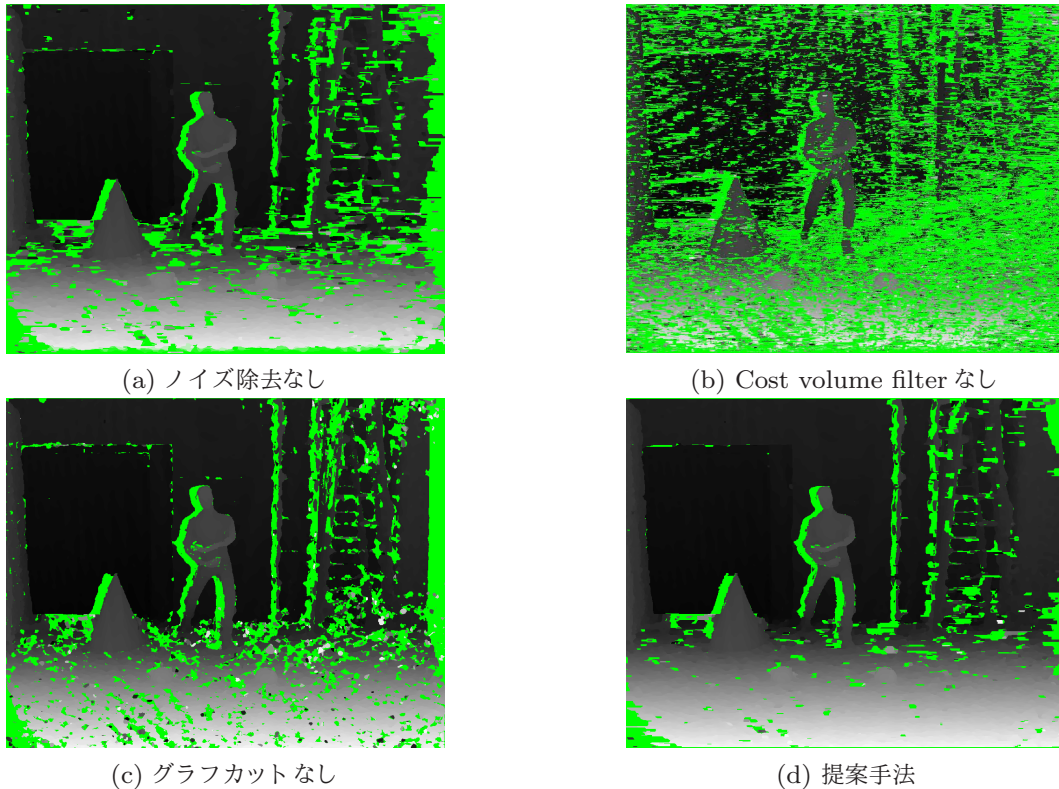


図 5.10 推定したデプスマップの比較

理するために必要なメモリ使用量を 1 としたときの、相対画素数と相対メモリ使用量との関係を示したものである。この図を見ると、2次元グラフカットで必要となるメモリ使用量は、3次元グラフカットに比べて、約 1/8 で済むことが確認できた。映像制作に応用するにあたっては、多数のフレームで構成された時系列のデプスマップを、高解像度で推定する必要があるため、処理時間の短縮とメモリ使用量の削減は大きな効果となる。

図 5.10 は、提案手法から各処理を除いてデプスマップを推定することで、各処理の効果を確認した視差画像である。(a) はノイズ除去処理を除いて推定したデプスマップである。(d) の提案手法と比較すると、左右視差チェックの結果、推定エラーと判定された画素が多いことから、Graphcut を用いたデプスマップ推定においても、赤外線画像の SN 比の低さが推定率に影響を与えていることが確認できた。(b) は Cost volume filter 処理を除いて推定したデプスマップである。画面全体に推定エラーと判定された画素が発生し、Graphcut 処理を用いる場合においても、その前処理として Cost volume filter が必要なことが確認できた。(c) は 2次元グラフカットを用いて推定したデプスマップである。このデータセットにおいても、射影変換の影響が大きい正対していない領域における推定率の改善が確認できた。

図 5.10 は、グラフカットを用いずに推定したデプスマップとグラフカットを用いて推定したデプスマップの推定率を比較したものである。グラフカットを用いて推定したデプスマップの推定率が、用いずに推定したものよりも高くなっており、このデータセットにおいても定量

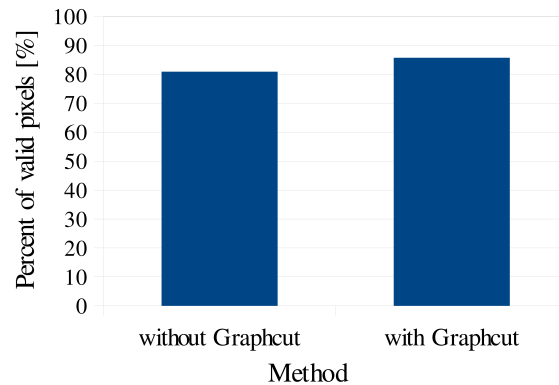


図 5.11 グラフカットを用いたときと用いなかったときの推定率の比較

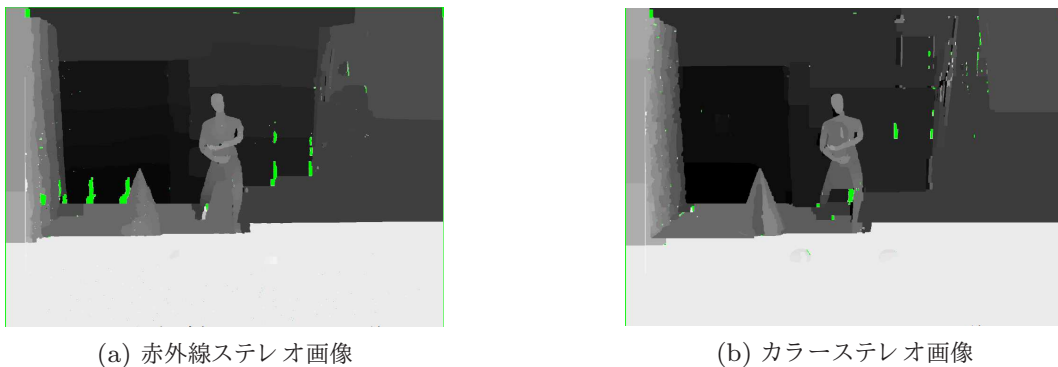


図 5.12 OpenCV のグラフカット手法を用いて推定した視差画像

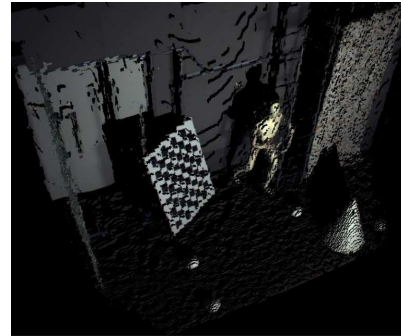
的な改善を確認することができた。

また他の手法との比較するために、(a) 赤外線ステレオ画像と (b) カラーステレオ画像から、それぞれ OpenCV に実装されているグラフカットを用いて推定した視差画像を図 5.12 に示す。赤外線、カラーいずれのステレオ画像から推定した視差画像においても、被写体境界付近のエッジが保存された視差画像を推定できているが、射影変換の大きい床の推定に失敗している。また、視差画像の階調が少なく、奥行き解像度が減少してしまっている。そのほか、梯子のような細い構造物で構成された被写体の推定に失敗している。このことから、提案手法は一般的なグラフカットによる手法より、有効であることが確認できた。

図 5.13 は、提案手法を用いて推定したデプスマップを、カメラパラメータを用いて 3次元モデルに変換したものを比較している。(a) はベースラインの短い”mannequin75” データセットから推定した 3次元モデルである。デプスマップの階調が低いため、3次元モデル上のキャリブレーションパターンや手前のコーン型のオブジェクトを見ると、奥行き方向の解像度に離散的であることが確認できる。一方、(b) はベースラインの長い”mannequin460” データセットから推定した 3次元モデルである。(a) と比べて、キャリブレーションパターンやコーンを見ると、奥行きの解像度が高いことが確認できる。本手法を用いることにより、ベースラ



(a) ベースラインが短いとき



(b) ベースラインが長いとき

図 5.13 ベースラインが長いときと、短いときの 3 次元モデルの比較

インを長くしても、表面がカメラに正対していない領域も含めて、安定して推定することができるようになり、(b) のような奥行き解像度の高い 3 次元モデルを生成することができるようになった。本手法は、カーブフィッティングやデプスマップ推定後に 2 次元エッジ保存型平滑フィルターなどによるサブピクセル精度の後処理を行っていない。これらを用いることにより、奥行き方向の滑らかさはさらに改善されると考えられる。

本手法の一つの特徴として、一般的なカラーカメラや赤外線カメラからデプスマップを推定できる点が挙げられる。これらのカメラは、近年非常に高解像度のものが入手可能となっている。そこで、高解像度のカメラを用いて、高解像度のデプスマップを推定する実験を行った。図 5.14 は、(a) 第 3 章で記載した赤外線カラー一体型カメラアレーと、(b) 2048×2048 画素の高解像度の赤外線カラー分離型カメラアレーで撮影した赤外線ステレオ画像とカラーステレオ画像から推定したデプスマップを比較したものである。(a) と比べて、(b) ははるかに高解像度のデプスマップを高精度に推定することができた。図 5.15 は、図 5.14 のデプスマップをカメラパラメーターを用いて生成した 3 次元モデルを上部から見た時の画像を比較したものである。(a) と比べて、(b) は高解像度の画像から推定したため、3 次元モデル上で水平方向にドットが密になっているほか、奥行き方向についても、左画像の各画素に対応する右画像の画素を密に探索できるため、密な点群を推定できている。映像制作の応用にあたっては、高い品質の映像を制作する必要があるため、空間的に高解像度の 3 次元モデルの生成が要求されることから、高解像度カメラを用いて、高解像度の 3 次元モデルが生成できることは大きなメリットとなる。

5.5 まとめ

第 4 章で述べた Cost volume filter 処理を導入した基本処理では、ベースラインが短いとき、高精度なデプスマップを推定することができた。しかし、ベースラインが短いと、奥行き解像度が低くなってしまふ。映像制作への応用を考慮した時系列 3 次元モデルを生成するた

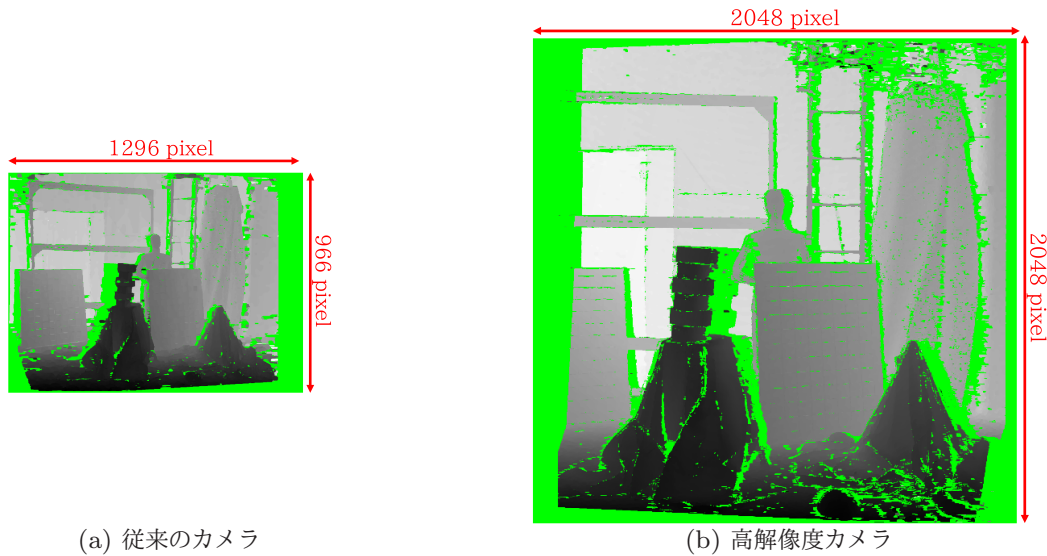


図 5.14 従来のカメラと高解像度カメラから推定したデプスマップ

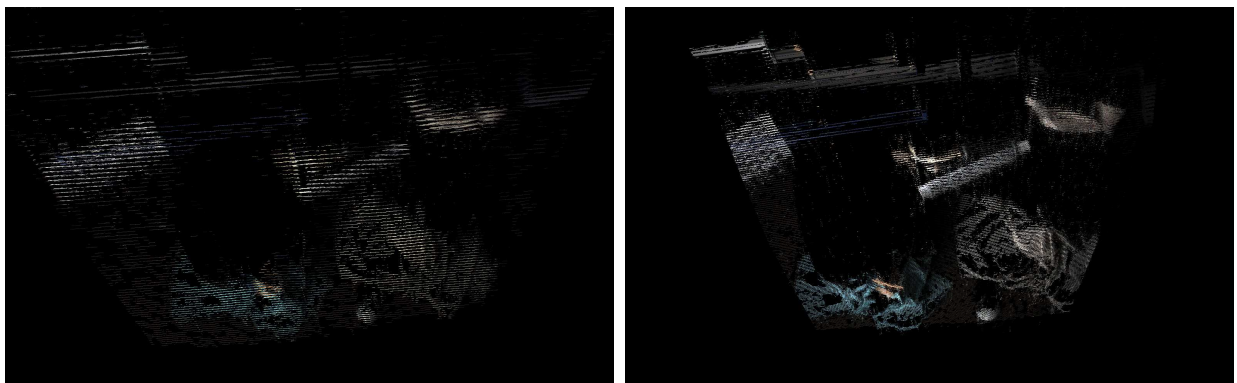


図 5.15 従来のカメラと高解像度カメラから推定した 3次元モデル

めには、ベースラインを長くして奥行き解像度を高くする必要があるが、ベースラインを長くすると、カメラに正対していない領域の精度低下が課題となっていた。これは、Cost volume filterを用いたデプスマップ推定手法が、同視差の近傍画素の Cost を共有化して精度の改善を図っているため、潜在的に被写体の表面がカメラに正対していることを前提としているために発生した。

本章では、赤外線カラステレオカメラの左右の撮影位置が異なるために生じる、射影変換による 2次元画像上の形状変化に対して、画素毎の頑強なマッチングを実現するために、2次元グラフカットを導入したデプスマップ推定手法を提案した。グラフカットを用いた一般的なデプスマップ推定手法は、すべての画素の全視差にノードを割り付ける 3次元グラフカットを用いるのが一般的であるが、処理時間とメモリ使用量が膨大であるため、複数フレームを処理する必要がある時系列デプスマップの生成には不向きである。本手法では、赤外線ステレオ画

像の SN 比が低いことから、グラフカットの前に Cost volume filter 処理を入れるとともに、Cost volume filter により垂直方向の連続性を確保した上で、スキャンライン毎に独立に 2次元グラフカットを実施した。これにより、スキャンライン毎に実施すると、通常懸念されるストリーキング現象も少なく、3次元グラフカットを用いた推定結果とほぼ同程度の精度を実現することができた。また、3次元グラフカットによる推定精度を維持したまま、処理時間で半分、メモリ使用量で約 7/8 の改善を図ることができた。

第 6 章

st-Cost Volume Filter による奥行き推定

6.1 はじめに

第 5 章では、図 6.1(a) のように、赤外線ステレオ画像とカラーステレオ画像から生成した Cost volume に対して、Cost volume filter 処理をしたのち、スキャンライン毎に 2 次元グラフカットを実行することにより、カメラ間の距離が離れていても、カメラに正対していない領域を含めて高精度なデプスマップを推定できる拡張処理を提案した。しかし、この手法はフ

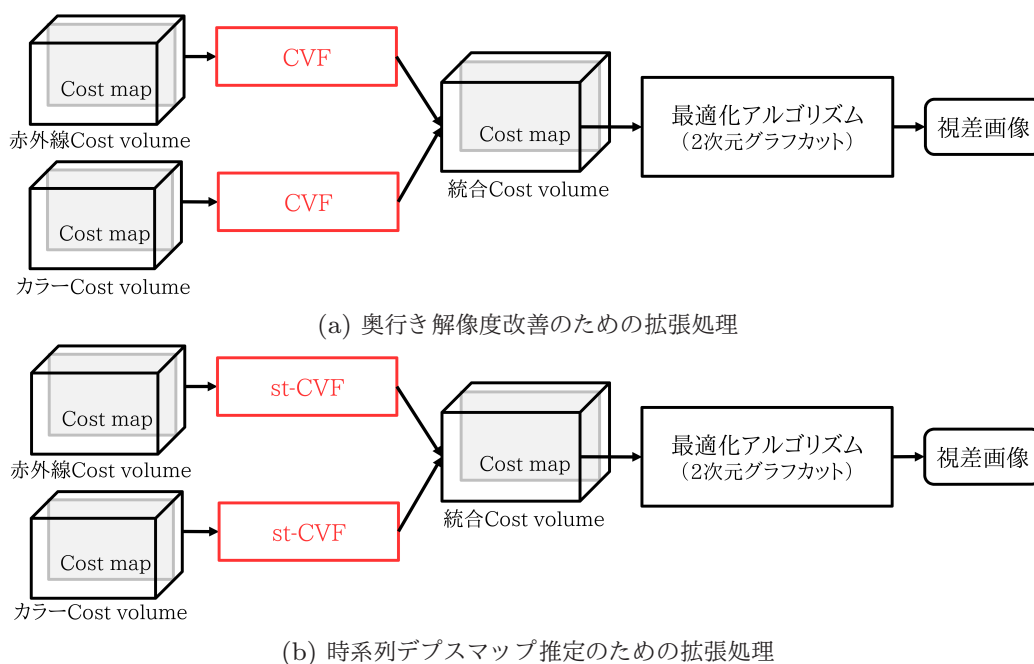
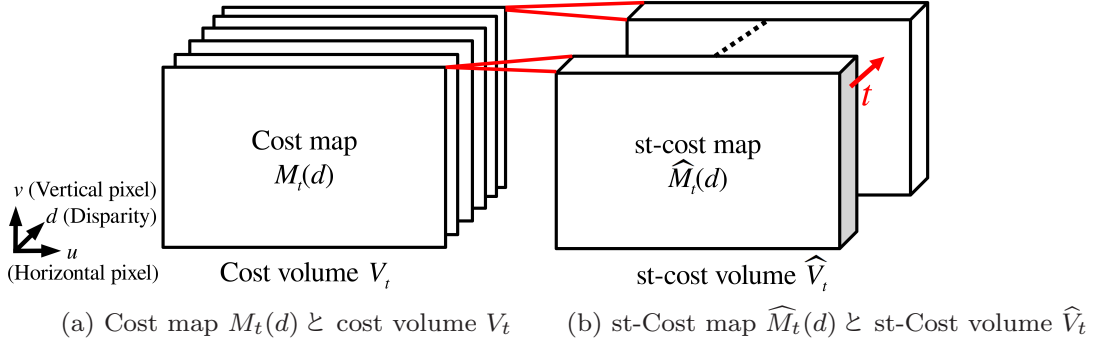


図 6.1 赤外線カラーステレオ画像からのデプスマップ推定手法


 図 6.2 Cost map $M_t(d)$ と Cost volume V_t の時間拡張

フレーム内の処理となり、時間方向の連続性が考慮されていない。そのため、複数フレームで構成された時系列のデプスマップを推定すると、フレーム毎に独立なデプスマップが生成されるため、時系列デプスマップ上でフリッカーを生じてしまう。本研究では、映像制作に応用可能な時系列 3 次元モデルを生成することが最終目標であることから、時間方向の連続性が考慮された、動画としての高品質の時系列デプスマップを生成することが必要である。

そこで、本章では Cost volume filter を時間方向に拡張した st-Cost volume filter (Spatial-temporal cost volume filter) を提案する [11][20]。Cost volume filter を時間方向に拡張するにあたり、はじめに 6.2 節では Cost volume を時間方向に拡張する。次に 6.3 節で、Cost volume filter に用いるエッジ保存型平滑フィルターの CLMF を、時間方向に拡張した st-CLMF (Spatial-temporal CLMF) を提案する。6.4 節では、st-Cost volume と st-CLMF を用いて st-Cost volume filter を提案し、奥行き推定に適用する。6.5 節では、st-Cost volume filter の有効性を示すため、実空間を撮影した赤外線カラーステレオ画像からデプスマップを推定する実験を実施し、6.6 節でまとめる。

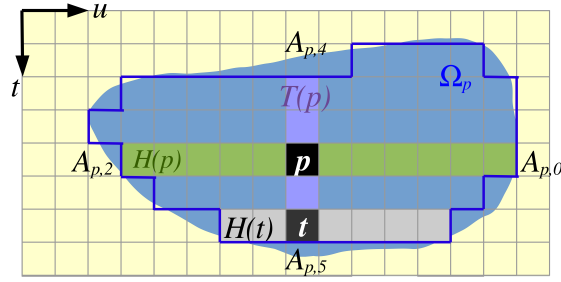
6.2 st-Cost Volume

本節では、st-Cost volume filter を実現するために、Cost map $M_t(d)$ を時間方向に拡張する。各 Cost map $M_t(d)$ は、図 6.2 のように、水平画素と垂直画素の 2 軸を持つ 2 次元配列である。これに時間軸を追加し、同一視差の複数フレームの Cost map を時間軸上に並べた 3 次元配列を生成し、st-Cost map (Spatio-temporal cost map) $\widehat{M}_t(d)$ を定義した。すなわち、st-Cost map $\widehat{M}_t(d)$ は式 6.1 で定義できる。

$$\widehat{M}_t(d) = \{C_\tau(u, v, d) \mid \tau \in [t - \zeta, t + \zeta], u \in [0, w], v \in [0, h]\} \quad (6.1)$$

ここで、 ζ は、時間軸方向に並べたフレーム数の半分である。また、 w と h は、それぞれ水平画素数と垂直画素数である。

Cost volume V_t は、この st-Cost map $\widehat{M}_t(d)$ を視差分グループ化することで、時間方向に


 図 6.3 3次元適応型カーネル Ω_p の水平断面

拡張する。すなわち、st-Cost volume \widehat{V}_t は式 6.2 で定義する。

$$\widehat{V}_t = \{\widehat{M}_t(d) \mid d \in [d_{min}, d_{max}]\} \quad (6.2)$$

ただし、視差の範囲は、 $[d_{min}, d_{max}]$ である。

6.3 st-CLMF

Cost volume filter は、各 Cost map を 2次元エッジ保存型平滑フィルターで平滑化処理を行うことで実施した。Cost volume filter を時間方向に拡張するにあたっては、この 2次元エッジ保存型平滑フィルターを 3次元に拡張する必要がある。そこで、本節では、2次元エッジ保存型平滑フィルターの一つである CLMF を時間方向に拡張した st-CLMF を提案する。図 6.3 は、st-Cost map の水平断面を示したものである。水平軸と垂直軸は、それぞれ水平画素と時間方向を示す。はじめに、図 4.13 の 2次元の適応型カーネルと同様に、中央画素 p から水平方向と垂直方向にアームを伸ばして、ガイド画像上で現在位置の画素値と、次の画素値との差が閾値以上となった点でアームを停止し、アーム長 $A_{p,0} \sim A_{p,3}$ を取得する。次に、図 6.3 のように、時間軸の後ろ方向にアームを伸ばし、同様にガイド画像上で現在位置の画素値と、次の画素値との差が閾値以上となった点でアームを停止し、アーム長 $A_{p,4}$ を取得する。その後、中央画素 p から時間軸の前方向にアームを伸ばし、ガイド画像上で現在位置の画素値と、次の画素値との差が閾値以上となった点でアームを停止し、アーム長 $A_{p,5}$ を取得する。これにより、図 6.3 上で、紺色の領域に対応した、青色の直線で囲まれた領域 Ω の時空間カーネルを抽出することができる。このとき、時間方向のアーム $T(p)$ は、式 6.3 で定義される。

$$T(p) = \{(u, v, t) \mid u = u_p, v = v_p, t \in [t_p - A_{p,4}, t_p + A_{p,5}]\} \quad (6.3)$$

ただし、 t_p は現フレームの番号である。この式 6.3 を用いて、時空間適応型カーネル $\widehat{\Omega}$ は式 6.4 で定義できる。

$$\widehat{\Omega}_p = \bigcup_{t \in T(p)} \Omega_p(t) = \bigcup_{t \in T(p)} \left(\bigcup_{q \in V(t)} H(q) \right) \quad (6.4)$$

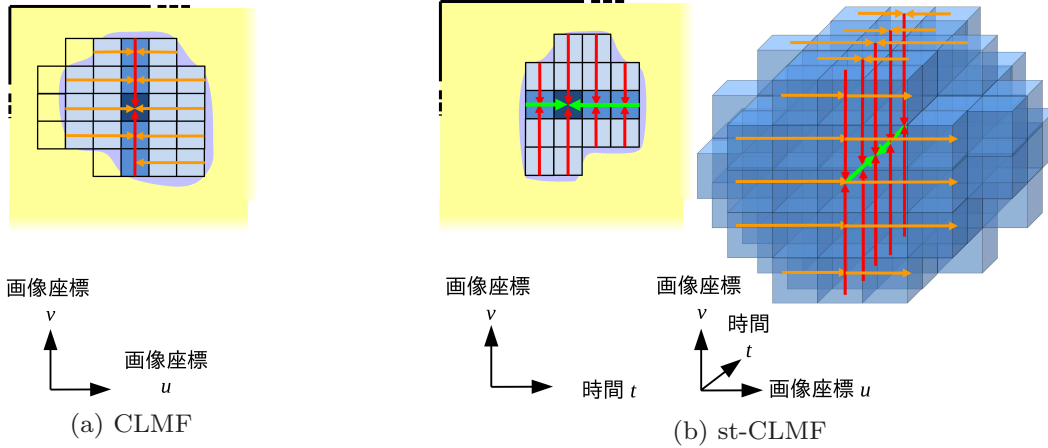


図 6.4 CLMF と st-CLMF の積分画像を用いた処理過程

式 6.4 の計算は、 a_k と b_k が式 6.4 と式 4.7 から計算できることから、平滑化後の出力画像 S は式 4.9 と 1 次元積分画像を 3 回用いて、効率的に計算することができる。

図 6.4 に、CLMF と st-CLMF の積分画像を用いた処理の概念図を示す。CLMF は (a) のように、適応型カーネル内の水平方向の Cost を “背骨” に集めたのち、中心画素に集めることによって、総 Cost を算出する。一方、st-CLMF は CLMF と同様に水平方向と垂直方向の Cost を中心画素に集めたあと、各フレームの中心画素に集められた総 Cost を時空間適応型カーネルの中心画素に集めることにより、総 Cost を算出する。この処理過程の概念を 3 次元的に図示したのが、(b) の右側の画像である。

6.4 st-Cost Volume Filter を用いた奥行き推定

st-Cost volume filter は、6.3 節で提案した st-CLMF を、各視差 d の st-Cost map $\widehat{M}_t(d)$ に対して処理することで実現する [11]。st-Cost volume filter 処理後、各画素について Cost が最小となる視差を選択することにより、デプスマップを推定することができる。時系列デプスマップを推定するにあたっては、この st-Cost volume filter を時間方向にスライドする窓処理を複数フレームに対して行う。

6.5 実験結果

Cost volume filter の時間方向への拡張の効果を確認するため、3.4 節のデータセットを用いて実験を行った。図 6.5 は、3 つのデータセット “man” と “toss”、 “mannequin460” から、(a)(b)(c) は Cost volume filter (CVF) を用いて、(d)(e)(f) は st-Cost volume filter (stCVF) を用いて推定したデプスマップである。これを見ると、st-Cost volume filter によるデプスマップの推定結果の方が、cost volume filter による推定結果に比べてやや改善しているが、大

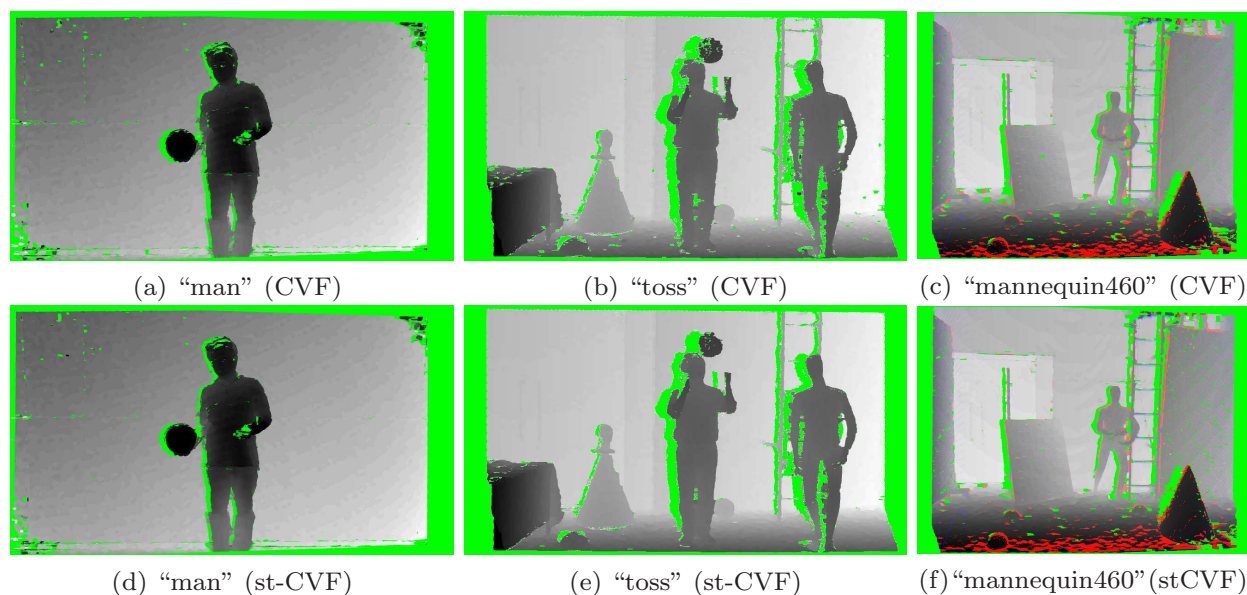


図 6.5 CVF(上段)と st-CVF(下段)によるデプスマップの比較

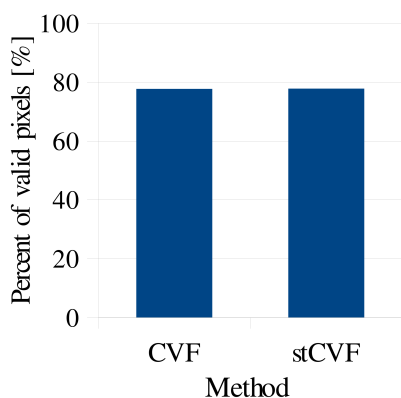


図 6.6 “mannequin460” から CVF と st-CVF で推定したデプスマップの推定率

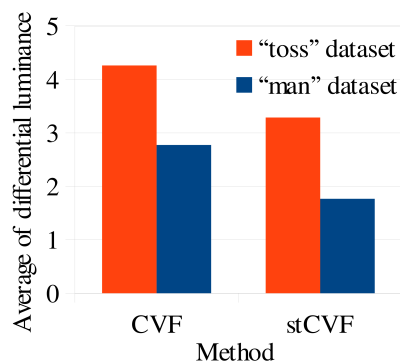


図 6.7 “toss” と “man” から CVF と st-CVF で推定したデプスマップの前後差分画像の平均輝度値

きな改善は見られない。

図 6.6 は、“mannequin460” から、cost volume filter と st-cost volume filter でデプスマップを推定したときの推定率を比較したものである。推定率は、すべての画素数に対する、左右視差チェックののち奥行きを推定できたと判定された画素数の割合である。この図を見ても、Cost volume filter を時間方向に拡張することで、多少の推定率の改善が見られるものの、改善は限定的である。

しかし、フレーム間の連続性は大幅に改善された。図 6.8 は、Cost volume filter と st-Cost volume filter で推定したデプスマップの隣り合う 2 フレーム間での差分画像を示したものである。st-Cost volume filter で推定したデプスマップの差分画像の方が、Cost volume filter

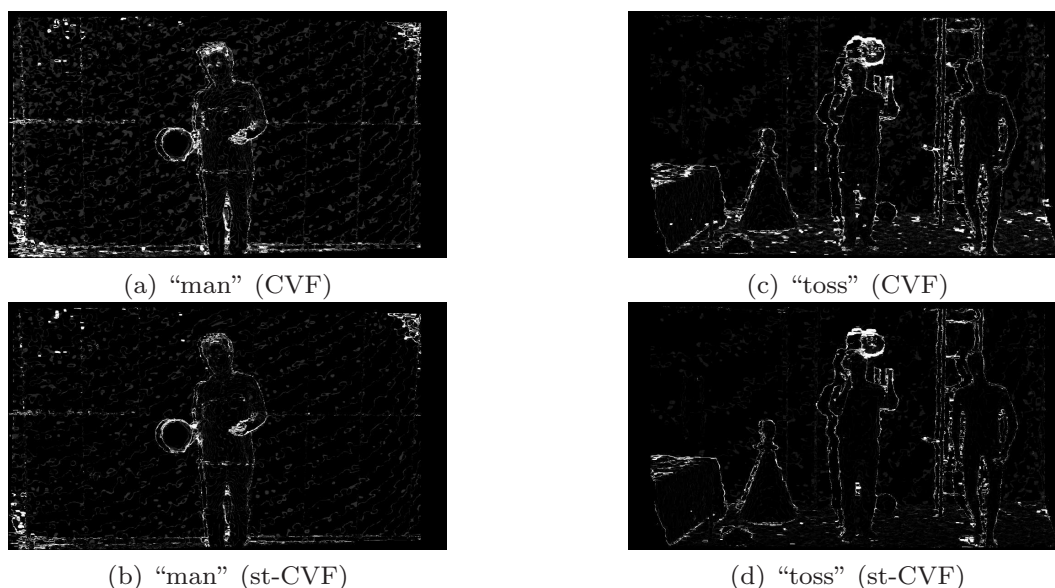


図 6.8 CVF(上段)と st-CVF(下段)で推定したデプスマップの差分画像の比較

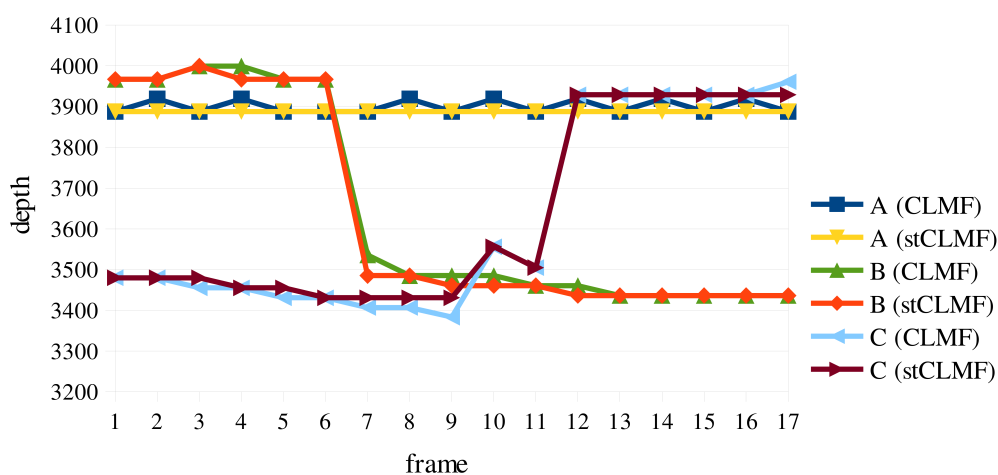


図 6.9 図 3.9 (c) の点 A、B、C について、“man” データセットから、CVF と st-CVF で推定した奥行き値

で推定したものよりも暗く、フレーム間の連続性があることが確認できる。

図 6.7 は、図 6.8 の平均輝度値を比較した定量的な実験結果である。これを見ても、“man” と “toss” のいずれのデータセットにおいても、st-Cost volume filter で推定したデプスマップのフレーム間差分画像の方が、Cost volume filter で求めたデプスマップのフレーム間差分画像よりも、平均輝度値が下がっており、フレーム間の変化量が少ないことがわかる。

図 6.9 は、“man” データセットから、Cost volume filter と st-Cost volume filter で推定したデプスマップのうち、図 3.9(c) 内の点 A、B、C における奥行き値の時間変化を 17 フレーム分、示したものである。点 A は動かない領域にあり、奥行き値は 17 フレームで同じ値をと

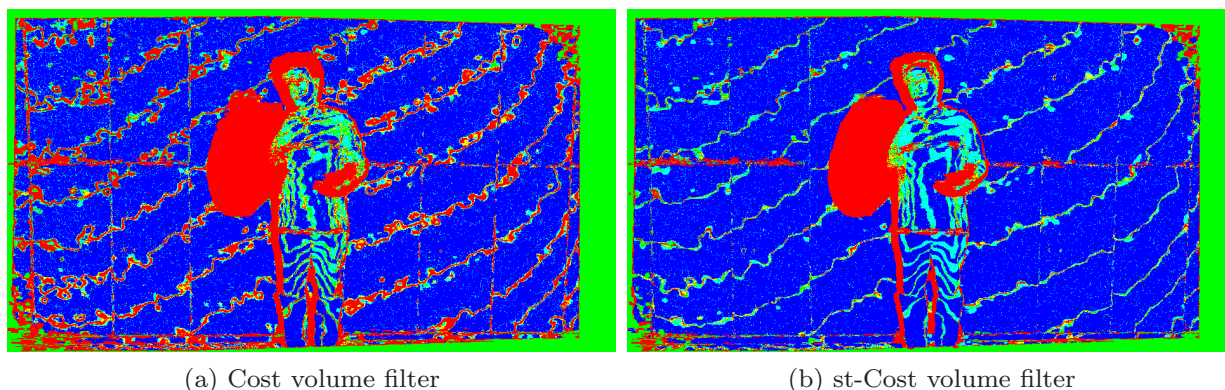


図 6.10 現フレームと次フレーム間の差分の絶対値を 17 フレーム分累積したヒートマップ

るはずである。しかし、Cost volume filter で推定した奥行き値は、フレーム毎に独立に推定しているため、時間方向に多少変化している。この変化は、時系列デプスマップを動画として再生するとフリッカーを発生させる。一方で、st-Cost volume filter で推定したデプスマップは、フレーム間の連続性を考慮しているため、17 フレーム間で同じ値をとっている。これによりフリッカーが抑制されることになる。点 B は、はじめ背景の壁の上であり、その後下がってきた腕の上にある。この場合、奥行き値は、壁の上から腕の上に変化する際、不連続な点を生じる。提案手法の st-Cost volume filter では、時間方向の連続性を考慮に入れつつ、時空間適応型カーネルにより、不連続点を検知できるため、この不連続な奥行きを正しく推定できている。点 C は点 B の逆で、はじめにボールの上であり、その後壁の上に奥行き値が変化する。st-Cost volume filter で推定した奥行き値は、不連続点においては Cost volume filter と同じ推定値をとっており、一方で時間方向に連続領域においては平滑化されており、時系列デプスマップの推定精度の改善を実現できている。

図 6.10 は、(a) Cost volume filter と (b) st-Cost volume filter で推定したデプスマップの現フレームと次フレームとの差分の絶対値を、画素毎に 17 フレーム分累積したものをヒートマップで示したものである。寒色の画素は累積差分が少ない点を示し、暖色の画素は、累積差分が多い点を示す。(a) では、背景の壁などの動きのない領域においても、暖色の点が存在する。一方、(b) では、動きのない領域における暖色の点は少なくなっており、時間方向の推定精度が改善している。これは、Cost volume filter で推定した時系列デプスマップは、フレーム間で独立に処理しているために、視差が次の視差に変わる境界領域で、特に高い頻度で奥行きが変化しており、フリッカーを生じさせていた一方、st-Cost volume filter では、フレーム間の連続性も考慮されているため、視差が次の視差に変わる境界領域においても、推定値の変化は少なく、フリッカーの発生が抑制されていることを示す。

これらのことから、st-Cost volume filter は、時間方向に拡張しつつも、モーションブラーを生じることなく、効果的にフリッカーを抑制できていることが確認された。映像制作への応用にあたっては、時系列デプスマップを統合して時系列 3 次元モデルを生成することから、フ

フリッカーの少ない時系列デプスマップの推定は重要である。

6.6 まとめ

第 5 章では、1 フレーム分の赤外線ステレオ画像とカラーステレオ画像から、Cost volume filter 処理ののちに、2 次元グラフカットをスキャンライン毎に実施することにより、カメラ間距離が離れて、射影変換による 2 次元画像上の変形が生じても、頑強に高精度のデプスマップを推定する手法を提案した。しかし、フレーム毎に独立に処理し、時間方向の連続性が考慮されていなかったため、時系列のデプスマップを推定するとフリッカーを生じてしまう課題があった。本研究の最終目標は、映像制作への応用が可能な時系列 3 次元モデルを生成することであるため、フリッカーの少ない時系列デプスマップを推定することが重要である。

そこで、本章では、Cost volume filter を時間方向に拡張した st-Cost volume filter を提案し、時間方向の連続性を考慮した時系列デプスマップの推定手法を提案した。そのために、はじめに Cost map と Cost volume を時間方向に拡張し、st-Cost map と st-Cost volume を定義した。次に、Cost volume filter で用いる 2 次元エッジ保存型平滑化フィルターである CLMF を時間方向に拡張し、st-CLMF を提案した。そのうえで、st-CLMF を用いて st-Cost volume filter を実施し、時系列デプスマップを推定する手法を提案した。実験結果より、st-Cost volume filter を用いた奥行き推定手法は、Cost volume filter に比べて、推定精度の大きな改善はなかったものの、フレーム間の連続性が考慮されたため、フリッカーの少ない時系列デプスマップを推定できることを確認できた。

第 7 章

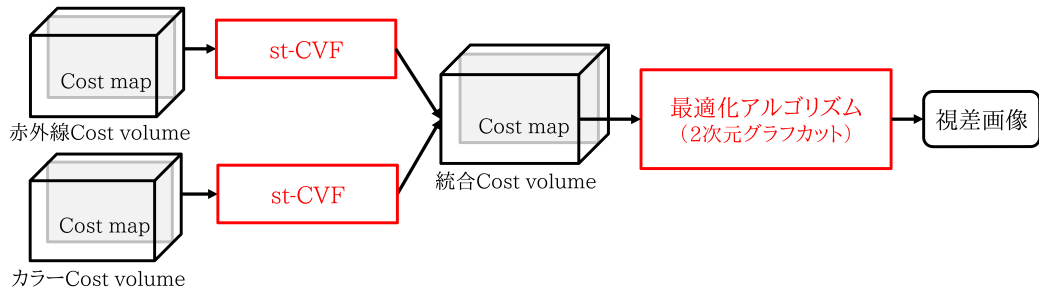
DP Cost volume filter による奥行き推定

7.1 はじめに

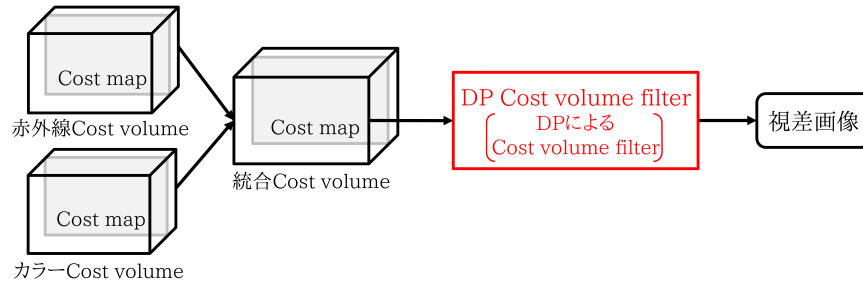
Cost volume filter 処理によるデプスマップ推定手法は、同一視差の近傍画素の Cost のみを共有し、潜在的に被写体表面がカメラに正対していることを前提にしていることから、第 5 章では最適化処理としてグラフカットを導入することにより、異なる視差の近傍画素の Cost も考慮するようにして、ベースラインが長い場合でもカメラに正対していない領域でも奥行きが正対できるような推定手法を提案した。グラフカットを導入することによって、射影変換による左右画像の変形にも対応する奥行き推定の精度改善を図った。しかし、通常の 3 次元グラフカットによる推定処理は、処理時間と使用コスト量が膨大であることから、2 次元グラフカットをスキャンライン毎に実施することで、精度を維持しつつ、処理時間と使用コスト量の削減をはかった。

しかし、2 次元グラフカットを用いても、グラフカット自体が多く処理時間とメモリ使用量を必要とする。そこで、本章ではさらなる処理時間の短縮とメモリ使用量の削減を図るために、図 7.1 のように、Cost volume filter において 2 次元平滑化フィルター処理を行う代わりに、Dynamic Programming(DP)を用いることにより、Cost volume filter 内に最適化アルゴリズムを組み込んだ [10]。

第 5 章ではスキャンライン毎の視差推定において、2 次元グラフカットを用いてデプスマップを推定した。しかし、2 次元の最適化問題としては DP の方がアルゴリズムがシンプルで計算時間が短いため一般的である。デプスマップ推定においては、スキャンライン毎に推定せず、画面全体でデプスマップで推定する場合には、XY の平面方向のほか、視差 d 方向も含めた 3 次元問題を扱う必要がある。DP は 2 次元問題のみしか扱えず、3 次元問題を扱えないことから、デプスマップの推定にはグラフカットが用いられるようになった経緯がある。

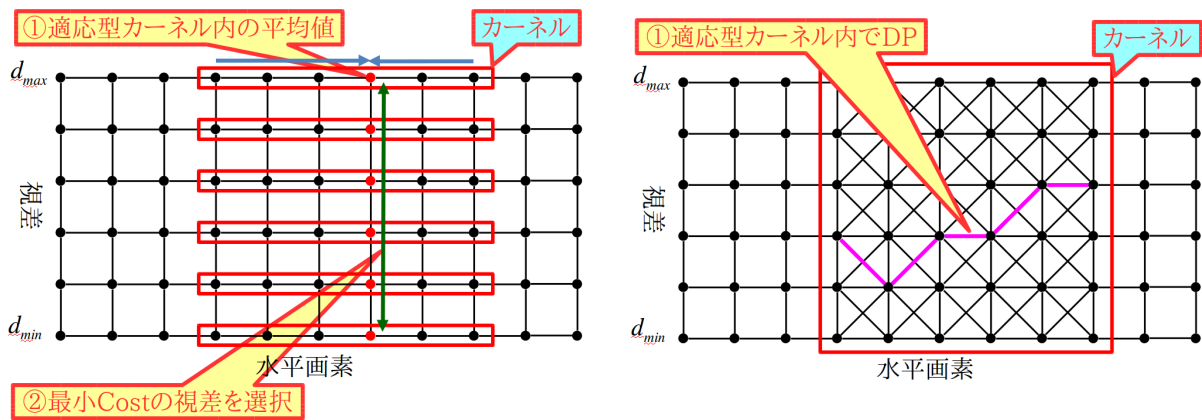


(a) 2次元グラフカットによるデプスマップ推定



(b) DP Cost volume filter によるデプスマップ推定

図 7.1 2次元グラフカットと DP Cost volume filter によるデプスマップ推定処理の比較



(a) Cost volume filter

(b) DP Cost volume filter

図 7.2 Cost volume filter と DP Cost volume filter の処理の比較

そこで、本章では最適化アルゴリズムとして、グラフカットの代わりに DP を用いた。図 7.2 のように、Cost volume filter 処理後に最適化アルゴリズムを実施するのではなく、Cost volume filter 内の 2 次元平滑化フィルターの代わりに DP を用いる DP Cost volume filter を提案し、推定精度を維持しつつ計算時間とメモリ使用量の改善を図った。

DP を用いたデプスマップ推定手法も、以前より長い期間研究されており、さまざまな手法が提案されている [4][24][37][39][40]。しかし、一般的な DP には、2 次元グラフカットと比べて、以下のような課題がある。

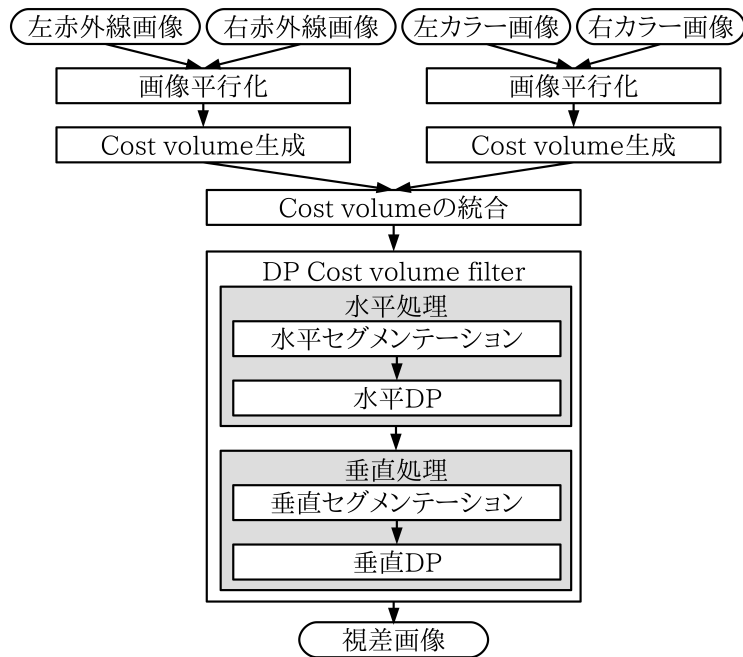


図 7.3 DP Cost volume filter を用いたデプスマップ推定手法の処理過程

1. 始点および終点が決まっている必要がある。
2. 左右の画素の順序が入れ替わることはできない。
3. 始点から終点までは連続でなければならない。

本章では、上記の課題に対応しながら DP をデプスマップ推定に適用する手法を検討した。

提案する DP を用いたデプスマップ推定手法の処理仮定を図 7.3 に示す。赤外線ステレオ画像とカラーステレオ画像は、はじめに画像平行化処理され、それぞれ Cost volume が生成される。これらの Cost volume をカーネルサイズに基づいて選択することにより、1 つの Cost volume に統合し、この Cost volume に対して、水平方向と垂直方向の 2 段階の DP を用いて、視差画像を推定する。この視差画像を左右両方について求めたあと、左右視差チェックで推定エラー画素を排除したのち、カメラパラメータを用いてデプスマップを生成する。本手法では、水平、垂直の各段階において左カラー画像を用いてセグメンテーション処理を行い、各セグメント内は連続した平面もしくは曲面と仮定することで、DP の 2 番目と 3 番目の課題を克服する。以降、本章ではセグメンテーションの方法と、DP の基本的な処理について述べたあと、水平方向／垂直方向の DP の実行方法について述べる。その後、実験結果を示して本手法の有効性を述べたあと、まとめる。

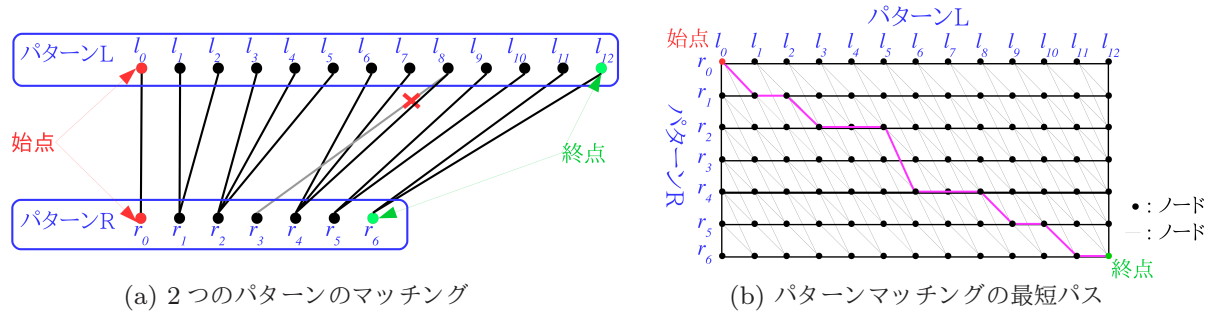


図 7.4 DP の原理

7.2 Dynamic Programming

Dynamic Programming (DP) は、2 つの 1 次元パターン間の最適な対応関係とパターン間の距離 (DP 距離) を求めることができる [56]。例えば、図 7.4(a) のような、パターン $L(L = l_1, l_2 \dots l_m)$ とパターン $R(R = r_1, r_2 \dots r_n)$ があつたとする。このとき、(b) のように、横軸にパターン L 、縦軸にパターン R を取り、各ノードに Cost を割りつけ、始点から終点までの Cost が最小となる最短パスを見つけると、そのパスが通る点が 2 つのパターンの最適マッチングとなるのが、DP である。DP は、伸長した 2 つの 1 次元パターンのマッチングが可能なることから、音声認識や文字認識において広く用いられてきた。ステレオマッチングに適用する際には、ステレオ画像上でスキャンライン毎に、2 つの輝度分布の 1 次元パターンのマッチング処理を行うことで推定することができる。ライン v のマッチングは、Cost $C(u, v, d)$ を用いて、式 7.4 のように定義されたエネルギー E を最小化するパスを求めることが推定できる。

$$E = \sum_{i=1}^m C(u_i, v, d_i + \delta), \quad -1 \leq \delta \leq 1 \quad (7.1)$$

このエネルギー E の最小値は、展開して整理すると、

$$\begin{aligned} \min E &= \min_{\substack{u_1 \dots u_m \\ -1 \leq \delta \leq 1}} \sum_{i=1}^m C(u_i, v, d_i + \delta) \\ &= \min_{\substack{u_2 \dots u_m \\ -1 \leq \delta \leq 1}} \left[\sum_{i=3}^m C(u_i, v, d_i + \delta) + \left\{ C(u_2, v, d_2 + \delta) + \min_{\substack{u_1 \\ -1 \leq \delta \leq 1}} C(u_1, v, d_1 + \delta) \right\} \right] \end{aligned} \quad (7.2)$$

ここで、右辺の $\{ \}$ の中を

$$\mu_i(u_i) = C(u_i, v, d_i + \delta) + \min_{\substack{u_{i-1} \\ -1 \leq \delta \leq 1}} C(u_{i-1}, v, d_{i-1} + \delta) \quad (7.3)$$

とおき、さらに整理すると、

$$\begin{aligned} \min E &= \min_{\substack{u_2 \dots u_m \\ -1 \leq \delta \leq 1}} \left[\sum_{i=3}^m C(u_i, v, d_i + \delta) + \mu_2(u_2) \right] \\ &= \min_{u_m} \mu_m(u_m) \end{aligned} \quad (7.4)$$

となり、このときの $\min E$ が DP 距離となる。したがって、スキャンライン上の各画素について、このメッセージ μ_i を求め、終点において μ_m が最小となるパスを終点から始点まで逆方向にトレースすることで、最短パスを見つけることができる。

しかし、実際のステレオマッチングにおいては、パターンの始点がわからない。そこで、提案手法では、各スキャンライン内で 1 次元のセグメンテーション処理 (水平セグメンテーション) をして、分割した領域の左端にある画素のすべての視差 $[d_{min}, d_{max}]$ が始点と考え、それぞれのメッセージ μ を計算する。また、終点についても不明であることから、右端の画素に辿り着いたすべての視差のメッセージ μ の中で最も DP 距離が小さい視差 d を終点と考え、そのメッセージ μ を終点から始点にトレースすることで最短パスを求める。

7.1 節で第 2 番目の課題として述べた通り、DP は、図 7.4(b) の左上から右下への最短パスを求めるアルゴリズムであることから、パターン L と R の順序は同じで入れ替えられない。そこで、本手法は、水平セグメンテーションで分割された各領域内は連続した平線もしくは曲線であるという前提をおく。

水平方向の DP を実行したあと、4.4.3 節の CLMF で用いる適応型カーネルに類似したアプローチ (Cross-based DP) で垂直方向についてもセグメンテーション処理を行って DP を用いて、水平と垂直の両方向で最短なパスを求めることで、2 次元のパターンマッチングで行う。

7.3 セグメンテーション

7.3.1 水平セグメンテーション

水平方向のセグメンテーションは、スキャンライン v 毎に、左端の画素から走査していき、現在の画素 p_j の画素値 $I(p_j)$ と一つ前の画素 p_{j-1} の画素値 $I(p_{j-1})$ との差が閾値 T_c 以上のとき、前のセグメントの端と判断し、現画素から次のセグメントとして記録する。この処理を右端の画素まで行うことでセグメンテーション処理を実施する。すなわち、

$$\begin{cases} \zeta_{p_j} = \zeta_{p_{j-1}}, & (|I(p_j) - I(p_{j-1})| \leq T_c) \\ \zeta_{p_j} \neq \zeta_{p_{j-1}}, & (|I(p_j) - I(p_{j-1})| > T_c) \end{cases} \quad (7.5)$$

ここで、 ζ_p は画素 p が属する水平セグメントの ID を示す。また、図 7.5 のように、各セグメントの左端の画素 p_l と右端の画素 p_r は、式 7.6 で定義する。

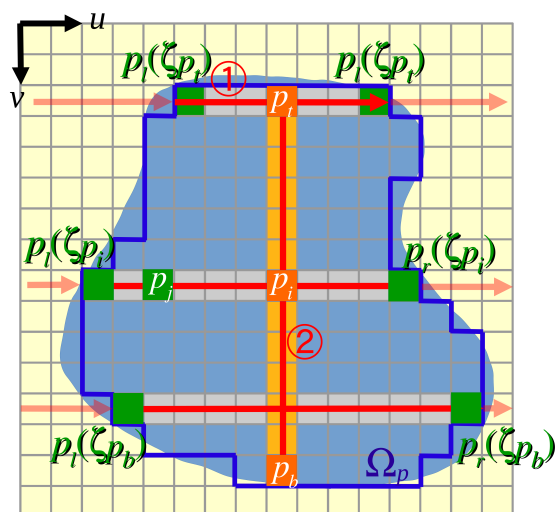


図 7.5 Cross-based セグメンテーション

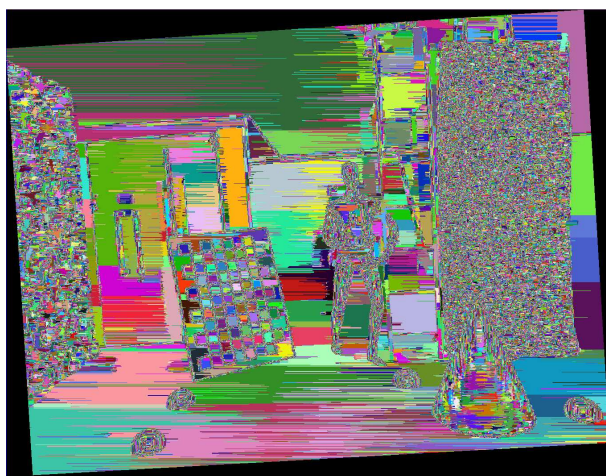
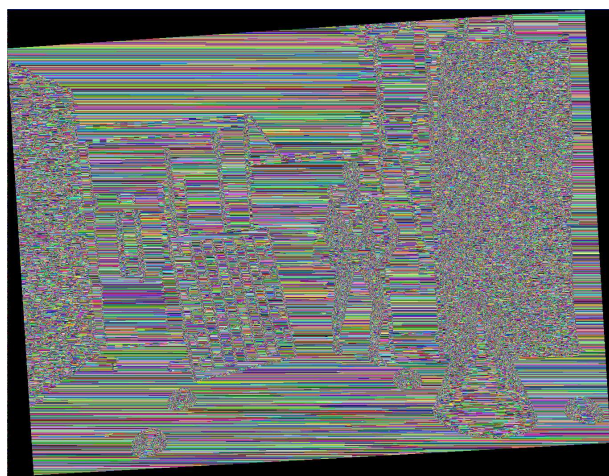


図 7.6 水平セグメンテーション後の各セグメント 図 7.7 垂直セグメンテーション後の各セグメント

$$\begin{aligned}
 p_l(\zeta) &= (u_l(\zeta), v_l(\zeta)) \\
 p_r(\zeta) &= (u_r(\zeta), v_r(\zeta))
 \end{aligned}
 \tag{7.6}$$

図 7.7 は、”mannequin460” データセットの左画像に対して、水平セグメンテーションを実施したあとの各水平セグメントをランダムな色で表示したものである。このように、水平セグメンテーションを実施することにより、スキャンライン毎の水平セグメントを抽出することができる。

7.3.2 垂直セグメンテーション

水平セグメンテーション処理ののち、垂直セグメンテーションを実施する。垂直セグメンテーションは、画像の左上から右下に向けて順次スキャンしていき、垂直セグメンテーションが完了していない画素 p_j を見つけたら、 p_j が属する水平セグメント ζ_{p_j} の中心画素 $p_t = (u_t, v_t)$ を取得する。そして、適応型カーネルに類似した形で、図 7.5 のように、中心画素 p_t から下方向にアームを下方向に伸ばし、現画素の画素値と次の画素の画素値との差が閾値 T_c 以上になった時点でアームを停止し、下端の画素 p_b の位置 (u_b, v_b) を取得する。この垂直アーム上の各画素 p_i の属する水平セグメント ζ_{p_i} 内の画素を統合することにより、2次元のセグメントを取得することができる。すなわち、セグメント Ω は式 7.7 となる。

$$\Omega = \bigcup_{p_t \leq p_i \leq p_b} \left(\bigcup_{p_l(\zeta_{p_i}) \leq p_j \leq p_r(\zeta_{p_i})} p_j \right) \quad (7.7)$$

垂直セグメンテーションの結果、取得した 2 次元セグメントを図 7.7 に示す。各セグメントは、ランダムに選択された色で表示されている。

7.4 Cross-based DP

DP は、2 次元上の 2 点の最短ルートを見つけるアルゴリズムであるため、水平、垂直、奥行き の 3 次元の最適化を一度に行うことはできない。そこで、本手法は水平方向と垂直方向の 2 段階の DP を実施する Cross-based DP を提案し、デプスマップを推定する。

7.4.1 水平方向の DP

水平方向の DP は、さらに 2 段階の DP で構成される。はじめに、図 7.8 のように水平セグメントの左端の画素から、その水平セグメントが属する 2 次元セグメントの垂直アームまでの DP を実施し、次に右端の画素から垂直アームまでの DP を実施する。左端から垂直アームまでの水平方向の DP は、図 7.9 のような、ノードとエッジで構成されるグラフ構造を用いて処理を行う。各ノードには、統合 Cost volume の Cost C^i を用いて、式 7.8 で定義される $C_{\zeta,h}(j, d)$ が付与される。

$$C_{\zeta,h}(j, d) = C^i(u - p_l(\zeta), v, d) \quad (7.8)$$

DP を実施にあたっては、グラフの左端を各水平セグメントの左端、グラフの右端は各水平セグメント上で垂直アームと交差する点とする。そして、グラフの左端から右端までの最短ルートを DP を用いて推定する。DP が最小化するエネルギー E は、式 7.9 で定義される。

$$E = \sum_{0 \leq j \leq N} \{ C_{\zeta,h}(j, d) + \min_{-1 \leq \delta \leq 1} (\mu_h(j - 1, d + \delta)) \} \quad (7.9)$$

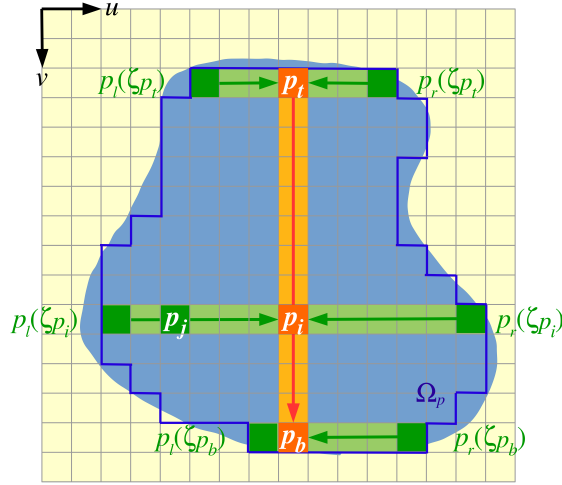


図 7.8 Cross-based DP

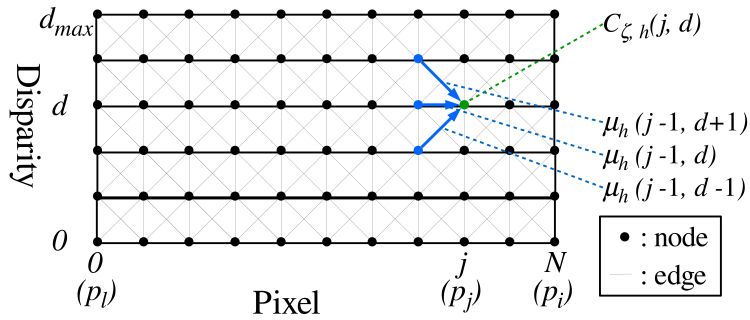


図 7.9 水平方向の DP

ここで、メッセージ μ_h は、式 7.10 で定義される。

$$\mu_h(j, d) = \min\{\mu_h(j-1, \delta), d-1 \leq \delta \leq d+1\} + C_{\zeta, h}(j, d) \quad (7.10)$$

本手法では、各水平セグメント内は連続した平線もしくは曲線であることを前提としていることから、 $\{\mu_h(j-1, d-1), \mu_h(j-1, d), \mu_h(j-1, d+1)\}$ の 3 つのメッセージのみが次の処理に影響を与えると仮定する。そのため、不連続点を検知する一般的な DP を用いたデプスマップ推定手法のように、一つ前のすべてのメッセージを参照する必要がないことから、処理時間の短縮をはかることができる。Cross-based DP の最終段階において、最小パスを逆にトレースできるように、右端までメッセージを計算した結果、メッセージ μ_h が最小となる一つ前のノードを $\theta_{\zeta, h}(j, d)$ に格納する。すなわち、

$$\theta_{\zeta, h}(j, d) = \arg \min_{d-1 \leq \delta \leq d+1} \{\mu_h(j-1, \delta)\} \quad (7.11)$$

DP が右端まで計算されたのち、右端の画素 p_i の最小メッセージ $\mu_h(N, d)$ は、DP の Cost

に用いるため、左側の水平メッセージ μ_l として格納される。すなわち、

$$\mu_l(v - p_t, d) = \mu_h(N, d) \quad (7.12)$$

垂直アームの左側の水平方向の DP の処理が終わると、次に右端の画素 p_r から、垂直アームの位置にある画素 p_i までを実行する 2 つ目の水平方向の DP が実施される。この処理は、1 つ目の DP と同様に処理され、垂直アームの位置にある画素 p_i における、最終的な N 番目のメッセージ $\mu_h(N, d)$ を、垂直方向の DP の Cost として用いるために、 μ_r として格納する。すなわち、

$$\mu_r(v - p_t, d) = \mu_h(N, d) \quad (7.13)$$

7.4.2 垂直方向の DP

垂直方向の DP は、水平方向の DP と同様の処理となるが、垂直アームの上端の画素から下端の画素までの 1 回だけ実施される。各エッジに付与する Cost C_v は、左側の水平方向の DP で取得した DP 距離 μ_l と、右側の水平方向の DP で取得した DP 距離 μ_r の足しあわせで定義する。すなわち、

$$C_v(i, d) = \mu_l(i, d) + \mu_r(i, d) \quad (7.14)$$

この Cost C_v を用いて垂直方向の DP を実施するが、水平方向の DP のときと同様に、あとでトレースバックできるように、このときにメッセージ μ_v が最小となる一つ前のノードは、 $\theta_v(i, d)$ に格納する。上端の画素から下端の画素まですべてのメッセージ μ_v が垂直方向の DP で計算されたのち、終点は、下端の画素 p_b において、メッセージ μ_v が最小となる視差 d_{end} を選択して決定する。つまり、 d_{end} は式 7.17 で定義される。

$$d_{end} = \arg \min_{0 \leq d \leq d_{max}} (\mu_v(N, d)) \quad (7.15)$$

7.5 DP Cost volume filter による奥行き推定

出力する視差画像の各画素 p_i の視差 d は、垂直方向の DP の結果、下端の画素でメッセージ μ_v が最小となる視差 d_{end} から、垂直方向にトレースバックすることで求めることができる。すなわち、

$$D(p_i) = \theta_v(p_{i+1} - p_t, D(p_{i+1})), \quad D(p_b) = d_{end} \quad (7.16)$$

これにより、垂直アーム上の各画素の視差 D を取得することができる。次に、垂直アーム上の各画素の視差 D から、左右水平方向への視差 D を求める。

水平方向の各 DP の終点における視差 $D(p_i)$ は、垂直方向の DP の結果、トレースバックして得られた視差 D とする。このようにして、左側の水平方向の各画素 p_j における視差 D

は、式 7.8 で取得することができる。

$$D(p_j) = \theta(p_{j+1} - p_l, D(p_{j+1})) \quad (7.17)$$

この処理を、垂直アームの右側についても同様に行うことで、セグメント内のあるスキャンライン v における視差マップを取得することができる。この水平方向のトレースバックを、垂直アーム上にあるすべての画素から、左右両サイドに実施することで、セグメント内全体の視差マップを取得することができる。さらに、この一連の水平と垂直方向の DP による奥行き推定処理を、すべてのセグメントについて処理すると視差画像を取得することができる。

7.6 実験結果

提案手法の有効性を示すため、実空間を赤外線カラー一体型カメラアレーで撮影した 3.4 節の”mannequin75”と”mannequin460”、”desk” データセットからデプスマップを推定した。また、この手法は一般的なカラーステレオ画像からのデプスマップ推定にも応用できることから、その有効性についても確認するため、Middlebury データセットのうち、”cones”と”motorcycle”を用いてデプスマップを推定した。

7.6.1 赤外線カラーステレオ画像からのデプスマップ推定結果

図 7.10 は、3つのデータセット”mannequin75”と”mannequin460”、”desk”について、Cost volume filter(CVF)、Cost volume filter と 2次元グラフカット (CVF+2D GC)、DP Cost volume filter(DP CVF)、Cost volume filter と DP Cost volume filter(CVF+DP CVF) で推定したデプスマップを比較したものである。Cost volume filter には、CLMF を 2次元平滑化フィルターとして用いた。上段は、ベースラインが短い”mannequin75” データセットを各手法で推定したデプスマップである。ベースラインが短いとき、手法の違いで大きな違いは表れなく、Cost volume filter でも、十分精度の高いデプスマップを取得することができた。

中段は、ベースラインが長い”mannequin460” データセットを各手法で推定したデプスマップを比較したものである。ベースラインが長いと、Cost volume filter 処理を用いて推定したデプスマップには、上段のベースラインが短いときに比べて、推定エラー画素が増えている。推定エラー画素は主に床など、表面がカメラに正対していない領域で発生しており、前述の通り、射影変換による左右画像の変形による精度の低下と思われる。2次元グラフカットを用いることにより、これらのエラーは軽減されており、床を含めて安定してデプスマップを推定できている。DP Cost volume filter を用いて推定したデプスマップを見ると、2次元グラフカットで推定したデプスマップに比べて、床を含めて推定エラー画素が現象しているが、ストリーキング現象が発生してしまっている。これは、DP Cost volume filter はスキャンライン間の垂直方向の連続性も考慮されているが、ノイズはスキャンライン毎に独立に発生している

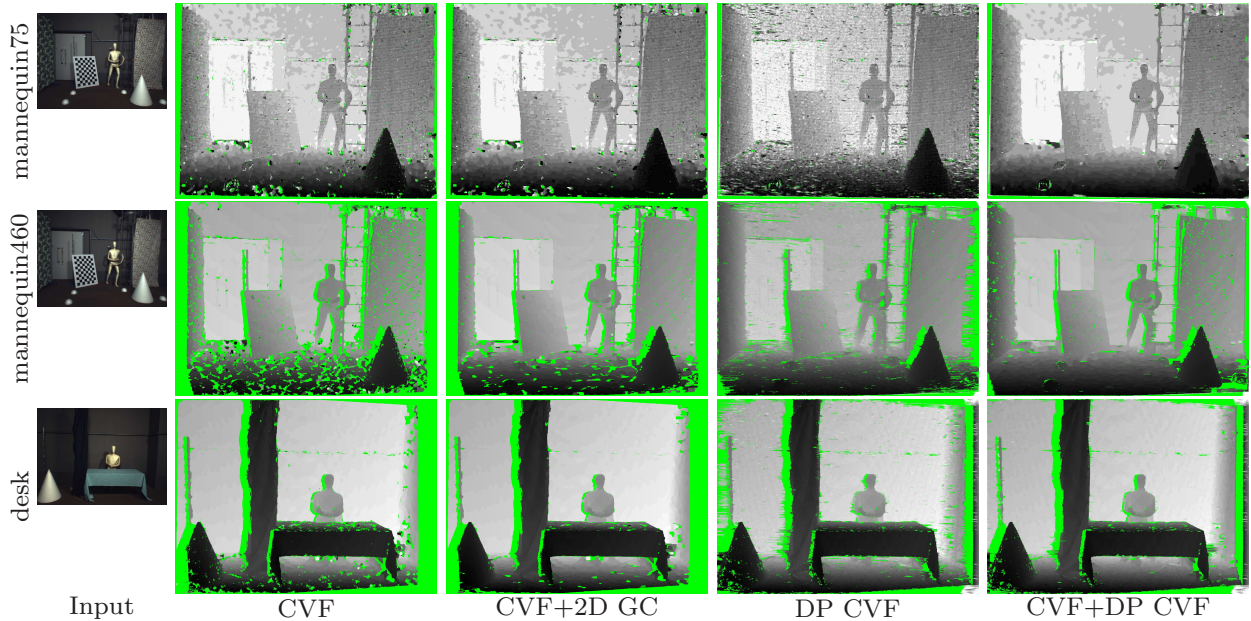
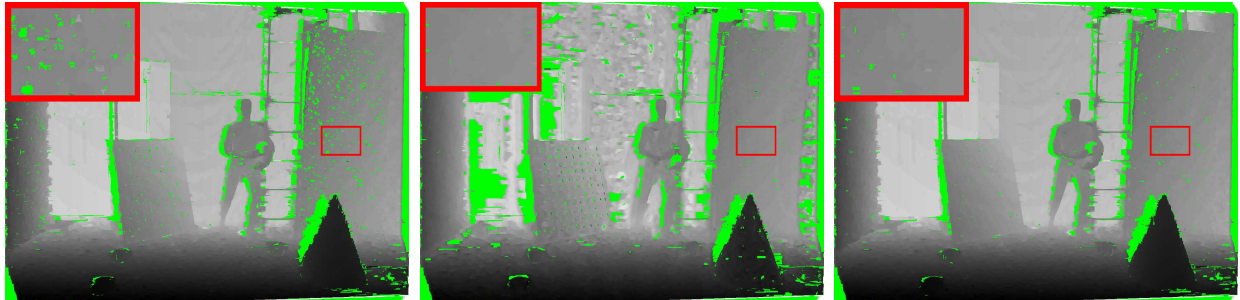


図 7.10 異なる手法で推定したデプスマップの比較

ため、推定結果がスキャンライン毎に異なってしまったためかと思われる。DP Cost volume filter の前処理として、Cost volume filter 処理を入れて推定したデプスマップを見ると、ストリーキング現象は発生していません。Cost volume filter と 2 次元グラフカットを用いて推定したデプスマップと同程度の高精度のデプスマップを推定できている。これは、Cost volume filter 処理により、ノイズの影響が除去されたため、ストリーキングの発生を抑制できたものと考えられる。

下段は、“desk” データセットを各手法で推定したデプスマップを比較したものであるが、“mannequin460” データセットを各手法で推定したデプスマップと同様な結果が得られている。すなわち、Cost volume filter 処理のみで推定したデプスマップでは、ベースラインの長いことに起因した射影変換の影響で推定エラー画素が多く発生している。また、Cost volume filter と 2 次元グラフカットを併用すると、これらの推定エラー画素は軽減する。DP Cost volume filter による推定では、Cost volume filter と 2 次元グラフカットを併用した推定結果に比べて、推定エラー画素は現象しているが、中段と同様にストリーキング現象が発生している。一方で、Cost volume filter と DP Cost volume filter を併用したデプスマップ推定手法では、推定エラー画素が少なく、ストリーキングのない高精度なデプスマップが推定できている。

図 7.11 は、入力画像の種類の違いによるデプスマップの推定結果を比較したものである。左上の画像は、右側の赤枠部分の拡大図である。(a) は、赤外線ステレオ画像のみから、提案手法を用いて推定したデプスマップである。全体的に安定して奥行きを推定できているが、テクスチャーの多い領域を中心に推定エラー画素が多い。これは、テクスチャーが多い領域は、



(a) 赤外線ステレオ画像のみ (b) カラーステレオ画像のみ (c) 赤外線とカラーのステレオ画像

図 7.11 入力画像の種類の違いによるデプスマップの推定結果の比較

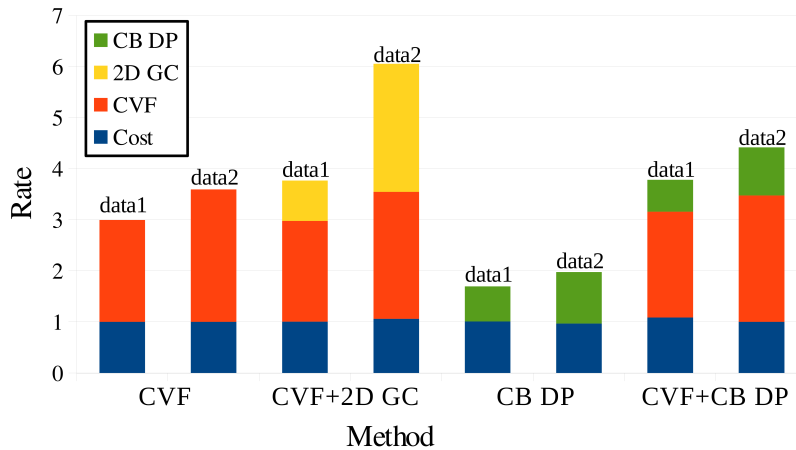


図 7.12 “mannequin75”(data 1) と “mannequin460”(data 2) の処理時間

適応型カーネルとセグメントが小さくなるために、ドットが十分含まれないために発生してしまっていると考えられる。(b) は、カラーステレオ画像のみから、提案手法を用いて推定したデプスマップである。テクスチャーの多い右側の領域では、安定して奥行きを推定できているが、テクスチャーの少ない白いドアの付近などで、推定に大きく失敗している。また、グレーの壁は一部奥行きを推定できているが、平面であるはずであるのに、デプスマップ上でスポット暗くなっている場所があり、正しく推定できていないことがわかる。(c) は、赤外線ステレオ画像とカラーステレオ画像の両方を入力して推定したデプスマップである。全体的に安定して奥行きを推定できている上、テクスチャーの多い右側の領域においても、推定エラー画像が軽減されていることがわかる。(a) の拡大図と比較しても、(c) の拡大図上の推定エラー画像は大幅に軽減されている。

図 7.12 は、“mannequin75”(data 1) と “mannequin460”(data 2) を各手法で処理したときの処理時間を比較したものである。各処理の内訳は、色で分けられている。処理時間は、画像サイズや推定する範囲の視差数に依存することから、各処理の処理時間は、Cost volume を生成する時間で正規化した。DP Cost volume filter の処理時間は、Cost volume filter の

処理時間より短いですが、赤外線画像のように SN 比が低い画像では、精度が低下してしまう。“mannequin75” データセットの場合、Cost volume filter と DP Cost volume filter を組み合わせた処理時間は、Cost volume filter と 2次元グラフカットと組み合わせた処理時間とほとんど変わらないが、“mannequin460” データセットの場合、全処理時間の 30% 削減することができた。これより、DP Cost volume filter を用いた奥行き推定手法は、推定する視差数が多いほど、効果が大きくなることがわかった。映像制作へ応用する場合には、奥行き方向の解像度も高い必要があることから、DP Cost volume filter は映像制作への応用時、特に効果が大きくなる。DP Cost volume filter 自体の処理だけに注目すると、2次元グラフカットと比べて、60% の処理時間の削減ができています。したがって、本章での提案手法は、特にベースラインが長く、視差数が多い場合、または画素数が多い場合、2次元グラフカットを用いた奥行き推定手法に比べて、推定エラー画素数を減らすことができたとともに、少ない処理時間で効率的にデプスマップを推定できることを確認することができた。

処理に必要なメモリ使用量については、Cost volume filter と 2次元グラフカットを組み合わせた奥行き推定手法と、Cost volume filter と DP Cost volume filter を組み合わせた奥行き手法とでは、Cost volume filter が必要とするメモリ使用量が大半を占めるため、ほとんど変わらなかったが、DP Cost volume filter のみを用いた奥行き推定手法では、Cost volume filter のメモリ使用量の半分で処理できることが確認できた。

7.6.2 カラーステレオ画像からのデプスマップ推定結果

本章の提案手法は、赤外線カラーステレオ画像からデプスマップを推定するために検討されたものである。そのためには、赤外線画像上のドットパターンがテクスチャーとして用いるには粗いため、大きなカーネルを用いる必要があり、大きなカーネルはベースラインが長いとき、射影変換の影響を大きく受け、左右の画像が大きく変形するのに対応するために検討された。しかし、この手法は、一般的なカラーステレオ画像からの奥行き推定にも利用することができ、その場合にもテクスチャーが少ない領域や、SN 比が低い暗い領域の推定精度の改善にも効果的である。図 7.13 は、Middlebury ステレオデータセット [31] のうち、“cones” と “motorcycle” から推定した視差画像を比較したものである。“cones” は、9枚の異なる位置から撮影した画像で構成されており、ステレオマッチングの実験では、一般的に 2番目と 6番目の 2枚の画像から視差画像を推定するが、我々はよりベースラインが長いステレオ画像からの視差画像の推定精度を確認するため、1番目と 9番目の画像から視差画像を推定した。

“cones” については、Cost volume filter で視差画像を推定した場合、図 7.13(c) 上の赤丸で示した 2つのコーンの推定に失敗している。2次元グラフカットや DP Cost volume filter を用いた奥行き推定手法では、近傍画素の異なる視差の Cost が考慮されるため、これらの領域の推定にも成功している。

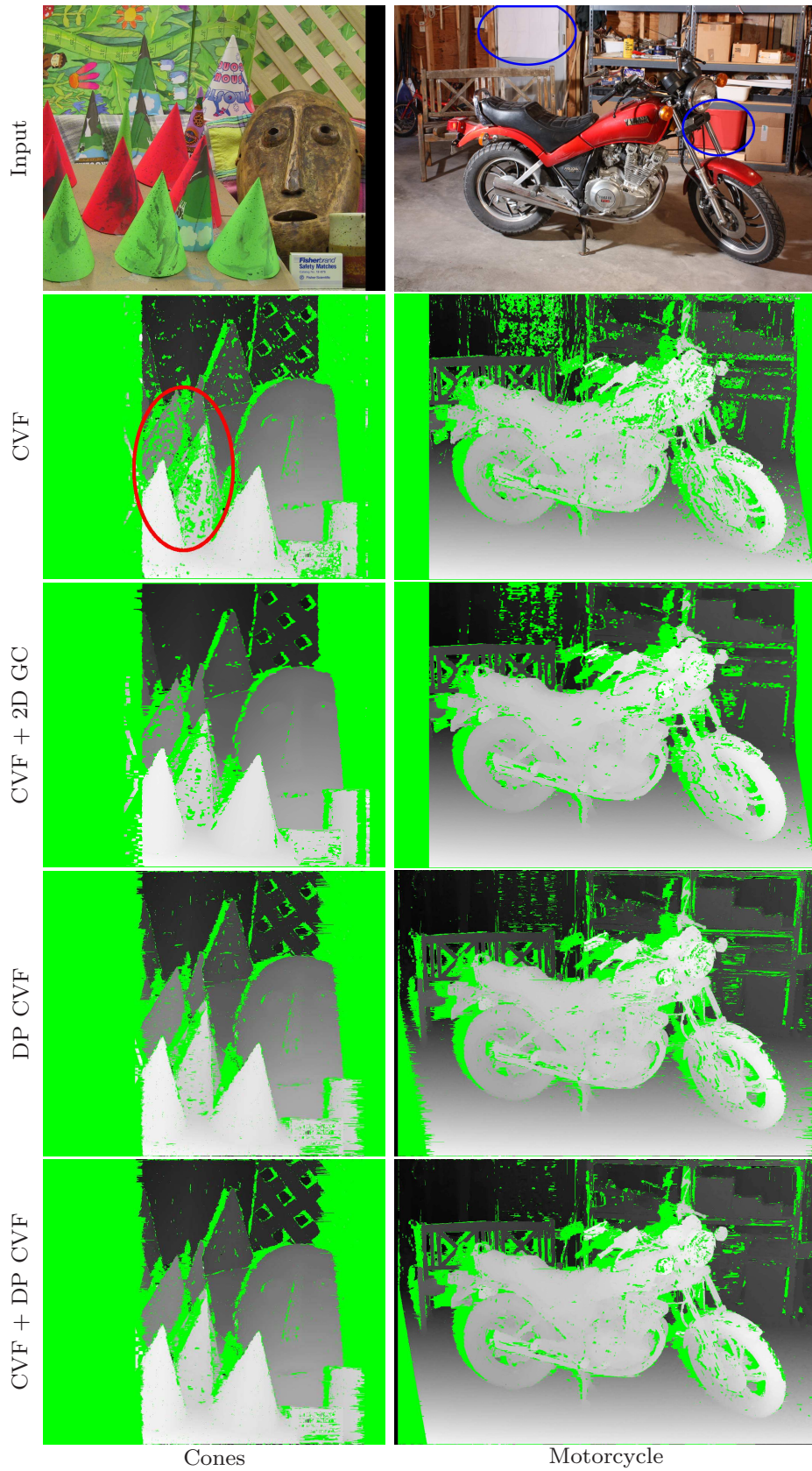


図 7.13 Middlebury ステレオデータセットから推定した視差画像

“motorcycle”については、DP Cost volume filter と、Cost volume filter 後に DP Cost volume filter を実施した方法では、テクスチャーの少ない白いポスターや、入力画像上で緑色の丸枠で示した赤い箱などの領域で、他の手法より推定率が改善した。推定率は、DP Cost volume filter は適応型カーネルを用いているため、境界付近の不連続性を適切に扱うことができ、近傍の画素の異なる視差の Cost を参照するために改善したと思われる。Middlebury のステレオデータセットは、提案手法にとって比較的狭いベースラインで撮影されており、提案手法の効果はさらにベースラインが長いステレオマッチングのときに効果はさらに大きくなると考えられる。もし、対象の被写体がステレオマッチングを行うのに最低限のテクスチャーがある場合、提案手法は赤外線ドットパターンを照射せずに、通常のカラーステレオ画像から効果的にデプスマップを推定することができる。

デプスマップが、赤外線カラーステレオカメラで撮影した画像から推定されるとき、赤外線画像は、SN 比が低いため、DP Cost volume filter を実施する前に、Cost volume filter を行うとさらに効果的に推定することができる。カラーステレオ画像の場合、カラー画像は SN 比が高いため、DP Cost volume filter は Cost volume filter なしでデプスマップを十分な精度でデプスマップを推定することができる。これにより、DP Cost volume filter は、Cost volume filter よりも短い処理時間で、より高い精度のデプスマップを推定することができることがわかった。

7.7 まとめ

本章では、1 台の赤外線プロジェクターと 2 台の赤外線カラーステレオカメラから、DP Cost volume filter を用いてデプスマップを推定する手法を提案した。提案手法は、映像制作への応用を考慮したデプスマップ推定手法であるため、ステレオカメラのベースラインは、奥行き解像度を高くするために、長くして撮影したステレオ画像からデプスマップを推定できるよう検討した。しかし、ベースラインが長くなると、射影変換の影響で、左右の画像が大きく変形するために、Winner-takes-all 的な視差選択の代わりに、最適化アルゴリズムを導入する必要があった。最適化アルゴリズムは、計算時間が長くなるほか、多くのメモリを使用する必要があった。そこで、本章では、2 次元グラフカットの代わりに、2 つの水平 DP と 1 つの垂直 DP を組み合わせた Cross-based DP による DP Cost volume filter を提案した。

この手法は、入力画像上で類似色の領域に分割し、領域内は局所的に連続な曲面もしくは平面であると想定して推定するものである。水平 DP と垂直 DP という 2 段階の 2 次元最適化アルゴリズムを用いることにより、2 次元グラフカットよりも、短い処理時間で同等の画質を推定することができた。また、提案手法は、一般的なカラーステレオ画像からのデプスマップ推定においても、特にテクスチャーの乏しい領域に効果的であることを示した。本提案手法は、Cost volume filter に DP を組み込んだ推定手法ということができる。この処理は、並列

処理にも適していることから、GPU(Graphics Processing Unit) 処理を実装することにより、処理時間をより高くすることができると思われる。

第 8 章

映像制作への応用

8.1 はじめに

本研究の最終的な目標は、映像制作に応用できる時系列 3 次元モデルを生成することである。時系列の 3 次元モデルを生成するには、ある撮影点からのデプスマップを推定できれば、複数撮影点からのデプスマップを推定して取得した複数のデプスマップを 1 つに統合することで、3 次元モデルを生成することができる。

映像制作に応用するには、テクスチャーのない領域も含めて安定してデプスマップを推定できる必要があるため、第 4 章から第 7 章では、赤外線ドットマップを照射して、赤外線ステレオ画像とカラーステレオ画像からデプスマップを推定する手法を提案した。また、映像制作に応用するためには、水平方向と垂直方向だけでなく、奥行き方向にも空間解像度の高いデプスマップを推定する必要がある。そのため、奥行き解像度を高くするために、カメラ間のベースラインを長くしても安定してデプスマップを推定できるように、2 次元グラフカットや DP などの最適化手法を導入してデプスマップを推定する手法を提案した。また、時系列 3 次元モデルを生成するにあたっては、フリッカーの少ないデプスマップを推定する必要があることから、Cost volume 時間方向に拡張して、フリッカーを抑制した時系列デプスマップを推定する手法を提案した。

本章では、これまで提案したこれらの手法を用いて推定したデプスマップを統合して、3 次元モデルを生成する手法について述べる。また、生成した 3 次元モデルを、2 次元映像制作および 3 次元映像制作へ応用する例について紹介する。

実空間のシーンの 3 次元モデルを生成する技術は、任意始点映像の生成 [12] や、特別なメガネを必要としない 3 次元テレビ [2] において、要素技術の一つとなる。これらへの応用にあたっては、ある 1 始点からの奥行き値を 2 次元に配列した 2.5 次元のデプスマップではなく、視点が移動しても、3 次元情報がない欠落領域が発生しない 3 次元モデルを生成する必要がある。この 3 次元モデルの生成は、応用分野が広いことから、コンピュータービジョンにおいて

早期より長期に渡って研究されており、多くの手法が提案されている。その多くの手法は、複数のカラーカメラで撮影した多視点画像からシーンの 3 次元モデルを生成する。これらの手法は、大きく 2 つに分類することができる。

1 つ目は、ステレオマッチングのアプローチを用いる手法である。このアプローチは、本研究でも採用しているもので、複数の位置から撮影した多視点画像から、2 枚の画像を選択して、それらの画像からデプスマップを推定する。そして、異なる複数の画像の組から、複数のデプスマップを推定し、それらを一つの 3 次元モデルに生成するアプローチである。

2 つ目は、多視点を用いた Volumetric なアプローチによる手法である。このアプローチでは、3 次元モデルを生成する領域を”ボクセル”と呼ばれる小さな 3 次元空間に分割し、各ボクセルについて、そのボクセルが被写体の表面である可能性を、3 つ以上の多視点画像をボクセルに投影して、その類似度から 3 次元モデルを推定する。

近年提案されている手法では、テクスチャが十分にあれば正確な 3 次元モデルを生成することができる。しかし、このアプローチでも、被写体表面に十分なテクスチャがないと 3 次元モデルの精度が低下してしまう課題がある。前述のように、本研究で提案している手法は、ステレオマッチングのアプローチを用いているが、赤外線カラーカメラアレーを用いる。

本章では、複数のデプスマップを統合する手法と、穴埋めをして、統合したモデルの Refinement を行う手法を提案する。本章で提案する手法は、はじめに複数のデプスマップを単純に重ね合わせて 3 次元モデルを生成したあと、そのモデルを各カメラに投影したときの投影画像は、同じであるはずであるという前提をおき、3 次元モデルの精度を改善するものである。本手法では、各デプスマップから 3 次元空間に変換した各点の色と、その点を投影したカラー画像上の投影点の色との類似度が最も高い変換点を、被写体表面と仮定し、その点とカメラの光学主点を結ぶ直線上で、手前にある点を削除することで、モデル精度の改善を図る。このようにして求めた表面の裏にある、残りの点については、映像制作に大きな悪影響を与えないことから、そのまま残す。

このようにして求めた 3 次元モデルは、撮影点から見えないオクルージョン領域や、各デプスマップ推定における推定エラー画素などの理由から 3 次元情報のない欠落領域が生じる。本手法では、このような欠落領域を埋める欠落領域補完処理も提案する。この処理は、色が類似した点は、類似した奥行き値であると仮定して、3 次元モデルを各撮影点に投影したデプスマップ上で、CLMF を用いた平滑化処理により行う。最終的な 3 次元モデルは、このようにして求めたデプスマップを統合して生成する。

8.2 従来手法

レーザーを用いたレンジセンサーを用いて、静止物のデプスマップを高精度に計測する技術は以前より確率していることから、静止物の複数のデプスマップを統合して 3 次元モデルを

生成する手法は、長期に渡って研究されている。Turk と Levoy[36] は、レンジスキャナーで取得した複数のデプスマップを統合し、冗長なポリゴンメッシュを削除しながら、残ったポリゴンをつなぎ合わせる手法を提案した。また、Hilton らは、被写体表面の連続性を考慮して、Volumetric なアプローチで 3 次元モデルを求める手法を提案した [9]。しかし、これらの手法は、時間方向に変調された TOF 方式で十分な観測時間を用いて取得した高精度のデプスマップを対象としており、静止した被写体にしか対応していない。

一方で、動物体の推定エラーを含む複数のデプスマップを、各点が各カメラに見えるか否かという情報である Visibility を用いたのアプローチで統合する手法も提案されている [26]。この手法は、ある頂点の Visibility の他の頂点への影響を考慮して、推定エラーの頂点を削除する手法である。この手法は、表面が単純な形状であるとき、良好に動作するが、表面形状が複雑になり、デプスマップ上のある画素の位置とその光学主点の延長線上に複数の表面があるとき、延長線上に一つの頂点しか存在しないと潜在的に仮定しているため、精度が低下する課題がある。

提案手法は、複数のデプスマップを、そのまま重ね合わせて生成した統合モデルを、各カメラに投影し、各頂点のカラーと、その頂点を投影したカラー画像上の投影点のカラーとの類似度を確認するため、各カメラから Visible な頂点のみをチェックするため、複雑な形状にも対応することができる。

8.3 複数デプスマップからの 3 次元モデル生成

本節では、これまで述べてきた赤外線カラーカメラアレーを用いて取得した複数のデプスマップを統合する手法について述べる。はじめに、本章で用いる表記について説明したあと、統合方法について述べる。

8.3.1 表記

本章では、Table 8.1 にリスト表示されている表記を用いる。以下、図 8.1 を用いて、各表記について定義する。

カメラ l 上の画素 i の位置は、 $p_{l,i}$ で定義する。 D_l は、カメラ l のデプスマップである。 $V(p_{l,i})$ は、デプスマップ D_l 上の画素の位置 $p_{l,i}$ の頂点である。 $P_m(\cdot)$ は、カメラ m の画素の位置を返す関数である。すなわち、 $p_{m,j} = P_m(p_{l,i})$ である。 $S_{m,j}$ は、 $p_{m,j}$ に投影される頂点群を示す。すなわち、 $S_{m,j} = \{V(p) | p_{m,j} = P_m(V(p))\}$ である。 $s_{m,j,r}$ は、 $S_{m,j}$ の r 番目の要素である。すなわち、 $s_{m,j,r} \in S_{m,j}$ である。 $K(\cdot)$ は、頂点のカラー値を返す関数である。 Z_m は、カメラ m のカメラ座標系における、頂点の z 座標を示す。

表 8.1 各変数の表記

表記	内容
$p_{l,i}$	カメラ l 上の i 番目の画素の位置
D_l	カメラ l のデプスマップ
$V(p_{l,i})$	デプスマップ D_l 上の画素 $p_{l,i}$ の頂点
$P_m(\cdot)$	カメラ m 上の投影点を返す関数
$S_{m,j}$	$p_{m,j}$ に投影される 頂点群
$K(\cdot)$	頂点のカラー値を返す関数
Z_m	カメラ m のカメラ座標系における頂点の Z 座標

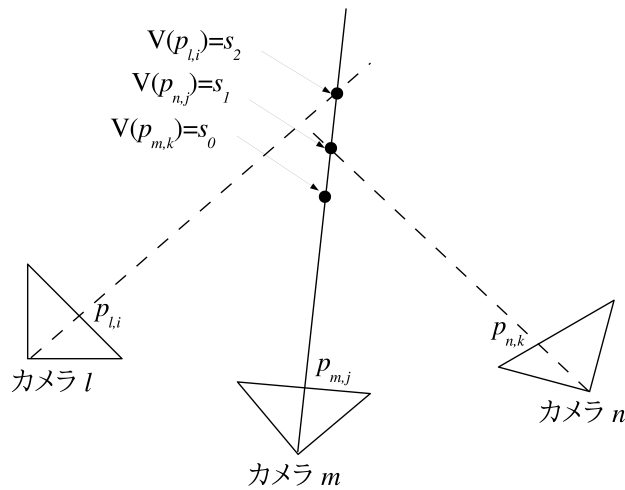


図 8.1 本章での各カメラと光軸上の点の表記

8.3.2 Visibility ベースのデプスマップ統合手法

提案手法は、はじめに、図 8.2 のように、本研究で求めた複数のデプスマップを 3 次元空間に重ね合わせて統合し、初期 3 次元モデル Q を生成する。次に、初期 3 次元モデル Q を各カメラに投影して、頂点カラー値とカラー画像上の投影点のカラー値との類似度から、推定エラーの頂点を削除する。

ここで、カメラ 1 に投影する場合を例に取って、推定エラーの頂点をどのように削除するのか説明する。頂点 $s_{m,i,r}$ のカラー値 $K(s_{m,i,r})$ と各画素 $p_{m,i}$ のカラー値 $I_m(p_{m,i})$ との類似度を Cost $E_{m,i,r}$ として式 8.1 で定義する。

$$E_{m,i,r} = |K(s_{m,i,r}) - I_m(p_{m,i})|. \quad (8.1)$$

提案手法では、カメラ m 上の位置 $p_{m,i}$ にある、最小 Cost E となる頂点 $T_{m,i}$ が、被写体表

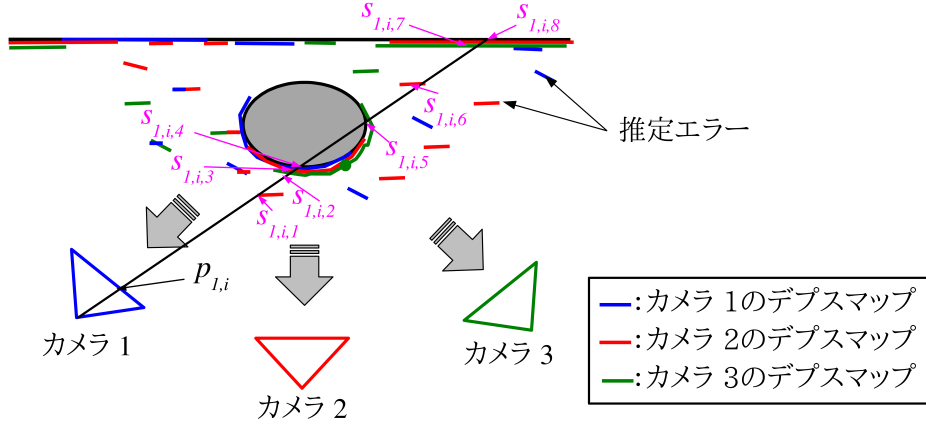


図 8.2 複数デプスマップを 3 次元空間に重ね合わせて統合した初期 3 次元モデル Q

面にあると仮定する。すなわち、

$$T_{m,i} = \{s_{m,i,w} | w = \operatorname{argmin}_r E_{m,i,r}\} \quad (8.2)$$

そして、被写体表面と推定された $T_{m,i}$ よりもカメラに近いすべての頂点 $Y_{m,i}$ を削除する。このとき、 $Y_{m,i}$ は、式 8.3 となる。

$$Y_{m,i} = \{s_{m,i,r} | Z_m(s_{m,i,r}) < Z_m(T_{m,i})\}. \quad (8.3)$$

この処理をすべてのカメラに行うことで、推定エラーの頂点を削除した統合 3 次元モデル U を取得することができる。

8.3.3 欠落領域の補完処理

統合 3 次元モデル U には、オクルージョンやデプスマップ推定時の推定エラー、統合処理上での推定エラー頂点の削除処理のために、表面が推定されなかった欠落領域が穴となって残ってしまう。本節では、このような欠落領域の補完方法について述べる。

補完処理は、8.3.2 節で述べた統合処理で取得した統合 3 次元モデル U を各カメラ m に投影した修正デプスマップ D'_m 毎に処理される。このときの修正デプスマップ D'_m は、式 8.4 で取得できる。

$$D'_m(P_m(g)) = Z_m(g), \quad g = \arg \min_q Z_m(q), \quad q \in Q \quad (8.4)$$

欠落領域の補完処理の概念を図 8.3 に示す。補完処理は、類似色の頂点は同様の奥行き値であると仮定して、4.4.3 節で述べた CLMF による平滑化処理で行う。このとき、カメラ m のカラー画像 I_m がガイド画像となり、修正デプスマップ D'_m が平滑化対象画像となる。この処理により、欠落領域の奥行き値は、CLMF の適応型カーネルの中央画素の色と類似している

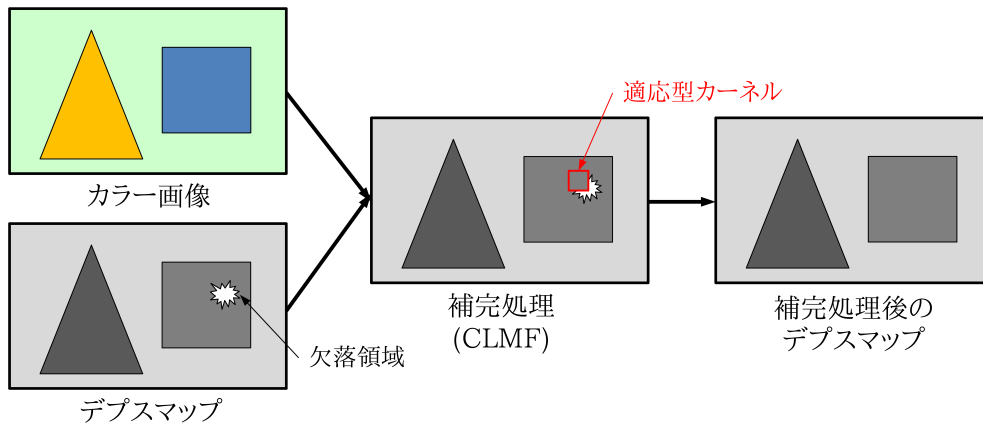


図 8.3 欠落領域の補完処理

近傍画素の奥行き値が割り当てられる。この処理を、カラーカメラ m に投影したすべてのデプスマップ D'_m について行う。最終的な統合 3 次元モデルは、これらのデプスマップを世界座標系の 3 次元空間中に変換して、重ね合わせることで取得することができる。

8.4 映像制作への応用

本節では、本研究で提案した奥行き推定手法を用いて推定した時系列デプスマップを統合して生成した時系列 3 次元モデルの映像制作への応用例について述べる。

8.4.1 2 次元映像制作

3 次元モデルを生成できると、撮影時と異なる視点からの映像を制作することができることから、さまざま用途に応用することができる。これまでに、我々は、HD カメラ 24 台で撮影した多視点映像で能を撮影し、その多視点画像から 3 次元モデルを生成して、ユーザーの好きな視点から能の演技を見ることのできる伝統舞踊アーカイブシステムを開発した [12]。このシステムは、図 8.4 のような画面上で、視点コントローラーを用いて、ユーザーが見たい視点を指定し、その視点からの映像を生成して提示するシステムである。このシステムは、能のような形状の変化する被写体を、3 次元モデルデータとして蓄えることができることから、無形文化財のアーカイブスとして効果的である。

また、放送番組への 3 次元モデルの応用についても、我々はこれまでにドラマへの応用をはかってきた [13][53]。図 8.5 の (a) は、一人の演者が 1 分間演じたシーンを、場所や角度、再生のタイミングを変えて配置して生成した群衆シーンである。このような群衆シーンの映像を制作するには、通常多くのエキストラを必要とするが、このような多視点映像から生成した 3 次元モデルのデジタルエキストラを用いると、効果的に制作することができる。(b) は、2009



図 8.4 伝統舞踊アーカイブシステム



(a) 3次元モデルを用いた群衆シーン


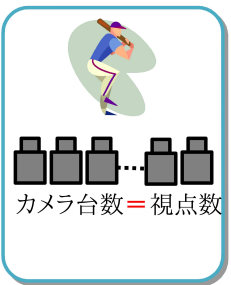
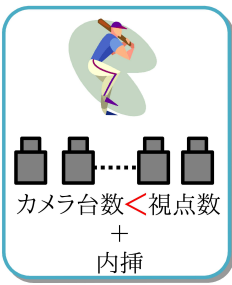
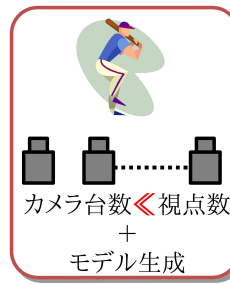

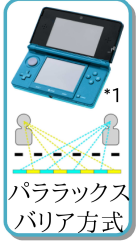
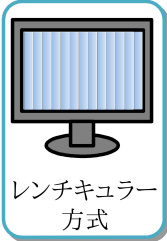

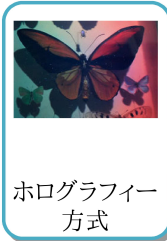


(b) ドラマ番組への応用

図 8.5 3次元モデルの2次元映像制作への応用例

年に放送されたスペシャルドラマ「坂の上の雲」で放送に用いられた3次元モデルを使用したシーンの一例である。背景で走り回っている兵士が、多視点映像から生成した時系列3次元モデルである。3次元モデルを利用すると、撮影時と異なる位置からの映像を生成することから、編集時に視点を自由に変更できるために、制作の柔軟性が高まるため効果的である。また、データ形式がCGと同じであり、既存のCGのワークフローを用いて、任意の視点からの映像を生成できるため、制作現場での親和性が高い。さらに、視点を自由に変更できることから、一度生成した3次元モデルをデータベースとして登録しておけば、その後も何度も自由に用いることができる。そのため、3次元モデルを2次元映像制作へ応用することによる効果は大きい。

しかし、これらの手法もテクスチャーのない領域については、3次元情報を取得することができないため、被写体に制約がかかってしまう課題がある。本研究で検討した手法を用いると、そのような領域も含めて安定的に3次元情報を取得することができるため、被写体への制約が緩和され、より実用的な2次元映像制作を可能とする。

	ステレオ(2眼)	多眼(2眼以上)			
撮影	 <p>現行の3D映画</p> <p>・観測値のみで高品質</p>	 <p>カメラ台数 = 視点数</p> <p>・観測値のみで高品質 ・視点に応じて台数増加</p>	 <p>カメラ台数 < 視点数 + 内挿</p> <p>・光線空間での内挿 ・比較的多くの台数</p>	 <p>カメラ台数 ≪ 視点数 + モデル生成</p> <p>・3次元幾何を用いた内挿 ・視点に比べ少ない台数</p>	
表示	 <p>メガネ方式</p>	 <p>パララックスバリア方式^{*1}</p>	 <p>レンチキュラー方式</p>	 <p>インテグラル方式</p>	 <p>ホログラフィー方式</p>

*1:文献 [1]

図 8.6 3次元映像制作への応用

8.4.2 3次元映像制作

3次元モデルは、3次元映像の制作にも応用することができる。図 8.6 のように、3次元映像は、ステレオによるものと多眼によるものと、大きく 2 つに分類することができる。ステレオによる 3次元映像はすでに実用化されており、映画やテレビ、ゲームなどに応用されている。撮影にあたっては、通常 2 台のカメラを用いる。表示にあたっては、2 台の映像をそれぞれ右目と左目に提示するために、視聴者は偏光メガネや液晶シャッターメガネなどの特殊なメガネをかける必要がある。そのため、視聴者は高解像度の 3次元映像を享受することができるが、メガネをかけることが負担となる。パララックスバリア方式のように、メガネをかけずに 3次元映像を享受できる方式も実用化されているが、焦点が合っている位置と、被写体が実際に 3次元空間に復元される位置とが異なるため、目の疲労が懸念されている。これは、メガネ方式でも同様である。

一方、多眼による 3次元映像は、実空間と同様の光線空間を再現するため、焦点が合っている位置と、被写体が 3次元空間に復元される位置が一致するため、ステレオ方式に比べて目の疲労が少ないと言われている。多眼方式の表示方法には、主にレンチキュラー方式、インテグラル方式、ホログラフィー形式などがある。レンチキュラー方式は、かまぼこ型のレンズを 1次元に配置したレンズアレーをディスプレイの前に配置して、水平視差のみを再現する方式で

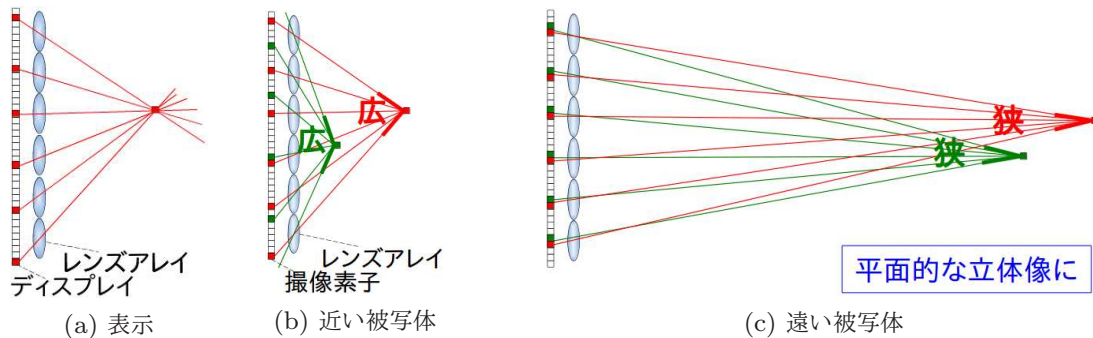


図 8.7 レンズアレーを用いた撮影方法

ある。垂直方向に解像度の高い 3 次元映像を再現できるが、水平視差のみであるため、頭を傾けると立体感が失われてしまう。一般的に、撮影はレンチキュラーレンズアレーを通して、高解像度のカメラで行う。インテグラル形式は、凸レンズを 2 次元配列に配置したレンズアレーをディスプレイの前に設置して、水平方向と垂直方向の両方を再現する方式である。レンチキュラー方式に比べて、垂直方向の解像度が減ってしまうが、頭を傾けても立体感が失われないため、家庭でさまざまな体勢で視聴するテレビへの応用に適している。撮影は、インテグラルレンズアレーを通して、高解像度のカメラで行う。ホログラフィー形式は、レーザー光などのコヒーレント光を 2 つに分離して、長さの異なる経路で被写体に物体光と参照光を照射して、その干渉縞を記録し、その記録された干渉縞に再度 1 つのコヒーレント光を照射することで、3 次元映像を再生する方式である。水平方向と垂直方向の視差を再現することができるが、挟ピッチの光変調器を必要とする。

レンチキュラー方式とインテグラル方式については、上記の撮影方法の他に、再現する光線方向の数だけカメラを配置して取得することも可能である。しかし、その台数は膨大になることから、映像制作現場での運用が困難である課題がある。一方で、3 次元モデルを用いると、仮想空間で仮想カメラを自由に配置でき、撮影方向以外の光線もシミュレーションできることから、撮影台数を飛躍的に削減することができる [21][48][55]。また、ホログラフィー方式による 3 次元映像の生成も、インテグラル方式の信号から変換することができる [27] ことから、少ないカメラ台数でホログラフィー方式の 3 次元映像も生成することができる。さらに、レンズアレー越しに撮影する方法の場合、図 8.7 のように、近い被写体は視差のある立体感を伴った 3 次元映像を生成することができるが、遠方の被写体は視差がつかないため、立体感が軽減してしまう課題がある。

本研究で提案する手法を用いると 3 次元モデルを生成できることから、3 次元モデルを自由に配置できたり、現実では困難な大きなレンズアレーやカメラを配置できることから、遠方の被写体であっても視差の十分ある 3 次元映像を制作することができる。3 次元モデルを用いて、インテグラル方式の 3 次元映像を制作する場合には、図 8.8(a) のように、実際のモデルの位置を測定するため、仮想空間内で、仮想のレンズアレーや仮想の高解像度カメラを設定し

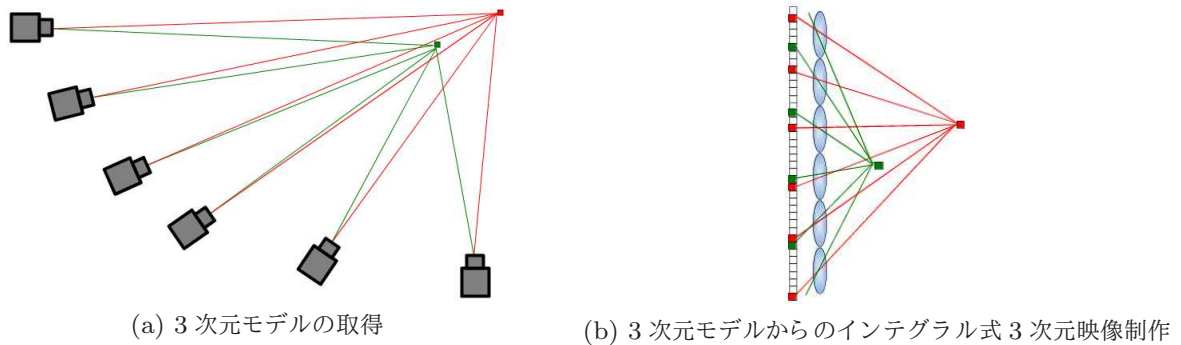


図 8.8 3次元モデルを用いた撮影方法

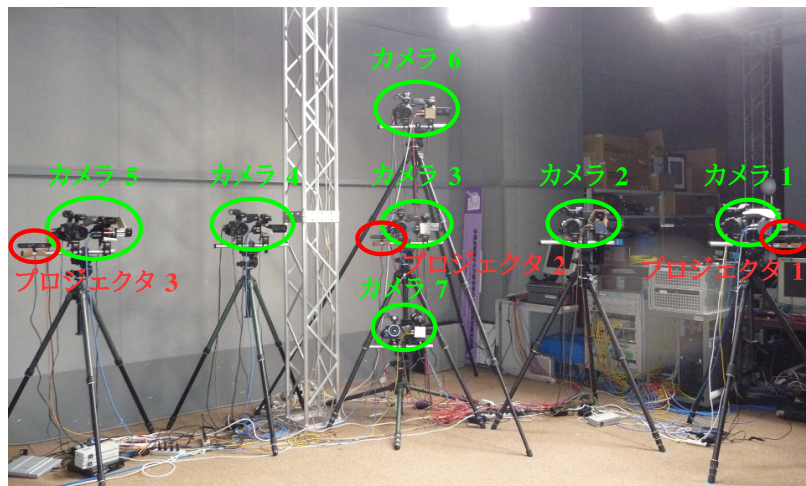


図 8.9 “comedy” データセット撮影時のカメラ配置

てインテグラル方式の3次元映像を生成する。そのため、(b)のように、実際の撮影では実現が困難な大きなレンズアレーやカメラをシミュレーションすることができ、遠方の被写体の被写体も十分視差のある立体像を再現することができる。

本章では、インテグラル方式の3次元映像生成への応用を例に、3次元モデルの3次元映像制作への応用について紹介する。

8.5 実験結果

本章で提案した統合手法の有効性を示すため、“comedy” データセットを用いて実験を行った。図 8.9 は、カメラ配置の様子を示したものである。3 台の赤外線プロジェクターと、6 台の赤外線カメラ、6 台のカラーカメラが水平方向と垂直方向に配置されている。緑色の丸枠は、赤外線カメラとカラーカメラを示しており、それらは隣り合うように配置されている。このうち、2 台の赤外線カメラと 2 台のカラーカメラを 1 組として選択して、合計 5 組のカメラアレーから 5 枚のデプスマップを推定した。各ステレオカメラ間の角度は、約 22 度である。



(a) カメラ 1 (b) カメラ 5 (c) カメラ 6 (d) カメラ 7

図 8.10 “comedy” データセットを撮影した最も外側のカメラの撮影画像 (上段は赤外線画像、下段はカラー画像)



(a) 初期 3 次元モデル (b) 統合 3 次元モデル (c) 最終統合モデル

図 8.11 デプスマップを統合した結果 (上段: 右上からの映像、下段: 上からの映像)

図 8.10 は、このカメラアレーのカバーしている領域を示すために、最も外側に位置するカメラの入力画像を示したものである。上段は赤外線画像で、下段はカラー画像である。

図 8.11 は、複数デプスマップを統合した 3 次元モデルである。(a) は、各デプスマップを 3 次元空間に重ね合わせて生成した初期 3 次元モデルである。各デプスマップにおいて、推定エラーの頂点が残っているために、モデルの右上の視点にモデルを投影した画像を見ると、赤丸内等に推定エラーの頂点が緑の背景などに写り込んでしまっている。これらの推定エラーの頂点は、推定時の撮影点以外の位置に投影すると、視覚的に大きな悪影響を与える。(b) は、初期 3 次元モデルを各カメラに投影して、頂点のカラー値と投影したカメラ画像上の投影点のカラー値との類似度が最も高い頂点よりも、カメラに近い頂点を削除した修正統合 3 次元モデルを示したものである。赤枠で示した緑の背景上に発生していた手の推定エラーの頂点がほとんど削除されており、投影画像上で視覚的に大きな改善が確認できる。しかし、推定エラー頂点の削除処理により、欠落領域が拡大している。(c) は、CLMF を用いて欠落領域の補完処理をした後の最終統合 3 次元モデルを示したものである。緑の背景や、床で発生していた欠落領域

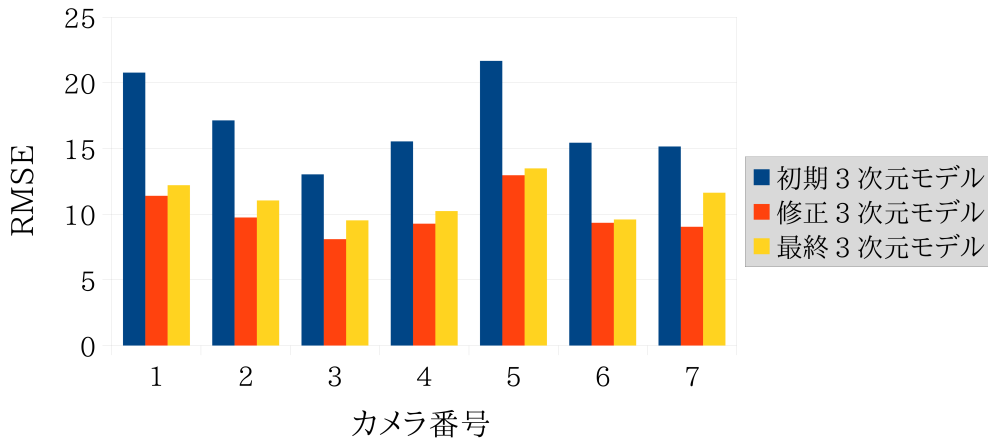


図 8.12 頂点のカラー値とカラー画像上の投影点のカラー値との RMSE

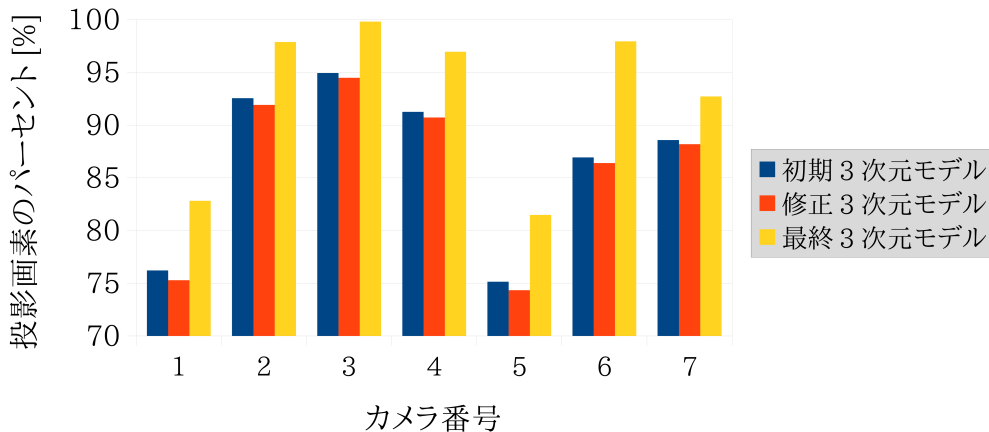


図 8.13 総画素数に対する投影点のパーセント

が補完されているのが確認できる。

図 8.12 は、最終統合 3 次元モデルの各頂点のカラー値と、各頂点を各カラーカメラに投影したときのカラー画像のカラー値との RMSE (Root Mean Square Error) を示したものである。また、図 8.13 は、最終統合 3 次元モデルを各カメラに投影したときの、各カメラの総画素数に対する投影点のパーセンテージを示したものである。RMSE は、推定エラー頂点の除去により減少していることから、視覚的な 3 次元モデルの改善を定量的に確認することができたが、補完処理により多少上がった。しかし、投影点のパーセントは補完処理によって増加し、最終統合 3 次元モデルの精度が大幅に改善した。このことから、補完処理による RMSE の上昇は限定的であり、総合的にはモデルの品質は改善ができたと考えている。

図 8.14 は、生成した 3 次元モデルを映像制作に応用した一例である。(a) のようなカラー画像と (b) のような赤外線画像からデプスマップを複数枚生成し、それらを本章で提案した統合方法で統合した 3 次元モデルから、右上の視点に投影した画像を (c) に示す。3 次元モデルを生成すると、カメラを配置した範囲内であれば、自由に視点を移動できることから、2 次元

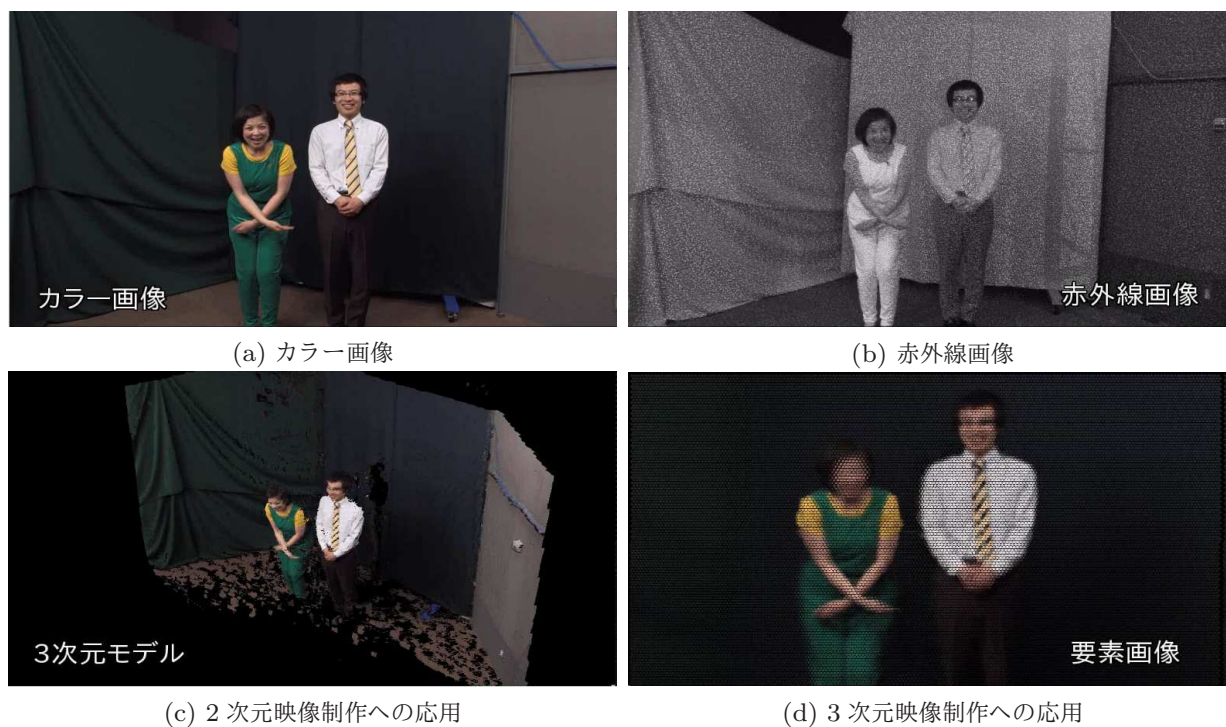


図 8.14 3次元モデルの映像制作への応用

映像制作に応用することができる。本手法では、赤外線ドットパターンを照射しているので、被写体のテクスチャーの有無に関わらず、安定して時系列のデプスマップを生成できることから、より実用性の高い2次元映像制作を実現することができる。

(d) は、3次元映像制作への応用の一例として、インテグラル式の立体像を再生するためにディスプレイ面に表示する要素画像群を示したものである。これは、レンズアレーの各レンズに対応する要素画像がレンズ数並んだ画像で、この画像を表示しているディスプレイ面から焦点距離離れた位置にインテグラルレンズアレーを設置することにより、特殊なメガネなしの立体像を再生することができる。図 8.15 は、ディスプレイ面から焦点距離離れた位置にレンズアレーを設置して、3次元映像を再生した様子を再撮影した画像である。立体像が再生されている様子が確認できる。このように、本手法は2次元映像だけでなく、3次元映像についても応用が可能である。

8.6 まとめ

本章では、本研究で提案したデプスマップ推定手法を用いて取得した複数のデプスマップを統合する手法を提案した。この手法は、はじめに推定した複数のデプスマップを3次元空間上で重ね合わせ、初期3次元モデルを生成する。次に、初期3次元モデルを各カメラに投影し、各頂点のカラー値と投影したカラー画像上の投影点のカラー値との類似度が最も高い頂点が、

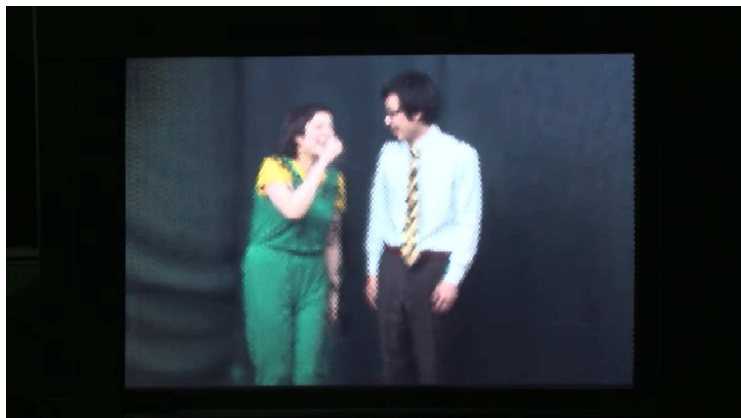


図 8.15 インテグラル式立体像の再撮影した画像

被写体の表面にあると仮定して、それよりもカメラ側にある頂点を削除することで、推定エラーの頂点を除去して修正 3 次元モデルを生成する。さらに、修正 3 次元モデルは、オクルージョンや左右視差チェック、推定エラー除去により、欠落領域が生じてしまうことから、修正 3 次元モデルを各撮影点からのデプスマップに変換し、当該カメラのカラー画像をガイドとした CLMF による平滑化処理を行うことにより、欠落領域の補完処理を行った。

“comedy” データセットを用いて 3 次元モデルの生成実験を行ったところ、推定エラー除去処理により、視覚的に大きく改善できることを確認できた。また、補完処理により、欠落領域についても視覚的に大きな改善を確認することができた。さらに、RMSE と推定率による定量的な実験結果から、推定エラー除去処理により RMSE が減少し、画質が改善されていることを確認することができた。また、欠落領域の補完処理により、推定率が上がったことにより、定量的にも 3 次元モデルの品質の改善を確認することができた。推定エラー除去処理により推定率は多少減少したが、補完処理との組み合わせにより、総合的な 3 次元モデルの品質改善が図られている。

本手法による統合処理により、本研究で提案したデプスマップ推定手法が、2 次元映像や 3 次元映像の制作に応用できることを示した。実験では、本研究で提案した奥行き推定手法を用いて取得した複数のデプスマップを統合した 3 次元モデルを、ある仮定の撮影点に投影して、撮影点以外の位置からの映像を生成した。これにより、任意視点映像の生成が可能なが示され、2 次元映像における制作への応用の可能性を示すことができた。また、本章の統合手法で生成した 3 次元モデルからインテグラル式立体像を生成する実験を実施し、本研究の 3 次元映像制作への応用の可能性も示すことができた。

第9章

結論

本研究では、映像制作に応用が可能なデプスマップ推定手法を検討した。映像制作への応用にあたっては、デプスマップが奥行き方向も含めて高解像度であること、推定精度が高いこと、3次元形状とともにテクスチャーを同時に取得できること、複数のデプスマップが同時に取得できること、動画に対応できること、被写体のテクスチャーの依存が少ないこと、被写体の形状に制約が少ないこと、被写体の素材に制約が少ないことなど、多くの要求条件を満たす必要がある。そこで、本研究では、赤外線プロジェクターでドットパターンを被写体に照射し、2台の赤外線カラーカメラで撮影した赤外線ステレオ画像とカラーステレオ画像から、ステレオマッチングのアプローチによりデプスマップを推定する手法を提案した。これにより、鏡面反射や半透明・透明といった素材の依存性を除くすべての要求条件を満たす奥行き推定手法を考案した。

第4章では、図9.1(a)のように、赤外線ステレオ画像とカラーステレオ画像から Cost volume を生成し、CLMF による Cost volume filter 処理をしたあと、カーネルサイズにより、1つの Cost volume に統合してデプスマップを推定する基本手法を提案した。これにより、SN 比の低い赤外線画像からも、高い精度でデプスマップを推定でき、被写体の境界付近の膨張減少のないデプスマップを推定できる手法を実現した。しかし、Cost volume filter は、同一視差上の近傍画素の Cost を共有することにより、推定精度を改善する手法であることから、潜在的に被写体の表面は、カメラに正対しているという仮定をおいている。そのため、被写体表面がカメラに正対していない領域の推定精度が低い課題が明らかになった。

そこで、第5章では、図9.1(b)のように、Cost volume filter ののち、各画素において Cost が最小となる視差を選択する Winner-takes-all 戦略ではなく、最適化アルゴリズムを用いて視差の異なる近傍画素の Cost を考慮に入れる手法を提案した。最適化アルゴリズムには、グラフカットを用いた。従来手法のグラフカットは、各画素とそれに対応する視差にノードを配置し、ノード間をエッジで結んだ上で、エッジに Cost をエネルギーとして付与した3次元グラフカットを実施するが、この手法では多くの計算時間とメモリ使用量を必要とする。映像制

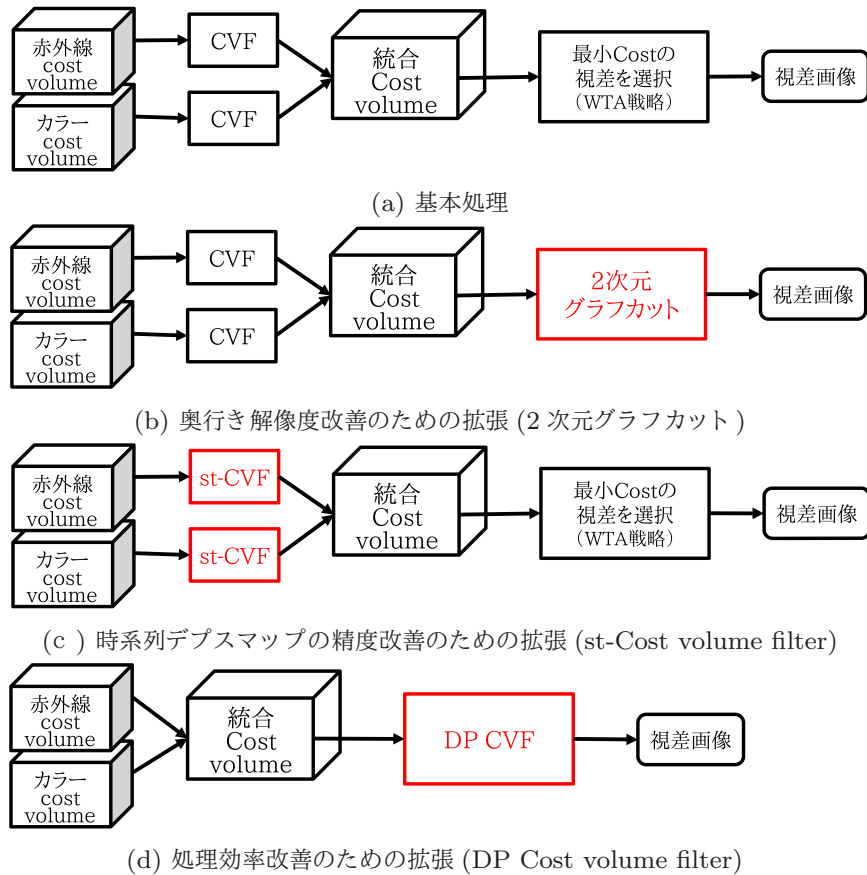


図 9.1 赤外線ステレオ画像とカラーステレオ画像からのデプスマップ推定

作への応用を考慮に入れると、画素数が多く、高い奥行き解像度を有したデプスマップを、多くのフレーム分推定して時系列デプスマップを取得する必要があることから、計算時間とメモリ使用量は極力少なく、現実的な計算時間とメモリ使用量で行う必要がある。そこで、提案手法では Cost volume filter で、垂直性の連続性はすでに考慮に入っており、左右の撮影点の違いによる画像の変形は水平方向にのみ生じることに注目し、スキャンライン毎にグラフカットを行う 2次元グラフカットを用いた。その結果、半分の計算時間、8分の1のメモリ使用量で、3次元グラフカットで推定したデプスマップとほぼ同品質のデプスマップを推定することができた。

しかし、この手法は、フレーム毎で独立に処理する手法であったため、時系列のデプスマップを生成すると、フレーム間の連続性が考慮していないため、デプスマップ上でフリッカーが生じる課題があった。本研究の最終目標は、映像制作への応用が可能な時系列 3次元モデルを生成することから、時系列デプスマップ上のフリッカーは大きな課題となることから、第 6 章では、図 9.1(c) のように、フレーム間の連続性を考慮した時系列デプスマップ推定手法を検討した。そのために、Cost volume と CLMF を時間方向に拡張し、st-Cost volume と st-CLMF を提案した。その上、これらを用いて Cost volume filter を時間方向に拡張して、

表 9.1 各手法の性能得失比較表

項目	CVF	2D GC	st-CVF	st-CVF+2D GC	DP CVF
フリッカー	×	×	○	○	× (時間拡張未対応)
非正対領域	×	○	×	○	○
長いカメラ間距離	×	○	×	○	○
処理時間	△	×	△	×	○
メモリ使用量	△	×	△	×	○

st-Cost volume filter を提案した。その結果、フリッカーを大きく抑制した時系列デプスマップを推定することができるようになった。しかし、グラフカットアルゴリズム自体が多く of 計算時間とメモリを必要とすることから、第 7 章では、図 9.1(d) のように、2 次元グラフカットよりも効率的に最適化問題を解くことができる DP を用いた奥行き推定手法を提案した。DP は、グラフカットよりも効率的に最適化問題を解くことができるが、2 次元の最適化問題にしか対応できないことから、水平方向と垂直方向の 2 段階の DP を用いることで、3 次元の最適化を行った。また、DP は、始点と終点が事前にわかっていないこと、左右の画素の順序が入れ替わることができないこと、始点から終点まで連続していないといけなないことなどの制約がある。そこで、はじめに入力画像のセグメンテーションを行い、各セグメント内は連続した平面もしくは曲面であるという仮定において、これらの制約に対応した。その結果、2 次元グラフカットを用いて推定したデプスマップとほぼ同品質を維持したまま、大幅な計算時間の削減を実現できた。特に、画素数や視差数が多くなると、処理時間の改善が大きくなることがわかり、映像制作への応用に向いていることが確認できた。この手法は一般的なカラーステレオ画像からデプスマップを推定することにも応用が可能である。その場合には、カラーカメラの SN 比が高いことから、Cost volume filter を併用する必要がないため、計算時間の大幅な短縮化ができることがわかった。

これら手法の性能の得失を比較すると、表 9.1 のようになる。基本処理である Cost volume filter では処理時間とメモリ使用量は許容範囲内であるが、フリッカーが発生し、カメラに正対していない領域の精度が低下する。2 次元グラフカットを用いた場合、正対していない領域やカメラ間隔を広げた際も、高い精度で推定することができたが、処理時間とメモリ使用量に課題がある。時間拡張した st-Cost volume filter はフリッカーが抑制され、時系列デプスマップの精度は改善したが、正対していない領域の精度や、カメラ間隔を広げた場合に精度が低下する。一方で、DP Cost volume filter は、まだ時間拡張をしていないため、フリッカーが発生するが、その他の課題は克服されている。時間拡張については、st-Cost volume filter と同様に、1 次元 DP を時空間方向に 3 回行えば実現できると考えられるため、本論文ではこの手法が最もよいと考えている。

第 8 章では、本研究で提案した奥行き推定手法を用いて推定した複数のデプスマップを統合して、3 次元モデルを生成する手法を提案した。この処理では、はじめに取得した複数のデプ

スマップを 3 次元空間上に重ね合わせて初期 3 次元モデルを生成する。次に、初期 3 次元モデルを各カメラに投影して、各頂点のカラー値と、カラー画像上の投影点のカラー値との類似度が最も高い頂点が被写体表面であると仮定して、それより前の頂点を削除することで推定エラーを除去して修正 3 次元モデルを生成する。この修正 3 次元モデルは、オクルージョンや左右視差チェック、推定エラー頂点の除去処理により、3 次元情報のない欠落領域を含む。そこで、修正 3 次元モデルを各カメラに投影して、各撮影点からのデプスマップを生成し、類似色の奥行きは類似した奥行き値であると仮定して、カラー画像をガイドとした、CLMF による平滑化処理をデプスマップに行って補完処理を行う。このようにして、補完した複数のデプスマップを 3 次元空間上に重ね合わせることで、最終 3 次元モデルを生成した。実験では、最終 3 次元モデルを用いて、撮影していない位置からの映像を生成し、本研究で提案したデプスマップ推定手法が 2 次元映像制作に応用が可能であることを示した。また、この最終 3 次元モデルから要素画像群を生成してディスプレイに表示し、焦点距離離れた位置にレンズアレーを設置して、立体像を再生する実験を行い、本手法が 3 次元映像制作にも応用が可能であることを示した。

このように、本研究では赤外線プロジェクターと複数の赤外線カラーカメラを用いて、映像制作への応用にあたっての要求条件のうち、被写体の素材の依存度が低くする点を除いたすべての条件を満たしたデプスマップを生成する手法を確立することができた。また、実際に本研究を用いて生成したデプスマップを用いて、2 次元映像および 3 次元映像制作への応用の可能性を示した。

本研究で提案した手法は、DP を用いた奥行き推定手法と、複数デプスマップの統合手法については、時間方向の拡張が行われていない。はじめに推定するデプスマップにおいて、時間方向の連続性が考慮されているため、フリッカーへの対応が行われているため、基本的な対応はできているが、これらの時間拡張によって、最終的な時系列 3 次元モデルの精度の改善が図れると思われる。この点については今後の課題としたい。

参考文献

- [1] 任天堂製 Nintendo3DS. <http://www.nintendo.co.jp/3ds>.
- [2] Jun Arai, Masahiro Kawakita, Takayuki Yamashita, Hisayuki Sasaki, Masato Miura, Hitoshi Hiura, Makoto Okui, and Fumio Okano. Integral three-dimensional television with video system using pixel-offset method. *Opt. Express*, Vol. 21, No. 3, pp. 3474–3485, Feb 2013.
- [3] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 26–33 vol.1, Oct 2003.
- [4] Yi Deng and Xueyin Lin. A fast line segment based dense stereo algorithm using tree dynamic programming. In Ale Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision ECCV 2006*, Vol. 3953 of *Lecture Notes in Computer Science*, pp. 201–212. Springer Berlin Heidelberg, 2006.
- [5] P.Felzenszwalb D.Huttenlocher. Efficient belief propagation for early vision. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 261–268, 2004.
- [6] キヤノン製 XHG1. <http://cweb.canon.jp/manual/dvcamera/pdf/xhg1-xha1.pdf>.
- [7] S.B. Gokturk, H. Yalcin, and C. Bamji. A time-of-flight depth sensor - system description, issues and solutions. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pp. 35–35, June 2004.
- [8] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 35, No. 6, pp. 1397–1409, June 2013.
- [9] A. Hilton, A.J. Stoddart, J. Illingworth, and T. Windeatt. Reliable surface reconstruction from multiple range images. Bernard Buxton and Roberto Cipolla, editors, *Computer Vision — ECCV '96*, 第 1064 卷 of *Lecture Notes in Computer Science*, pp. 117–126. Springer Berlin Heidelberg, 1996.
- [10] K. Hisatomi, M. Kano, K. Ikeya, M. Katayama, T. Mishina, and 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. In *3D Vision (3DV), 2015 International Conference on*, pp. 580–588, Oct 2015.
- [11] Kensuke Hisatomi, Kensuke Ikeya, Miwa Katayama, Yuichi Iwadate, and Kiyoharu Aizawa. Depth estimation by cost volume with spatial-temporal cross-based local multipoint filter using projecting infrared patterns. In *Proceedings of the 10th European Conference on Visual Media Production, CVMP '13*, pp. 10:1–10:8, New York, NY, USA, 2013. ACM.
- [12] Kensuke Hisatomi, Miwa Katayama, Kimihiro Tomiyama, and Yuichi Iwadate. 3d archive system for traditional performing arts. *International Journal of Computer Vision*, Vol. 94, No. 1, pp. 78–88, 2011.
- [13] Kensuke Hisatomi, Kimihiro Tomiyama, Miwa Katayama, Yuichi Iwadate, Koji Matsunaga, Yoshiyuki Ito, and Wataru Ishihara. A method of video production using dynamic 3d models and its application to making scenes of a crowd. *Motion Imaging Journal, SMPTE*, Vol. 118, No. 7, pp. 29–36, Oct 2009.
- [14] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 35, No. 2, pp. 504–511, Feb 2013.
- [15] Katsushi Ikeuchi, Kazuhide hasegawa, Atsushi Nakazawa, Jun Takamatsu, Takeshi Oishi, and Tomohito Masuda. Bayon digital archival project. *10th International Conference on Virtual Systems and Multimedia (VSMM)*, 2004.
- [16] James Imber, Jean-Yves Guillemaut, and Adrian Hilton. *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II*, chapter Intrinsic Textures for Relightable Free-Viewpoint Video, pp. 392–407. Springer International Publishing, Cham, 2014.
- [17] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11*, pp. 559–568, New York, NY, USA, 2011. ACM.
- [18] JAI 製 AD-130GE. <http://www.jai.com/en/products/ad-130ge>.
- [19] 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. In *3D Systems and Applications (3DSA)*, pp. S1–3, 2013.

-
- [20] K.Hisatomi M.Kano K.Ikeya M.Katayama T.Mishina Y.Iwadate K.Aizawa. Depth estimation using an infrared dot projector and an infrared color stereo camera. In *IEEE Transactions on Circuit and Systems for Video Technology (TCSVT) To be appeared*, 2016.
- [21] Miwa Katayama and Yuichi Iwadate. A method for converting three-dimensional models into auto-stereoscopic images based on integral photography, 2008.
- [22] KinectV1. <http://www.xbox.com/en-us/kinect>.
- [23] KinectV2. <http://www.xbox.com/ja-jp/kinect?xr=shellnav>.
- [24] Cheng Lei, J. Selzer, and Yee-Hong Yang. Region-tree based stereo using dynamic programming optimization. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, pp. 2378–2385, 2006.
- [25] Jiangbo Lu, Keyang Shi, Dongbo Min, Liang Lin, and M.N. Do. Cross-based local multipoint filtering. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 430–437, June 2012.
- [26] P. Merrell, A. Akbarzadeh, Liang Wang, P. Mordohai, J.-M. Frahm, Ruigang Yang, D. Nister, and M. Pollefeys. Real-time visibility-based fusion of depth maps. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, Oct 2007.
- [27] Tomoyuki Mishina, Makoto Okui, and Fumio Okano. Generation of holograms using integral photography, 2004.
- [28] Andreas Nüchter, Kai Lingemann, Joachim Hertzberg, and Hartmut Surmann. 6d slam—3d mapping outdoor environments: Research articles. *J. Field Robot.*, Vol. 24, No. 8-9, pp. 699–722, August 2007.
- [29] Georgia Sandbach, Stefanos Zafeiriou, Maja Pantic, and Lijun Yin. Static and dynamic 3d facial expression recognition: A comprehensive survey. *Image Vision Comput.*, Vol. 30, No. 10, pp. 683–697, October 2012.
- [30] K. Sande. A practical setup for voxel coloring using off-the-shelf components. *University of Amsterdam Bachelor thesis*, 2004.
- [31] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, Vol. 47, No. 1-3, pp. 7–42, 2002.
- [32] S. Seitz and C. Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision (IJCV)*, Vol. 35, No. 2, pp. 151–173, 1999.
- [33] Noah Snavely, StevenM. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, Vol. 80, No. 2, pp. 189–

- 210, 2008.
- [34] SwissRanger. <http://www.mesa-imaging.ch/>.
- [35] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, Vol. 3, No. 4, pp. 323–344, August 1987.
- [36] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pp. 311–318, New York, NY, USA, 1994. ACM.
- [37] O. Veksler. Stereo correspondence by dynamic programming on a tree. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 2, pp. 384–390 vol. 2, June 2005.
- [38] Paul Viola and MichaelJ. Jones. Robust real-time face detection. *International Journal of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.
- [39] Liang Wang, Miao Liao, Minglun Gong, Ruigang Yang, and D. Nister. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pp. 798–805, June 2006.
- [40] Liang Wang, Ruigang Yang, Minglun Gong, and Miao Liao. Real-time stereo using approximated joint bilateral filtering and dynamic programming. *Journal of Real-Time Image Processing*, Vol. 9, No. 3, pp. 447–461, 2014.
- [41] Xtion. http://www.asus.com/multimedia/xtion_pro/.
- [42] K.Hisatomi K.Tomiyama M.Katayama Y.Iwadate. 3d reconstruction using graph cut with view-dependent polygon texture blending. In *IEE 5th European Conference on Visual Media Production (CVMP 2008)*, 2008.
- [43] K.Hisatomi K.Tomiyama M.Katayama Y.Iwadate. Method of 3d reconstruction using graph cuts, and its application to preserving intangible cultural heritage. In *IEEE workshop on eHeritage and Digital Art Preservation*, pp. 923–930, 2009.
- [44] K.Ikeya K.Hisatomi M.Katayama Y.Iwadate. Depth estimation from three cameras using belief propagation. In *8th European Conference On Visual Media Production (CVMP 2011)*, pp. 118–125, 2011.
- [45] Ke Zhang, Jiangbo Lu, and G. Lafruit. Cross-based local stereo matching using orthogonal integral images. *Circuits and Systems for Video Technology, IEEE Transactions on*, Vol. 19, No. 7, pp. 1073–1079, July 2009.
- [46] Li Zhang, B. Curless, and S.M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *3D Data Processing Visualization*

- and Transmission, 2002. Proceedings. First International Symposium on*, pp. 24–36, 2002.
- [47] キヤノン製 XF105. <http://cweb.canon.jp/prodv/lineup/xf105/index.html>.
- [48] 岩館祐一, 片山美和. 斜投影によるインテグラル立体像の生成手法 (立体映像技術一般). 映像情報メディア学会技術報告, Vol. 34, No. 43, pp. 17–20, oct 2010.
- [49] 久富健介, 加納正規, 池谷健佑, 片山美和, 三科智之, 相澤清晴. 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 ME2014-109, Vol. 38, No. 51, pp. 17–20, 2014.
- [50] 久富健介, 池谷健佑, 片山美和, 岩館祐一. 複数距離画像からのインテグラル立体像生成手法. *3DIT2012-34 IDY2012-19*, Vol. 36, No. 24, pp. 7–10, 2012.
- [51] 久富健介, 富山仁博, 片山美和, 岩館祐一. グラフカットを用いた多視点映像からの3次元形状復元法の検討. 2008年映像情報メディア学会冬季大会講演予稿集, pp. 6–7, 2008.
- [52] 久富健介, 富山仁博, 片山美和, 岩館祐一. 動的3次元映像におけるブレンディングを用いた視点依存型面テクスチャ手法の検討. 第11回画像の認識・理解シンポジウム論文集 (MIRU 2008), 第IS4-37巻, pp. 1352–1357, 2008.
- [53] 久富健介, 富山仁博, 片山美和, 岩館祐一, 松永孝治, 井藤良幸, 石原渉. デジタルエキストラを用いた映像制作手法. 映像情報メディア学会誌, Vol. 66, No. 11, pp. J426–J433, 2012.
- [54] 久富健介, 富山仁博, 片山美和, 岩館祐一, 松永孝治, 井藤良幸, 石原渉. 動的3次元モデルのドラマへの応用: 映像制作の新たな試み. 映像情報メディア学会年次大会講演予稿集, No. 2010, pp. 13-1-1-13-1-2, aug 2010.
- [55] 池谷健佑, 久富健介, 片山美和, 岩館祐一. 多視点映像からのインテグラル立体像生成手法. 映像情報メディア学会誌.
- [56] 内田誠一, 石川博. 特徴照合. 電子情報通信学会「知識ベース」, pp. 9–21, 2012.

発表文献

【 本論文を構成する論文】

学術雑誌論文

- [1] Kensuke Hisatomi, Masanori Kano, Kensuke Ikeya, Miwa Katayama, Tomoyuki Mishina, Yuichi Iwadate, Kiyoharu Aizawa: “Depth Estimation Using an Infrared Dot Projector and an Infrared Color Stereo Camera” IEEE Transactions on Circuit and Systems for Video Technology (TCSVT), To be appeared, 2016

国際会議

- [2] Kensuke Hisatomi, Masanori Kano, Kensuke Ikeya, Miwa Katayama, Tomoyuki Mishina, Kiyoharu Aizawa: “Depth Estimation Based on an Infrared Projector and an Infrared Color Stereo Camera by Using Cross-Based Dynamic Programming with Cost Volume Filter” 3D Vision (3DV), p. 580-588, 2015
- [3] Kensuke Hisatomi, Kensuke Ikeya, Miwa Katayama, Yuichi Iwadate, Kiyoharu 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), pp. 10, 2013
- [4] Kensuke Hisatomi, Kensuke Ikeya, Miwa Katayama, Yuichi Iwadate, Kiyoharu 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 (3DSA), S1-3, 2013

国内研究会・シンポジウム等

- [5] 久富健介, 加納正規, 池谷健佑, 片山美和, 三科智之, 相澤清晴: “赤外線ドットパターンと赤外線カラ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, ME2014-51, vol. 38, pp. 17-20, 2014

展示

- [6] Kensuke Hisatomi: “インテグラル立体テレビ” NHK 技研公開 2014 展示資料, 展示 9, 2014

【その他の論文】

学術雑誌論文

- [7] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “多視点映像からのインテグラル立体像生成手法 スポーツシーンの3次元モデル生成と立体映像変換” 映像情報メディア学会誌, Vol. 67, No. 7, pp. J229-J240, 2013
- [8] 久富健介, 富山仁博, 片山美和, 岩館祐一, 松永孝治, 井藤良幸, 石原渉: “デジタルエクストラを用いた映像制作手法” 映像情報メディア学会誌, Vol. 66, No. 11, pp. J426-J433, 2012
- [9] Kensuke Hisatomi, Kimihiro Tomiyama, Miwa Katayama, Yuichi Iwadate: “3D Archive System for Traditional Performing Arts” International Journal of Computer Vision (IJCV), Vol. 94, No. 1, pp. 78-88, 2011
- [10] Kensuke Hisatomi, Kimihiro Tomiyama, Miwa Katayama, Yuichi Iwadate, Koji Matsunaga, Yoshiyuki Ito, Wataru Ishihara: “A method of video production using dynamic 3d models and its application to making scenes of a crowd” SMPTE Motion Imaging Journal, Vol. 118, No. 7, pp. 29-36, 2008
- [11] 青木義満, 久富健介, 橋本周司: “分散協調処理を用いたロバストな能動的人物追跡システム” 画像電子学会誌, Vol. 28, No. 5, pp. 596-604, 1999
- [12] 横山太郎, 田中克明, 久富健介, 八木康史, 谷内田正彦, 原文雄, 橋本周司: “正面顔画像からの輪郭特徴抽出” 映像情報メディア学会誌, Vol. 53, No. 11, pp. 1605-1614, 1999

国際会議

- [13] Kensuke Ikeya, Kensuke Hisatomi, Miwa Katayama, Tomoyuki Mishina, Yuichi Iwadate: “Bullet Time Using Multi-Viewpoint Robotic Camera System” 11th European Conference on Visual Media Production (CVMP), 2014
- [14] Hitoshi Hiura, Tomoyuki Mishina, Jun Arai, Kensuke Hisatomi, Yuichi Iwadate, Takayuki Ito: “A Study on Accommodation Response and Depth Perception in Viewing Integral Photography” 5th International Conference on 3D Systems and Applications (3DSA), P2-2, 2013
- [15] Kensuke Ikeya, Kensuke Hisatomi, Miwa Katayama, Yuichi Iwadate: “Depth esti-

- mation from three cameras using belief propagation” 8th European Conference On Visual Media Production (CVMP), pp. 118-125, 2011
- [16] Kensuke Hisatomi, Kimihiro Tomiyama, Miwa Katayama, Yuichi Iwadate: “Method of 3d reconstruction using graph cuts, and its application to preserving intangible cultural heritage” IEEE workshop on eHeritage and Digital Art Preservation, pp. 923-930, 2009
- [17] Kensuke Hisatomi, Kimihiro Tomiyama, Miwa Katayama, Yuichi Iwadate: “3D Reconstruction using Graph Cut with View-dependent Polygon Texture Blending” IEE 5th European Conference on Visual Media Production (CVMP), 2008
- [18] K. Tomiyama, K. Hisatomi, M. Katayama, Y. Iwadate, T. Hayashida, H. Maruyama, T. G. Etoh, T. Yoshida: “Advanced Video Image Technologies for Sports TV Productions” NAB Proceedings, pp. 193-198, 2008
- [19] Yoshimitsu Aoki, Kensuke Hisatomi, Shuji Hashimoto: “Robust and Active Human Face Tracking Vision using Multiple Information” Proc. of World Multiconference on Systems, Cybernetics and Informatics (SCI), Vol.5, 28-33, 1999
- [20] Soo Hyun Cho, Kensuke Hisatomi, Shuji Hashimoto: “Cracks and Strain Feature Extraction of the Concrete Block Surface Using Image Processing Technique” 5th Korea-Japan Joint Workshop on Computer Vision (FCV), pp. 68-73, 1999
- [21] Soo Hyun Cho, Kensuke Hisatomi, Shuji Hashimoto: “Cracks and Displacement Feature Extraction of the Concrete Block Surface” IAPR Workshop on Machine Vision Applications (MVA), pp. 246-249, 1998
- [22] Soo Hyun Cho, Kensuke Hisatomi, Shuji Hashimoto: “Cracks Feature Extraction and Strain Measurement of the Concrete Block Surface Using Image Processing Technique” Proc. of QCAV’98, 1998

国内研究会・シンポジウム等

- [23] 久富健介, 池谷健佑, 片山美和, 岩館祐一: “複数距離画像からのインテグラル立体像生成手法” 3DIT2012-34 IDY2012-19, vol. 36, pp. 7-10, 2012
- [24] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “弱校正多視点ロボットカメラシステムにおける方向制御手法” 映像情報メディア学会冬季大会講演予稿集, 2012
- [25] 岩館祐一, 片山美和, 久富健介, 池谷健佑, 相澤清晴, 苗村健, 広田光一, 趙志鎬, 久保友香, 山崎眞見, 及川道雄, 小池崇文, 宇都木契: “「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, pp. 64-69, 2012
- [26] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “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), vol. IS4-30, pp.1486-1493, 2011
- [27] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “被写体の動きをパンフオロー可能な多視点ロボットカメラシステム” 映像情報メディア学会冬季大会講演予稿集, pp. 9-3, 2011
- [28] 岩館祐一, 相澤清晴, 苗村健, 広田光一, 山崎眞見, 及川道雄, 小池崇文, 片山美和, 久富健介, 池谷健佑, 趙志鎬, 久保友香, 宇都木契: “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” 3DIT2011-79 IDY2011-49 IST2011-58, vol. 35, pp. 1-4, 2011
- [29] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “多視点ロボットカメラシステムを用いた映像表現手法” 映像情報メディア学会年次大会講演予稿集, pp.21-1, 2012
- [30] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “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” 3DIT2011-80 IDY2011-50 IST2011-59, vol. 35, pp. 5-8, 2011
- [31] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “信頼度伝搬法による多視点画像からの距離推定手法” 映像情報メディア学会年次大会講演予稿集, pp.4-6, 2011
- [32] 久富健介, 富山仁博, 片山美和, 岩館祐一, 松永孝治, 井藤良幸, 石原渉: “動的3次元モデルのドラマへの応用” 映像情報メディア学会年次大会講演予稿集, pp. 13-1, 2010
- [33] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “階層型信頼度伝搬法による視差推定実験” 映像情報メディア学会年次大会講演予稿集, pp. 12-3, 2010
- [34] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “階層型信頼度伝搬法による視差推定に関する一検討” 3DIT2010-40 HI2010-88, vol. 34, pp. 55-58, 2010
- [35] 池谷健佑, 久富健介, 片山美和, 岩館祐一: “多視点ロボットカメラの方向制御手法に関する検討” 映像情報メディア学会冬季大会講演予稿集, pp. 8-5, 2010
- [36] 岩館祐一, 片山美和, 久富健介, 池谷健佑, 岡野文男: “超臨場感立体コンテンツ生成技術の基盤構築” NICT 超臨場感コミュニケーションシンポジウム, 2009
- [37] 相澤清晴, 岩館祐一, 内海章, 山崎俊彦, 折原豊, 片山美和, 富山仁博, 久富健介, 山添大丈, 韓昇龍, 李琳珊, 葛西大佑, 中川卓, 柴山悠毅, 山田健太郎: “伝統舞踊の3次元映像アーカイブ生成・圧縮・表示・活用技術” 第6回文部科学省・知的資産のための技術基盤プロジェクト展示会, pp. 17-26, 2009
- [38] 久富健介, 富山仁博, 片山美和, 岩館祐一: “視点依存型面テクスチャにおけるブレンディングに関する検討” 映像情報メディア学会年次大会講演予稿集, pp. 6-5, 2008
- [39] 久富健介, 富山仁博, 片山美和, 岩館祐一: “グラフカットを用いた多視点映像からの3次元形状復元法の検討” 映像情報メディア学会冬季大会講演予稿集, pp. 6-7, 2008
- [40] 久富健介, 富山仁博, 片山美和, 岩館祐一: “動的3次元映像におけるブレンディングを用いた視点依存型面テクスチャ手法の検討” 第11回画像の認識・理解シンポジウム論文集

(MIRU 2008), vol. IS4-37, pp. 1352-1357, 2008

- [41] 片山美和, 富山仁博, 久富健介, 岩館祐一: “3次元モデルを用いた伝統舞踊アーカイブ” 2008年アート・ドキュメンテーション学会年次大会シンポジウム, pp. 36-41, 2008
- [42] 富山仁博, 片山美和, 久富健介, 岩館祐一: “多視点映像生成技術の研究とアプリケーション” IE2007-294 PRMU2007-278, vol. 107, pp. 207-214, 2008

展示

- [43] Yuichi Iwadate, Miwa Katayama, Kimihiro Tomiyama, Kensuke Hisatomi: “3D Video System for Archiving Japanese Traditional Performing Art” 12th IEEE International Conference on Computer Vision 2009 (ICCV), pp. D-13, 2009
- [44] K. Aizawa, Y. Iwadate, A. Utsumi, T. Yamasaki, M. Katayama, K. Tomiyama, K. Hisatomi, H. Yamazoe, S.R. Han, N.S. Lee, D. Kasai, S. Nakagawa: “Archiving Performances of Japanese Traditional Dramatic Arts with a Dynamic 3D Model - 3D Video Archiving of Traditional Performing Arts -” The 3rd Pacific-Rim Symposium on Image and Video Technology (PSIVT), pp. D-I-2, 2009
- [45] 相澤清晴, 岩館祐一, 内海章, 片山美和, 山崎俊彦, 折原豊, 富山仁博, 久富健介, 山添大丈, 徐建鋒, 韓昇龍, 前田崇, 李琳珊, 葛西大佑, 中島寛, 浜崎佑樹: “伝統舞踊の3次元映像アーカイブ生成・圧縮・表示・活用技術” 第70回情報処理学会全国大会サテライトイベント「わくわく IT@あきば 2008 セッション 6」, 2008
- [46] Yuichi Iwadate, Miwa Katayama, Kimihiro Tomiyama, Kensuke Hisatomi: “3D Video System for Archiving Japanese Traditional Performing Art” The 14th International Multimedia Modeling Conference (MMM), pp. CC-2D, 2008

謝辞

本論文の作成ならびに構成する研究には、多くの方々のご支援とご指導を賜りました。

指導教官の相澤清晴教授におかれましては、2007年から文部科学省委託研究プロジェクト「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、2009年からはNICT委託研究プロジェクト「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」において共同研究をさせていただき、5年間に渡って貴重なご助言をいただきながら、協力して研究を進めさせていただきました。博士過程入学後も、研究の進め方や方向性について熱心にご指導と激励をいただき、本研究を有意義な成果に導いてくださいました。時にはご多忙な中、休日にも打合せや投稿論文へのコメントをいただきました。共同研究から長期に渡って変わらぬご厚情、心より感謝申し上げます。

同じく研究室の指導教官である山崎俊彦准教授におかれましても、文部科学省委託研究プロジェクトならびにNICT委託研究プロジェクトにおいて、共同研究させていただきました。共同研究の打合せにあたって、常に積極的に研究に対するアドバイスを多くいただき、それらのコメントは博士過程に入学するにあたっての礎となりました。博士過程入学後も変わらず熱心にご指導いただき、研究室での打合せでの多くの貴重なアドバイスから、研究に対する姿勢について多く学びました。

本論文をまとめるにあたり、論文審査委員の先生方からも多大なご指導とご鞭撻を賜りました。佐藤洋一教授におかれましては、本研究の赤外線カラーステレオカメラからのデプスマップ推定についてアドバイスをいただき、研究の位置づけや方向性について確認することができました。心より感謝申し上げます。苗村健教授におかれましては、3次元映像への応用の観点を含んだ本研究の位置づけについてご助言いただき、論文の完成度を高めることができました。厚く御礼申し上げます。大石岳史准教授におかれましては、3次元モデル生成の観点を含む本研究の特徴についてご助言いただき、本研究の新規性、有効性について確認することができました。厚く御礼申し上げます。また、池内克史名誉教授にはアドバイザーとして、研究の過程において本研究の特質についての的確なご助言をいただきました。心より感謝申し上げます。

本研究の機会を与えてくださり、ご指導、ご鞭撻を賜った日本放送協会の多くの方々へ心よ

り感謝いたします。日本放送協会元技師長の久保田啓一氏には、放送技術研究所に配属となる以前よりご指導いただき、配属後も所長時代に公私に渡って大変お世話になりました。心より感謝申し上げます。また、現所長の黒田徹氏も研究環境の整備に日頃よりご尽力いただき、本研究の機会を与えていただき、感謝いたします。

岩館祐一氏には、研究所に配属以降、長きに渡り公私に渡って、最も近いところで研究の姿勢や進め方の基本からご指導いただきました。お忙しい中でも、いつでも相談に乗っていただき、多くの経験の場を与えていただき、多くのことを学ぶことができました。厚く御礼申し上げます。三科智之氏には、日頃より研究環境の整備にご尽力いただいた上、多くの経験の場を与えていただきました。厚く御礼申し上げます。片山美和氏には、3次元映像を中心に多くのご助言をいただくとともに、実験にも多大なご協力をいただきました。心より感謝申し上げます。富山仁博氏には、同じ研究グループの先輩として、配属直後から3次元形状復元や3次元映像処理を中心に、ご指導いただき専門的な知識の基礎を教えてくださいました。厚く御礼申し上げます。池谷健佑氏、加納正規氏には同じグループとして、多くの議論や共同作業をさせていただき、また実験等さまざまなサポートをいただきました。心から感謝申し上げます。澤島康仁氏には、博士課程中、また本論文を仕上げるにあたり、貴重なご意見をいただきました。心より感謝申し上げます。

今井亨部長、菊池宏部長には、本研究を進めるにあたって、業務の配慮をいただくなど、多大なサポートをいただきました。厚く御礼申し上げます。木村恭子氏、中戸川剛氏、金子浩之氏、澤谷郁子氏にも、多大なサポートをいただきました。心より感謝申し上げます。

東京大学相澤・山崎研究室の皆様には研究室の打ち合わせの際、また予備審査や本審査にあたって、本研究へのご助言、ご意見を賜りました。また、秘書の赤羽彩子氏と松林真幸氏には事務手続きを中心にサポートしていただきました。厚く御礼申し上げます。

最後に、大学院まで学ぶ機会を与えてくれた両親、そして働きながら大学に通うということで少なからず負担がかかったにもかかわらず、応援してくれた妻と息子をはじめとした家族に感謝いたします。社会人博士課程ということで、家族のサポートなしに本論文をまとめることはできませんでした。

2015年12月11日