

視点依存の輝度変動に頑健な
マルチベースラインステレオ法

Multiple-baseline Stereo Robust against
Brightness Variation Depending on
Viewpoints

2015 年 2 月

坂 本 静 生

Shizuo Sakamoto

視点依存の輝度変動に頑健な
マルチベースラインステレオ法

Multiple-baseline Stereo Robust against
Brightness Variation Depending on
Viewpoints

2015年2月

早稲田大学大学院 基幹理工学研究科
情報理工学専攻 画像情報研究

坂本 静生

Shizuo Sakamoto

〈概要〉 異なる視点から撮影した画像をもとに三次元情報を得ることは、コンピュータビジョンの重要なテーマのひとつである。三角測量の原理に基づく三次元情報の取得には、光パターンを投射するアクティブな手法と、投射しないパッシブな手法(ステレオ法)がある。アクティブな手法は一般に光パターンの投影とともに複数回の撮影が必要で高精度な計測には適切であるが、より自然な環境下での計測ではステレオ法が重要である。これまで多くのステレオ法が研究され、二眼・三眼・マルチベースラインの各ステレオや移動カメラのシステムへ適用されてきた。一般的に三次元情報推定精度を上げるには、カメラ間距離を大きくすべきである。しかし距離を大きくすることは対応点探索を困難にする。現実には撮影対象となる物体は完全散乱反射特性ではないなど視点依存の輝度変動があり、視点間距離が大きくなるほど影響が大きい。

この論文では、まずアクティブな手法として著者らが提案する、複数の視点からの幾何拘束を利用する多眼正弦波格子位相シフト法を述べる。これによりマルチスリット光投影法などよりも高精度・高密度かつ高速な計測が実現できることを示すとともに、著者らが開発した本方式に基づく人間の顔及び頭部全周をそれぞれ計測するシステムについて説明する。

次にパッシブな手法として、視点依存の輝度変動に頑健な二つのマルチベースラインステレオ法を提案する。提案法では新しくステレオ法に近傍カメラの概念を導入し、近傍カメラ間比較を総和するふたつのコスト関数を定義する。ひとつめは視点位置に対し線形の輝度変動を許容するモデルであり、従来のマルチベースラインステレオ法を上回る精度であることを実験的に示す。しかし完全散乱仮定の下で、直接の比較対象である近傍カメラで構成するマルチベースラインステレオ法と同程度の精度しか発揮できないことを理論的に明らかにする。

これを踏まえて輝度変動許容モデルの制約を強めることにより、完全散乱仮定の下で同一のベースライン長をもつ二眼ステレオと同等な三次元情報推定精度であることを理論的に証明可能な手法を提案する。さらに視点依存の輝度変動をもつ実際の画像セットを用いることにより、二眼ステレオだけでなく従来のマルチベースラインステレオと同等あるいは優れる精度をもつことを実験的に示す。

<Summary> Acquisition of three-dimensional information through images captured from different viewpoints is one of the most important themes to be solved in computer vision. There are two categories to obtain three-dimensional information based on the triangulation principle. One is an active method that actively projects light pattern(s) onto an object, and the other is passive (or called *stereo*) method that employs only passive devices, i.e. cameras. Generally, light pattern projection and multiple-time image acquisition are necessary to the active method and then it is appropriate to be applied for highly-precise three-dimensional information. However, the passive method still has a significant role on applications under natural environments. Lots of stereo algorithms have been studied, and they have been applied to two-camera, trinocular, multiple-baseline stereos and moving-camera systems. The larger distance between cameras should basically be required for better depth accuracy. However, this often makes the correspondence issue more difficult. Brightness variation depending on viewpoints can be considered as one of the reasons in real scenes, which may be caused by non-Lambertian reflectance property of an observed object.

The paper firstly shows a sinusoidal pattern projection technique with phase shift under geometrical constraints using multiple cameras, which is proposed by the author group. This technique enables more precise, denser and faster measurement than ordinal ones, e.g., multiple-slit projection. Then, rangefinder systems for human faces and its heads, which were developed by the author group, are described respectively.

The paper also shows two kinds of novel multiple-baseline stereo algorithms robust against brightness variation depending on viewpoints. They introduce new concept of neighbor-camera and define cost functions summarizing neighbor-camera comparisons respectively. One of them allows linear brightness variation through neighbor cameras and reduced three-dimensional errors are shown in comparison with an existing multiple-baseline stereo algorithm experimentally. However, the author proves its theoretical accuracy under Lambertian assumption is comparable to a multiple-baseline stereo with a shorter baseline length equal to neighbors explained above.

Hence, the author nextly proposes the other algorithm that has the stronger constraint than previous proposal, which is theoretically comparable to a

two-camera stereo algorithm that has the same baseline length. The experimental results by using real image sets that include brightness variation depending on viewpoints indicate it can show comparable or better accuracies not only to a two-camera stereo but also to a multiple-baseline stereo.

内容

第1章 序論	6
第1節 本論文の背景	6
第1項 アクティブな三角測量の原理に基づく三次元情報推定手法	6
第2項 パッシブな三角測量の原理に基づく三次元情報推定手法	8
第2節 本論文の目的	15
第3節 本論文の構成	16
第2章 複数視点からの幾何拘束を用いる多眼正弦波格子位相シフト法	19
第1節 アクティブな手法	19
第2節 アクティブな三次元情報の取得手法	20
第1項 スポット光・スリット光投影法	20
第2項 マルチスリット光投影法	21
第3項 空間コード化パターン光投影法	22
第4項 モアレトポグラフィ	23
第5項 正弦波格子位相シフト法	24
第3節 多眼正弦波格子位相シフト法	27
第4節 顔用レンジファインダと頭部全周レンジファインダの試作	31
第3章 視点依存の輝度変動に頑健なマルチベースラインステレオ法の提案	42
第1節 ステレオ法における視点依存の輝度変動対応	42
第2節 視点依存の輝度変動に頑健なマルチベースラインステレオ法の提案	46
第3節 提案手法の実験的評価	53
第4章 視点依存の輝度変動に頑健なマルチベースライン法の改良と検証	56
第1節 完全散乱反射特性の下での理論的な RMS 評価	56
第2節 視点依存の輝度変動を含む画像を用いた RMS 誤差実験評価 その1	67
第3節 視点依存の輝度変動に頑健なマルチベースラインステレオ法の改良	72
第4節 完全散乱反射特性の下での改良版マルチベースラインステレオ法の理論的な RMS 評価	74
第5節 視点依存の輝度変動を含む画像を用いた RMS 誤差実験評価 その2	81
第5章 まとめ	92
参考文献	93
業績リスト	101
査読付き論文	101

国際発表	102
特許	104
総説	106
国内発表・講演・セミナー.....	107
その他著作	111

第1章 序論

第1節 本論文の背景

異なる視点から撮影した画像をもとに三次元情報を得ることは、コンピュータビジョンの重要なテーマのひとつである[1][2][3][4][5]。その中でも三角測量の原理に基づく手法は現在でも多くの論文が活発に出される分野であり、既知である複数の位置から計測対象位置を観測し、交点を計算することで三次元情報を取得するものである。この原理に基づく手法には、光パターンを投射するアクティブな手法と、投射しないパッシブな手法（ステレオ法）がある[5]。この二種類の手法についてそれぞれ背景を説明する。

第1項 アクティブな三角測量の原理に基づく三次元情報推定手法

アクティブな手法では光プロジェクタから容易に判別がつく光情報を投影するとともに、計測対象物体表面上に照射された光情報をカメラを用いて観測方向を同定することで三角測量を行う。ここで投影する光情報の種類によって計測の効率や計測点の密度、計測精度が大きく左右される。光情報の種類とは、すなわち光プロジェクタから計測対象物体表面上の各位置に対する投射方向が、カメラで撮影した画像あるいは画像群から推定できるようにする情報である。

この光情報としては大まかに二つの流れで研究が進められてきた。一つはスリット光の投影[6]から始まり、マルチスリット光[7]や空間コード化法[8]へ続く手法である。スリット光投影法は線状の光パターンを物体に照射し、カメラで線を検出して三角測量する処理を繰り返す手法であり、効率化のため複数本のスリット光を投影する、あるいは空間全体を複数の二値パターンでコード化する手法などへ進化してきた。しかしスリット光は完全な線に絞れるわけではないため、カメラの全ての画素単位で精密かつ稠密に計測することが困難であり、かつ計測密度を上げるのに比例して撮影回数が増加する課題がある。マルチスリット光では一度に投射するスリットの本数分だけ効率化することが可能である。しかし、スリット光の幅に起因する計測精度の限界や、画素単位での稠密な計測が困難であるオリジナルのスリット光投影法の原理的課題に加え、それぞれのスリット光を弁別しなければ正確な三次元情報は得られないためにスリット光の数は無暗に増やすことができず大幅な効率化はできない。空間コード化法は空間全体を効率的にコード化可能な光パターンを利用する方法であり、スリット光よりも撮影回数を相対的に抑えることはできるが、各コードにはある程度の幅をもたせる必要であるため計測密度の問題は解決することはできない。またこれらとは別に、投影する光パ

ターンをさらに工夫することで、画素単位での稠密な計測を可能にする手法も提案されてきてはいたが、計測精度が高いとは言えなかった[10][11]。

もう一つの流れはモアレトポグラフィである[12]。この手法には格子パターン照射法と格子パターン投影法の二種類がある。格子パターン照射法はひとつの大きな格子パターンを光プロジェクタとカメラで共用し、計測物体の直前に置く簡単な構成である。これに対し格子パターン投影法は光プロジェクタとカメラの前にそれぞれ小型の格子パターンを配置する構成であり、小型化が容易で設計や計測の自由度、柔軟性を高めることができ適用範囲も広く、主流となってきた。

この格子パターン投影型モアレトポグラフィでは、格子を通して計測対象物体へ投影された光パターンが対象物体の形状に応じて変形する。変形したパターンをカメラ側の格子を通して撮影することで、等高線情報を取り出せる。しかし、単に三次元情報の獲得を目的とするならばモアレパターンを発生させる必要はなく、さらに等高線情報から三次元情報への再構成を余儀なくさせてしまうことが逆に自動計測を困難としているとも考えることができる。そこで研究が進展するに従ってカメラ側の格子パターンを廃するとともに、感度を上げる方法である縞走査あるいは位相シフト法[13]を組み合わせることで、等高線から三次元情報への再構成処理を省きながら計測精度を向上させる手法である、正弦波格子位相シフト法が提案された[14][15][16][17]。

正弦波格子位相シフト法は計測時間が短く(最小 3 回の格子パターン投影と撮影だけでよい)画素単位での計測が可能な特徴がある。この正弦波格子位相シフト法の投射光は、計測空間全体に正弦波状の強度分布で広がるパターンであり、位相をシフトさせながら複数回投影と撮影を行う。格子パターン投影装置からみたとき、計測対象となる空間を正弦波の初期位相値でコード化することと等価である。しかもスリット光投影法や空間コード化法と異なり、位相値を空間的に徐々に変化させて投影することから、位相値を画素単位の情報だけから計算することが可能である。従って大幅な計測密度を達成しつつ、計測密度を上げるために計測回数を増やす必要がない点が特徴である。また得られる位相値は、原理的に物体の反射特性及び環境光に影響されない。

しかし正弦波は 2π の周期関数であるために、計算して得られる位相値は 2π モジュロ(2π 毎に折りたたまれる)となる。よって位相値から三次元情報を得るには、画像上に亘って位相値全体が滑らかとなるよう各画素の位相値に 2π の整数倍を適宜足し込む位相接続処理が必要である[18][19]。この位相接続処理は、干渉計の分野でさまざまな研究がなされてきているが、接続結果は計測対象の形状やノイズに大きく左右され、必ずしも安定した結果を得ることができなかった。

近年、廉価で手軽に三次元情報を得ることができる民生用 RGB-D カメラとして Microsoft 社 Kinect[20]や ASUS 社 Xtion Pro LIVE[21]などが発表されている。これらの詳細なアルゴリズムは公開されていない。しかし、いずれもジェスチャなどによるコンピュータの操作(ナチュラルユーザインタフェース)を目的としており、計測精度を犠牲にしてリアルタイム性を重んじたバランスの設計となっていることが知られている。実際、文献[22]では Kinect の計測精度評価を行ったところ、奥行方向の計測精度は 2cm 弱と報告されており、一般的なレンジファインダと比較して 2 桁程度悪いことがあることが示されている。

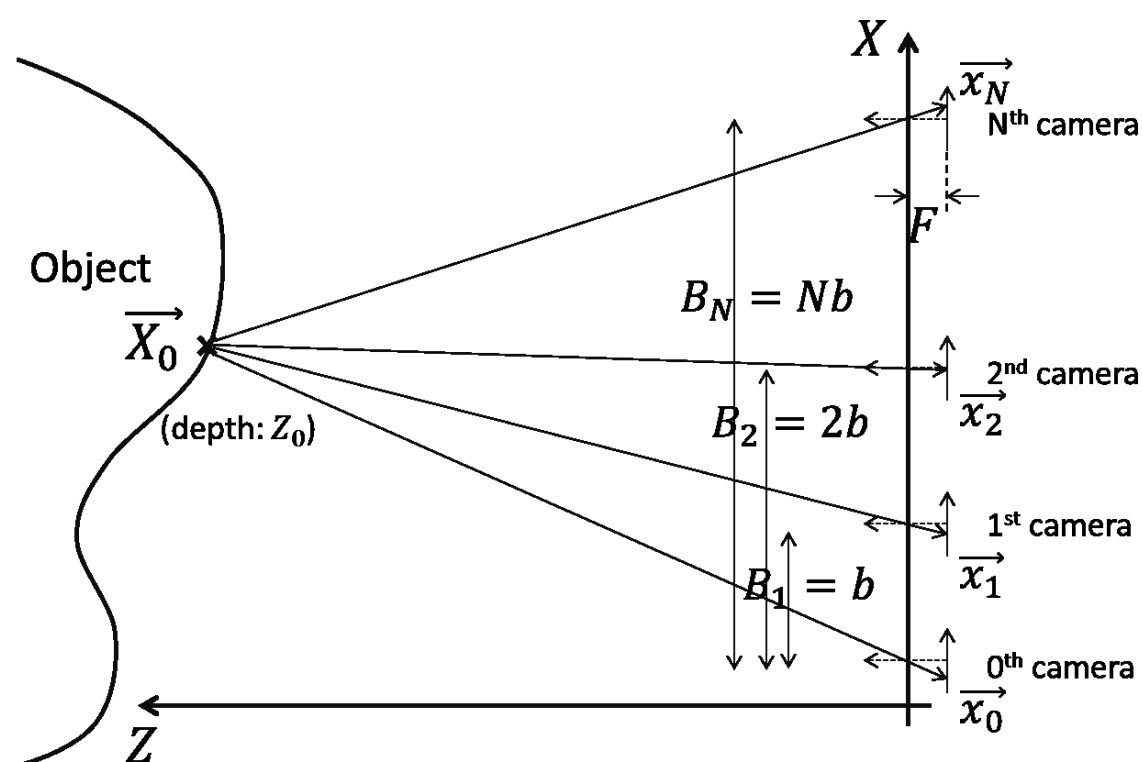


図 1 マルチベースラインステレオの設定

第 2 項 パッシブな三角測量の原理に基づく三次元情報推定手法

アクティブな手法は一般的に、計測対象へ光パターンの投影と画像撮影を複数回繰り返す必要がある。従って、例えば自然な環境下での人の活動などを計測の対象とするときには、光パターンの投射がなく、一度の撮影タイミングで撮影する画像だけから、高い精度で三次元情報を得ることができるステレオ法は非常に重要である。ステレオ法は多くの研究者により研究開発が進められるとともに、二眼・三眼・マルチベースラインの各ステレオや移動カメラのシステムへ適用されてきた[1][2][3][4]。

図 1 の通り $N + 1$ 台の共通の焦点距離 F をもつカメラが、全て平行の光軸をもち光軸と垂直な直線上に等間隔 b で並んでいるとする¹。0 番目と i 番目のカメラ間距離を B_i と表記しベースラインと呼び、図 1 の配置では $B_i = ib$ となる。各カメラで撮影される画像には x - y 座標系が定義されており、各 x ・ y 軸はすべてのカメラでそれぞれ平行とする。またベースラインとカメラの x 軸は平行とする。光軸方向に撮影対象物体があり、その物体上の観察点 \vec{X}_0 とカメラ間距離を Z_0 (奥行き) とする。観察点 \vec{X}_0 が 0 番目と i 番目のふたつのカメラで撮影された画像上に投影された座標は、ベースラインと垂直である y 軸方向は全て同一であるけれども x 軸方向で差が生じる。この投影された座標位置である \vec{x}_0 と \vec{x}_i の x 軸座標値である x_0 と x_i の差を視差と呼び $d_i(x_0)$ とする。これらパラメータ間には式 (1) の関係があることが知られている。

$$\begin{aligned} d_i(x_0) &= x_0 - x_i \\ &= \frac{B_i F}{Z_0} \\ &= \frac{ibF}{Z_0} \end{aligned} \tag{1}$$

ステレオ法の目的は画像中の画素を処理し投影点間の対応を自動的に探して視差を求めることである。式 (1) から明らかな通り視差 $d_i(x_0)$ はベースライン $B_i = ib$ に比例する。視差が有限の画素位置で量子化されていることを考慮すると、奥行き推定精度を上げるには i を大きくする、即ち 0 番目のカメラからできるだけ遠いカメラを選択しベースラインを長くするべきである。

式 (1) では奥行き Z_0 が分母に現れるため、奥行きの計測精度を分析するにはその逆数の導入が便利である [23][24]。

$$d_i(x_0) = ibF\zeta_0(x_0) \tag{2}$$

ここで $\zeta_0(x_0) = 1/Z_0$ であり、視差 $d_i(x_0)$ と比例する値であることからステレオ法の結果との関係が直接的に説明することが可能である。文献 [23][24] では、輝度値を直接比較するステレオ法を分析対象として、奥行きの逆数 $\zeta_0(x_0)$ の推定精度が画素に重畳するノイズに比例し、ベースラインの距離に反比例することが理論的に示された。

¹ カメラ台数が一般的な N ではなく $N + 1$ としたのは、後の理論的な評価の際に数式がやや簡略な形式となることが理由である。

ステレオ法の歴史的文献のひとつである[25]では、画像中から安定に検出できるコーナー状の輝度変動をもつ地点を見つけて対応付けを行う手法が提案されている。同様な特徴点検出は[26][27]などでも提案されており、当初のステレオ法はこのような画像中に疎に存在する、確実に対応付けが可能な特徴点をまず見つけた後、特徴的な点の対応付けを行う手法から構成されていた。

このような特徴点はコーナー状の地点だけではなく、ブロップ状の領域にも適用できる SIFT (Shift Invariant Feature Transform) や GLOH (Gradient Location-Orientation Histogram) などが提案されており[28][29]、さらに一般的な画像データベースから学習により精度を高めた特徴[30]などもある。また特徴点の対応付けとカメラ位置や向きなどを最適化するバンドル調整[31]の手法があり、大量の不確かなカメラ位置から撮影された画像群から 3D シーンを復元する問題などに適用されている[32][33][34][35]。但しよい初期値をあらかじめ与えないと局所解に陥ることが知られていることから、文献[36]ではよい推定ができる画像の部分集合を選択して最適化を進めた後に、他の画像を追加する手法を提案している。

これらに対して、画素ごとの密な対応付けを行う手法があり近年研究が盛んに行われている。疎な対応付けでは特徴点以外の場所における三次元情報は基本的に補間に頼らざるを得ないのに対し、密な対応付けでは何らかの画像特徴あるいは滑らかさなどの拘束条件を活用して画素毎に三次元情報を推定する。ステレオ法に関する文献は膨大であるが、以下では密なステレオ法をおおまかに次のステップに分類して概観する：

- (ア) 画素/領域類似度計算
- (イ) 最適化と視差値計算

また視差値を画素単位以下に改善するポスト処理を適用することがある。

ステレオ法では、(イ)までで通常画素単位の離散的な視差を求める。得られた離散的な視差に対して曲面を当てはめることなどで画素単位以下の精度で奥行きを推定することができる[37]。しかし当てはめる小領域が、同一の物体に属する表面に正しく対応していなければならないなどの問題が知られている。

(ア) 画素/領域類似度計算

最も単純な画素/領域の類似度は差分の絶対値や二乗値である。単一の画素値を比較する場合もあるが、多くは注目画素位置を中心とした小領域におけ

る積算値である SAD (Summation of Absolute Differences)[38]や SSD (Summation of Squared Differences)[23]を利用することも多い。

例えば図 1 において、0番目とN番目のカメラにおける \vec{x}_0 地点と \vec{x}_N 地点の輝度値、 $i_0(\vec{x}_0)$ と $i_N(\vec{x}_N)$ の類似度を $\rho(i_0(\vec{x}_0) - i_N(\vec{x}_N))$ とする。計測対象物体が完全散乱特性である場合には $i_0(\vec{x}_0)$ と $i_N(\vec{x}_N)$ は一致すべきである。しかし通常画像センサに熱雑音が存在していることから正規分布に従うランダムなノイズが重畳すると仮定するのが自然である。このとき類似度が最大となる地点を最尤推定で求めるとすると、 $\rho()$ は二次関数となる。

一方で現実の画像センサではショットノイズといった大きなノイズが加わることがある。またステレオ法ではオクルージョンにより対応付けができない、即ち類似度が高い画素がない場合に二次関数を適用すると、類似度の大きな低下を招いて推定結果が劣化することが知られている。よって輝度差が大きいときに類似度が二次関数よりも緩やかに低下する関数にすることで改善を可能であることが知られている。代表例として差の絶対値が用いられる。他に輝度差が小さい場合は二次関数だがある差以上の場合に一定値にクリップする、増加量を緩やかにする関数(例えば Geman-McClure 関数)などが提案・検討されてきた[39]。

上記の類似度は、注目画素 \vec{x}_0 地点と \vec{x}_N 地点付近の小領域において総和することで、小領域の類似度としてそのまま適用できる。例えば \vec{x}_0 及び \vec{x}_N を中心とした ± 1 画素をそれぞれの小領域とすると、類似度は $\sum_{\vec{\varepsilon}=(-1,-1),(-1,0),(-1,1),\dots,(1,1)} \rho(i_0(\vec{x}_0 + \vec{\varepsilon}) - i_N(\vec{x}_N + \vec{\varepsilon}))$ のように計算できる。ここで $\rho()$ が二次関数の場合には SSD、絶対値の場合 SAD といった代表的な小領域の類似度になる。他に異なる形式の代表的な小領域間類似度として、正規化相関(NCC)が挙げられる[25]。SSD や SAD と異なり、正規化相関は小領域の輝度値が全体に係数倍されていても不変であることが知られている。輝度値にオフセットが存在する場合にはいずれの類似度も影響を受けるが、小領域の平均値を差し引いた ZSAD、ZSAD、ZNCC により回避することができる。

小領域の形状としては注目画素を中心とした矩形とするだけでなく、正規分布関数による重みづけを行う手法[40]、異なる画素位置を起点とする複数の小領域を利用する手法[41]、可変サイズの小領域を利用する手法[42]などのバリエーションがある。

また小領域を対象として抽出した特徴を利用する手法もある。小領域の輝度勾配を用いる手法(一次微分は[43]、二次微分は[45])、正規分布関数の0次～三次微分までを組み合わせて用いる手法[46]などが知られている。しかしこれらの手法で得られる画像特徴は原画像に比較してぼけており、各画素位置で正確なディスパリティを得ることが難しくなる。この課題に対して ZSAD

などのような小領域の平均値との差分ではなく、小領域に対するバイラテラルフィルタで得られた画素値との差分を用いる手法がある[47]。同様に小領域から得られる特徴としては、順序尺度[48]、ランク変換やセンサス変換[49]、輝度分布に基づく相互情報量[50]なども挙げられる。

以上の特徴量について網羅的に調査した文献[51]では、輝度に関する様々な変動を付与した画像や、照明条件を大きく変更した画像を用いて評価を行って、センサス変換が最も頑健であったと報告した。特に、カメラのゲインやバイアスの違い、レンズによる周辺減光などの厳密な校正[52][53]ができない場合における有効性を示したと言える。しかしいずれの特徴も、何らかの意味で輝度分布がカメラによらず安定であること、また小領域がカメラに正対する平板とみなせること、あるいはベースラインが小さくカメラから小領域の画像上での形状がほぼ不変であることを仮定している。文献[54]では、カメラから観測する小領域の形状が変化すると、小領域をもとにした比較が困難となる指摘がなされている。

カメラ特性の違いについては前処理として、ヒストグラム平坦化[55]や相互情報量に基づく手法[56]で画像全体でのゲインやバイアスを補正する方法が提案されている。またカメラ特性に限らず画像上で滑らかに変化する照明条件他を補正する手法として、0番目と N 番目のカメラにおける x_0 地点と x_N 地点の輝度値、 $i_0(\vec{x}_0)$ と $i_N(\vec{x}_N)$ の類似度を $\rho(\kappa_{0,N}(\vec{x}_0) \cdot i_0(\vec{x}_0) - i_N(\vec{x}_N))$ とし、画像ペアの画素ごとに定めるべきパラメータ $\kappa_{0,N}(\vec{x}_0)$ を、近傍パラメータとの差分が小さくなるような正則条件を付与して、視差と同時に求める手法が提案されている[57][58]。但し、パラメータ $\kappa_{0,N}(\vec{x}_0)$ の値を自由に決めることができると、こういった画素ペアであっても類似度を最も高くさせることができるため、本パラメータに対する正則条件は非常に強くしなければならない。即ち一般的に複雑な形状である観測対象物体の非完全散乱成分を記述・補正することは困難である。

この他には、観測対象物体が空間的に高い周波数のテクスチャから構成される場合に、サンプリング位置により画素の輝度値が異なってしまう問題に対処するため、線形補間した輝度値との比較を行う手法が提案されている[59]。

マルチベースラインステレオ法はステレオ法の拡張として定義される。ステレオ法での類似度を $\rho(i_0(\vec{x}_0) - i_N(\vec{x}_N))$ とすると、マルチベースラインステレオ法では一般 $\sum_{i=1}^N \rho(i_0(\vec{x}_0) - i_i(\vec{x}_i))$ となる。ステレオ法と同様に小領域の類似度である SSD を利用する手法[22]などもある。マルチベースラインステレオ法固有のオクルージョン対策としては、中央値である $\text{median}[\rho(i_0(\vec{x}_0) - i_1(\vec{x}_1)), \rho(i_0(\vec{x}_0) - i_2(\vec{x}_2)), \dots, \rho(i_0(\vec{x}_0) - i_N(\vec{x}_N))]$ を類似度とする手法や、 $i_i(\vec{x}_i)$

($i = 0, \dots, N$)の輝度値分布がもつエントロピーを類似度とする手法[60]があるが、いずれも全輝度値は同一であることを仮定している。

一方で非完全散乱に対する研究は多くない。文献[61]では非完全散乱の影響を避けるカメラと物体の配置について議論したが、画素/領域類似度計算に関しては何ら提案を行っていない。陰影からの形状復元に非完全散乱成分の検出を提案する手法[62][63]やその複数画像への応用[64]があるが、一般に陰影からの形状復元は局所解に陥りやすく、現実の画像を対象とするには取り扱いが難しい。文献[65]では、まず信頼できる地点に対しステレオ法を適用した後、陰影からの形状復元とステレオ法を繰り返す手法が提案されているが、事前に物体の反射特性を計測しておく必要がある。他には複数種類の照明条件下で撮影されたステレオ画像から画素毎に輝度値の比を計算した後にステレオ法を適用する手法[66][67]も提案されているが、もはやパッシブな手法ではなく、複数回に亘って撮影することが必要となる。

(イ) 最適化と視差値計算

続く最適化と視差値計算では、(ア)で算出した類似度が最も高い対応画素または視差を求める手法や、何らかの意味で計測対象物体の表面の滑らかさ仮定を用いて最適化を行う手法がある。単純な類似度計算を用いる場合には単純な前者の手法ではよい視差が得られないことが多いため、小領域の類似度を利用するマルチベースラインステレオ法により、複数の画像の輝度情報を利用することで誤対応の発生を抑制する手法が提案されている[23]。

良好な視差を推定するために、近年では何らかの大域的最適化が行われる。図1において、観測対象物体上の観察点 \vec{X}_0 が0番目と N 番目のふたつのカメラで撮影された画像上の投影座標位置間の視差 d_N を求めるとき、画像全体で式(3)を最小化する視差のセット、あるいは視差マップ $\{d_N\}$ を求めることとする[68][69]。

$$\Phi^{\text{total}}(\{d_N\}) = \Phi^{\text{data}}(\{d_N\}) + \lambda \Phi^{\text{smoothness}}(\{d_N\}) \quad (3)$$

第一項 $\Phi^{\text{data}}(\{d_N\})$ は視差マップ $\{d\}$ が与えられたときの画像間マッチングエネルギーであり、ステップ(a)により計算された類似度を全画素に亘って加算したものである。また第二項 $\Phi^{\text{smoothness}}(\{d_N\})$ は滑らかさエネルギーであり、画像上における近傍画素との視差の差分に応じたエネルギーを全画素に亘って加算したものである。 λ は両エネルギーを按分するパラメータである。

式(3)の最小化問題は二次元画像上における最適化であり、一般的に NP 困難であることが知られている。しかし視差が生じる座標方向だけに限定した

一次元的な大域的最適解であれば、動的計画法により得られることが知られている。この動的計画法に基づき画素毎に密な視差推定を行う手法としては、文献[55][70][71]などがある。文献[55][71]では動的計画法は次元の最適解を求める手法であるが、直前に推定した $y-1$ 座標値の視差との差が小さくなるようにすることで、二次元的に滑らかとする工夫も取り込んでいる。動的計画法による視差推定結果は、次に述べる最適化には精度には及ばないが処理は高速であることが知られている。

様々なアルゴリズムにより、式(3)の解を得ることができる。物性物理学のアナロジーである古典的な焼きなまし法[72]は、現実的な処理時間にはならないけれども十分ゆっくりと温度を下げることにより大域的な最適解を得ることができることを示した。より現実的な処理時間となるように平均場近似を用いる焼きなまし法[73]や、非常に単純な各変数を順に更新する反復条件付モード[74]など、局所的な最適解を得る手法も利用されてきた。

これらの物理学から導かれた最適化アルゴリズムは、より良好な解が得られる信念伝播法やグラフカットなどに代表されるアルゴリズムによって、近年取って代わられてきている。

信念伝播法はベイジアンネットワークによる推論問題を解く手法として提案された[75]。ループあり信念伝播法により、循環ループが存在する場合に拡張されたことにより、ステレオ法などに適用することが可能となった[76][77]。文献[78]では画像上の各行・各列でそれぞれメッセージを伝播させることで、高速な伝播を可能とする手法を提案した。また文献[79]で提案されたツリー再重みメッセージパスや、この手法を逐次拡張した手法[68][80]なども提案されている。信念伝播法は次数が大きくなるにつれて効率が大きく悪化することが知られており、次数に比例するオーダーに削減する手法[81]などの研究が進められている。

グラフカットとは、最小切断あるいは最大フロー問題に対する解法を指す。二値の場合の多項式時間で大域最適化を行う手法[82]として提案されたのち、ステレオ法にも適用されるようになった[83][84][85][86]。ステレオ法では多値である視差の最適化を行わなければならないため、 α - β 交換あるいは α 拡張と呼ばれる二値最適化を繰り返し適用する手法[84]が提案された。 α - β 交換とは、まず二つのラベル α 及び β (ステレオ法の場合は視差)を選択して二値最適化を行うことを繰り返す手法である。また α 拡張とはひとつのラベル α を選択して、現在のラベルのままとするか α をとるか二値最適化を行うことを繰り返す手法である。二値最適化を行う手法として、より一般的なエネルギー関数で近似解が得られる手法[87][88]などを始め、効率的なアルゴリズムが開発され続けている。

第2節 本論文の目的

本論文では、著者らはまずアクティブな手法として、複数の視点からの幾何拘束を利用する多眼正弦波格子位相シフト法を提案した。正弦波格子位相シフト法において、同一位相値をもつ投射光パターンをそれぞれ仮想的なスリット光パターンと考え、各画素位置における 2π モジュロな位相値それぞれに対応する複数の三次元座標位置を求めた後、もうひとつのカメラへ投影した画素の位相値を検証することにより絶対位相値を決定する。これにより画像に稠密な絶対位相値が得られることとなり、位相接続処理の正確性に大きく頼らない結果を得ることができる。この結果、正弦波格子位相シフト法が本来保持していた、一般的なマルチスリット光投影法などよりも高精度・高密度かつ高速な計測が容易に実現可能となる。

この提案した多眼正弦波格子位相シフト法を、著者らは人間の顔及び頭部全周をそれぞれ計測するシステムへと適用した[89][90]。両システムともに、画素単位での稠密な計測を可能とするとともに、顔計測システムでは撮影時間0.6秒と奥行き計測精度0.2mm以下を、頭部全周計測システムでは撮影時間1秒と奥行き計測精度0.5mm以下をそれぞれ達成した。

しかしアクティブな手法により高精度に三次元情報を推定するには、計測にある程度の時間が必要となる。そこで基本的に一回の撮影タイミングで得られる画像から推定が可能なステレオ法が期待される。だが、例えば動いている人間を精度よく計測する用途を想定したときに必要な技術が十分検討されてきたとは言い難い。実際、人間の肌は真皮と表皮の層構造をなしていることから、完全散乱から外れる反射特性をもっていることがよく知られている[91][92]。他の物体であっても、短いベースラインでは問題にならないが高い奥行き推定精度を求めてベースラインを長くすると、完全散乱から外れる、あるいは各カメラからのぞむ物体表面が各画像上で同一形状でない[54]などの問題が生じることから、ステップ(ア)の従来研究では対応できず、ステップ(イ)では対応の範囲外であることが明らかである。

よって本論文ではマルチベースラインステレオ法をベースとし、近傍カメラの概念を初めて導入することによって非完全散乱反射特性を含む、視点依存で輝度に変動する場合に対応できるアルゴリズムを提案する[93][94]。文献[93]では、ある視差に対応する各カメラの画素値を全て直接比較するのではなく、近傍カメラ毎に該近傍内の平均値と比較する手法を提案する。これにより近傍内における視点依存の輝度変動が線形近似可能な場合に全く影響がないようアルゴリズムを構成することが可能になる。同時に、大きな視点依存

の輝度変動を含むステレオ画像データセットに提案アルゴリズムを適用し、従来手法に比較して良い奥行き推定結果が得られることを示す。

しかし、文献[93]では実験的に奥行き推定精度がよかったことを示しただけであった。そこで文献[94]では、文献[23][24]に準じる理論的な三次元情報の推定精度評価を行うことで、文献[93]の提案手法では近傍カメラと同一のベースラインをもつステレオ法と同じオーダーの奥行き推定精度しかもたないことを示す。また実験的な評価と分析を加えることで、当初期待していた精度が得られていないことも実データにより示すことで、近傍カメラ内での線形の変動を許容するモデルではベースラインに応じた三次元情報推定精度が得られないことを明らかにする。

そこで近傍カメラ内だけでの直接比較という枠組を保持しながら、線形の輝度変動に対してペナルティが課されるようアルゴリズムを構成することで、最大ベースラインをもつ二眼ステレオ法と同じオーダーの奥行き推定精度をもたせることができることを示す。また、大きな視点依存の輝度変動を含むステレオ画像データセットにこの改良したアルゴリズムを適用し、従来の二眼ステレオだけでなくマルチベースラインステレオ法に比較しても同等あるいは良い奥行き推定結果が得られることを示す。

第3節 本論文の構成

第2章では光パターンを投影するアクティブな三次元情報取得方法として、新しい多眼正弦波格子位相シフト法を提案するとともに、人間の顔及び頭部全周をそれぞれ計測対象とするレンジファインダについて述べる。第1節では三角測量の原理に基づく画像による三次元情報計測について、光パターンを投影するアクティブな手法と、投影しないパッシブな手法であるステレオ法を説明した後、第2節でアクティブな手法について概観する。第3節ではアクティブな手法のひとつである、投射する光パターンを正弦波で輝度変調するとともにその位相値によって投射角度がわかるようエンコードする方法をもとにして、著者らが提案した多眼正弦波格子位相シフト法を述べる[89]。

正弦波による輝度変調方式では画像上で観測される位相値が 2π モジュロであるために、絶対位相値へ変換しなければ三次元情報を求めることができない。従来はノイズや計測対象物体の形状によって動作が不安定となる位相接続処理に結果が依存していた。著者らは複数位置に配置された光パターン投影装置及びカメラの幾何学的な拘束条件を利用することにより、安定に絶対位相が決定できることを示す。本提案手法によりトレード・オフの関係にある、短い画像撮影時間と高精度かつ高密度な計測の両方を高いレベルで両立することが可能となる。

第 4 節において著者らが構築した人間の顔面を 0.6 秒の撮影時間・0.2 mm の奥行き推定精度で計測するシステム[89]及び、より多くの光プロジェクタ及びカメラを組み合わせ、人間の頭部全周を 1 秒の撮影時間・0.5 mm 奥行き推定精度で計測するシステムについて説明する[90]。

第 3 章ではパッシブな手法であるステレオ法の検討を進め、新しいマルチベースラインステレオ法を提案する[93]。第 1 節では単一の校正カメラを移動させながら撮影した、正しい視差が合わせて計測された公開ステレオ画像データセットを検証する。この、原論文[93]時点では存在しなかった Middulebury ステレオデータ[95][96]を利用することにより、ベースラインが短い場合には対応点間の輝度値変動は小さいものの、長くなると輝度変動が大きくなることを示し、人間の顔でないものであってもベースラインが長くなると完全散乱から外れていく実例を説明する。

これを踏まえ、第 2 節で著者はマルチベースラインステレオ法に近傍カメラの概念を新しく導入することで、視点依存の輝度変動に頑健なアルゴリズムを提案する[93]。このアルゴリズムでは対応点探索の際あらかじめ定義する近傍カメラ毎に、近傍中央のカメラの対応画素の輝度値を該近傍内の対応画素の輝度平均値とだけ直接比較する。これにより近傍内における輝度変動がカメラ位置に線形で近似できる場合に悪影響がないよう構成する。第 3 節では、第 1 節で説明した画像データセットを用いて評価実験を行う。対照実験として完全散乱反射特性を仮定する、従来の全カメラの画素値を直接比較するマルチベースラインステレオの例としてを説明するとともに、大きな輝度変動が存在する公開ステレオ画像データセットに適用する。各ステレオ法をできるだけ客観的に比較評価するために、各ステレオ法に複数の最適化手法を組み合わせ、さらに正則化パラメータを網羅的に探索して得た最良の RMS 誤差により比較評価し、優位性を示す。

第 4 章では第 3 章の手法を改良した、さらに新しいアルゴリズムを提案する[94]。まず第 1 節において従来の二眼ステレオ法及びマルチベースラインステレオ法とともに、第 3 章で提案したアルゴリズムに対して完全散乱反射特性を仮定した理論的な三次元情報推定精度評価を行う。これにより第 3 章提案のアルゴリズムが、近傍カメラがもつベースライン長に比例する推定視差の RMS 誤差しか得られないことを示す。第 2 節ではこれら三つのステレオ法に対して別の実験データによって評価実験を行うことで、提案アルゴリズムが従来の二眼ステレオに及ばないことを示し、近傍カメラ内での線形変動の許容は、緩和しすぎた条件であったことを明らかにする。本結果に基づき第 3 節では、あるカメラの画素輝度値を比較する際、当該カメラを中央とする近傍内対応画素の輝度平均値とともに、その前後の近傍に属するカメラ

の対応画素の輝度平均値とも比較する、ふたつめの改良したアルゴリズムを提案する。この改良したアルゴリズムは近傍内における線形な輝度変動はもはや許容しないものの、依然として遠方のカメラとは輝度を直接比較しないことから、第3章で提案のアルゴリズムと従来のマルチベースラインステレオとの中間的な手法にあたる。第4節では第1節と同様に完全散乱反射特性の仮定のもとで理論的な精度評価を行い、提案した改良アルゴリズムは、最大のベースライン長をもつ従来の二眼ステレオ法と同等の **RMS** 誤差であることを示す。さらに最大ベースライン長と近傍カメラの台数を固定しながら全カメラ数を増やしていくと、最大ベースライン長の二眼ステレオ法の **RMS** 誤差を途中で逆転することを示す。この場合全カメラ台数増加は近傍カメラのベースライン長を小さくすることを意味しており、即ち視点依存の輝度変動にはその大きさに応じてカメラ台数を増やすことで影響を削減できることを示唆する。第5節では提案した改良アルゴリズムの実験的評価を行い、完全散乱を仮定した理論的な評価では同等程度であった従来の二眼ステレオ、及び劣っていた従来のマルチベースラインステレオに対して、テクスチャが比較的豊富ではあるが輝度変動が存在する実画像に適用したところ、前者には優位な結果が、後者にはほぼ同等な結果がそれぞれ得られることを示す。またテクスチャに乏しく輝度変動が大きい物体では、従来のマルチベースラインステレオに比較して半分程度の **RMS** 誤差と大きく改善したことを示す。最後に第5章では、本研究の成果と意義、今後の課題をまとめる。

第2章 複数視点からの幾何拘束を用いる多眼正弦波格子位相シフト法

第1節 アクティブな手法

三角測量の原理は一般的な距離測定に古くから利用されてきた。計測対象物体を既知の異なる位置から目測し、各々の視線方向を定めることにより、対象までの距離が容易に計算できる。図2にこの説明図を記す。互いに位置関係がわかっている地点 x_0 及び x_N から点 \vec{X}_0 が、それぞれ視線 0 及び N の方向に観測されたとき、その交点位置として点 \vec{X}_0 が得られる。

様々な物体を対象として安定かつ精度よく計測することを目的とし、片方の視点位置に何らかの光等を物体に投射する光プロジェクタを配置したアクティブな手法がある。第2節ではアクティブな手法について、文献[5]に基づき投影される光のパターンにより分類して概観する。続いて第3節において、著者らが提案した多眼正弦波格子位相シフト法[89]を説明した後、第4節で提案手法に基づいた人間の顔[89]及び頭部全周を対象とする三次元情報取得装置[90]について述べる。

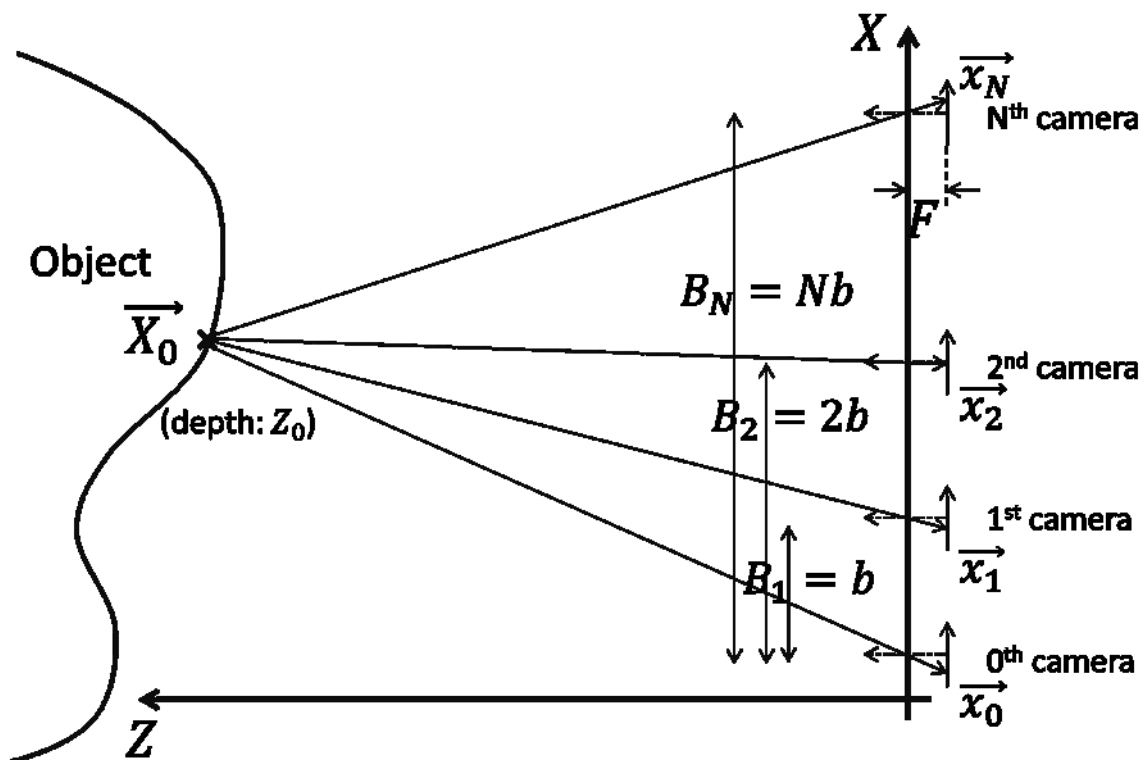


図2 三角測量の原理に基づく三次元情報の取得(図1の再掲)

第2節 アクティブな三次元情報の取得手法

本節ではアクティブな三次元情報の取得方法について、文献[5]に基づき投影される光のパターンにより分類して概観する。

第1項 スポット光・スリット光投影法

スポット光投影法は、レーザービームを対象物体へと投影してできる輝いたスポットを、異なった角度からカメラで捉えてスポットの三次元位置を求める手法であり、三角測量の原理に最も忠実なアクティブレンジファインダである。スポット光を走査ミラーなどによって計測対象全面に亘り走査する必要がある、撮影時間は非常に長くなる。

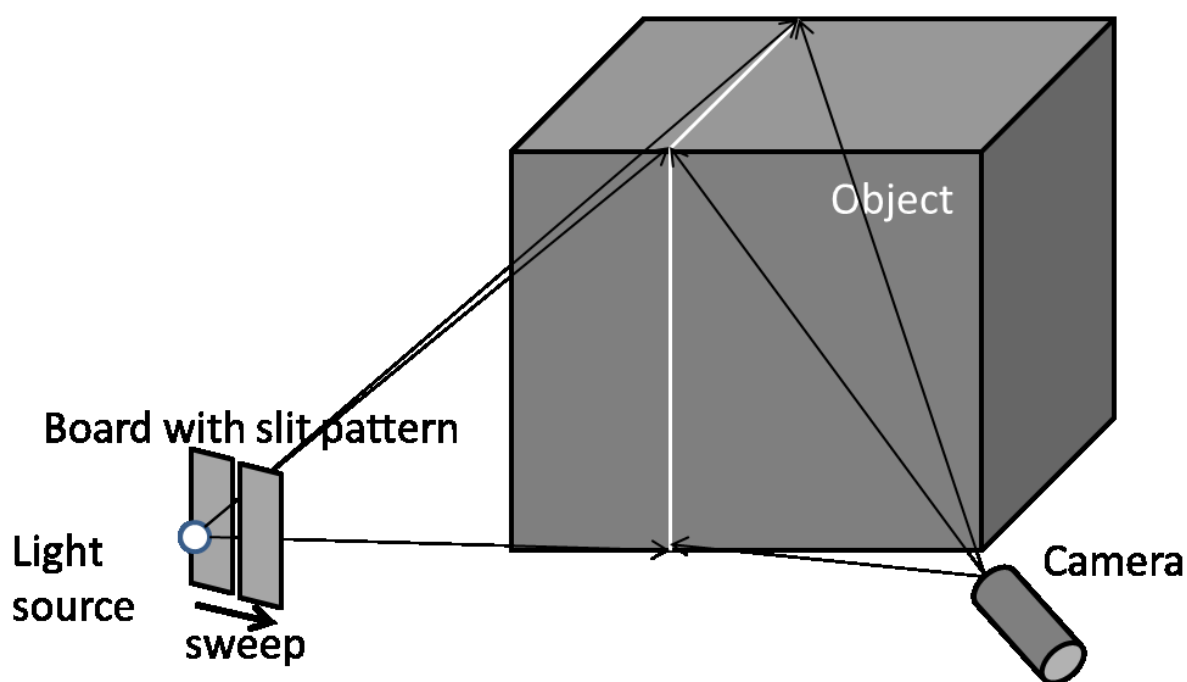


図3 スリット光投影法説明図

スリット光投影法は、別名光切断法とも呼ばれる最も知られた三次元画像撮影法である[6]。スポット光投影法に比較して、一回の撮像で一本のスリット光が照射された帯状の地点の三次元情報を得ることができる。原理図を図3に示す。計測時間はスポット光投影法に比べて一ライン分の走査時間を節約することができるが、空間分解能に比例してスリット光投射と撮影を繰り返す必要がある、計測時間を長引かせる原因となる。高速な計測には空間解像度を下げるか、特殊なセンサを活用するなどの工夫が必要である。またスポット光・スリット光は理想的な点や線状にはならず広がるため、空間解像度には限界がある。光源としてレーザーを用いるのが普通であるが、レーザ

一は位相が揃っているために、スペックルなどが発生して計測誤差が生じることがある。

第2項 マルチスリット光投影法

スリット光投影法で課題である長い撮影時間を、複数のスリット光を一度に投影することで、撮影時間をスリット光の数で除算しただけ短縮するのがマルチスリット光投影法である。しかし一般に、画像上に複数本のスリットが観測された場合、それぞれを見分けることができず三次元位置を一意に決定することができない。

ひとつの解決方法として、カメラをもう一台併用する手法を図4に示す[7]。

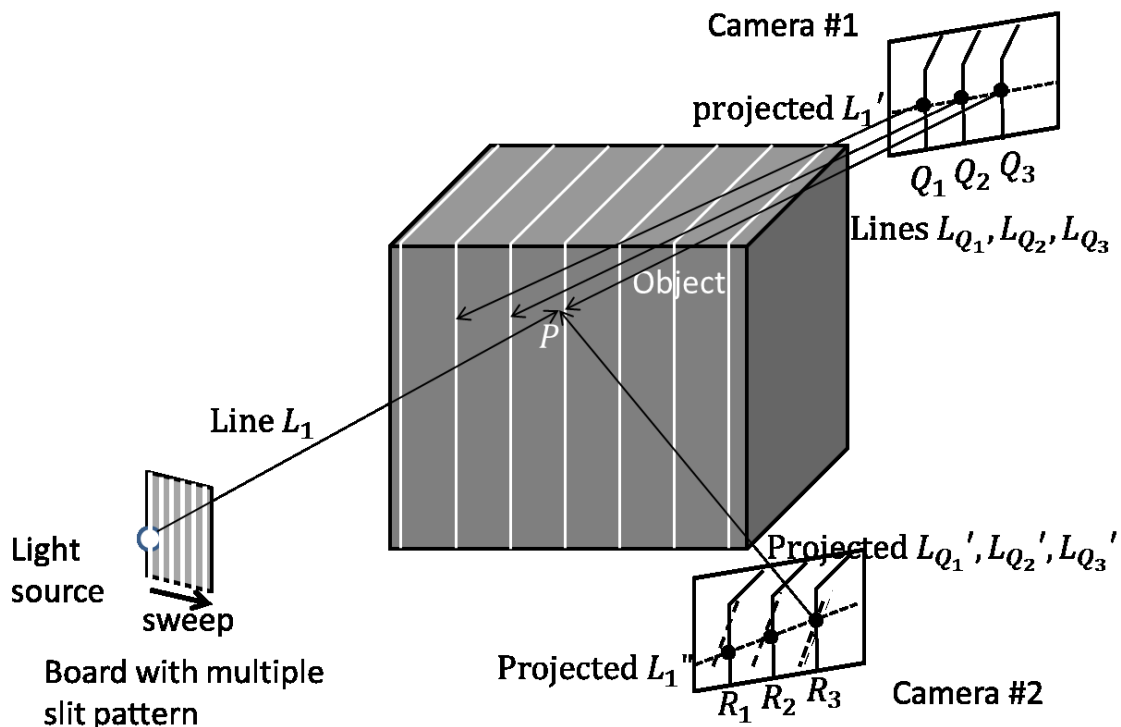


図4 マルチスリット法と三眼視法の組み合わせによる曖昧性除去

この手法では、マルチスリット光源から投射されたあるスリット像上にある点 P が、三次元空間上のどこに位置するか、カメラ1で撮影された画像からまず候補を算出する。光源から点 P へ伸ばした直線 L_1 はカメラ1だけを用いる場合、その画像上では投影された直線 L_1' として観測されており、点 P の候補はスリット像との交点である点 Q_1 から点 Q_3 までのいずれかであるとしかわからず、一意に定めることができない。そこで直線 L_1 と、カメラ1から点 Q_1 から点 Q_3 を結ぶ直線 L_{Q_1} から直線 L_{Q_3} を、それぞれカメラ2で撮影された画像上に投影し

なおすことで、直線 L_1'' 及び直線 L_{Q_1}' から直線 L_{Q_3}' を求める。カメラ2の画像上では直線 L_1'' とスリット像の交点として、点 P の候補が点 R_1 から点 R_3 までのいずれかとして得られるが、カメラ1に基づく候補である点 Q_1 から点 Q_3 に対応する直線 L_{Q_1}' から L_{Q_3}' と合致するのは、直線 L_{Q_3}' と交わる点 R_3 のみであると判定でき、点 P の正しい座標点を定めることができる。

但し対応付けが一意に決定可能であるためには、スリット光のピッチが十分粗い、また計測対象物体の形状が滑らかであるという拘束がつくことに注意が必要である。特にピッチを粗くすることは、計測密度を保ったまま計測時間を長くする、あるいは計測密度を低めて計測時間は短く抑えることに繋がることになる。

第3項 空間コード化パターン光投影法

投影光のパターンを空間的に変調することで、スリット光投影法よりも少ない投影回数で三次元画像を得る方法が提案されている。

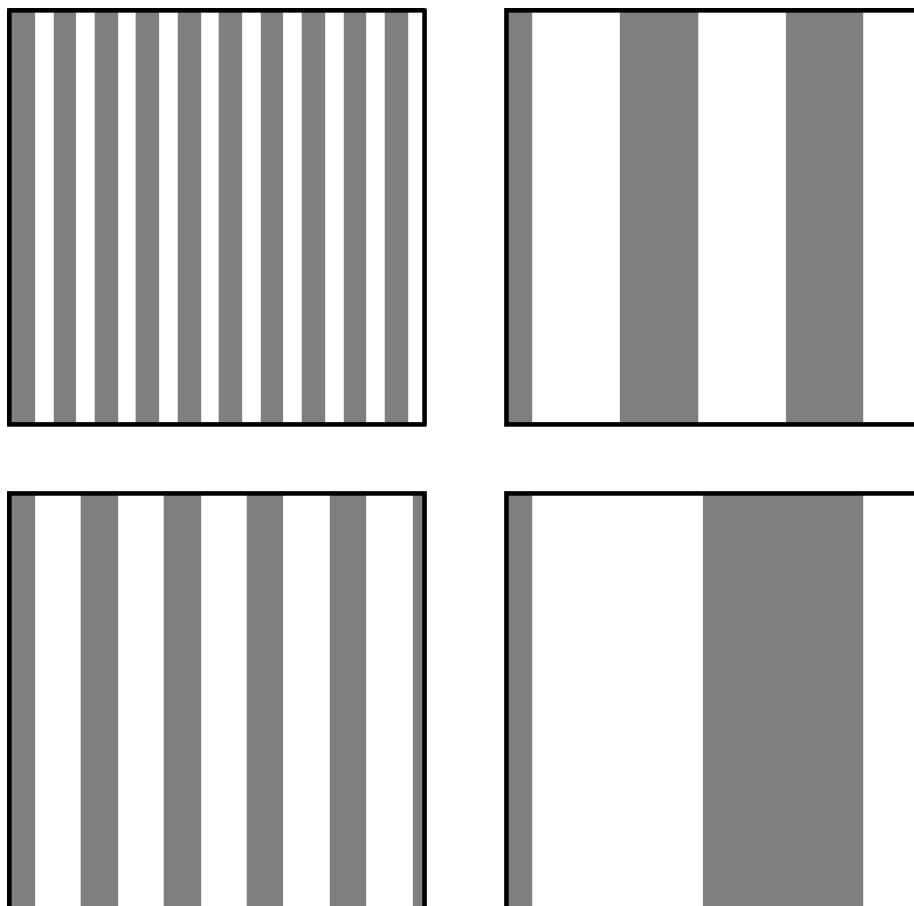


図5 多重コード化パターン光投影法で用いられる光パターン例

単板コード化法では、一種類の投影パターンにより空間をコード化する。

パターンとして、色彩や明暗の情報、あるいはスリットの幅や長さを変化させる。例えば文献[8]では、黒いスリットで分割された4色から構成されるスリット群を、色の順が繰り返しにならないよう配置することにより、光源からの投影方向が一意に定まるよう工夫している。。

多重コード化パターン光投影法は、複数のコード化パターンを投影することで空間をコード化する[9]。図3におけるスリット光投影法におけるスリット付ボードの代わりに、図5に示す構造の複数ボードに順次入れ替えることで、投影パターンの輝度値を空間的だけでなく時間的にも変調させるのが特徴である。しかし、この図からわかるように必要なボードの枚数は比較的多くなるとともに、空間解像度にも限界がある。

一方、画素単位の高い空間解像度をもつものとしては、レインボーレンジファインダ[10]や濃度傾斜パターン光投影法[11]などの、空間的に連続的な分布をなすパターンを用いる手法がある。しかし計測対象物体の反射率によりS/N比が低下して計測精度が劣化する課題があることが知られている。

第4項 モアレトポグラフィ

古くから種々の手法が提案されているモアレトポグラフィ[12]には格子パターン生成法に応じて、大きく格子パターン照射法と格子パターン投影法とがある。

図6左に示す格子パターン照射型モアレトポグラフィは、一つの格子パターンを光投射用及び撮影用に共用する方法である。光源から投射された光が共用の格子を通過した後、反射光がうまく格子を潜り抜けてカメラに到着するのはある奥行きを満たすときに限られ、等高線のパターン即ちモアレパターンが画像上に現れる。格子パターン照射型モアレトポグラフィは、比較的簡単な光学系で高い感度のモアレパターンを得ることができる。しかし、物体の直前に大きな格子マスクを置く必要があるために観測対象が制限される問題がある。図6右に示す格子パターン投影型モアレトポグラフィは、光照射用と撮影用の格子をそれぞれ別に用意する方法であり、小型化が容易で設計・計測の自由度や柔軟性が高めることができ、適用範囲も広い。

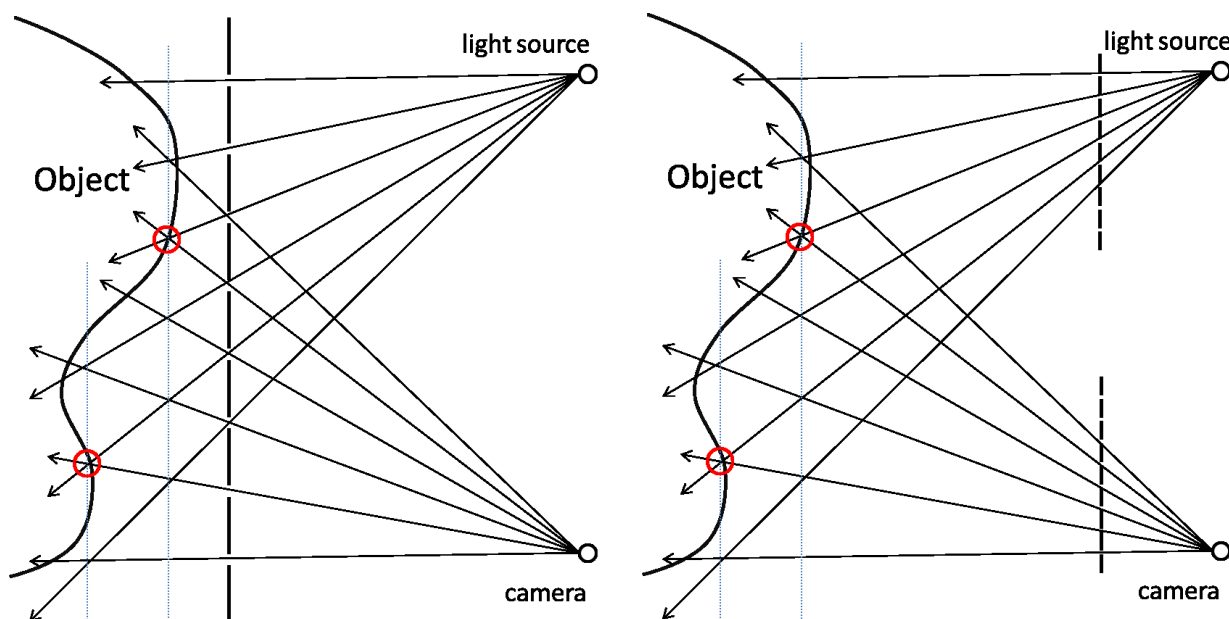


図6 格子パターン照射型モアレトポグラフィ(左)と格子パターン投影型モアレトポグラフィ(右) 赤い位置が等高線上の位置で明るく撮像される

第5項 正弦波格子位相シフト法

ここで、格子パターン投影型モアレトポグラフィとマルチスリット光投影法の考え方を合わせ、カメラ側の格子を省くとともに、位相シフト法[13]を組み合わせた格子パターン投影型モアレトポグラフィが提案された[14][15][16][17]。

格子を通して物体へと投射された光パターンは、物体の表面形状に応じて変形する。図6右に示した格子パターン投影型モアレトポグラフィでは別の格子を通過させて撮影することで、等高線情報を取り出す。しかし単に三次元情報を得ることが目的であれば、モアレパターンを生じさせる必要はない。さらに言えば、モアレトポグラフィからは直接的に三次元位置が得られない等高線情報だけしか求めることができないことが、逆に自動計測を困難としているとも考えられる。そこで変形格子パターン像をカメラで直接撮影することで、三次元形状を直接求める考え方が現れた。これはマルチスリット光投影法への接近と解釈できる。

さらに、モアレトポグラフィと位相シフト法[13]を組み合わせることにより、(1)凹凸判定などの面倒な処理を省く、(2)計測精度を向上する、(3)対象物体の表面反射率の影響を低減する方法が提案されてきた。

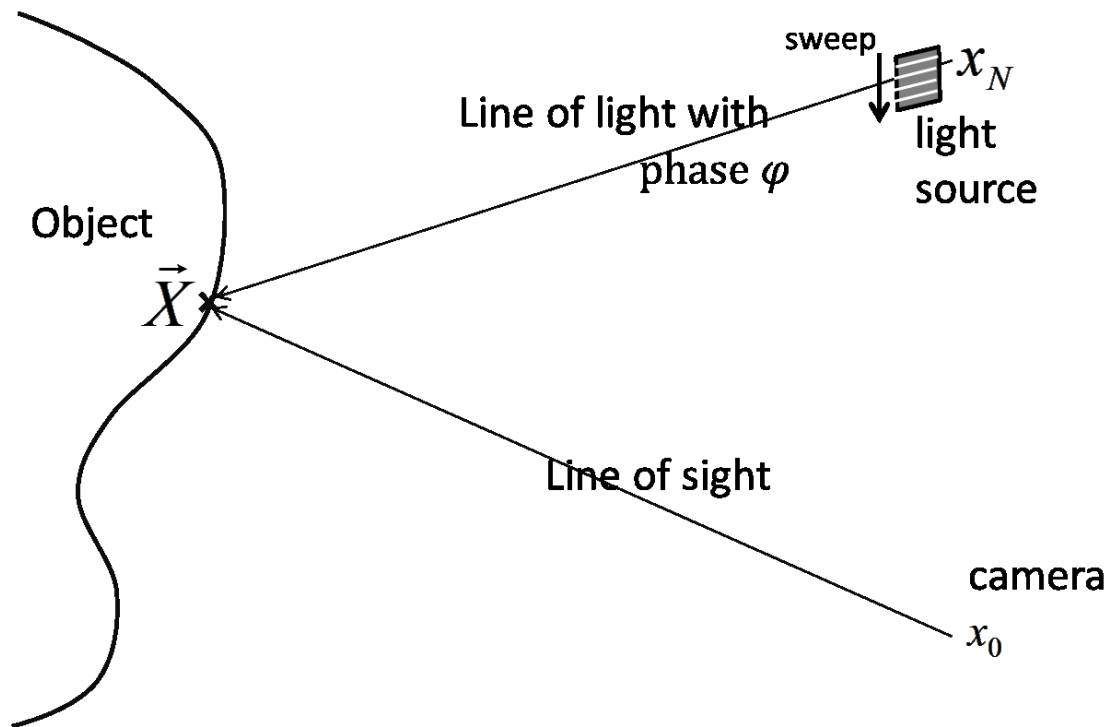


図 7 正弦波格子位相シフト法説明図

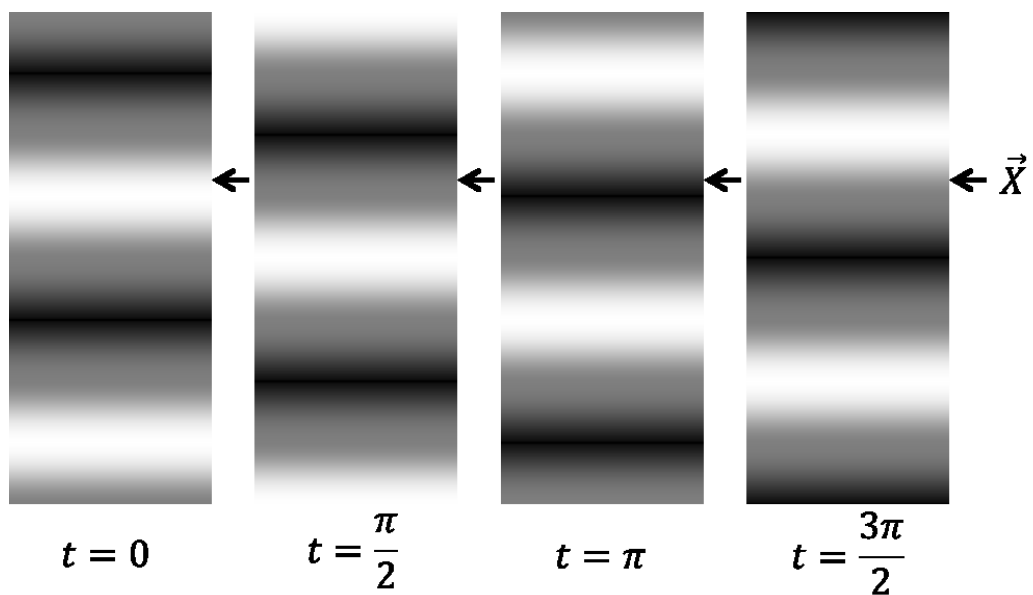


図 8 正弦波格子の移動と投影輝度

文献[14]は投影用格子を動かしながら計測し、得られた輝度値から干渉計の分野で知られる手法により位相を求める手法である。この手法では複数回のデータ計測が必要となるが、計測値は原理的に物体表面反射特性には依存しておらず、撮影用格子が原理的に不要で測定点毎に密な計測が可能となる特

徴がある。文献[15]は、計測デバイスにCCDを用いることで、画面全面での計測が可能なることを示した。さらに文献[16][17]で完成度が高められ、透過率が正弦波状の格子を動かす正弦波格子位相シフト法となった。模式図を図7に示して以下にその動作を説明する。

光源の前に置かれた正弦波格子を通し、正弦波状に強度が変化する光パターンを物体へ投射する。さらに正弦波格子を、波の方向へ動かしながら画像を撮影すると、観測対象物体上の点 \vec{X} には図8に示す位置を通過する光が投射されることとなる。従って、カメラにより観測される点 \vec{X} の輝度値 $I_{\vec{X}}(t)$ は式(4)で表される。

$$I_{\vec{X}}(t) = I_{\vec{X}}^B + A_{\vec{X}} \cos(t + \varphi_{\vec{X}}) \quad (4)$$

式(4)において、 $A_{\vec{X}}$ は正弦波を通した投射光の光強度と点 \vec{X} の反射率の積、 t は時間、 $I_{\vec{X}}^B$ は環境光強度と投射光と点 \vec{X} の反射率の積によるバイアス項である。また $\varphi_{\vec{X}}$ は $t = 0$ のときの位相値であり、光源から正弦波格子を通して照射する光パターンの照射角と対応する値である。

未知数は $I_{\vec{X}}^B \cdot A_{\vec{X}} \cdot \varphi_{\vec{X}}$ の三つであるので、最小3回の輝度値観測で位相値 $\varphi_{\vec{X}}$ 即ち点 \vec{X} の座標値を得ることができる。また、4回以上計測する場合には最小二乗法を用いて、より精度を高めて未知数を求めることができる。例えば図8の通り時刻 $t = 0 \cdot t = \pi/2 \cdot t = \pi \cdot t = 3\pi/2$ の4回輝度値を観測し $I_{\vec{X}}(0) \cdot I_{\vec{X}}(\pi/2) \cdot I_{\vec{X}}(\pi) \cdot I_{\vec{X}}(3\pi/2)$ を得るとすると、最小二乗法により $\varphi_{\vec{X}}$ の推定値である $\hat{\varphi}_{\vec{X}}$ は式(5)を用いて計算することができる。

$$\begin{aligned} \hat{\varphi}_{\vec{X}} &= \tan^{-1} \left[-\frac{\sum_{t=0, \pi/2, \pi, 3\pi/2} I_{\vec{X}}(t) \cdot \sin t}{\sum_{t=0, \pi/2, \pi, 3\pi/2} I_{\vec{X}}(t) \cdot \cos t} \right] \\ &= \tan^{-1} \left[\frac{I_{\vec{X}}(\pi/2) - I_{\vec{X}}(3\pi/2)}{I_{\vec{X}}(0) - I_{\vec{X}}(\pi)} \right] \end{aligned} \quad (5)$$

さらに一般には式(6)により $\hat{\varphi}_{\vec{X}}$ を得ることができる。よってより計測精度を高めたい場合には位相シフトの回数を増やすことによって実現することが可能である。

$$\hat{\varphi}_{\bar{x}} = \tan^{-1} \left[-\frac{\int_0^{2\pi} I_{\bar{x}}(t) \cdot \sin t}{\int_0^{2\pi} I_{\bar{x}}(t) \cdot \cos t} \right] \quad (6)$$

これらの式より、位相値 $\varphi_{\bar{x}}$ の推定値は計測対象物体の反射特性及び環境光に対して影響がないことが明らかである。光源、物体表面、カメラの位置関係は計測中不変であるため、ほとんどのステレオ法が仮定する完全散乱反射特性である必要もない。鏡面反射成分があったとしても、センサーのダイナミックレンジを超えなければ影響がないことがわかる。

但し得られる位相値 $\hat{\varphi}_{\bar{x}}$ は、三角関数が周期関数のため 2π モジュロ(2π 毎に折りたたまれる)となり、画像上に亘って位相値全体が滑らかとなるよう 2π の整数倍を適宜足し込む位相接続処理が必要である。この位相接続処理は、干渉計の分野でさまざまな研究がなされている[18][19]。また本手法は、レインボーレンジファインダ[9]などと同様に連続値で空間をコード化していることから、画素毎に密な三次元情報を得ることができることも優れた点である。但し、後述するように位相接続処理は必ずしも安定な結果が得られないことに加え、マルチスリット光投影法と同様に位相の絶対的な値即ち投射角度を一意に定める必要があるが、これまで解決されてこなかった。

第3節 多眼正弦波格子位相シフト法

正弦波格子位相シフト法は他の手法に比較して次の点で優れている：

- ✓ 計測時間が短い(最小3回のパターン投影と撮影)
- ✓ 画素単位で計測が可能
- ✓ 原理的に物体の反射特性に影響されない
- ✓ 原理的に環境光(但し計測中に安定していることが必要)に影響されない

しかし直接計算できる位相値には 2π モジュロであるために、段差が存在するような物体を計測する際に、位相接続処理で問題が生じることが知られている。

図9左に示す物体に対し、矢印方向から正弦波格子の光パターンを位相シフトさせながら投射するとする。投射する絶対位相値が図9右上で示す分布であるとき、計測される位相値は図9右下となり、○で示した領域では見かけ上ほぼ同一の計測位相値が得られる。位相接続とは、計測対象が滑らかであることを仮定して 2π の整数倍を加算あるいは減算することで、隣接する画

素が滑らかな位相値へ変換する処理である。従ってこの領域では位相接続処理により誤った加算あるいは減算をしてしまい、誤った形状を求めてしまう。このようにある程度以上複雑な形状の物体について、誤りなく位相接続処理を行うことは困難であることが知られている[18][19]。

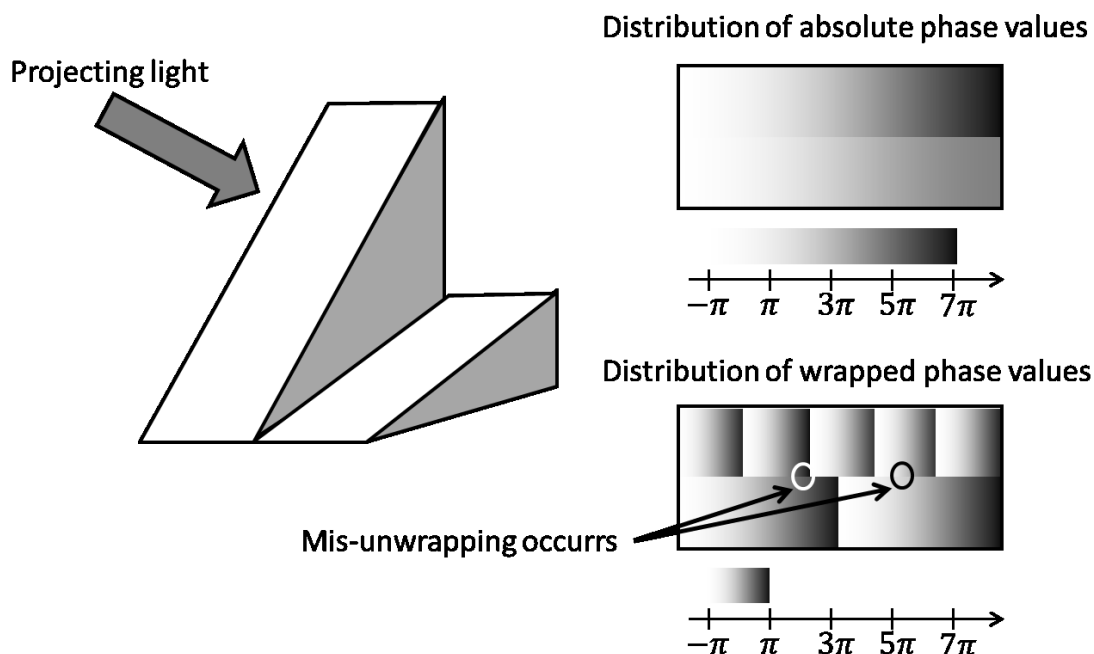


図 9 段差をもつ物体に対する位相接続処理の課題

また位相の絶対的な値(以下では絶対位相と呼ぶ)、即ち投射角度を一意に定める必要がある。図 10 に絶対位相についての説明図を示す。カメラより視線 #1 方向・#2 方向それぞれに対して計測位相値が φ (但し $-\pi \leq \varphi < \pi$) であり、二つの視線間で位相値が滑らかに 2π だけ変化する面が観測されたとする。もし絶対位相値が不明である場合、計測した面は図中の面 $A-B$ ・面 $A'-B'$ ・面 $A''-B''$ のいずれかであるかが決定できない。それぞれは大きさだけでなく角度も互いに異なることは明らかであり、物体表面を構成する各部分表面で同様な現象が生じることから、結果として相似でない複数の三次元形状が候補として得られてしまうだけとなる。

以上のふたつの問題は、計測位相値から直接絶対位相値との間のオフセットを得ることにより解決する。ここで著者らは複数のカメラと光源間の幾何的拘束を利用した絶対位相値の決定手法である多眼正弦波格子位相シフト法を提案する。これにより従来法では不可能であった画像上の非常に多くの点で絶対位相値を得られるようになり、不安定な位相接続処理に頼らない実用的なレンジファインダを構成することができる[89]。

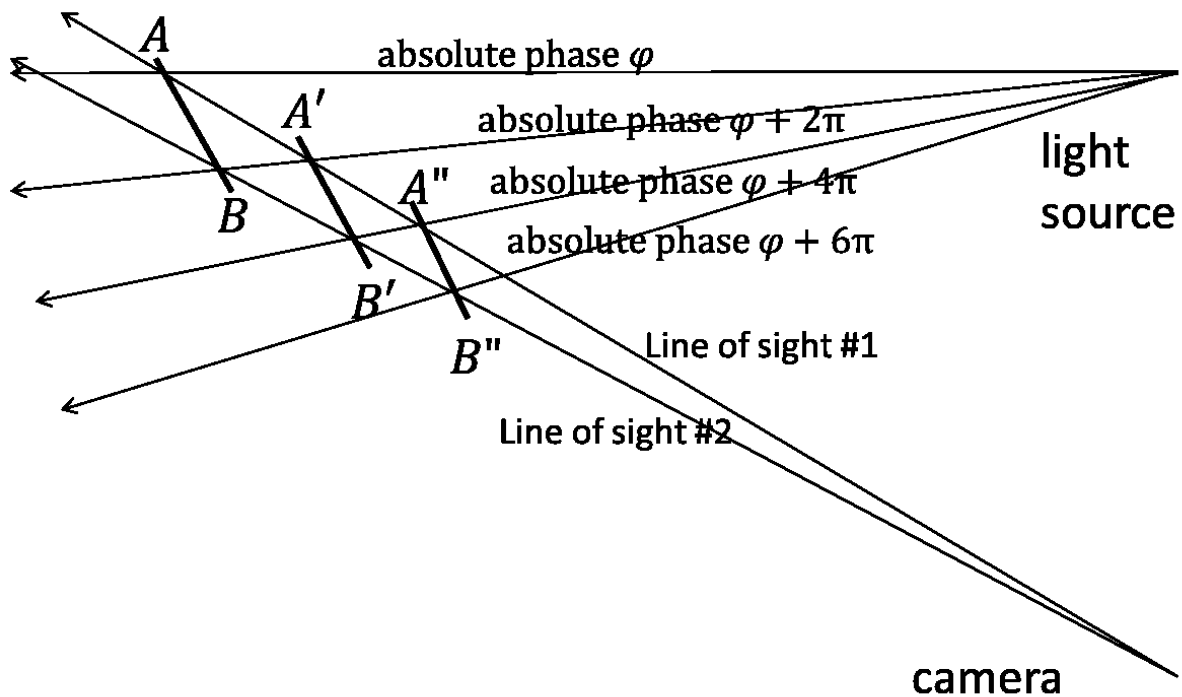


図 10 位相の絶対値不定性と三次元情報

絶対位相を得るための構成説明図である図 11 は、図 10 のカメラにもう一台カメラ(サブカメラと呼ぶ)を追加したものである。カメラの視線#1 方向の計測位相値が φ であるとき、対応して計測領域内に点 A・点 A'・点 A'' の複数の三次元座標候補点を得られる。これら候補点に対するサブカメラからの視線方向 #a・#b・#c を逆に計算し、各視線方向の計測位相値が同じ φ であるか否かを検定する。もし一点を除いた他の全ての候補点の位相値が φ であることが棄却される時、合致した点に対応点の三次元座標位置となり、さらに絶対位相値を計算することが可能となる。

この提案手法である多眼正弦波格子位相シフト法によって絶対位相値が決定可能であるためには、計測対象の形状の滑らかさに応じて位相の繰り返し周期が十分広くなければならない。即ち不特定の物体に対して必ず絶対位相値が得られるわけではない。しかし計測対象として顔に限定するなど計測対象のバラエティを絞る、より多くのカメラ・光源間の幾何学的関係を利用して同一位相値検定の信頼性を向上させるなどの工夫により、安定した計測が実現可能であると考えられる。

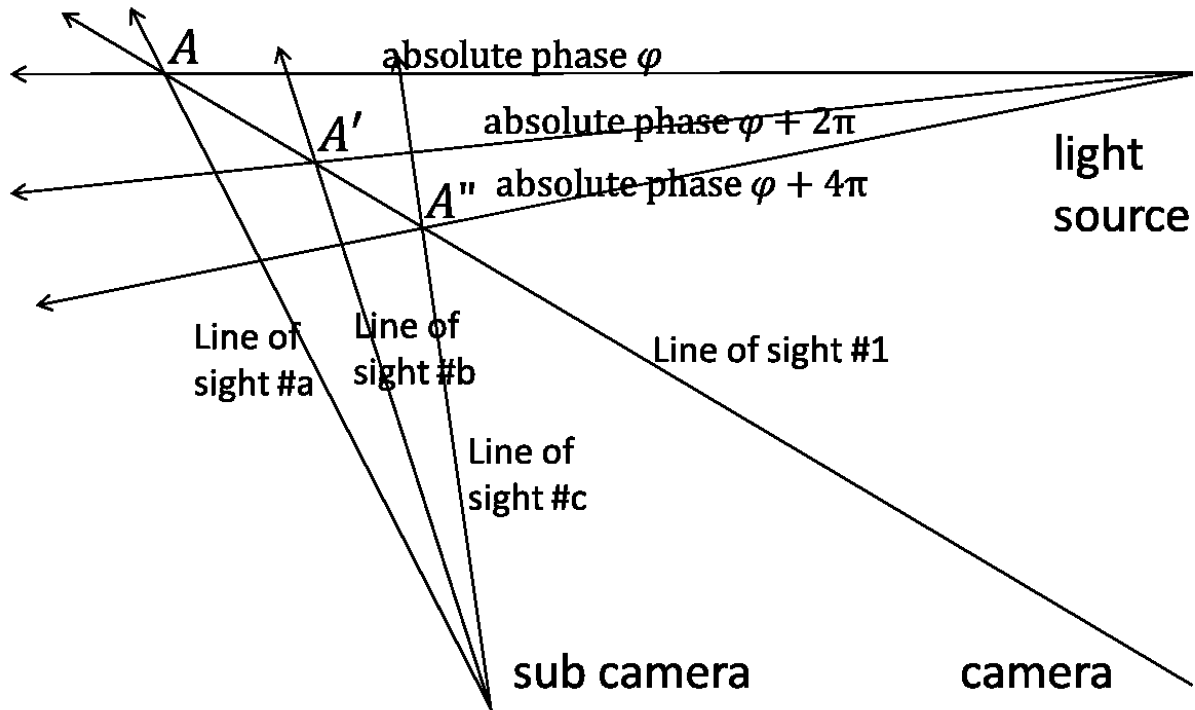


図 11 多眼正弦波格子位相シフト法による絶対位相値の決定

具体的な手順は次の通りである：

1. カメラ及びサブカメラで撮影した画像の全画素について、位相値と振幅を求める。位相値は式 (6) により、また振幅は $\sqrt{\left[\int_0^{2\pi} I_{\bar{x}}(t) \cdot \sin t\right]^2 + \left[\int_0^{2\pi} I_{\bar{x}}(t) \cdot \cos t\right]^2}$ によりそれぞれ計算する。得られた位相値は 2π モジュロである。
2. カメラ及びサブカメラで撮影した画像各画素の振幅に対して閾値を適用し、閾値以下の画素は振幅 0 とし利用しない。
3. カメラで撮影した画像について以下の処理を全画素繰り返す
 - (ア) 振幅が 0 なら次の画素へ処理を移す
 - (イ) 画素で定まる視線方向と、位相値に 2π の整数倍を足しこむことで定まる光源からの角度との交点を求め、複数の三次元座標候補を得る。
 - (ウ) 三次元座標候補をそれぞれサブカメラへ投影し、振幅が 0 ではなく位相値の 2π モジュロ表現値の差が閾値以内となるかを検証する。条件を満たす三次元座標候補点の一つに定まるとき、その点を計測三次元座標点と確定し、対応する絶対位相値を求める。
 - (エ) 次の画素へ処理を移す。

複数のカメラによる拘束を利用する手法は、マルチスリット光投影法に適用した文献[7]などにも見られる。しかしこの文献ではスリット位置でしか計測ができないのに比較して、本提案手法では全画素位置毎に計測位相値が得られるため、はるかに多くの地点で形状計測が可能なことが特徴である。

第4節 顔用レンジファインダと頭部全周レンジファインダの試作

著者らが提案した多眼正弦波格子位相シフト法をもとに、人間の顔及び頭部全周を計測するレンジファインダを試作した[89]。

図12に顔用レンジファインダの外観を示す。本体には左右両端に、カメラと光パターン投射機のペアをそれぞれ縦に配置しており、顔の左右側面を両耳まで計測できるよう構成している。顔の左(右)側面を計測するときには、反対側の右(左)側のカメラが絶対位相値を決定するためのサブカメラとして稼働する。位相シフト量は $\pi/2$ とし各カメラで4回の撮影を行う。

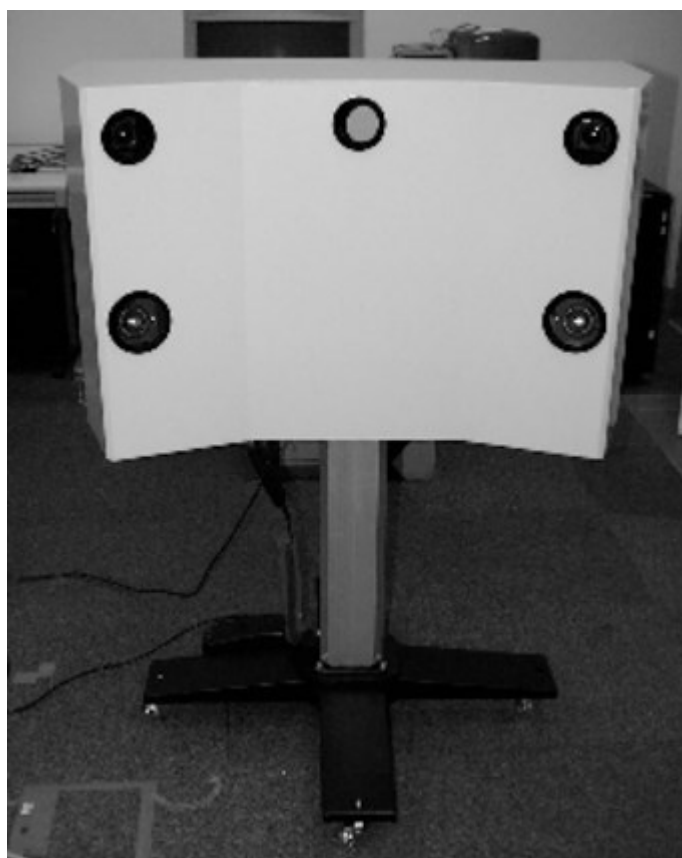


図12 顔用レンジファインダ外観図

本試作機では計測に約2.5秒必要とするが、画像撮影は左右合わせて計8回だけであることから、ハードウェアの改善により更なる短縮が可能である。

実際、この試作機の後改良して NEC が製品化した機種では、同一計測精度を保ちながら 0.6 秒での計測を実現している。

図 13 に、本論文で計測対象とした石膏像を示す。図 14 は正弦波格子を位相シフトさせながら石膏像に投射し、図 12 に向かって左側のカメラで撮影した画像を示す。

図 15 左は、図 14 の画像群から求めた位相値画像である。 $-\pi$ から π を黒から白に割り当てて示している。図 15 右は、向かって右側のカメラで撮影した画像群から求めた位相値画像である。なおこのとき正弦波の振幅である $A_{\bar{x}}$ (式(4)参照)をしきい値処理することで、正弦波パターンが照射されていない背景における位相値を除去している。



図 13 顔用レンジファインダの実験用石膏像

図 15 から求めた絶対位相値画像を図 16 左に、また絶対位相値が得られなかった領域については形状の滑らかさを仮定し位相接続した結果を図 16 右に示す。ここで 2π の整数倍のオフセット量に、それぞれ固有の色を割り当てて表示している。絶対位相値が得られなかった画素については位相接続処理を行わざるを得ない。しかし、図 16 左からわかるように位相接続が不得意とする不連続面において非常に多くの画素で絶対位相値が得られていることから、

全体を位相接続することと比較して難易度が下がっていることがわかる。

図 13 の石膏像の計測結果を図 17 に示す。左上はワイヤースケッチ、その他は照明を設定し、スムーズシェーディングにより生成した画像である。

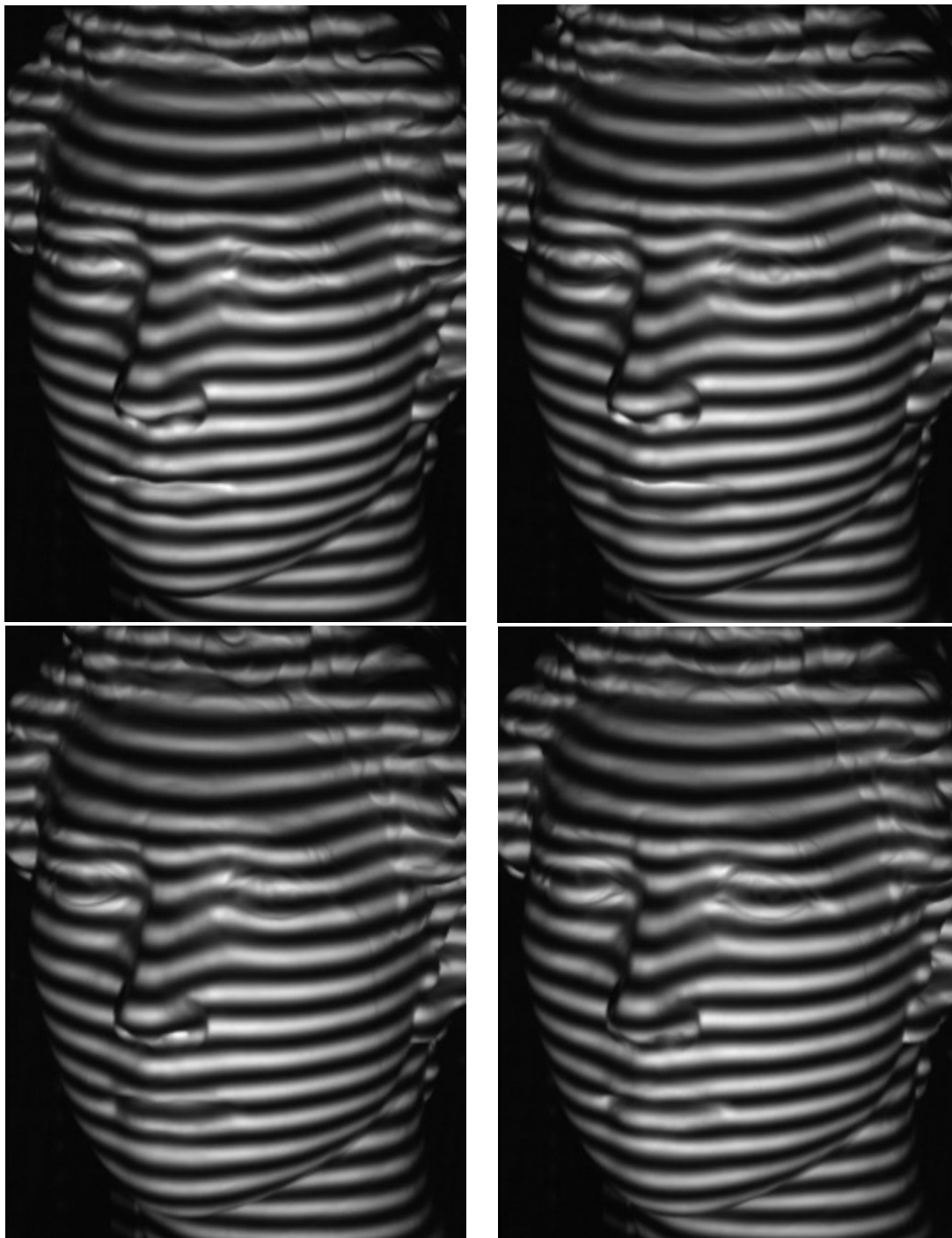


図 14 位相シフトした正弦波格子光パターンを投射した画像

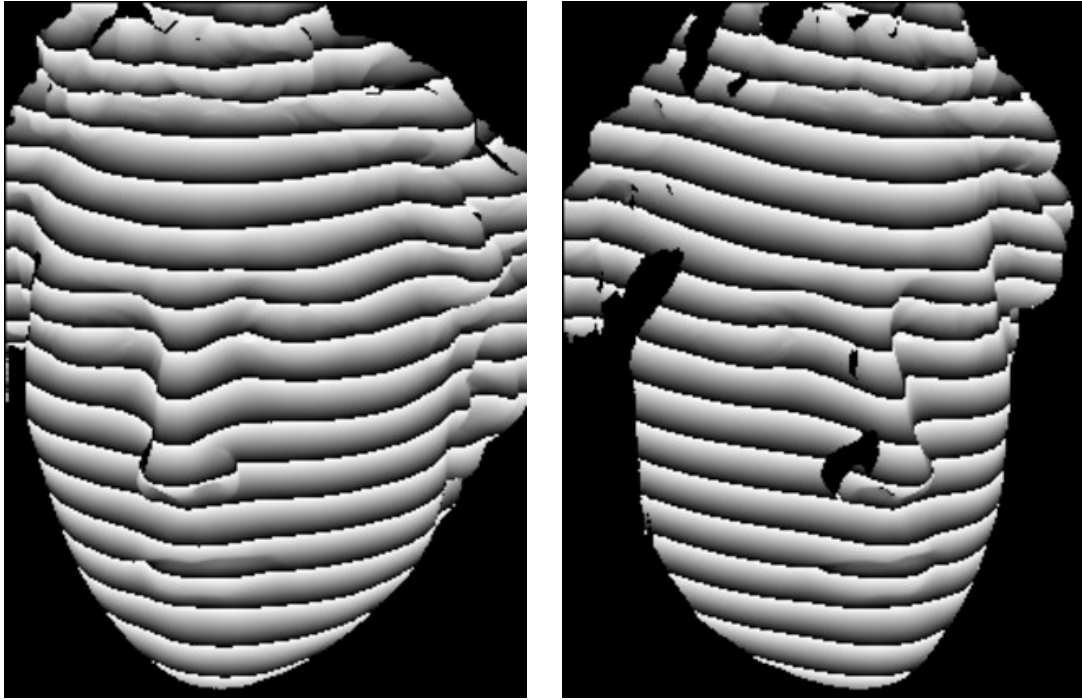


図 15 位相値画像($-\pi$ から π を黒から白にマップ)
左:向かって左側カメラ / 右:向かって右側のカメラ

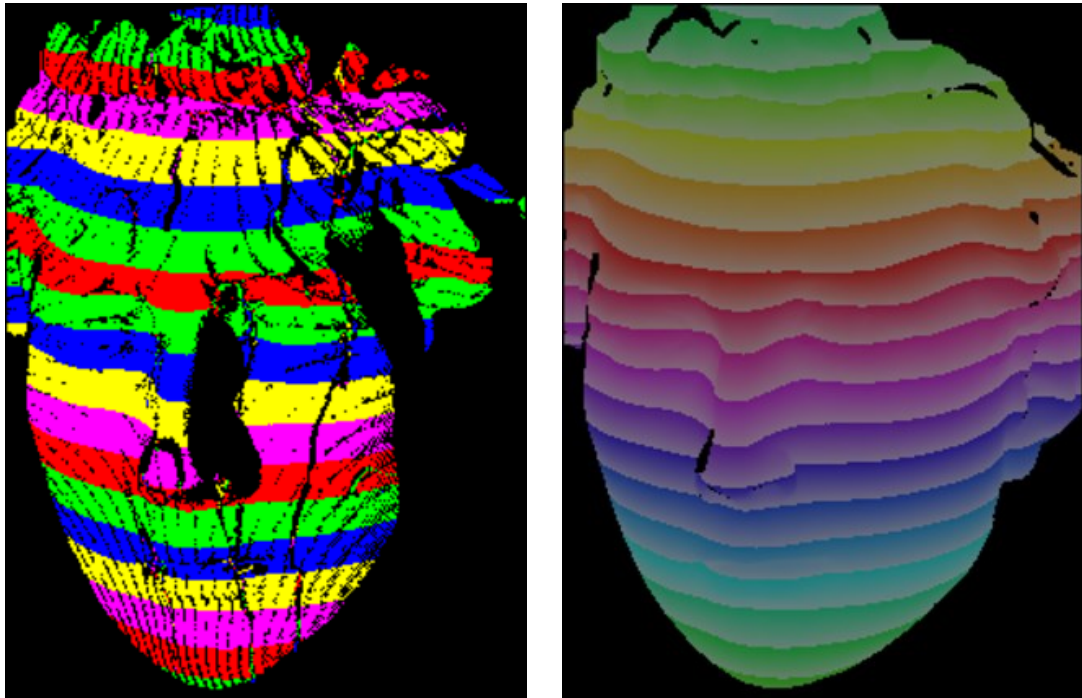


図 16 絶対位相値画像 左: 2π 整数倍値に対し色をマップ /
右:位相接続処理により補間した絶対位相値

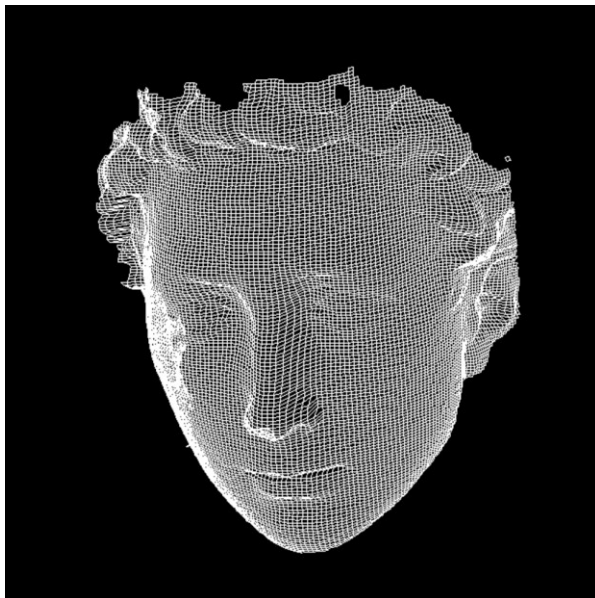




図 17 石膏像(図 13)の三次元形状計測結果 左上: ワイヤフレーム表示 / 右上・下: シェーディング表示

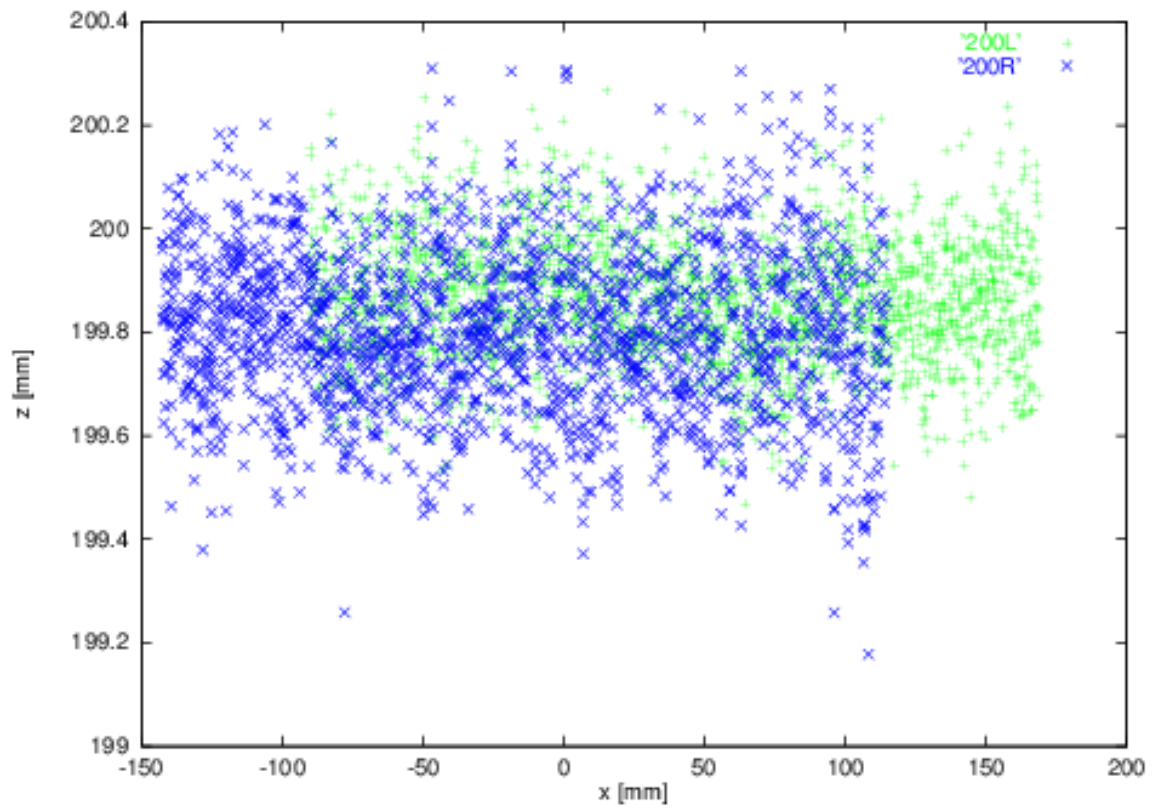


図 18 平板の計測値分布(x-z プロット) 平板設置奥行き位置 200mm 地点
青: 向かって右カメラの計測値 / 緑: 左カメラ

表 1 平板の計測値と標準偏差(誤差)

Depth position of the plane (mm)	Estimated depth position of the measured plane (mm)	Standard deviation (mm)
99.5	99.444	0.127
101.0	100.916	0.129
150.5	150.429	0.144
152.0	151.949	0.145
198.5	198.364	0.137
200.0	199.847	0.145
200.0	199.855	0.143
200.0	199.871	0.141
200.0	199.868	0.140
200.0	199.868	0.141

奥行き方向の計測精度を評価するために、平板を設置して測定した。結果を図 18 及び表 1 に示す。正弦波の振幅である A_x (式(4)参照)が、カメラのダイナミックレンジをほぼ使い切るよう手動で光源の光量を調整して平板計測した後、測定値に対して最小自乗法による平面をあてはめた。平板の位置は精密な自動テーブルを用いて定めており、その位置決め精度はカタログ上 0.1mm である。推定された平面と計測値との差を、ある位置でスライスして表示したのが図 18 である。この図は左カメラと右カメラそれぞれの計測結果を重ねて示しており、ほぼ重なることや顕著な分布構造が見えないことから、よく校正されており、主にノイズに基づく統計的誤差だけが現れていることがわかる。

奥行き位置として 10 地点を選んで求めた誤差値を表 1 に示す。同表後半の平板設置奥行き位置 200mm が図 18 で示すデータであり、5 回の計測で推定位置にほとんど差がないこと、標準偏差がいずれも 0.14mm と小さいことなどがわかる。他のデータも合わせると、測定誤差の標準偏差は、0.12~0.16mm 程度であり、レーザー光を用いた一般的な顔程度の大きさの物体を計測するスリット光投影法のレンジファインダが 0.5mm 程度の誤差であることを考えると、より高い精度を得ていることがわかる。

さらに計測地点 200.0mm で 5 回の計測を繰り返した結果がほぼ一致していることから、安定した計測が可能であると言える。

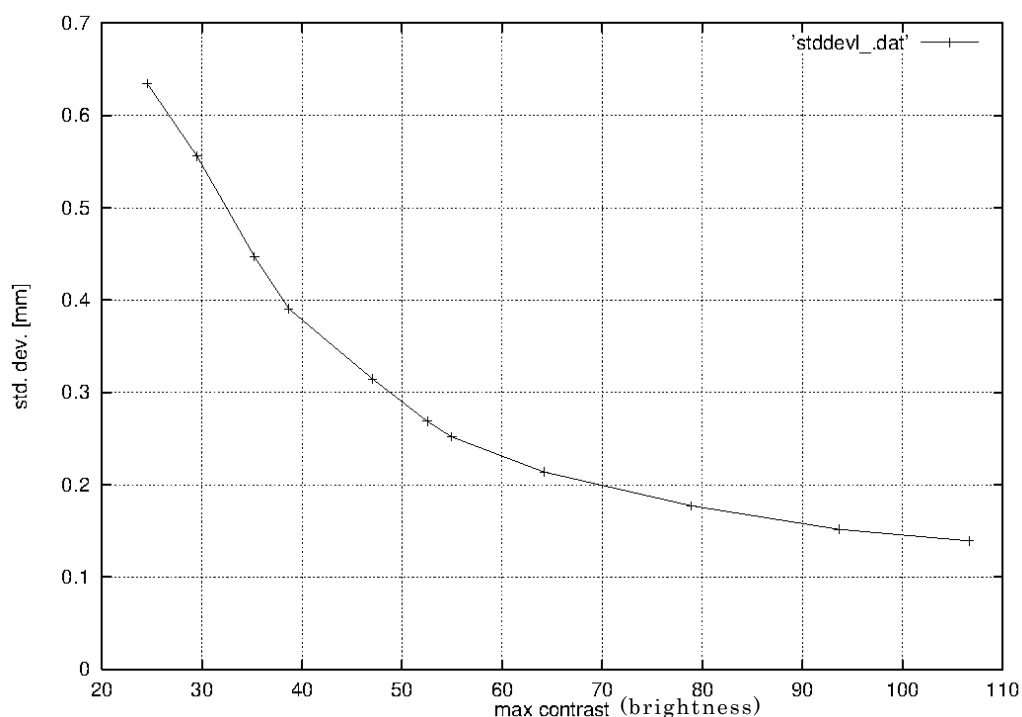


図 19 正弦波振幅最大値 $A_{\bar{x}}$ と標準偏差(奥行き測定誤差)

図 19 で、画像上で最も大きい正弦波振幅 $A_{\bar{x}}$ と奥行き測定誤差である標準偏差との関係を示している。振幅 $A_{\bar{x}}$ が小さいと S/N が劣化するために測定誤差が徐々に増加してゆくのがわかる。しかし、例えば振幅が輝度値レベルで 70 程度であっても標準偏差は 0.2mm に収まっていることから、一般的な室内環境であっても輝度値の値域である 0 から 255 におさめることができ、十分高精度な計測が可能であることを示唆している。

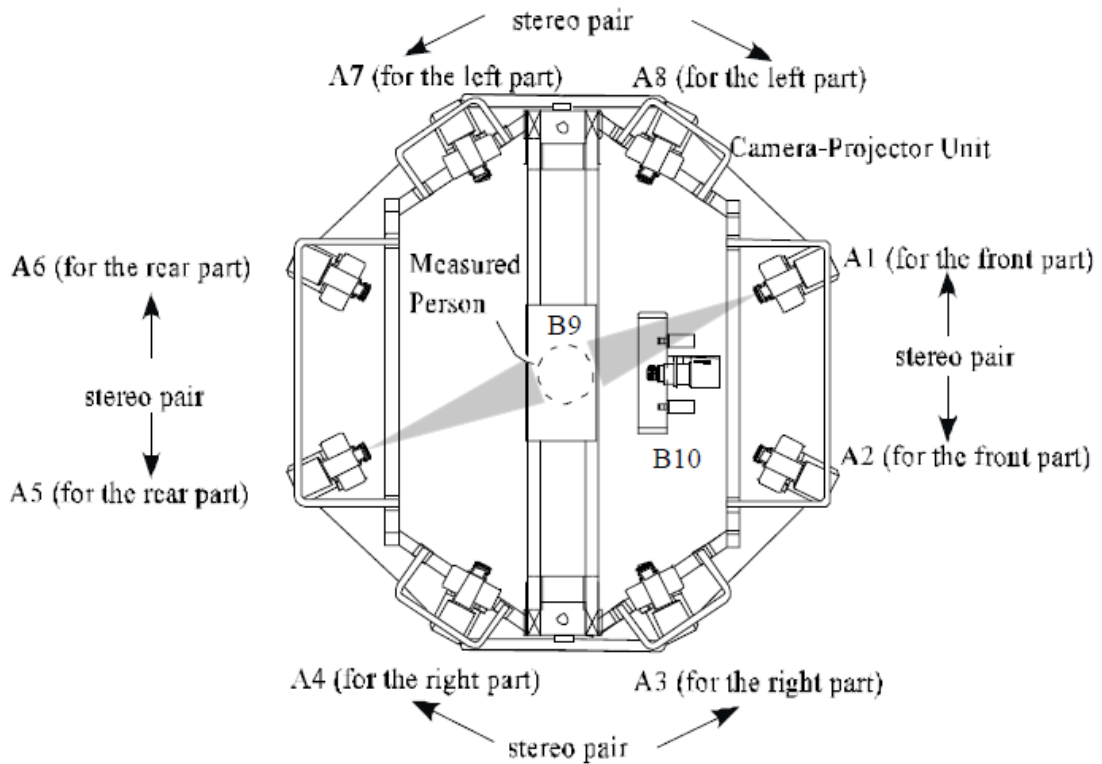


図 20 頭部全周レンジファインダ 平面構成図

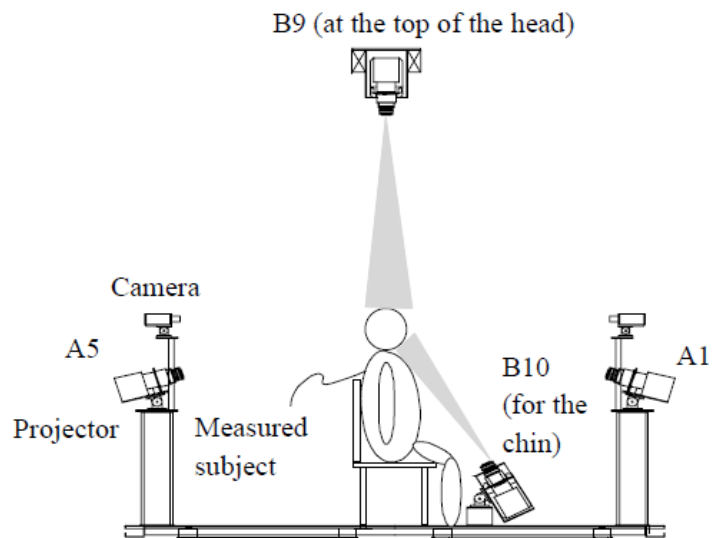


図 21 頭部全周レンジファインダ 側面構成図

次に、頭部全周レンジファインダの開発について説明する[90]。図 20 はシステムの平面構成図、図 21 は横からの構成図である。基本的に顔用レンジファインダを 4 セット平面に設置することで、頭部側面全面をカバーするとともに、残された頭頂部と顎下を計測するユニットから構成されている。正弦波の光パターン投影は、干渉しないように単独投影することが基本であるが、

本レンジファインダの場合中央に頭部が位置することから対角線上の光パターンは頭で遮蔽されることを考慮し、同時発光させることで撮影時間を短縮している。

図 22 にシステムのブロック図を示す。正弦波格子の位相シフトとカメラの画像撮影タイミングは専用の制御ボードによりフレーム単位で制御する。撮影された画像から位相値を、分散された PC1 で並列に処理した後、PC2 へデータを送信して頭部全体の計測値としてまとめる。

図 23 は、開発した頭部全周レンジファインダの外観である。図 24 左は評価に用いた実験用マネキンであり、彫りの深い老人を想定して人間工学研究センターが作成したものである。同図右はその計測結果であり、よく再現できていることがわかる。撮影時間は約 1 秒・計測精度は 0.5mm を達成することができ、産業技術総合研究所で研究用として稼働した。

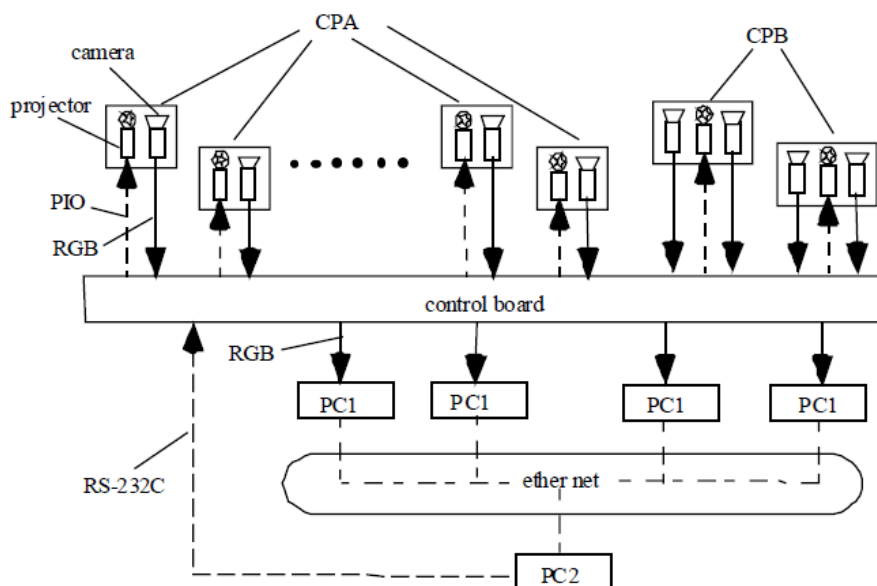


図 22 頭部全周レンジファインダ システム構成図



図 23 頭部全周レンジファインダ外観図

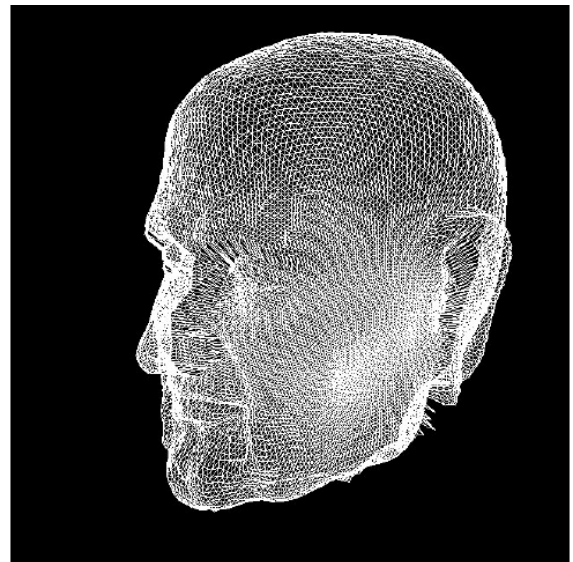


図 24 頭部全周レンジファインダの実験用マネキン(左)と計測結果(右)

第3章 視点依存の輝度変動に頑健なマルチベースラインステレオ法の提案

第1節 ステレオ法における視点依存の輝度変動対応

ステレオ法は、光パターンを投射する必要はなく一度の撮影タイミングで得られる画像群だけから三次元情報を推定することができることから、多くの研究者により研究開発が進められるとともに、二眼・三眼・マルチベースラインの各ステレオや移動カメラのシステムへ適用されてきた[1][2][3][4]。

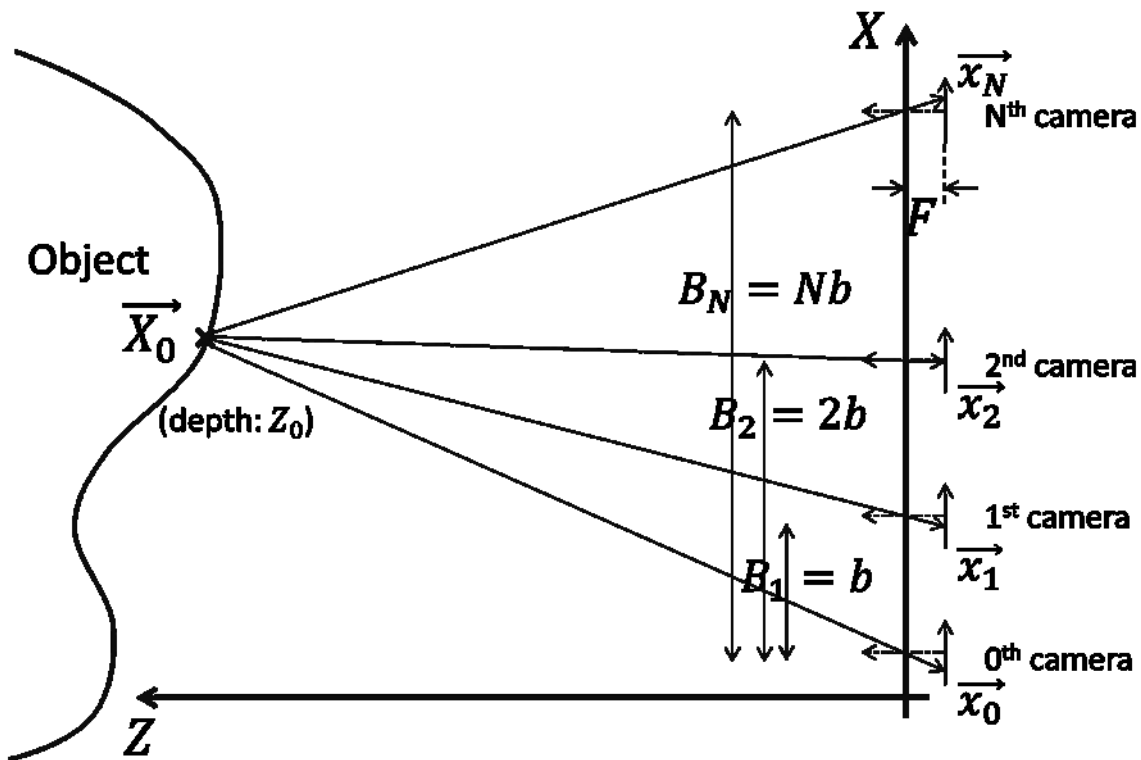


図 25 マルチベースラインステレオの一般的な設定(図 1 を再掲)

図 25 の通り $N+1$ 台の共通の焦点距離 F をもつカメラが、全て平行の光軸をもち光軸と垂直な直線上に等間隔 b で並んでいるとする。0番目と i 番目のカメラ間距離を B_i と書いてベースラインと呼び、図 25 の配置では $B_i = ib$ となる。各カメラで撮影される画像には x - y 座標系が定義されており各軸はすべてのカメラで平行とする。またベースラインとカメラの x 軸が平行であるとする。光軸方向に撮影対象物体があり、その物体上の観察点 \vec{X}_0 は、カメラからの奥行き距離が Z_0 であるとする。観察点 \vec{X}_0 が 0番目と i 番目のふたつのカメラで撮影された画像上に投影された座標は、ベースラインと垂直である y 軸方向は全て同一であるけれども x 軸方向で差が生じる。この投影された座標位置である x_0 と

x_i の差を視差と呼び $d_i(x_0)$ とする。これらパラメータ間には式(7)の関係があることが知られている(式(1)の再掲)。

$$\begin{aligned}d_i(x_0) &= x_0 - x_i \\ &= \frac{B_i F}{Z_0} \\ &= \frac{ibF}{Z_0}\end{aligned}\tag{7}$$

ステレオ法の目的は、画像中の画素を処理し投影点間の対応を自動的に探して視差を求めることである。式(7)から明らかな通り視差 $d_i(x_0)$ はベースライン $B_i = ib$ に比例する。視差が有限の画素位置で量子化されていることを考慮すると、奥行き推定精度を上げるには i を大きくする、即ち0番目のカメラからできるだけ遠いカメラを選択しベースラインを長くするべきである。しかし距離を大きくすることは対応点探索を困難にする。

もし焦点距離・幾何学歪・周辺減光や、画素毎のバイアスや利得など全てのカメラパラメータを校正したとしても[40][41]、ステレオ法には依然として次の課題がある：

- (a) 対応点探索範囲が広がることで、対応点候補がひとつ以上現れる可能性が高くなる。
- (b) もし対象物体が完全散乱反射特性でない場合、正しい対応点であっても異なる輝度をもつ可能性が高くなる。
- (c) 対象物体表面の法線ベクトルがカメラ光軸と並行していない場合、正しい対応点であっても異なる輝度をもつ可能性が高くなる。

課題(a)はマルチベースラインステレオ法により誤対応の発生を抑制する手法[23]や、動的計画法によるスキャンライン毎の最適化[55][70][71]、信念伝播法[68][76][77][78][79][80][81]やグラフカット[83][84][85][86]など進んだ最適化手法によって大きく改善されてきた。また SSD[23]や SAD[38]、NCC[25]、輝度勾配[43][44][45][46]、順序尺度[48]、センサス変換[49]など、小領域での比較による手法によっても誤対応を減らしてきた。しかしステレオ法は、物体の反射特性として完全散乱を仮定する、あるいは各カメラからのぞむ物体表面が各画像上で同一形状であること[54]を仮定し続けており、課題(b)および(c)は残されたままである。よって、高い精度の三次元情報を得るために新しい手法が求められている。

ここで実際の画像を材料として検討を進める。

Middlebury ステレオデータセット 2006 の中から lampshade1 を例示する [95][96]。カメラを水平移動する架台に設置し、図 25 に示したように一定間隔で撮影した 7 枚で構成されている。画像サイズは 3 種類あり、そのうちの一番小さな 433 x 370 画素と 1/3 の大きさにダウンサンプリングした画像を用いる。左から 2 番目と 6 番目の画像間の正しい視差マップは、文献 [96] に記された構造をもつ光を照射する手法で求めた。この視差マップには原画像における視差が格納されているため、実験に利用した画像に対する視差の精度は 1/3 画素である。

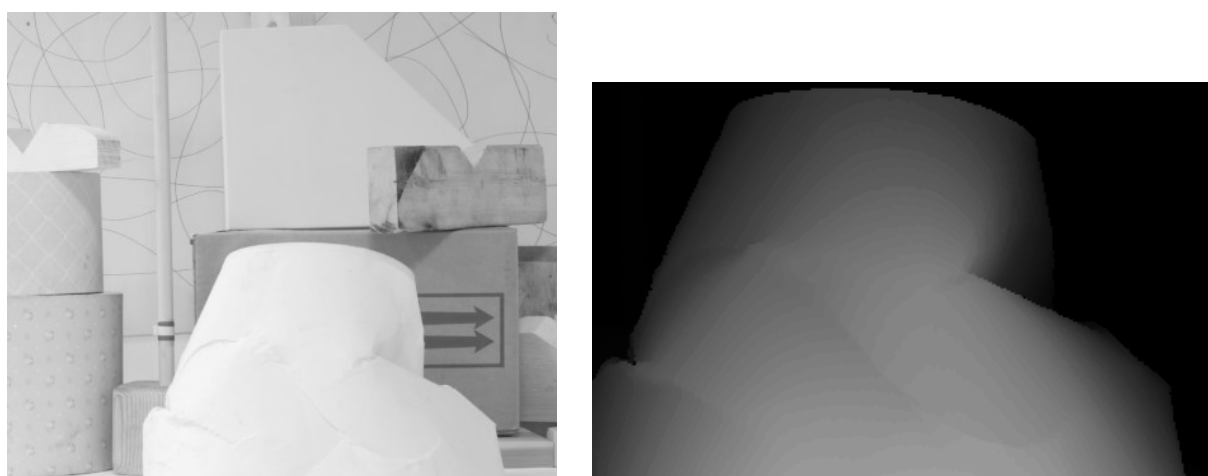


図 26 Middlebury データセット 2006 内の lampshade1 画像 [94][95]。左：左端位置から撮影した画像 / 右：左から 2 番目の位置から撮影した画像から同 6 番目の画像に対する視差画像 (最前部に位置するランプシェードだけを切り出して強調表示)

図 26 に配布された画像の一部を示す。図 26 左は左端位置から撮影した画像である。図 26 右は、左から 2 番目の位置から撮影した画像から同 6 番目の画像に対する視差画像であり、最前部に位置するランプシェードだけを切り出して強調表示した画像である。

検討の前に、まず図 26 右の視差画像を、左端撮影の画像における右端の画像との間の視差画像へ変換する。図 26 右の視差画像の各画素には 2 番目の画像から 6 番目の画像への視差 (ベースライン $4b$) が格納されていることから、左端画像状の対応画素は画像に向かって右に視差の半分だけ移動した位置に相当する (ベースライン b)。この対応画素へ格納すべき右端画像から左端画像 (ベースライン $6b$) への視差は 1.5 倍の値となる。次に変換した視差画像上で左右に隣り合った画素の視差値を比較し、向かって右の画素の視差が大きい場合

は左の画素位置を起点にして、右がその視差の幅だけ左方向に隠蔽されると判定する。

なお Middlebury データセットは左右両端画像に対する視差画像が与えられていないことが主要因と思われるが、この最大ベースラインでのステレオ法評価を行った過去研究を見つけることはできなかった。視点依存の大きな輝度変動が存在することから、従来のステレオ法研究に資することが難しいことが、視差画像を用意しなかった要因のひとつではないかと推測される。

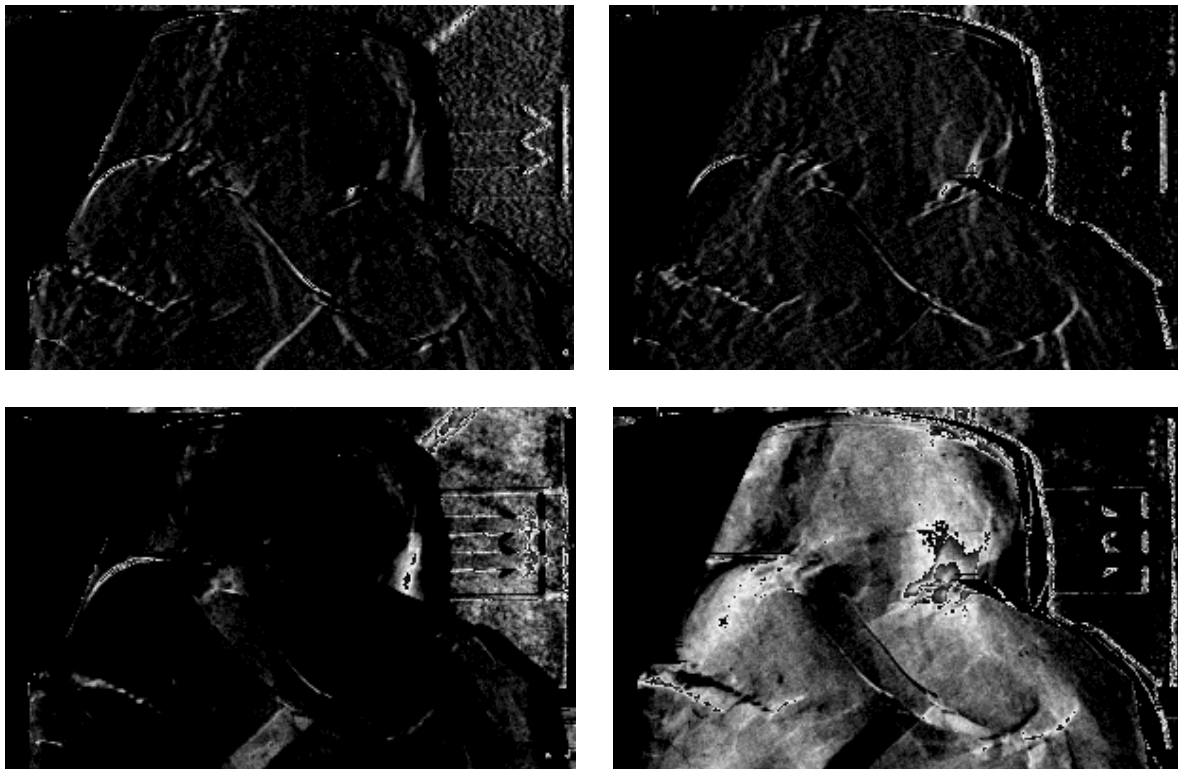


図 27 左端画像輝度を基準とした対応点間の輝度差、上段: 2 番目の画像との差 / 下段: 右端画像との差 / 左: 輝度差+ / 右: 輝度差- (実際の輝度差を 20 倍表示)

図 27 に左端画像を基準とした対応点間の輝度差を示す。図 27 上段は 3 番目の、図 27 下段は右端の画像との輝度差であり、それぞれ左は輝度差が+・右は-をそれぞれ意味する(輝度差は 20 倍に強調)。隣接した画像間では輝度差はほとんどないものの、最も離れた画像間では輝度が暗くなっていることがわかる。

このように視点が大きく移動するに伴い輝度変動が実際に発生しており、より精度の高い三次元情報を得るには対策が必要不可欠である。

第2節 視点依存の輝度変動に頑健なマルチベースラインステレオ法の提案

本節において、視点依存の輝度変動に頑健なマルチベースラインステレオ法を提案する[93]。この文献[93]は、文献[71]を拡張した、動的計画法に基づき最適化を行うアルゴリズムである。

図25の通り $N+1$ 台のカメラが間隔 b で X 軸上に並んでいるとする。全てのカメラ光軸は Z 軸に並行であり、同一サイズの画像を出力する。各画像の x 軸はベースラインに並行とし、視差は x 軸にだけ現れるとする。物体上の、奥行き Z_0 である観察点 \vec{X}_0 は、カメラ N が撮影される画像上で \vec{x}_i に投影される。このとき x_0 との視差 $d_N(x_0)$ は、式(7)に従い $B_N F/Z_0 = N b F/Z_0$ で得ることができる。また他のカメラ i における視差 $d_i(x_0)$ は $B_i F/Z_0 = i b F/Z_0 = i/N \cdot d_N(x_0)$ となり、 $d_N(x_0)$ を決めれば自動的に定まる。

文献[93]では最尤推定の枠組を採用する。最大ベースラインを構成するカメラ0とカメラ $N+1$ の二眼ステレオを基本としており、他の中間位置のカメラは補助的にもちいる。 $Z_{0,N}(x_0, x_N) = Z_{0,N}(x_0, x_0 - d_N(x_0))$ を、カメラ0上の座標 x_0 とカメラ N 上の座標 x_N が対応する事象とする。三次元座標地点 \vec{X}_0 が与えられたとき、事象 $Z_{0,N}(x_0, x_N)$ の尤度 $\Lambda(Z_{0,N}(x_0, x_N) | \vec{X}_0)$ を、式(8)で定義する。なおカメラ s に対して、カメラ N 上の座標 x_N が対応する座標 x_s は、式(7)より $x_s = x_0 - s b F/Z_0 = x_0 - s/N \cdot (x_0 - x_N) = x_0 - s \cdot d/N$ で与えられる。

$$\begin{aligned} \Lambda(Z_{0,N}(x_0, x_N) | \vec{X}_0) &= \left(\frac{1 - P_D}{\phi} \right)^{N \delta_{0,N}(x_0, x_N)} \\ &\cdot \left\{ P_D \cdot p(i_0(x_0) | \vec{X}_0) \cdot P_D \cdot p(i_N(x_N) | \vec{X}_0) \right. \\ &\cdot \left. \prod_{s=1}^{N-1} \left(\frac{1 - P_D}{\phi} \right)^{\delta_s(x_s)} \cdot [P_D \cdot p(i_s(x_s) | \vec{X}_0)]^{1 - \delta_s(x_s)} \right\}^{1 - \delta_{0,N}(x_0, x_N)} \end{aligned} \quad (8)$$

ここで $\delta_{0,N}(x_0, x_N)$ は、 \vec{X}_0 がカメラ0上の座標 x_0 とカメラ N 上の座標 x_N 両方へ隠ぺいなく投影されているときに0、片方あるいは両方が隠ぺいにより投影されていないときに1をとる変数である。 $\delta_s(x_s)$ ($s = 1, \dots, N-1$)は同様な変数であり、 \vec{X}_0 がカメラ s 上の座標 x_s へ投影されているときに0、隠ぺいにより投影されていないときに1をとる。 ϕ はカメラの視野、 P_D は \vec{X}_0 が画像上に投影される確率であり、ともにあらかじめ与えるパラメータである。このとき $(1 - P_D)/\phi$ は設置したカメラで撮影されない、即ち隠ぺいが発生する確率となる。

$p(i_s(x_s)|\vec{X}_0)$ ($s = 0, \dots, N$)は \vec{X}_0 がカメラ s の座標 x_s において輝度 $i_s(x_s)$ で観測される確率分布関数であり、正規分布関数である式(9)で与えられるとする²。

$$p(i_s(x_s)|\vec{X}_0) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \cdot \exp\left[-\frac{(i_s(x_s) - \bar{i}_s(x_s))^2}{2\sigma^2}\right] \quad (9)$$

式(9)中の $\bar{i}_s(x_s)$ は輝度の真値であって知ることはできない。文献[71]ではすべてのカメラで共通の輝度の真値があることを仮定し、式(10)で推定している。

$$\bar{i}_s(x_s) \approx \hat{i}_s(x_s) = \frac{i_0(x_0) + i_N(x_N)}{2} \quad (10)$$

即ち視点依存の輝度変動は存在しないことを仮定している。

本論文ではカメラ s を中心とした $s \pm M$ で定義されるカメラ近傍の概念を導入し、輝度 $\bar{i}_s(x_s)$ を $\hat{i}'_s(x_s)$ として式(11)で推定することを提案する。

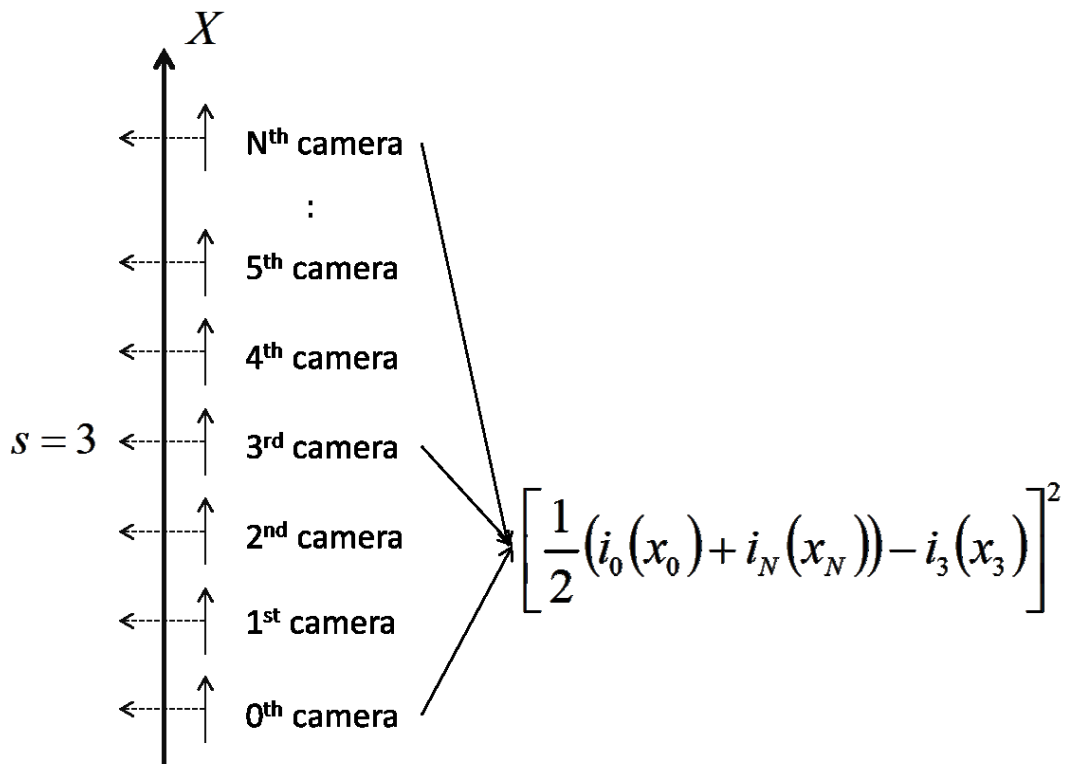
$$\bar{i}_s(x_s) \approx \hat{i}'_s(x_s) = \frac{1}{2M+1} \cdot \sum_{t=j_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+t}(x_{s+t}) \quad (11)$$

$$j_{\min}^{(s)} = \begin{cases} -s & \text{if } s \leq M \\ N - 2M - s & \text{else if } s > N - M \\ -M & \text{otherwise} \end{cases}$$

$$j_{\max}^{(s)} = \begin{cases} 2M - s & \text{if } s \leq M \\ N - s & \text{else if } s > N - M \\ M & \text{otherwise} \end{cases}$$

オクルージョンが存在しない、すなわち $\delta_{0,N}(x_0, x_N)$ 及び $\delta_s(x_s)$ がともに0であるときの尤度 $\Lambda(Z_{0,N}(x_0, x_N)|\vec{X}_0)$ である式(8)は、式(9)で定義される正規分布の積となる。図28に従来法である式(9)及び式(10)による尤度の計算ステップを示す。カメラ3の尤度は該カメラ3の輝度とカメラ0とカメラNの平均輝度との差に基づいた正規分布関数で計算する。従って物体の反射特性が完全散乱でないときには、たとえ対応する画素同士であっても輝度差の期待値は0にならない。

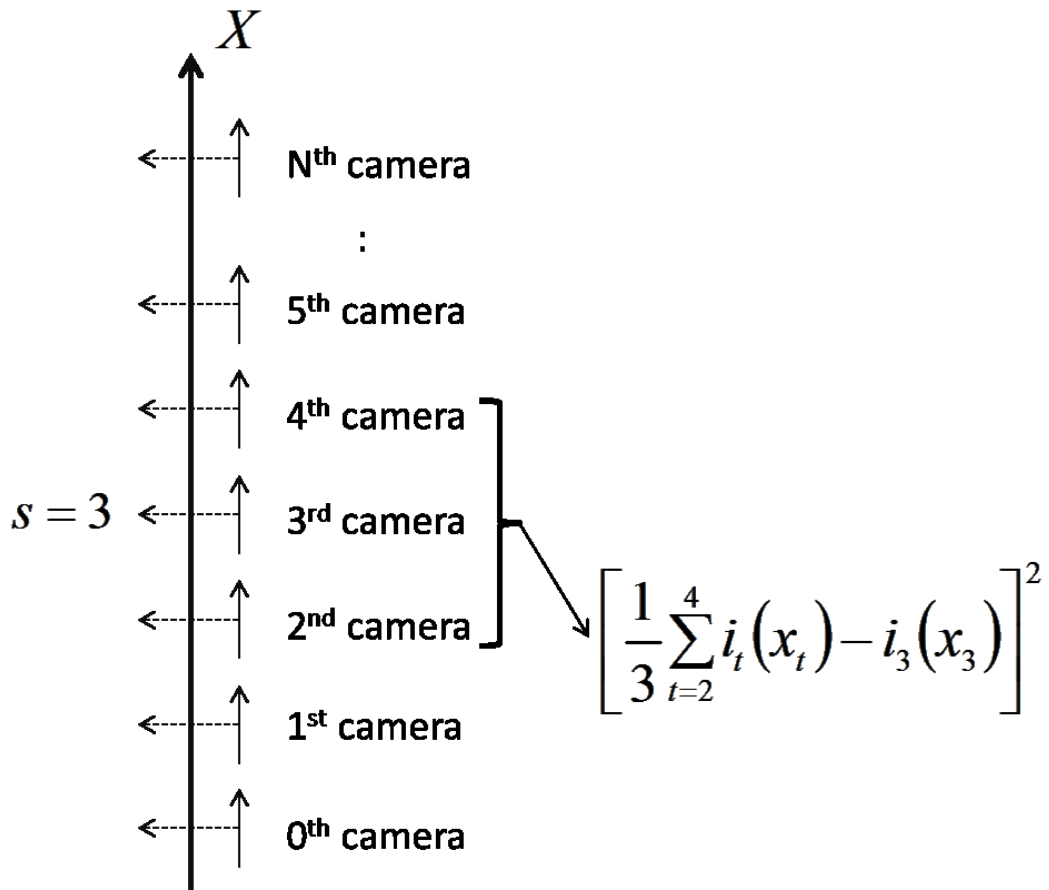
² ここでは $i_s(x_s)$ を輝度と記述しているが、他の画像特徴量を用いても構わない。



Summation regarding s

図 28 従来のマルチベースライン法(文献[71])の尤度計算ステップ
(式(8)/(9)/(10))

図 29 に提案手法である、式(9)及び式(11)による尤度の計算ステップを具体的に示す。 $M = 1$ のとき、カメラ3の尤度は該カメラ3の輝度と、カメラ2からカメラ4の平均輝度との差に基づいた正規分布関数で計算する。続いてカメラ毎の尤度を全て乗算することで、全体の尤度である $\Lambda(Z_{0,N}(x_0, x_N) | \bar{X}_0)$ を得る。従ってカメラ近傍における輝度変動が直線的である場合、尤度 $\Lambda(Z_{0,N}(x_0, x_N) | \bar{X}_0)$ には影響を及ぼさないことがわかる。物体が完全散乱反射ではない場合であっても、輝度変動がカメラ近傍で十分線形で近似できるように設定することで、影響をなくすことが可能となる。



Summation regarding s

図 29 輝度変化に頑健なマルチベースライン法の尤度計算ステップ
($M = 1 \cdot$ 式(8)/(9)/(11))

$\{d_N\}$ を各カメラ画像の中の全ての画素に対する視差の組み合わせのうちの一つのセット、 D をそのとり得る全ての組とすると、 $\{d_N\} \in D$ で表すことができる。ここで尤度 $\Lambda(Z_{0,N}(x_0, x_N) | \bar{X}_0)$ を $\{d_N\}$ で表せる全組分だけかけあわせて対数をとった、対数尤度 $J(\{d_N\})$ を考える。 D の中で $J(\{d_N\})$ を最小化する $\{d_N\}$ は、式(12)により得ることができる。

$$\begin{aligned}
\min_{\{d\} \in D} J(\{d\}) &= \min_{\{d\} \in D} \left\{ \sum_{\{d_N\}} \delta_{0,N}(x_0, x_N) \right. \\
&\quad \cdot \ln \frac{P_D \cdot \phi}{(1 - P_D)(2\pi\sigma^2)^{\frac{1}{2}}} + \frac{1 - \delta_{0,N}(x_0, x_N)}{N} \left[\frac{(i_0(x_0) - \hat{i}'_0(x_0))^2}{2\sigma^2} \right. \\
&\quad \left. + \frac{(i_N(x_N) - \hat{i}'_N(x_N))^2}{2\sigma^2} \right. \\
&\quad \left. + \sum_{s=1}^{N-1} \left(\delta_s(x_s) \cdot \ln \frac{P_D \cdot \phi}{(1 - P_D)(2\pi\sigma^2)^{\frac{1}{2}}} \right. \right. \\
&\quad \left. \left. + (1 - \delta_s(x_s)) \frac{(i_s(x_s) - \hat{i}'_s(x_s))^2}{2\sigma^2} \right) \right] \left. \right\} \\
&= \min_{\{d\} \in D} \left\{ \sum_{\{d_N\}} \delta_{0,N}(x_0, x_0 - d) \right. \tag{12}
\end{aligned}$$

$$\begin{aligned}
&\quad \cdot \ln \frac{P_D \cdot \phi}{(1 - P_D)(2\pi\sigma^2)^{\frac{1}{2}}} + \frac{1 - \delta_{0,N}(x_0, x_0 - d)}{N} \left[\frac{(i_0(x_0) - \hat{i}'_0(x_0))^2}{2\sigma^2} \right. \\
&\quad \left. + \frac{(i_N(x_0 - d) - \hat{i}'_N(x_0 - d))^2}{2\sigma^2} \right. \\
&\quad \left. + \sum_{s=1}^{N-1} \left(\delta_s \left(x_0 - \frac{s \cdot d}{N} \right) \cdot \ln \frac{P_D \cdot \phi}{(1 - P_D)(2\pi\sigma^2)^{\frac{1}{2}}} \right. \right. \\
&\quad \left. \left. + \left(1 - \delta_s \left(x_0 - \frac{s \cdot d}{N} \right) \right) \frac{\left(i_s \left(x_0 - \frac{s \cdot d}{N} \right) - \hat{i}'_s \left(x_0 - \frac{s \cdot d}{N} \right) \right)^2}{2\sigma^2} \right) \right] \left. \right\}
\end{aligned}$$

式(12)の最適化にあたっては、被撮影対象物体が滑らかであったとしてもそのような形状は事前知識として取り込まれていない。従って、式(12)を解いたとしても物理的に自然な形状となるとは限らない。そこで本論文では文献[93]

に従って、次の仮定の下で動的計画法により最適化を行う。

- (ア) 唯一性：左右の画素は互いに唯一の画素と対応する
- (イ) 単調順序性：対応画素の画像上での順序が入れ替わらない

図 30 に説明図を示す。この図には主体となるカメラ0及びカメラNのみを横軸及び縦軸に示しており、補助的な中間のカメラ s ($s = 1, \dots, N - 1$)は省略している。(a)・(c)・(e)の領域ではカメラ0の座標値が大きくなるにつれて対応するカメラNの座標値も増加している。しかし(f)のように減少するような対応は禁止されている(単調増加性)。また(b)・(d)の領域はそれぞれ片側のカメラでオクルージョンが発生しており、対応は一对多の関係にならない(唯一性)。

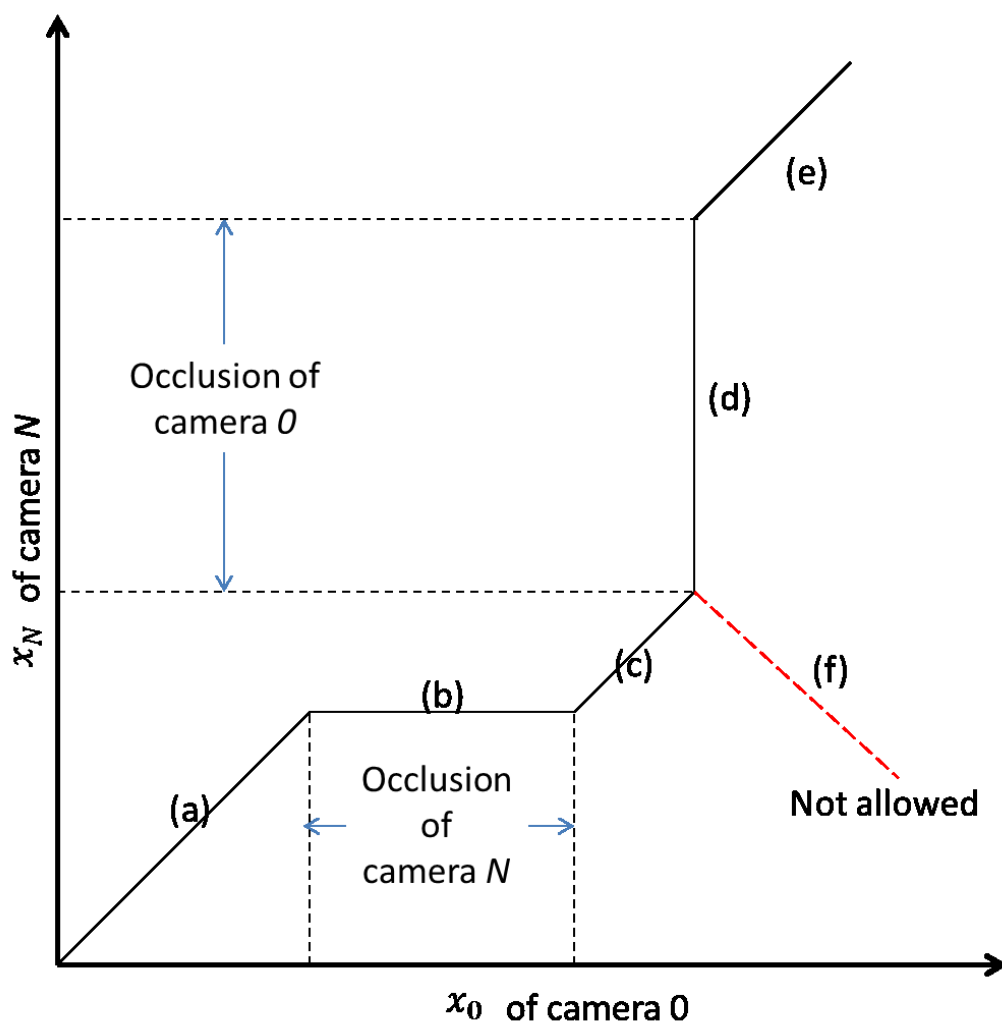


図 30 唯一性及び単調順序性仮定の下での動的計画法による最適化説明図
(文献[71] 図 1 から引用)

さらに最大の視差値を制限して計算量を抑えるとともに画像上の x 軸方向と、 y 軸方向の直前のスキャンラインとの差があるかどうかのペナルティ項を加えることで、物理的に自然な形状が得られるようにしている[93]。

また、提案のカメラ近傍の概念を取り入れたマルチベースラインステレオ法は、現代的なエネルギー最小化の枠組でも実装することができる。

カメラ0に対するカメラ N における視差のセット、あるいは視差マップ $\{d_N\}$ を推定することとする。この推定では式(13)の通りエネルギー最小化の枠組を採用する[68][69]。

$$\Phi^{\text{total}}(\{d_N\}) = \Phi^{\text{data}}(\{d_N\}) + \lambda \Phi^{\text{smoothness}}(\{d_N\}) \quad (13)$$

第一項 $\Phi^{\text{data}}(\{d_N\})$ は視差マップ $\{d_N\}$ が与えられたときの画像間マッチングエネルギーであり、画素毎のマッチングエネルギーである $\phi(\vec{x}_0, d_N)$ をカメラ0の画像上全画素に亘って式(14)の通り加算したものである。

$$\Phi^{\text{data}}(\{d_N\}) = \sum_{\vec{x}_0} \phi(\vec{x}_0, d_N) \quad (14)$$

ここで先ほどと同様に $\pm M$ で定義するカメラ近傍の概念を導入し、 $\phi(\vec{x}_0, d_N)$ を式(15)の通り定義する。なお以下では、 \vec{x}_0 から視差が発生せず本質的でない y 座標値を省略して単に x_0 と、合わせて画素毎のマッチングエネルギーを $\phi(x_0, d_N)$ とそれぞれ表記することとする。

$$\begin{aligned} \phi(x_0, d_N) &= \sum_{s=0}^N \left(\frac{1}{2M+1} \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+j}(x_{s+j}) - i_s(x_s) \right)^2 \\ &= \sum_{s=0}^N \left(\frac{1}{2M+1} \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+j} \left(x_s + \frac{j \cdot d_N}{N} \right) - i_s(x_s) \right)^2 \\ j_{\min}^{(s)} &= \begin{cases} -s & \text{if } s \leq M \\ N - 2M - s & \text{else if } s > N - M \\ -M & \text{otherwise} \end{cases} \end{aligned} \quad (15)$$

$$j_{\max}^{(s)} = \begin{cases} 2M - s & \text{if } s \leq M \\ N - s & \text{else if } s > N - M \\ M & \text{otherwise} \end{cases}$$

式(15)の第二項 $\Phi^{\text{smoothness}}(\{d_N\})$ は滑らかさコストであり、画像上における近傍画素との視差の差分に応じたコストを全画素に亘って加算したものである:

$$\begin{aligned} \Phi^{\text{smoothness}}(\{d_N\}) \\ = \sum_{x,y} [\rho(d_N(x,y) - d_N(x+1,y)) + \rho(d_N(x,y) - d_N(x,y+1))] \end{aligned} \quad (16)$$

ここで $\rho()$ は式(17)で定義する単調増加関数、 ρ^{\max} は滑らかさコストを制限するパラメータである [69]:

$$\rho(x) = \min(|x|, \rho^{\max}) \quad (17)$$

第 3 節 提案手法の実験的評価

第 1 節で説明した Middlebury ステレオデータセット 2006 の lampshade1 を評価に用いる。図 26・図 27 に示した通り大きな視点依存の輝度変動が含まれておりかつテクスチャがほとんどない、ステレオ法にとっては視差を推定することが非常に困難な画像である。

エネルギー最小化の枠組に基づくマルチベースラインステレオ法による結果を示す。エネルギー最小化には <http://vision.middlebury.edu/MRF/code/> で公開されるライブラリを利用する [69]。対照実験としては前述の実験と同様に、従来のマルチベースラインステレオ法である文献 [71] をそれぞれベースとし、動的計画法による最適化をエネルギー最小化に置き換えて実装する。最適化法は、ライブラリで利用できる全て、即ち反復条件付モード [74]・グラフカットの α 拡張及び α - β 交換 [84][85][86]・逐次信念伝播・最大積和信念伝播 [78]・逐次ツリー再重みメッセージパス [79][80] を用いる。式(13)におけるパラメータ λ は 1 から 20 まで 1 ステップで網羅的に探索する。パラメータ ρ^{\max} はエネルギー最小化ライブラリに付属したサンプルステレオプログラムの初期値である 2 を適用する。

推定した視差は正解との RMS 誤差を計算し、全ての最適化法・パラメータ λ を通して最小値を各ステレオ法の代表値とする。また提案手法における M は 1 に固定する。

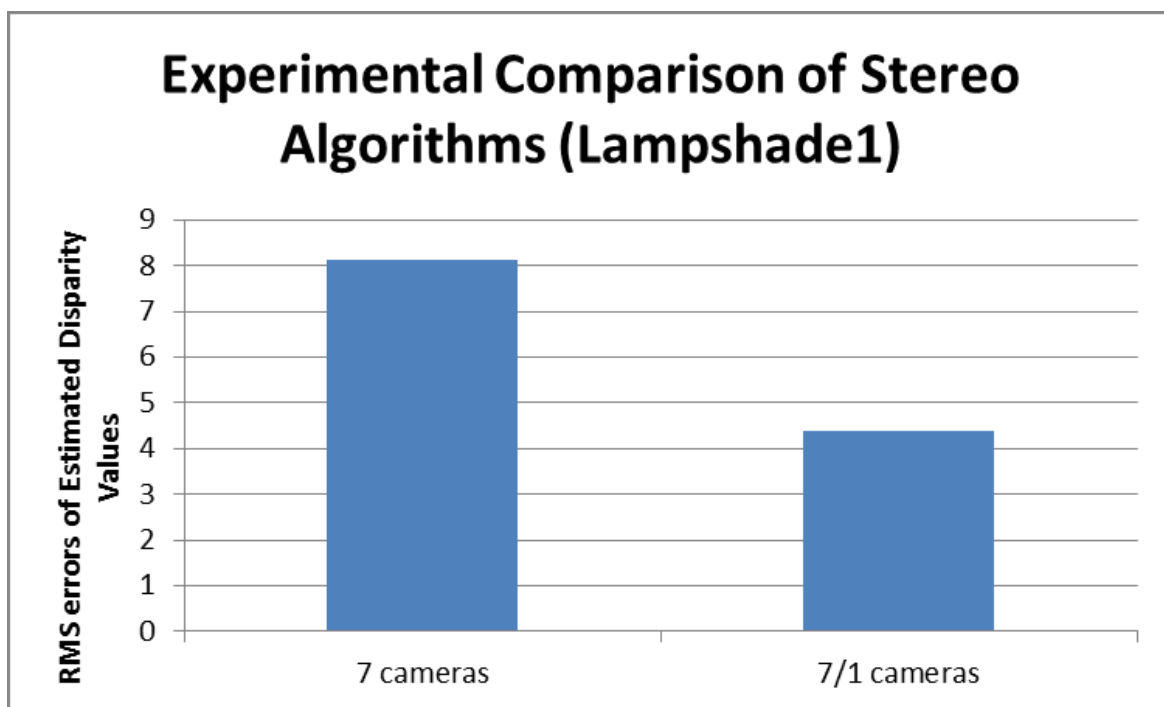


図 31 エネルギー最小化に基づく従来手法(7 cameras)と提案手法(7/1 cameras)の RMS 誤差

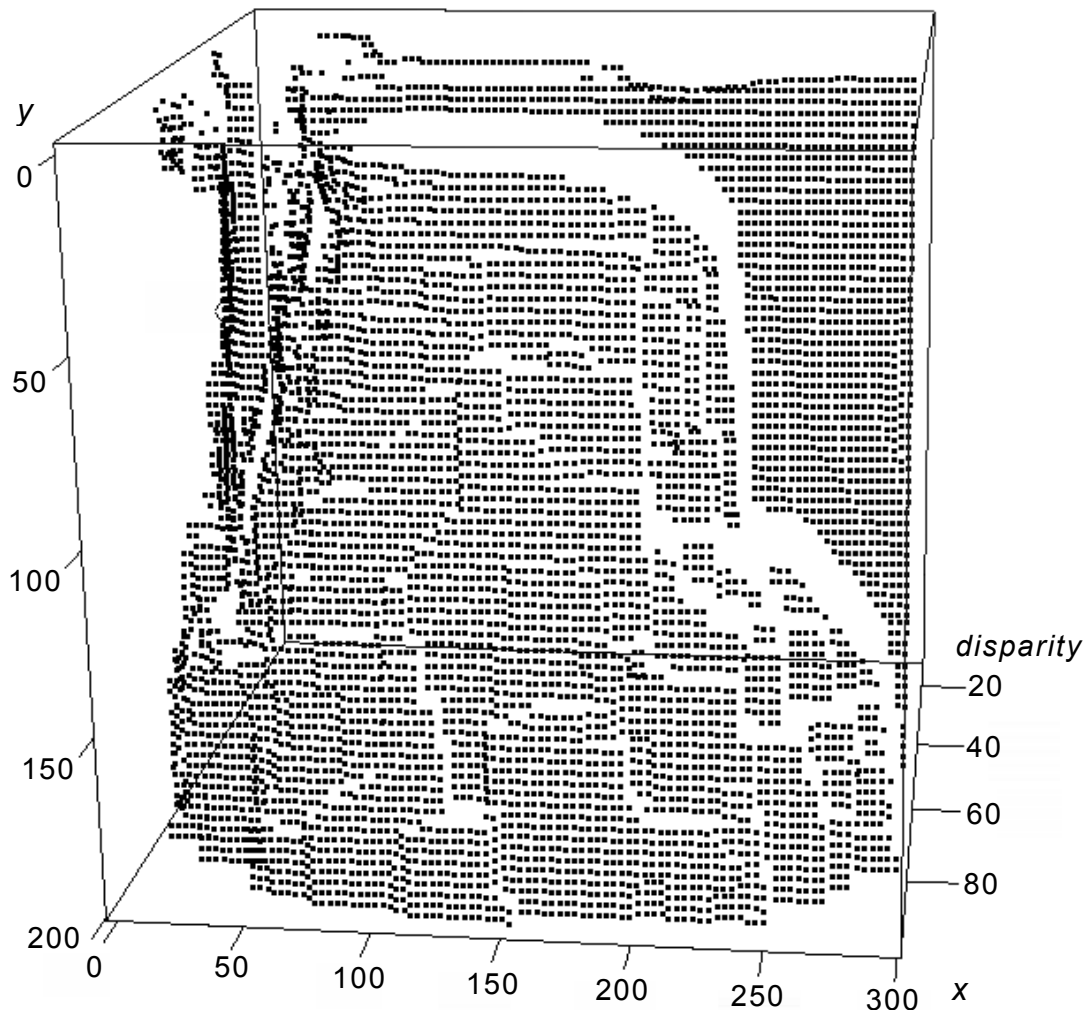


図 32 提案手法であるエネルギー最適化に基づく 7/1 cameras で推定した最良の推定視差($\lambda = 1$ 、逐次信念伝播)

図 31 に、従来手法[71]をエネルギー最小化の枠組で書き直したアルゴリズム(7 cameras と表記、7はカメラ数)と提案手法(7/1 cameras と表記、7はカメラ数・1は近傍を定義する M)の RMS 誤差を示す。これにより狙い通り大幅に改善していることがわかる。図 32 に最良の視差推定結果を示す。このとき $\lambda = 1$ 、最適化法は逐次信念伝播である。

第4章 視点依存の輝度変動に頑健なマルチベースライン法の改良と検証

第1節 完全散乱反射特性の下での理論的な RMS 評価

第3章では視点依存の輝度変動に頑健なマルチベースラインステレオ法を提案した。この提案手法を、実際に変動をもつデータに適用し、従来のマルチベースラインステレオ法と比較して良好な三次元情報推定精度が得られたことを示した。しかしこの提案手法が一般にどの程度の精度となるかについては明らかでなかった。

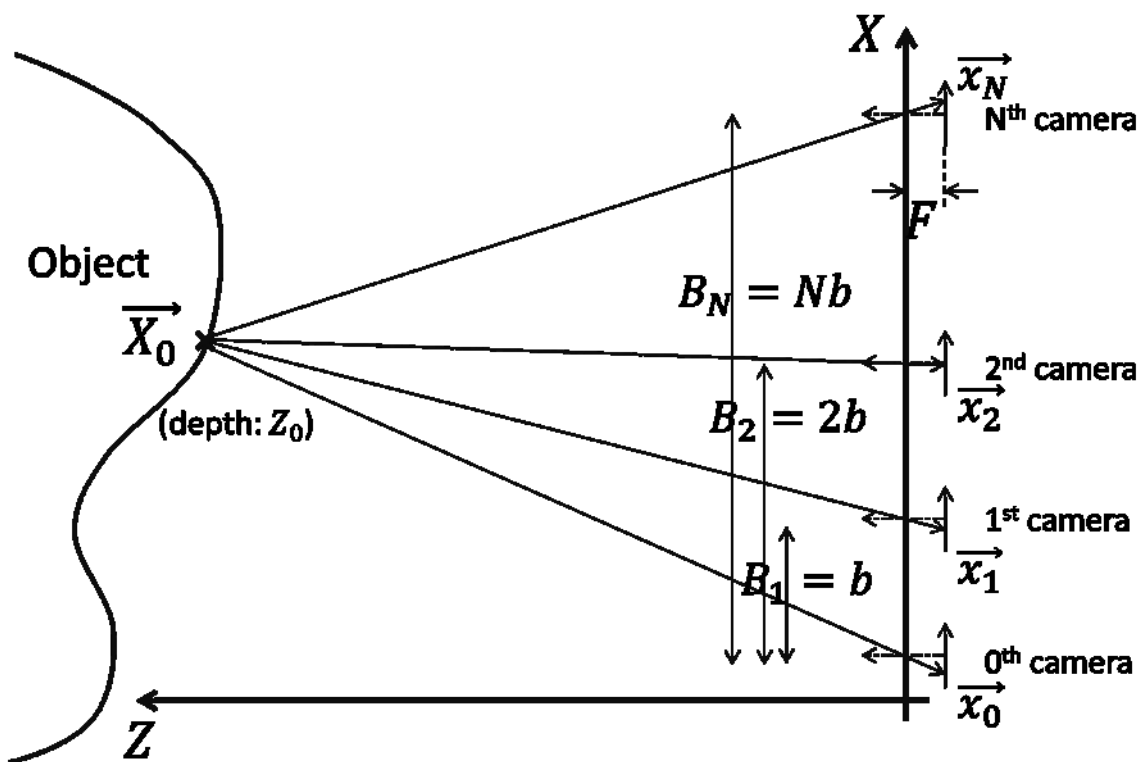


図33 マルチベースラインステレオの一般的な設定(図1を再掲)

そこで一般的な三次元情報推定精度を評価するために、本節では文献[23][24]に基づいて完全散乱反射特性の下での理論的な RMS を求める。提案手法に対する対照評価として従来法を二種類、文献[55]の二眼ステレオ法、文献[71]のマルチベースラインステレオ法を選択する。対照評価対象となるこれらは動的計画法に基づく最適化が特徴的なアルゴリズムである。しかし文献[23][24]に基づく理論的な評価では、第1章において著者が分類したステレオ法の内部処理のうち、画素/領域類似度計算部分だけを利用することになる(第3章におけるエネルギー最小化の枠組では式(15)の $\phi(x_0, d_N)$ に相当)ため、それ

ぞれの手法のうち画素単位での輝度値の差の二乗和処理だけに着目する。よって、本節における評価結果は文献[55][71]個別に対するものにとどまるわけではなく、画素単位での比較処理を採用するステレオ法全般の性質を評価していることに他ならない。また後にも述べるように、SSDなどの領域類似度を採用するステレオ法は基本的に画素単位の比較を領域の中で繰り返すことから、画素単位の比較処理を採用するステレオ法に対し統計的に領域を構成する画素数の平方根に反比例してRMSが改善する。

なお理論的な三次元情報推定精度評価にあたっては、奥行 Z_0 を直接評価するのではなく、画像処理上のエラーが表れる視差 $d_N(x_0)$ に比例する、奥行き Z_0 の逆数 $\zeta_0(x_0)$ を対象とすることで、数学的に取り扱いやすい形で評価を進める。

また第3章で提案した視点依存の輝度変動に頑健なマルチベースラインステレオ法については、RMSを計算するにあたって $N \geq 4M$ の条件下でおこなうこととした。この後の評価計算の途中で述べるように、具体的な計算を進めるには条件を定める必要が発生する。そのため、本論文で目的とする視点依存の輝度変動頑健性に合目的である、最大ベースラインに比較して近傍カメラ範囲が狭いことを条件として絞ることとした。なお計算の過程においては表記の簡略化のために、評価結果に影響のないy座標は無視することとする。

さらに以下では各アルゴリズムを簡潔に参照するため、文献[55]の手法を“2 cameras”、文献[71]の手法を“ $N+1$ cameras”、第3章で提案した手法を“ $N+1/M$ cameras”とする。なお $N+1$ 及び M は全カメラ数及び近傍カメラ定義パラメータである。

まず文献[55]の2 camerasの評価を行う。

図33において二台のカメラを最大幅のベースライン B_N を構成するよう配置する。対象物体上の奥行き Z_0 に位置する観察点 \vec{X}_0 は、カメラ0及びカメラ N の画像上におけるx座標値でそれぞれ x_0 及び x_N に投影される。 x_0 及び x_N の視差 $d_N(x_0)$ は、式(18)で計算することができる。

$$\begin{aligned} d_N(x_0) &= x_0 - x_N \\ &= NbF\zeta_0(x_0) \end{aligned} \quad (18)$$

ここで F 及び $\zeta_0(x_0)$ はカメラの焦点距離及び奥行き Z_0 の逆数である。カメラ0画像の座標 x_0 における輝度値 $i_0(x_0)$ を式(19)で定義する。

$$\begin{aligned} i_0(x_0) &= f_0(x_0) \\ &= f(x_0) + n_0(x_0) \end{aligned} \quad (19)$$

$f_0(x)$ 及び $f(x)$ はそれぞれカメラ0において観察された輝度分布関数及び、理想的な状況における輝度分布関数である。また、 $n_0(x)$ はカメラ0において、観測の度に重畳するノイズをモデル化した、平均0及び分散 σ^2 の正規分布関数である。もし物体の反射特性が完全散乱であれば、カメラ N 画像の座標 x_N における輝度値を一般的な形で $i_N(x_N) = f_N(x_N)$ と与えるとき、物体表面の反射特性が完全散乱であれば式(19)から次の式を与えることができる。

$$\begin{aligned} i_N(x_N) &= f_N(x_N) \\ &= f(x_N + NbF\zeta_0(x_0)) + n_N(x_N) \\ &= f(x_0) + n_N(x_N) \end{aligned} \quad (20)$$

$n_N(x)$ はカメラ N において、観測の度に重畳するノイズをモデル化した、平均0及び分散 σ^2 の正規分布関数である。オクルージョンの発生を無視すると、文献[55]の2 cameras におけるマッチングコストは画素単位での輝度値の差の二乗和となる。このとき視差 d_N の代わりに比例する数値である奥行き ζ_0 を用いることとし、その周りである $\zeta - \zeta_0$ での振る舞いを式(21)として示す。この式は、式(15)における $\phi()$ に対応する。

$$\begin{aligned} \phi(x_0, \zeta) &= \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_0(x_0) \right)^2 + \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_N(x_N) \right)^2 \\ &= \frac{1}{2} [f_0(x_0) - f_N(x_N)]^2 \\ &= \frac{1}{2} [f(x_0) + n_0(x_0) - f(x_0 - NbF(\zeta - \zeta_0)) - n_N(x_N)]^2 \end{aligned} \quad (21)$$

$\zeta - \zeta_0$ が十分小さい場合、 $f(x)$ は x_0 のまわりで線形関数により近似することができ、式(21)は式(22)と変形できる。なお以降では、正規分布関数 $n_0(x)$ 及び $n_N(x)$ を n_0 及び n_N と簡略表記する。

$$\begin{aligned} \phi(x_0, \zeta) &\cong \frac{1}{2} \{ [NbFf'(x_0)]^2 (\zeta - \zeta_0)^2 + 2NbFf'(x_0) (\zeta - \zeta_0) (n_0 - n_N) \\ &\quad + (n_0 - n_N)^2 \} \end{aligned} \quad (22)$$

式(22)を ζ で微分し0とおくことで、奥行き ζ_0 の推定値 $\hat{\zeta}_0$ を式(23)により

得ることができる。

$$\hat{\zeta}_0 = \zeta_0 - \frac{n_0 - n_N}{NbFf'(x_0)} \quad (23)$$

式(23)の分子は正規分布関数の差であるため偏差は存在せず、よって推定値 $\hat{\zeta}_0$ が不偏推定量であることがわかる。またその分散は式(24)と導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{2\sigma^2}{(Nb)^2[Ff'(x_0)]^2} \quad (24)$$

Nb はベースライン長 B_N と合致することから、2 cameras[55]の精度はベースライン長の自乗に反比例することがわかる。SSD を対象とした文献[23][24]ではこの評価値から、さらに領域を構成する画素数だけ小さくなっており、統計的に両者の結果はコンシステントである。なお一般にカメラから観測する物体の表面の小領域はカメラ毎に異なっており、視差が大きくなるにつれてその影響は大きくなるため、実際に得られる推定値が領域を構成する画素数だけ改善するとは限らない[54]。

次に文献[71]で示された $N+1$ cameras の評価を行う。

図 33 において $N+1$ 台のカメラを間隔 b で、座標 $X_0 \cdot X_1 \cdot \dots \cdot X_N$ に配置する。物体上の奥行き Z に位置する観察点 \vec{X} は、カメラ $0 \cdot 1 \cdot \dots \cdot N$ の画像上でそれぞれ $x_0 \cdot x_1 \cdot \dots \cdot x_N$ に投影される。オクルージョンの発生を無視すると、文献[71]の $N+1$ cameras におけるマッチングコストは式(25)で与えることができる。

なお文献[71]ではカメラは N 台であったが本論文では一貫して $N+1$ 台として取り扱っている。これは本節における理論的な評価で、数式がやや簡略化することができるのが理由である。

$$\phi(x_0, \zeta) = \sum_{s=0}^N \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_s(x_s) \right)^2 \quad (25)$$

式(25)を式(19)・(20)により式展開した後に、 $\zeta - \zeta_0$ が十分小さいことを仮定することで $f(x)$ を x_0 のまわりで線形関数により近似することが可能である。その結果として式(26)を得る。なお、正規分布関数 $n_0(x)$ 、 $n_s(x)$ 及び $n_N(x)$ を n_0 、 n_s 及

び n_N と簡略表記した。

$$\begin{aligned} \phi(x_0, \zeta) \cong & \frac{1}{4} \sum_{s=0}^N \left\{ [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \sum_{s=0}^N (N - 2s)^2 \right. \\ & - 2bFf'(x_0)(\zeta - \zeta_0) \sum_{s=0}^N (N - 2s)(n_0 + n_N - 2n_s) \\ & \left. + \sum_{s=0}^N (n_0 + n_N - 2n_s)^2 \right\} \end{aligned} \quad (26)$$

$\sum_{s=0}^N (N - 2s) = 0$ であることに注意するとともに整数列に関する和の公式[97]を利用することで、式(26)は式(27)の通り変形できる。

$$\begin{aligned} \phi(x_0, \zeta) \cong & \frac{N(N+1)(N+2)}{12} [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \\ & + bFf'(x_0)(\zeta - \zeta_0) \sum_{s=0}^N (N - 2s)n_s + \frac{1}{4} \sum_{s=0}^N (n_0 + n_N - 2n_s)^2 \end{aligned} \quad (27)$$

式(27)を ζ で微分し0とおくことで、奥行きの変数の推定値 $\hat{\zeta}_0$ を式(28)により得ることができる。

$$\hat{\zeta}_0 = \zeta_0 - \frac{6 \sum_{s=0}^N (N - 2s)n_s}{(N+1)(N+2)} \cdot \frac{1}{NbFf'(x_0)} \quad (28)$$

式(28)はやや複雑な形をしているが、分子は正規分布関数の和算であることから、推定値 $\hat{\zeta}_0$ が不偏推定量であることは明らかである。また、その分散は式(29)と導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{12N}{(N+1)(N+2)} \cdot \frac{\sigma^2}{(Nb)^2 [Ff'(x_0)]} \quad (29)$$

よって $N+1$ cameras[71]は 2 cameras[55]に比較し、誤差の分散が $6N/(N+1)(N+2)$ 倍小さい。この係数は $1/N$ に比例することから統計的にカメラの台数分だけ精度が向上すると解釈できる。また文献[23][24]の結果とは、画素値の比較関数が若干異なることが原因で生じる係数の違いに加え、文献[23][24]では SSD を評価対象としているため、これら文献の評価値では領域を構成する画素の数だけ小さくなっていることを含めてコンシステントである。

本節の最後に第 3 章で提案した $N+1/M$ cameras を評価する。第 3 章では視点に依存する輝度変化に対応するために、近傍カメラの概念を新しく導入したが、これまでこのような手法に対する理論的な評価が行われたことはなかった。

カメラは文献[71]に対する評価時と同じ設定とする。オクルージョンの発生を無視すると、 $N+1/M$ cameras におけるマッチングコストは式(30)で与えることができる(式(15)の再掲、但し視差 d_N を奥行の逆数である ζ に置き換え)。

$$\phi(x_0, \zeta) = \sum_{s=0}^N \left(\frac{1}{2M+1} \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+j}(x_{s+j}) - i_s(x_s) \right)^2 \quad (30)$$

式(30)を式(19)・(20)を利用して式展開した後、 $\zeta - \zeta_0$ が十分小さいことを仮定して $f(x)$ を x_0 のまわりで線形関数により近似することで、式(31)を得る。

$$\begin{aligned} \phi(x_0, \zeta) \cong & \frac{1}{(2M+1)^2} \sum_{s=0}^N \left\{ [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \left(\sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} j \right)^2 \right. \\ & - 2bFf'(x_0) (\zeta - \zeta_0) \left(\sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} j \right) \left[\sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} (n_{s+j} - n_s) \right] \\ & \left. + \left[\sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} (n_{s+j} - n_s) \right]^2 \right\} \quad (31) \end{aligned}$$

ここでまず N が $2M$ より大きいことを仮定するとともに、整数列の和の公式[97]を利用することで、式(31)は式(32)の通り変形できる。

$$\begin{aligned}
\phi(x_0, \zeta) \cong & \frac{1}{(2M+1)^2} \\
& \cdot \left[\frac{M(M+1)(2M+1)^3}{3} [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \right. \\
& - (2M+1)bFf'(x_0)(\zeta - \zeta_0) \\
& \cdot \left. \left\{ \sum_{k=0}^M [2(2M+1)k - M(3M+1)]n_k \right. \right. \\
& + M(M+1) \sum_{k=1}^M (n_{M+k} - n_{N-2M+k-1}) \\
& \left. \left. - \sum_{k=0}^M [2(2M+1)(M-k) - M(3M+1)]n_{N-M+k} \right\} \right. \\
& \left. + \sum_{s=0}^N \left[\sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} (n_{s+j} - n_s) \right]^2 \right]
\end{aligned} \tag{32}$$

式(32)を ζ で微分し0とおくことで、奥行きの変数の推定値 $\hat{\zeta}_0$ を式(33)により得ることができる。

$$\begin{aligned}
\hat{\zeta}_0 = \zeta_0 + & \left\{ \sum_{k=0}^M [2(2M+1)k - M(3M+1)]n_k \right. \\
& + M(M+1) \sum_{k=1}^M (n_{M+k} - n_{N-2M+k-1}) \\
& \left. - \sum_{k=0}^M [2(2M+1)(M-k) - M(3M+1)]n_{N-M+k} \right\} \\
& \cdot \frac{3}{(M+1)(2M+1)^2} \cdot \frac{1}{2MbFf'(x_0)}
\end{aligned} \tag{33}$$

式(33)は複雑な形をしているが第二項の分子は正規分布関数の和演算しかないことから、推定値 $\hat{\zeta}_0$ が不偏推定量であることは明らかである。さらに N が $4M$ より大きいと仮定することで、その分散を式(34)の通り導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{6M(5M^2 + 5M + 2)}{(M+1)(2M+1)^3} \cdot \frac{\sigma^2}{(2Mb)^2 [Ff'(x_0)]^2} \tag{34}$$

また N が $4M$ に等しい場合には、その分散を式(35)の通り導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{12M(5M^3 + 6M^2 + 3M + 1)}{(M+1)(2M+1)^4} \cdot \frac{\sigma^2}{(2Mb)^2 [Ff'(x_0)]^2} \tag{35}$$

式(34)及び式(35)ともに複雑な形をしているが両方ともカメラ台数に関する N に依存せず、第一因数を無視すればベースライン $2Mb$ の二眼ステレオ法と同程度であることが明らかとなった。つまり $N+1$ 台のカメラを用いても、アルゴリズム上直接輝度値を比較する対象である、近傍カメラだけのステレオ法と同程度の精度しか得られないことがわかる。これは画素での輝度値比較からSSDなど領域での比較へ変更しても状況を変えることはできない。従って、最大ベースラインに応じた三次元情報推定精度を得るためには、アルゴリズムの変更が必要であることが明らかとなった。

なお第一因数は $1/M$ に比例することから、近傍で直接比較するカメラ台数分だけ精度が向上すると解釈できる。

表2に評価対象である3種類のステレオ法のコスト関数を、表3にそれぞれ

れの奥行き逆数の推定値 $\hat{\zeta}_0$ の標準偏差をまとめるとともに、以下で総括する。

表 3 よりから次を述べる事ができる: $N + 1$ cameras は 2 cameras に比較し、RMS 誤差が $\sqrt{6N/(N+1)(N+2)}$ 倍小さい。この係数は $1/\sqrt{N}$ に比例しており、統計的にカメラ台数分だけ精度が向上すると解釈でき、また文献[23][24]の結果とコンシステントである。

$N + 1/M$ cameras の RMS 誤差は複雑であるものの、カメラ台数に関する N に依存せず、第一因数を無視すればベースライン $2Mb$ の 2 cameras と同等であることがわかる。つまり $N + 1$ 台のカメラを用いても、直接輝度値を比較する近傍カメラだけのステレオ法と同程度の精度しか得られないことが明らかである。なお第一因数は $1/\sqrt{M}$ に比例することから、アルゴリズム上近傍で直接比較するカメラ台数分だけ 2 cameras より統計的に精度が向上すると解釈できる。

表 2 RMS 評価対象となるステレオ法のコスト関数

Stereo algorithms	Cost functions
2 cameras[55]	$\left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_0(x_0)\right)^2 + \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_N(x_N)\right)^2$ <p style="text-align: right;">...Eq.(21)</p>
$N + 1$ cameras[71]	$\sum_{s=0}^N \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_s(x_s)\right)^2$ <p style="text-align: right;">...Eq.(25)</p>
$N + 1/M$ cameras (Chapter 3)	$\sum_{s=0}^N \left(\frac{1}{2M+1} \sum_{j=J_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+j}(x_{s+j}) - i_s(x_s)\right)^2$ <p style="text-align: right;">...Eq.(30)</p>

表 3 完全散乱仮定の下での理論評価比較結果 (奥行き逆数の RMS 誤差、 $N + 1/M$ cameras は $N \geq 4M$ 条件下)

Stereo algorithms	Theoretical RMS errors of inversed depth values without a constant term $\sigma/Ff'(x_0)$
2 cameras[55]	$\frac{\sqrt{2}}{Nb}$ <p style="text-align: right;">...Eq.(24)</p>
$N + 1$ cameras[71]	$2 \sqrt{\frac{3N}{(N+1)(N+2)}} \cdot \frac{1}{Nb}$ <p style="text-align: right;">...Eq.(29)</p>
$N + 1/M$ cameras (Chapter 3)	<p>If $N > 4M$,</p> $\sqrt{\frac{6M(5M^2 + 5M + 2)}{(M+1)(2M+1)^3}} \cdot \frac{1}{2Mb}$ <p style="text-align: right;">...Eq.(34)</p> <p>else if $N = 4M$,</p> $\sqrt{\frac{12M(5M^3 + 6M^2 + 3M + 1)}{(M+1)(2M+1)^4}} \cdot \frac{1}{2Mb}$ <p style="text-align: right;">...Eq.(35)</p>

図 34 は $N + 1$ が 9、 M が 1 または 2 の場合における、9 cameras で正規化した RMS 誤差値である。この図から 9/1 cameras は 9 cameras よりも 4 倍、2 cameras よりも 3 倍劣ることがわかる。9/2 cameras は 9/1 cameras より改善するものの、依然として 2 cameras に劣っている。

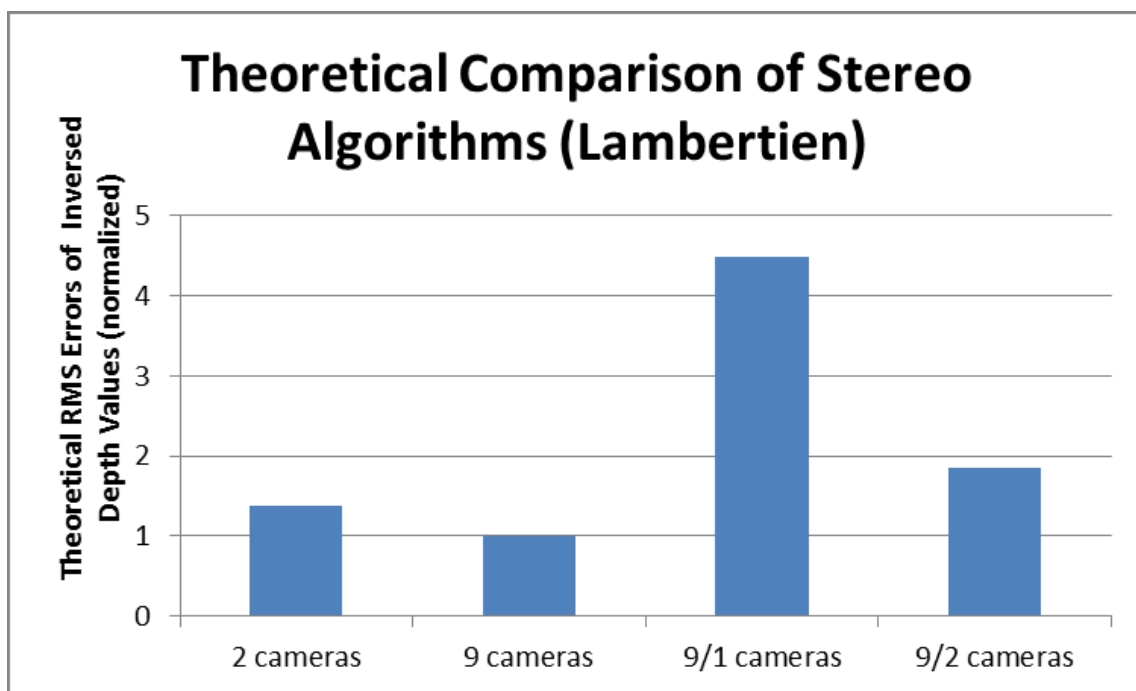


図 34 完全散乱仮定の下での理論評価値の比較結果(奥行き逆数の RMS 誤差、9 cameras[71]で正規化した値)

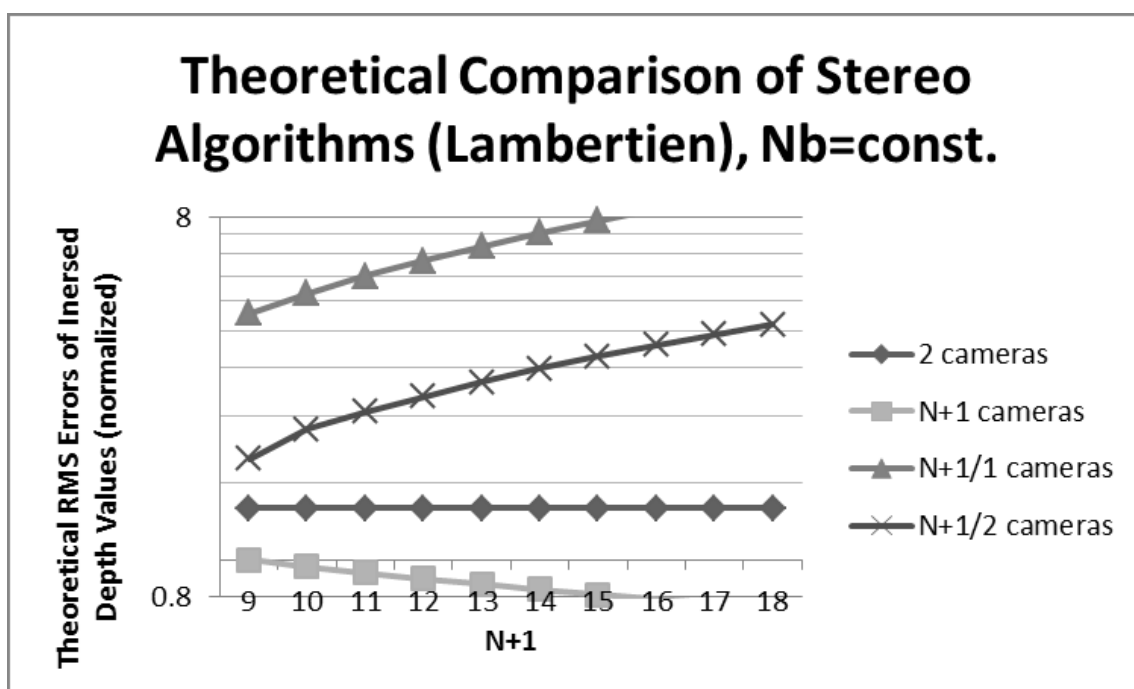


図 35 完全散乱仮定の下での理論評価値の比較結果(奥行き逆数の RMS 誤差、9 cameras[71]で正規化した値、横軸は全カメラ数)

ベースライン Nb を固定、かつ M を 1 または 2 即ち近傍カメラ数を固定した条件で、全カメラ数による RMS 誤差依存性を図 35 に示す。図 34 と同様に値は 9

cameras で正規化した。 Nb が固定なので2 cameras の RMS 誤差は一定となる。全カメラ数 $N+1$ を増やすことで、近傍カメラだけを考えたときのベースライン $2Mb \propto 2M/N$ が相対的に小さくなることから、第3章で提案するマルチベースラインステレオ法に対する視点依存の輝度変動の影響は小さくなっていくはずであるが、実際には $N+1/1$ cameras 及び $N+1/2$ cameras の理論的な RMS 誤差の評価結果は劣化していくことが明確となった。

以上の理論的解析を通じ、第3章で提案した視線依存の輝度変動に頑健なマルチベースラインステレオ法はよい奥行き推定精度をもたないことが明らかとなった。従ってベースライン Nb に従った精度が得られるように、何らかのアルゴリズムの改良を行うことが必要である。

第2節 視点依存の輝度変動を含む画像を用いた RMS 誤差実験評価 その1

ここで更に分析を進めるために、第3章で用いた Middlebury データセット 2006 よりも枚数が多く(9枚)、テクスチャが豊富で最適化の手法による結果への影響が小さい画像セットを利用する。

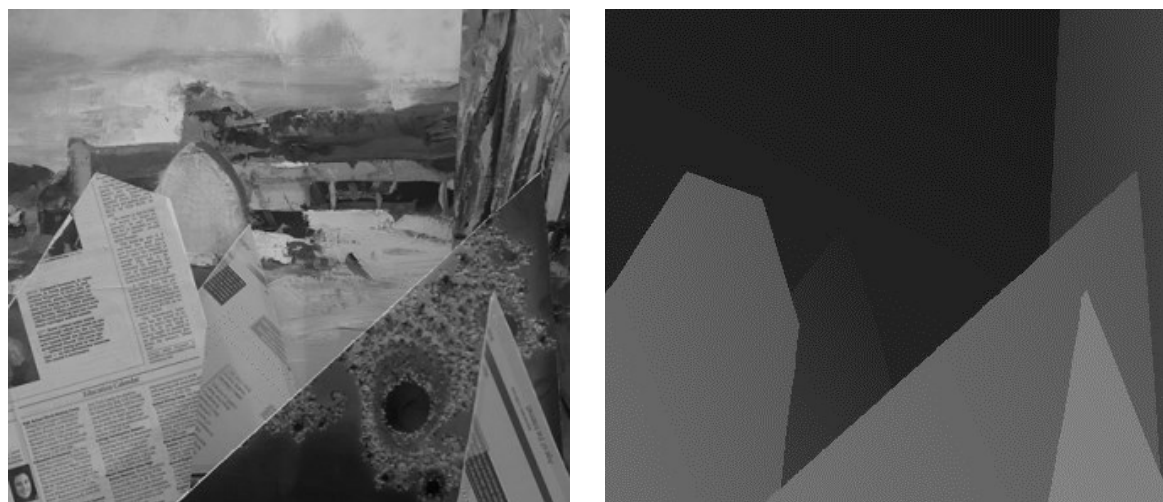


図36 Middlebury データセット 2001 内の barn1 画像[3]。左: 左端位置から撮影した画像 / 右: 左から3番目の位置から撮影した画像から同7番目の画像に対する視差画像

文献[3]に詳述された Middlebury ステレオデータセット 2001 の中から、barn1 を用いる。この画像セットはカメラを水平移動する架台に設置し、図33に示したように一定間隔で撮影した9枚で構成されている。高解像度デジタルカメラはマニュアル露出・マニュアルフォーカスに設定されており、撮影画像は 512×384 画素にダウンサンプリングされた。画像には平面物体が写っており、正しい視差マップは手作業でラベル付して求められた。図36に

配布された画像の一部を示す。図 36 左は左端位置から撮影した画像、図 36 右は左から 3 番目の位置から撮影した画像から同 7 番目の画像に対する視差画像であり、1/8画素精度をもつ。

評価の前に、まず図 36 右の視差画像を、左端撮影の画像における右端の画像との間の視差画像へ変換する。図 36 右の視差画像の各画素には 3 番目の画像から 7 番目の画像への視差(ベースライン4b)が格納されていることから、左端画像状の対応画素は画像に向かって右に視差の半分だけ移動した位置に相当する(ベースライン2b)。この対応画素へ格納すべき右端画像から左端画像(ベースライン8b)への視差は 2 倍の値となる。

次に変換した視差画像上で左右に隣り合った画素の視差値を比較し、向かって右の画素の視差が大きい場合は左の画素位置を起点にして、右がその視差の幅だけ左方向に隠蔽されると判定する。また被撮影平板の境界は右端画素位置に正確に対応しないため、左右隣り合った画素の視差値の差が閾値以上の場合には取り除いた。変換して得た視差画像を、図 36 右と同じ輝度となるよう正規化して図 37 左に示す。図 37 右は平板の傾きがわかるよう強調したものである。

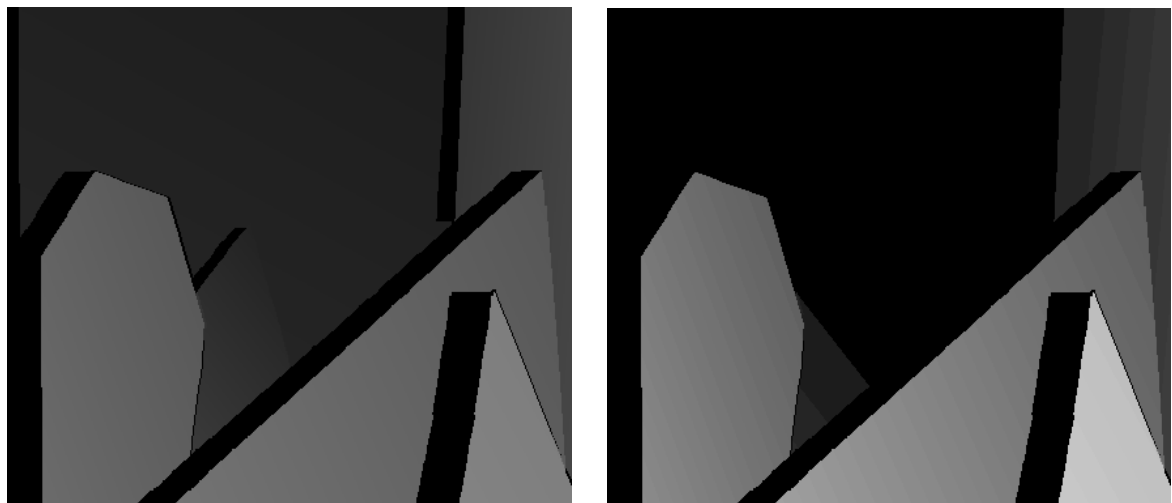


図 37 左端画像を基準とした右端画像に対する視差画像
(黒は右端画像ではオクルージョンで観測できない、あるいは平板境界で評価に適しないとして除去した画素)
左: 視差画像 / 右: 視差画像を平板の傾きがわかるよう強調

求めた視差値により画像間の対応点が正しく定まる。この正しい対応点間の輝度変動を評価する。図 38 に左端画像を基準とした対応点間の輝度差を示す。図 38 上段は 3 番目の、図 38 下段は右端の画像との輝度差であり、それぞれ左は輝度差が+・右は-をそれぞれ意味する(輝度差は 20 倍に強調)。全体に明るく・暗くなる傾向や完全に位置を合わせ込むことが困難な細かなテク

スチャ部分の他に、完全散乱では説明できない部分的な明るく・暗くなる場所があることがわかる。

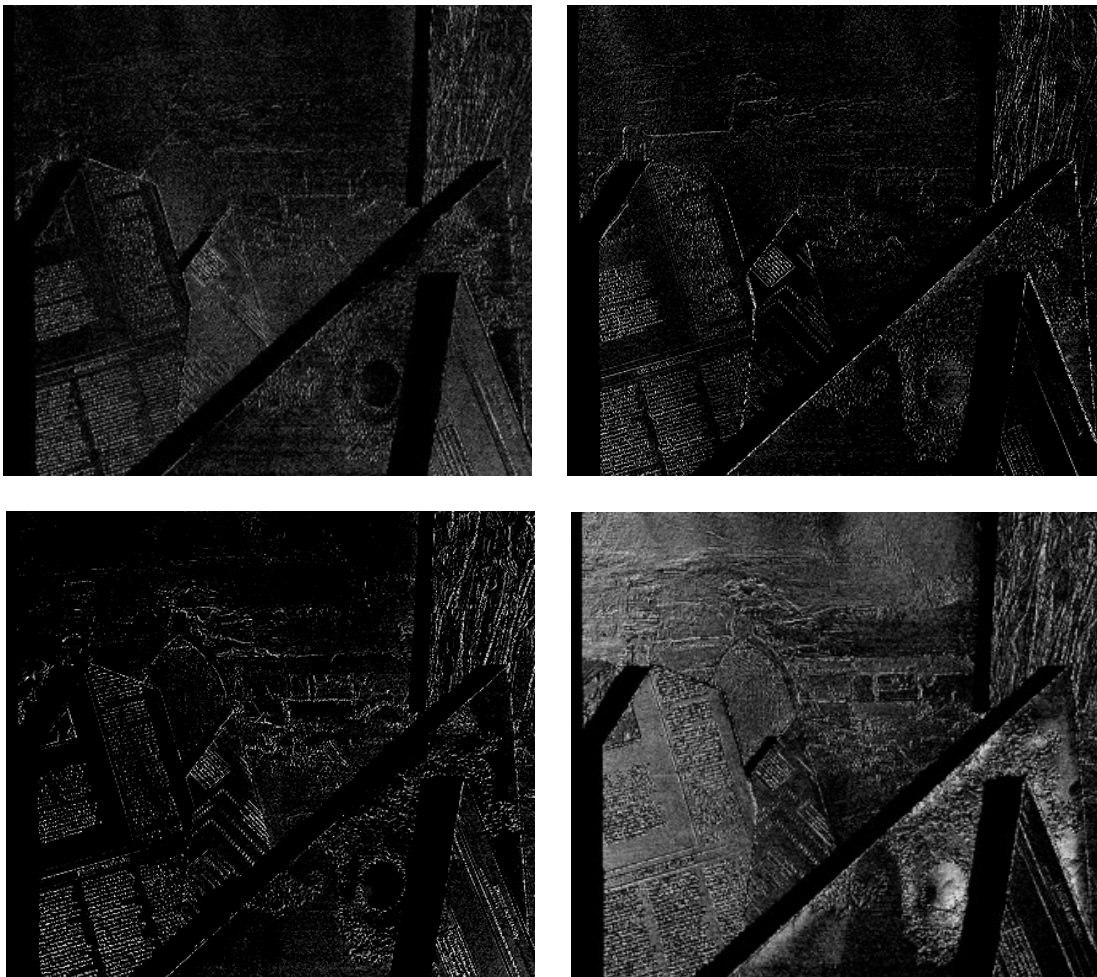


図 38 左端画像輝度を基準とした対応点間の輝度差、上段: 3 番目の画像との差 / 下段: 右端画像との差 / 左: 輝度差+ / 右: 輝度差- (実際の輝度差を 20 倍表示)

この barn1 に対して 2 cameras[55]・9 cameras[71]・9/1 cameras・9/2 cameras(第 3 章提案)をそれぞれ適用する。エネルギー最小化には <http://vision.middlebury.edu/MRF/code/>で公開されるライブラリを利用する[69]。最適化法は第 3 章における評価実験と同様に、ライブラリで利用できる全て、即ち反復条件付モード[74]・グラフカットの α 拡張及び α - β 交換[84][85][86]・逐次信念伝播・最大積和信念伝播[78]・逐次ツリー再重みメッセージパス[79][80]を用いる。式(6)におけるパラメータ λ は 2 cameras・9 cameras で 5 から 600 まで 5 ステップ、9/1 cameras・9/2 cameras で 5 から 400

まで5ステップで網羅的に探索する。パラメータ ρ^{\max} はエネルギー最小化ライブラリに付属したサンプルステレオプログラムの初期値である2を適用する。

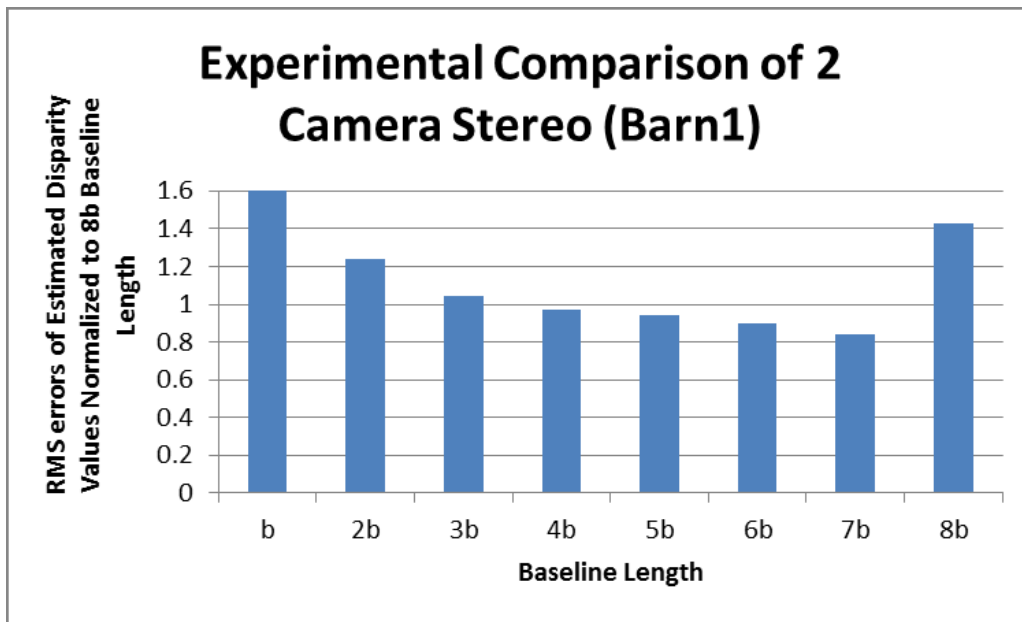


図 39 ベースライン b から $8b$ における 2 cameras[55] による推定視差の RMS 誤差、最大ベースライン $8b$ で正規化

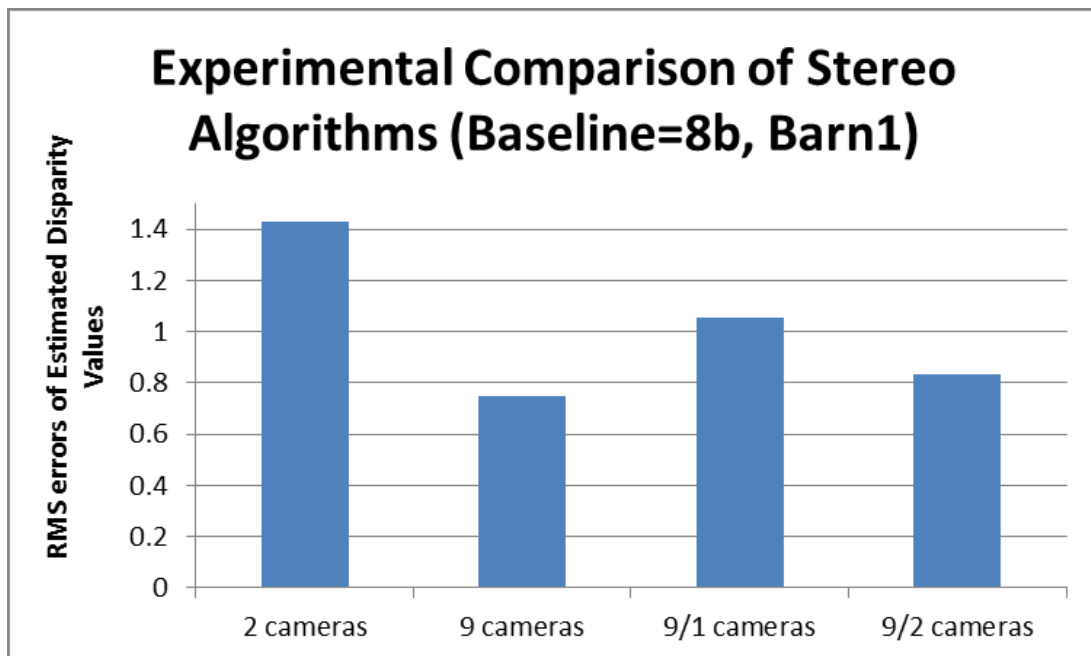


図 40 ベースライン $8b$ における複数のステレオ法による推定視差の RMS 誤差

推定した視差は正解との RMS 誤差を計算し、全ての最適化法・パラメータ λ を通して最小値を各ステレオ法の代表値とすることで、ステレオ法それ自体の能力を最大限評価する。なお視差は奥行きに比例することが式(18)から明らかである。従って、理論評価の奥行き逆数の RMS 誤差と実験評価の推定視差の RMS 誤差は相対値で比較可能である。同様に式(18)より、視差はベースライン長に比例することから、異なるベースラインの RMS 誤差を直接比較可能となるように最大ベースライン $8b$ 相当にそれぞれ変換する。

図 39 に 2 cameras で推定した視差の RMS 誤差を示す。ベースラインが b から $7b$ へ大きくなるとともに RMS 誤差は小さくなっていく。しかし $8b$ になると対応点の推定が困難となり RMS 誤差は増大する。Barn1 の撮影対象はテクスチャに富んでおり多少の輝度変動が存在してもある程度の正しい視差推定は可能であるが、正しい視差位置であってもマッチングコストが十分下がらないために探索範囲が広がることで誤対応が増加すると考えられる。

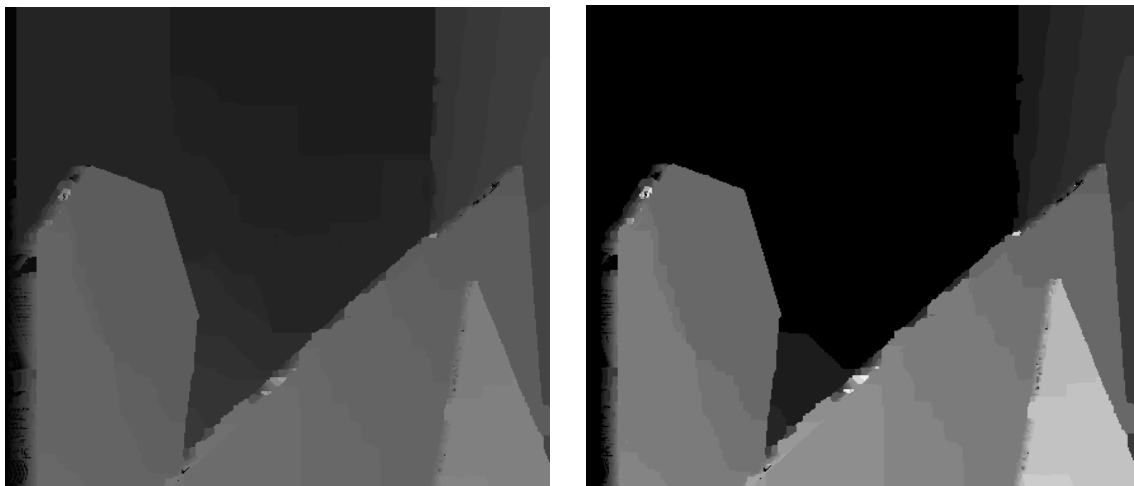


図 41 最良の推定視差結果 9/2 cameras ($\lambda = 25$ 、グラフカット α 拡張)、左: 視差画像 / 右: 視差画像を図 39 右と同じパラメータで強調

図 40 にベースライン $8b$ の際の各手法の結果を示す。図 40 の左ふたつは 2 cameras 及び 9 cameras の RMS 誤差である。文献[23]で述べられる通り、マルチベースラインステレオにより誤対応が低減できる効果が現れていると考えられる。

図 40 の次のふたつは 9/1 cameras 及び 9/2 cameras の RMS 誤差である。第 1 節で述べた理論的評価結果の通り、これらふたつの結果は 9 cameras に劣っている。しかし理論的評価値に比較して差は大きく狭まっていることが

わかる。これは前述の通り、現実の画像は理論的評価が仮定した完全散乱から外れていることによる影響と考えることができる。

図 41 に、ベースライン $8b$ のときに $9/2$ cameras で推定した最良の視差を示す。最適化手法はグラフカット α 拡張で $\lambda = 25$ である。図 41 右は差が確認できるよう強調した画像であり、図 37 と比較すると全体的にはよく推定できているものの、平板がなめらかとなっていないことがわかる。

第 3 節 視点依存の輝度変動に頑健なマルチベースラインステレオ法の改良

前節までの結果を踏まえて、カメラ近傍での輝度変動拘束を強めるよう新たに改良したマルチベースラインステレオ法を提案する。本節でアルゴリズムの提案を行った後、その理論的な三次元情報推定精度を第 4 節で、また実験的な同精度を第 5 節でそれぞれ示していく。

カメラ設定は第 3 章と全く同じである。図 33 の通り $N+1$ 台のカメラが等間隔 b で X 軸上に並んでいるとする。全てのカメラ光軸は Z 軸に並行であり、同一サイズの画像を出力する。各画像の x 軸はベースラインに並行とし、視差は x 軸にだけ現れるとする。物体上の、奥行き Z である観察点 \vec{X} は、カメラ i が出力する画像上で x_i に投影される。

ここでカメラ 0 に対するカメラ N における視差のセット、あるいは視差マップ $\{d_N\}$ をエネルギー最小化の枠組で推定する [68][69]。第 3 章のエネルギー最小化の枠組における提案手法の式 (15) に代えて、 $\phi(x_0, d_N)$ を式 (36) の通り定義する。

$$\phi(x_0, d_N) = \sum_{s=0}^N \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} \left[\frac{1}{2M+1} \sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} i_{s+j+l}(x_{s+j+l}) - i_s(x_s) \right]^2 \quad (36)$$

$$j_{\min}^{(s)} = \begin{cases} -s & \text{if } s \leq M \\ N - 2M - s & \text{else if } s > N - M \\ -M & \text{otherwise} \end{cases}$$

$$j_{\max}^{(s)} = \begin{cases} 2M - s & \text{if } s \leq M \\ N - s & \text{else if } s > N - M \\ M & \text{otherwise} \end{cases}$$

この式(36)は、第3章の式(15)と同様に $\pm M$ で定義する近傍カメラ内でマッチングコストを計算する。しかし式(15)では、近傍における輝度変動が直線的である場合全体のマッチングコストには影響しないよう定式化してあった。

図42は式(36)における計算ステップ例である。 $M=1$ のとき、カメラ3のマッチングコストは該カメラ3の輝度に対して、カメラ2からカメラ4の平均輝度とだけでなく、カメラ1からカメラ3の平均輝度、カメラ3からカメラ5の平均輝度の自乗差も加えて計算する。従ってカメラ位置に対し線形の輝度変動には影響を受けるようになったが、 M すなわち近傍の範囲が十分小さければ輝度変動に対する頑健性はもち続けるといえる。言い換えると、カメラ位置に対し線形の輝度変動は条件を緩和し過ぎが原因で奥行き推定精度に悪影響が出たと考えることができ、輝度変動に対してより制約の強い条件を課すアルゴリズムの提案をここに行うのである。

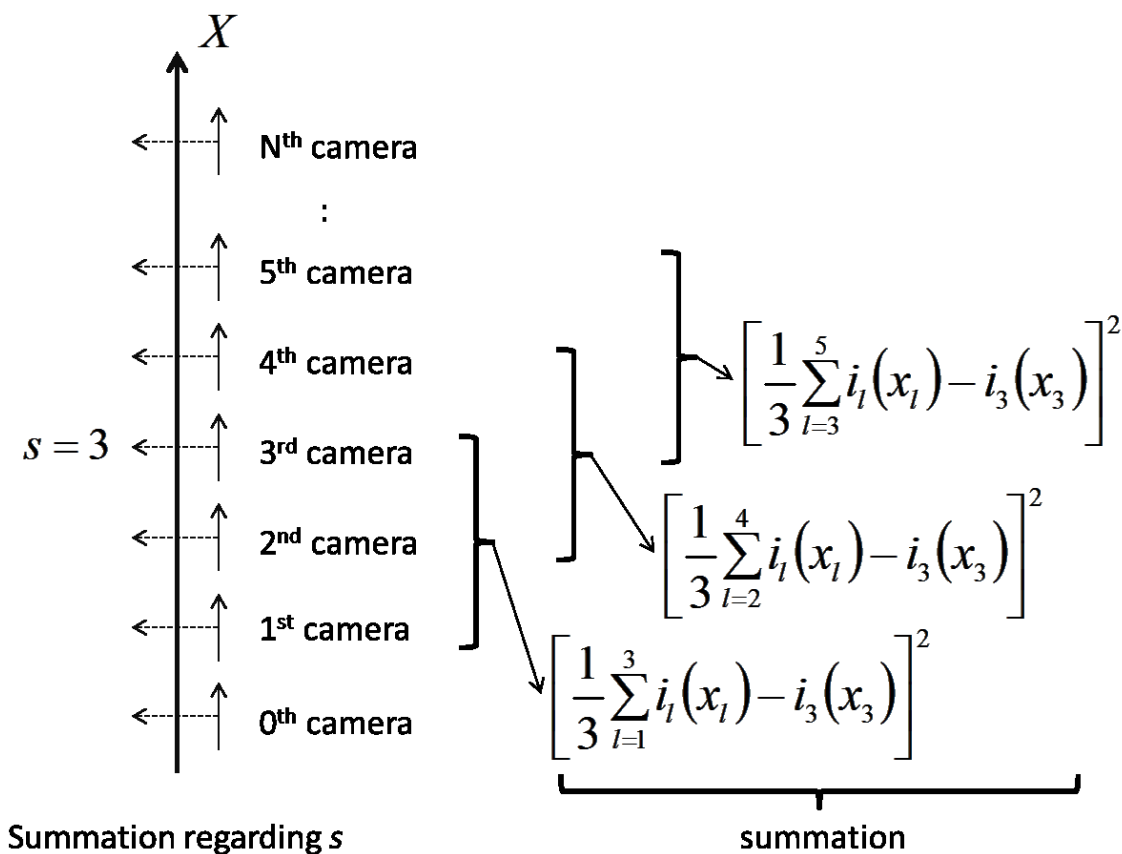


図42 提案する輝度変化に頑健なマルチベースライン法の計算ステップ ($M=1$ ・式(36))

また、式(36)では輝度の二乗差を類似度として採用しているが、近傍カメラのアイデアを生かしながら他の類似度へと変更することもできる。本論文で

は、第 1 章でまとめた類似度の中からオクルージョンやショットノイズなどの大きなノイズに対してより頑健であるとされる輝度差の絶対値を、追加で実装する。

第 4 節 完全散乱反射特性の下での改良版マルチベースラインステレオ法の理論的な RMS 評価

カメラは第 2 節の $N + 1$ cameras 及び $N + 1/M$ cameras を評価したときと同じ設定とし、類似度は解析が容易な式(36)に定義した輝度の二乗差とする。また以下では記述を簡略化するため、本章提案の手法を $N + 1/M$ cameras (rev) とする。オクルージョンの発生を無視すると、提案する $N + 1/M$ cameras (rev) におけるマッチングコストは式(37)となる。ただしここで視差 d_N の代わりに、比例する奥行の逆数である ζ に置き換えた。

$$\phi(x_0, \zeta) = \sum_{s=0}^N \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} \left[\frac{1}{2M+1} \sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} i_{s+j+l}(x_{s+j+l}) - i_s(x_s) \right]^2 \quad (37)$$

第 2 節と同様に式(37)を式(19)・(20)により式展開した後に、 $\zeta - \zeta_0$ が十分小さいことを仮定し $f()$ は x_0 のまわりで線形関数により近似することで、式(38)を得る。なお、規分布関数 $n_0(x)$ 、 $n_s(x)$ 及び $n_N(x)$ などを n_0 、 n_s 及び n_N と簡略表記した。

$$\begin{aligned}
\phi(x_0, \zeta) &\cong \frac{1}{(2M+1)^2} \\
&\cdot \sum_{s=0}^N \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} \left\{ [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \left[\sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} (j+l) \right]^2 \right. \\
&- 2bFf'(x_0) (\zeta - \zeta_0) \left[\sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} (j+l) \right] \left[\sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} (n_{s+j+l} - n_s) \right] \\
&\left. + \left[\sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} (n_{s+j+l} - n_s) \right]^2 \right\} \quad (38)
\end{aligned}$$

まず N が $2M$ より大きいことを仮定するとともに、整数列の和の公式[97]を利用することで、式(38)は式(39)の通り変形できる。

$$\begin{aligned}
\phi(x_0, \zeta) &\cong \frac{1}{(2M+1)^2} \\
&\cdot \left\{ \frac{M(M+1)(2M+1)^3(N-M+1)}{3} \cdot [bFf'(x_0)]^2 (\zeta - \zeta_0)^2 \right. \\
&+ 2(2M+1)^2 bFf'(x_0) (\zeta \\
&- \zeta_0) \left[M \sum_{k=0}^{2M} (M-k)n_k \right. \\
&+ \sum_{s=M}^{N-M} \sum_{k=s-M}^{s+M} (s-k)n_k + M \sum_{k=N-M+1}^N (N-M-k)n_k \left. \right] \\
&+ \sum_{s=0}^{M-1} \left[\sum_{l=0}^{2M} (n_l - n_s) \right]^2 \\
&\left. + \sum_{s=M}^{N-M} \left[\sum_{l=s-M}^{s+M} (n_l - n_s) \right]^2 + \sum_{s=N-M+1}^N \left[\sum_{l=N-2M}^N (n_l - n_s) \right]^2 \right\} \quad (39)
\end{aligned}$$

式(39)を ζ で微分し0とおくことで、奥行きの変数の推定値 $\hat{\zeta}_0$ を式(40)により得ることができる。

$$\hat{\zeta}_0 = \zeta_0 + \left[M \sum_{k=0}^{2M} (M-k)n_k + \sum_{s=M}^{N-M} \sum_{k=s-M}^{s+M} (s-k)n_k + M \sum_{k=N-2M}^N (N-M-k)n_k \right] \cdot \frac{3}{M(M+1)(2M+1)} \cdot \frac{1}{(N-M+1)bFf'(x_0)} \quad (40)$$

表4 式(40)の括弧[]内第二項における、 s 及び k のインデックスがついた正規分布関数と、それらの係数。 $N \geq 4M$ のとき影つきセルは加算すると0になる

s	k						
	$s-M$	$s-M+1$	$s-M+2$...	$s+M-2$	$s+M-1$	$s+M$
M	n_0	n_1	n_2	...	n_{2M-2}	n_{2M-1}	n_{2M}
$M+1$	n_1	n_2	n_3	...	n_{2M-1}	n_{2M}	n_{2M+1}
$M+2$	n_2	n_3	n_4	...	n_{2M}	n_{2M+1}	n_{2M+2}
\vdots	\vdots	\vdots	\vdots	...	\vdots	\vdots	\vdots
$3M-2$	n_{2M-2}	n_{2M-1}	n_{2M}	...	n_{4M-4}	n_{4M-3}	n_{4M-2}
$3M-1$	n_{2M-1}	n_{2M}	n_{2M+1}	...	n_{4M-3}	n_{4M-2}	n_{4M-1}
$3M$	n_{2M}	n_{2M+1}	n_{2M+2}	...	n_{4M-2}	n_{4M-1}	n_{4M}
$3M+1$	n_{2M+1}	n_{2M+2}	n_{2M+3}	...	n_{4M-1}	n_{4M}	n_{4M+1}
$3M+2$	n_{2M+2}	n_{2M+3}	n_{2M+4}	...	n_{4M}	n_{4M+1}	n_{4M+2}
\vdots	\vdots	\vdots	\vdots	...	\vdots	\vdots	\vdots
$N-3M-2$	n_{N-4M-2}	n_{N-4M-1}	n_{N-4M}	...	n_{N-2M-4}	n_{N-2M-3}	n_{N-2M-2}
$N-3M-1$	n_{N-4M-1}	n_{N-4M}	n_{N-4M+1}	...	n_{N-2M-3}	n_{N-2M-2}	n_{N-2M-1}
$N-3M$	n_{N-4M}	n_{N-4M+1}	n_{N-4M+2}	...	n_{N-2M-2}	n_{N-2M-1}	n_{N-2M}
$N-3M+1$	n_{N-4M+1}	n_{N-4M+2}	n_{N-4M+3}	...	n_{N-2M-1}	n_{N-2M}	n_{N-2M+1}
$N-3M+2$	n_{N-4M+2}	n_{N-4M+3}	n_{N-4M+4}	...	n_{N-2M}	n_{N-2M+1}	n_{N-2M+2}
\vdots	\vdots	\vdots	\vdots	...	\vdots	\vdots	\vdots
$N-M-2$	n_{N-2M-2}	n_{N-2M-1}	n_{N-2M}	...	n_{N-4}	n_{N-3}	n_{N-2}
$N-M-1$	n_{N-2M-1}	n_{N-2M}	n_{N-2M+1}	...	n_{N-3}	n_{N-2}	n_{N-1}
$N-M$	n_{N-2M}	n_{N-2M+1}	n_{N-2M+2}	...	n_{N-2}	n_{N-1}	n_N
coefficient	M	$M-1$	$M-2$		$-(M-2)$	$-(M-1)$	$-M$

式(40)は複雑な形をしているが第二項の分子は正規分布関数の和演算しかないことから、推定値 $\hat{\zeta}_0$ が不偏推定量であることは明らかである。

具体的な精度評価を行うためにさらに $N \geq 4M$ を仮定し、式(40)の第二項の括弧[]内の、さらに第二項に注目する。各正規分布関数の係数をまとめると表4となることから、この第二項は式(41)の通り整理することができる。

$$\begin{aligned} & \sum_{s=M}^{N-M} \sum_{k=s-M}^{s+M} (s-k)n_k \\ &= \sum_{i=0}^{2M-1} \left[\frac{M(M+1)}{2} - \frac{(M-i-1)(M-i)}{2} \right] n_i \\ & \quad - \sum_{i=N-2M+1}^N \left[\frac{M(M+1)}{2} - \frac{(N-M-i+1)(N-M-i)}{2} \right] n_i \end{aligned} \quad (41)$$

N が $4M$ より大きい場合には、その分散を式(42)の通り導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{6(7M^2 + 7M + 1)}{5M(M+1)(2M+1)} \cdot \frac{\sigma^2}{[(N-M+1)b]^2 [Ff'(x_0)]^2} \quad (42)$$

また N が $4M$ に等しい場合には、式(41)の括弧[]第一項及び第三項で重複項 n_{2M} が存在することに注意することで、その分散を式(43)の通り導くことができる。

$$\text{Var}(\hat{\zeta}_0) = \frac{6(14M^4 + 20M^3 + 30M^2 + 10M + 1)}{5M(M+1)^2(2M+1)^2} \cdot \frac{\sigma^2}{[(N-M+1)b]^2 [Ff'(x_0)]^2} \quad (43)$$

式(42)及び式(43)はともに複雑な形をしているが、両方ともに N に依っており、第一因子を無視すればベースライン $(N-M+1)b$ の2 camerasと同じであることがわかる。視点移動の輝度変動に頑健にするには N に対して M を十分小さく設定することが合目的であることから、ほぼベースライン Nb の2 camerasと同等といえる。第一因子は $1/M$ に比例し、第3章提案の $N+1/M$ camerasと同様に近傍で直接比較するカメラ台数分だけ2 camerasより精度が向上すると解釈できる。

表 5 評価対象となるステレオ法のコスト関数リスト (表 2 に本章提案手法を追加)

Stereo algorithms	Cost functions
2 cameras[55]	$\left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_0(x_0)\right)^2 + \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_N(x_N)\right)^2$ <p style="text-align: right;">...Eq.(21)</p>
$N + 1$ cameras[71]	$\sum_{s=0}^N \left(\frac{i_0(x_0) + i_N(x_N)}{2} - i_s(x_s)\right)^2$ <p style="text-align: right;">...Eq.(25)</p>
$N + 1/M$ cameras (Chapter 3)	$\sum_{s=0}^N \left(\frac{1}{2M + 1} \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} i_{s+j}(x_{s+j}) - i_s(x_s)\right)^2$ <p style="text-align: right;">...Eq.(30)</p>
$N + 1/M$ cameras (rev)	$\sum_{s=0}^N \sum_{j=j_{\min}^{(s)}}^{j_{\max}^{(s)}} \left(\frac{1}{2M + 1} \sum_{l=j_{\min}^{(s+j)}}^{j_{\max}^{(s+j)}} i_{s+j+l}(x_{s+j+l}) - i_s(x_s)\right)^2$ <p style="text-align: right;">...Eq.(37)</p>

表 5 に、第 1 節で評価した手法に加え本章で提案する $N + 1/M$ cameras (rev) を加えた全ての手法のコスト関数を、表 6 にそれらの奥行きの変換の理論評価結果をそれぞれまとめる。なお視点依存の輝度変動に頑健なふたつの手法は、 $N \geq 4M$ の下で具体的な評価式を計算した。本論文では視点依存の輝度変動頑健性が目的であることから、近傍カメラ範囲が狭い領域で評価することは合目的である。

表 6 から $N + 1/M$ cameras (rev) の RMS 誤差は N に依っており、第一項を無視すればベースライン $(N - M + 1)b$ の 2 cameras と同等であることがわかる。輝度変動に頑健とするには N に対して M を十分小さく設定することが合目的であることから、ほぼベースライン Nb の 2 cameras と同等といえる。第一項は $1/\sqrt{M}$ に比例し、 $N + 1/M$ cameras と同様に近傍で直接比較するカメラ台数分だけ 2 cameras より統計的に精度が向上すると解釈できる。

表 6 完全散乱仮定の下での理論評価比較結果 (奥行き逆数の RMS 誤差、 $N + 1/M$ cameras¹⁰⁾及び $N + 1/M$ cameras (rev)は、 $N \geq 4M$ 条件下、表 3 に本章提案手法を追加)

Stereo algorithms	Theoretical RMS errors of inversed depth values、 without a constant term $\sigma/Ff'(x_0)$
2 cameras[55]	$\frac{\sqrt{2}}{Nb}$ <p style="text-align: right;">...Eq.(24)</p>
$N + 1$ cameras[71]	$2\sqrt{\frac{3N}{(N+1)(N+2)} \cdot \frac{1}{Nb}}$ <p style="text-align: right;">...Eq.(29)</p>
$N + 1/M$ cameras (Chapter 3)	<p>If $N > 4M$,</p> $\sqrt{\frac{6M(5M^2 + 5M + 2)}{(M+1)(2M+1)^3} \cdot \frac{1}{2Mb}}$ <p style="text-align: right;">...Eq.(34)</p> <p>else if $N = 4M$,</p> $\sqrt{\frac{12M(5M^3 + 6M^2 + 3M + 1)}{(M+1)(2M+1)^4} \cdot \frac{1}{2Mb}}$ <p style="text-align: right;">...Eq.(35)</p>
$N + 1/M$ cameras (rev)	<p>If $N > 4M$,</p> $\sqrt{\frac{6(7M^2 + 7M + 1)}{5M(M+1)(2M+1)} \cdot \frac{1}{(N-M+1)b}}$ <p style="text-align: right;">...Eq.(42)</p> <p>else if $N = 4M$,</p> $\sqrt{\frac{6(14M^4 + 20M^3 + 30M^2 + 10M + 1)}{5M(M+1)^2(2M+1)^2} \cdot \frac{1}{(N-M+1)b}}$ <p style="text-align: right;">...Eq.(43)</p>

図 43 は $N + 1$ が9、 M が1または2の場合における、9 cameras で正規化した RMS 誤差値である。この図から9/1 cameras (rev)は2 cameras には劣るものの9/1 cameras 及び9/2 cameras に比較して大きく改善していることがわ

かる。また、 $9/2$ cameras (rev)は2 cameras とほぼ同等であるといえる。

ベースライン Nb を固定、かつ M を1または2即ち近傍カメラ数を固定した条件で、全カメラ数による RMS 誤差依存性を図 44 に示す。図 43 と同様に値は9 cameras で正規化した。 Nb が固定なので2 cameras の RMS 誤差は一定となる。一方提案法である $N + 1/1$ cameras (rev)は、第二因子の分母が Nb と固定値であることから一定となる。つまり $N + 1/1$ cameras (rev)はカメラ数を増やすに従い、奥行き推定精度を保ったまま輝度変動により頑健になっていく。さらに $N + 1/2$ cameras (rev)は徐々に精度が上昇し、 $N + 1 = 15$ の時点で2 cameras を逆転することがわかる。

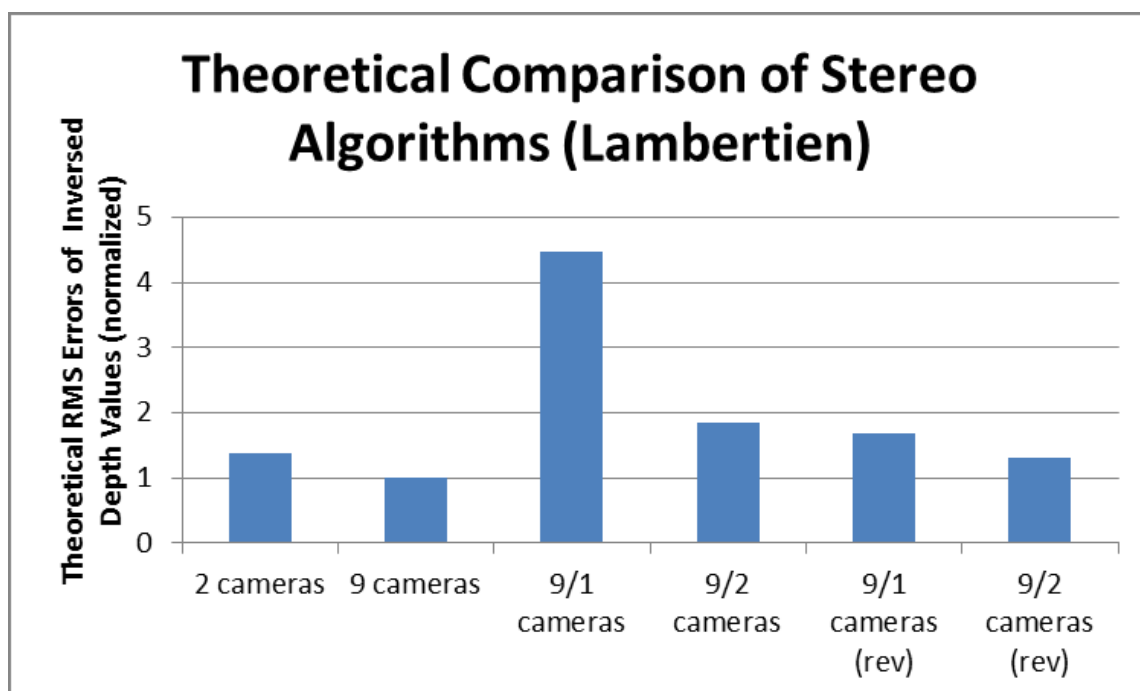


図 43 完全散乱仮定の下での理論評価値の比較結果 (奥行き逆数の RMS 誤差、9 cameras[71]で正規化した値)、図 34 に $9/1$ cameras (rev)・ $9/2$ cameras (rev)を追加

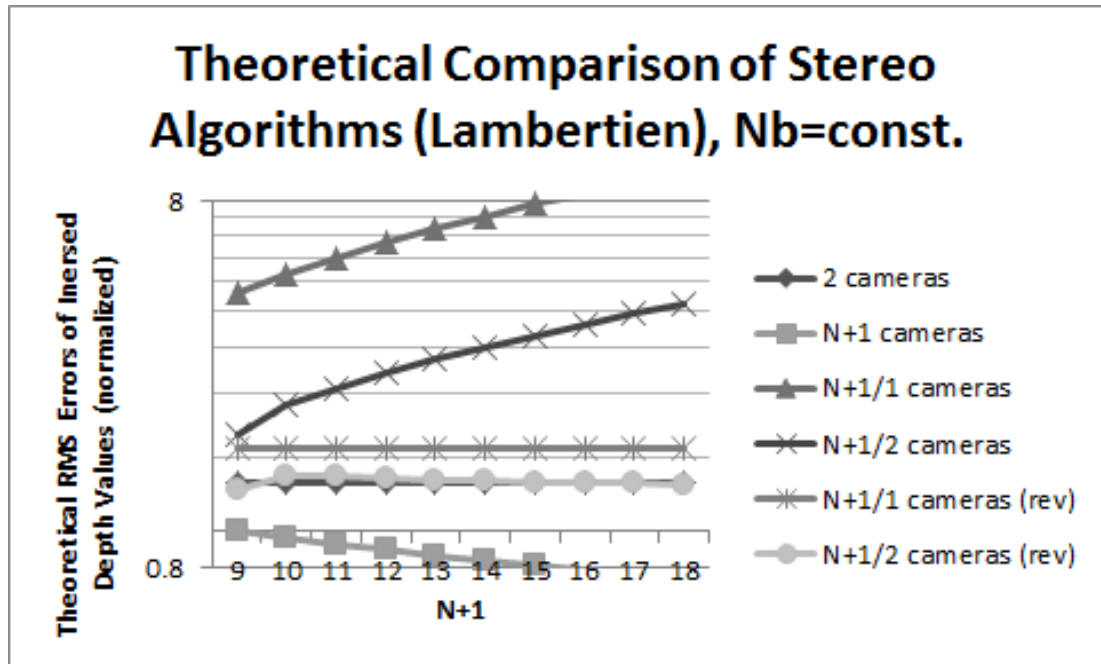


図 44 完全散乱仮定の下での理論評価値の比較結果（奥行き逆数の RMS 誤差、9 cameras[71]で正規化した値、横軸は全カメラ数）、図 35 に9/1 cameras (rev)・9/2 cameras (rev)を追加

第 5 節 視点依存の輝度変動を含む画像を用いた RMS 誤差実験評価 その 2

第 3 節と同様に Middlebury ステレオデータセット 2001 の中から、まず barn1 を用いて評価する [3]。

この barn1 に対して 9/1 cameras (rev)・9/2 cameras (rev)を適用する。エネルギー最小化には <http://vision.middlebury.edu/MRF/code/>で公開されるライブラリを利用する [69]。最適化法は第 3 節における評価実験と同様に、ライブラリで利用できる全てを用いる。式(13)におけるパラメータ λ は 5 から 400 まで 5 ステップで網羅的に探索する。パラメータ ρ^{\max} はエネルギー最小化ライブラリに付属したサンプルステレオプログラムの初期値である 2 を適用する。

また類似度は、式(37)の通り輝度値の二乗差を用いる。

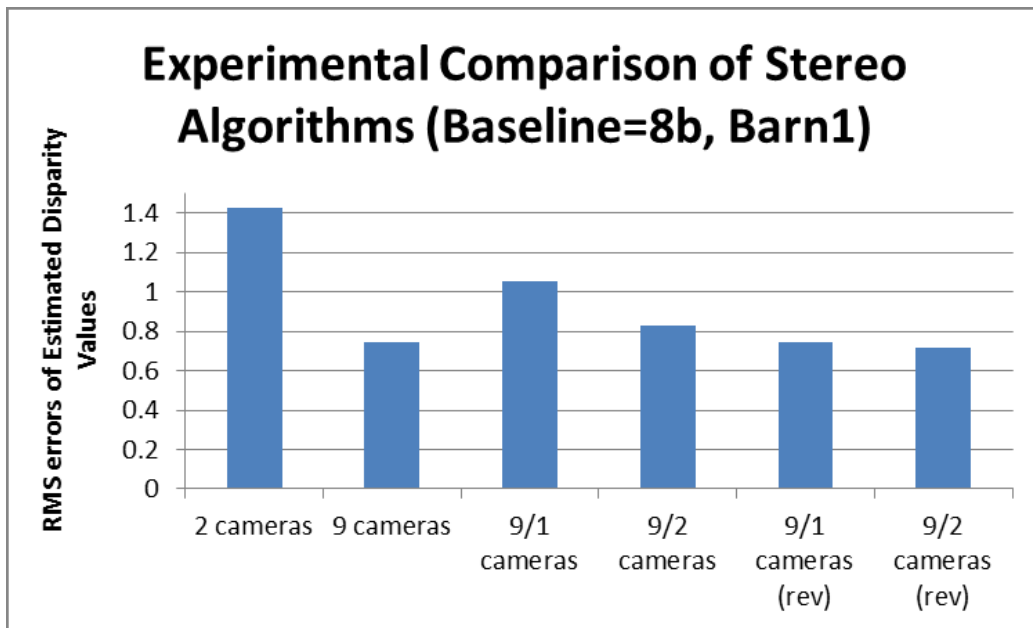


図 45 ベースライン8bにおける複数ステレオ法による推定視差の RMS 誤差、図 40 に9/1 cameras (rev)・9/2 cameras (rev)を追加

図 45 にベースライン8bの際の各手法の結果を示す。9/1 cameras (rev)及び9/2 cameras (rev)は、9 cameras の結果とほとんど同じかやや良好である。これらは前述の通り、現実の画像は理論的評価が仮定した完全散乱から外れていることによる影響と考えることができる。

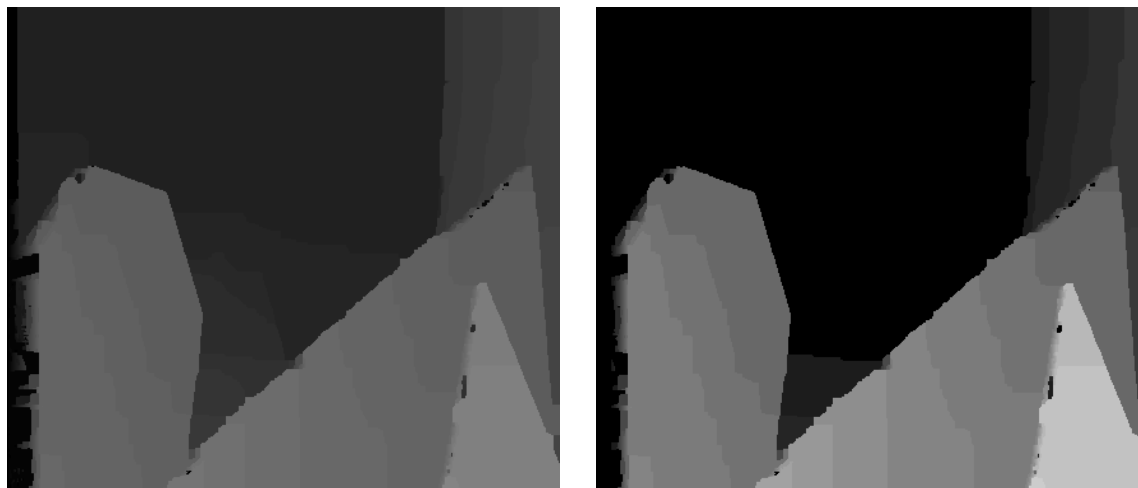


図 46 最良の推定視差結果 9/2 cameras (rev) ($\lambda = 140$ 、グラフカット α - β 交換) / 左: 視差画像 / 右: 視差画像を図 37 右と同じパラメータで強調

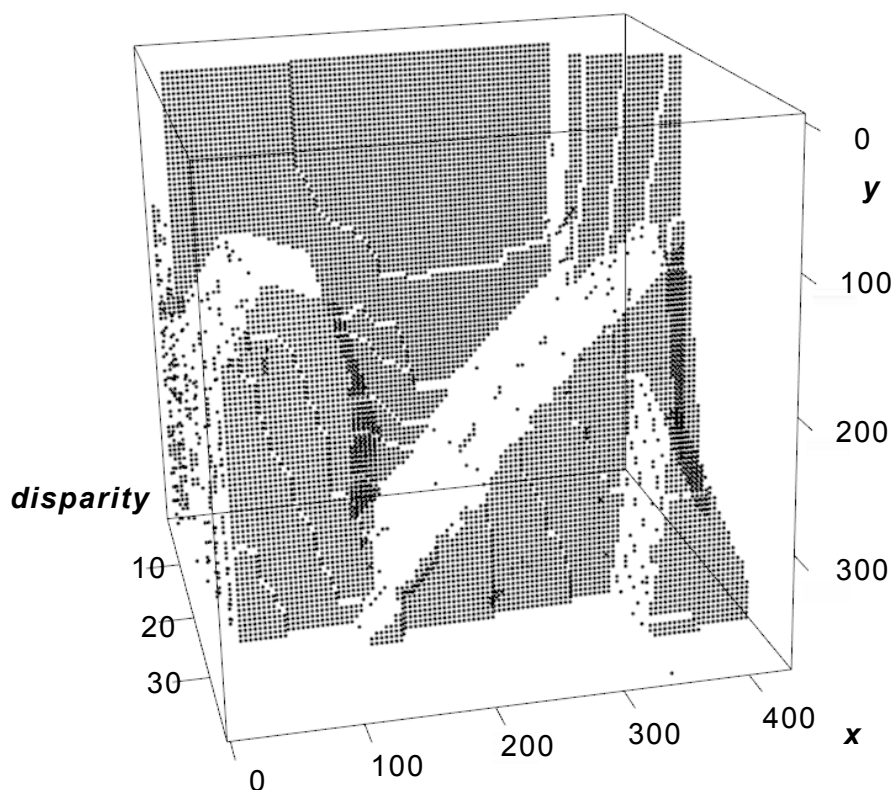


図 47 提案手法である 9/2 cameras (rev) で推定した最良の推定視差 ($\lambda = 140$ 、グラフカット α - β 交換)

図 46 に、ベースライン 8b のときに 9/2 cameras (rev) で推定した最良の視差を示す。最適化手法はグラフカット α - β 交換で $\lambda = 140$ である。図 46 右は差が確認できるよう強調した画像であり、図 39 の第 3 章提案手法の推定結果と比較して、本章で提案する手法がなめらかな形状の平板をより良く推定できていることがわかる。

図 47 に 9/2 cameras (rev) で推定した最良の推定視差結果を示す。このときの最適化手法はグラフカット α - β 交換、 $\lambda = 140$ である。

次に第 3 章でも用いた Middlebury ステレオデータセット 2006 から lampshade1 を評価に用いる。

図 48 に 7 cameras 及び 7/1 cameras、7/1 cameras (rev) による推定視差の RMS 誤差を示す。評価は第 3 章と同様に全ての最適化法とパラメータ λ に探索し最小値を各ステレオ法の代表値とした。テクスチャがなく視線依存の輝度変動があるこのような画像では、barn1 で最も精度が良かった従来のマルチベースライン法とともに、第 3 章で提案した視点依存の輝度変動に頑健な

マルチベースラインステレオ法の精度を、提案手法が大きく上回ったといえる。

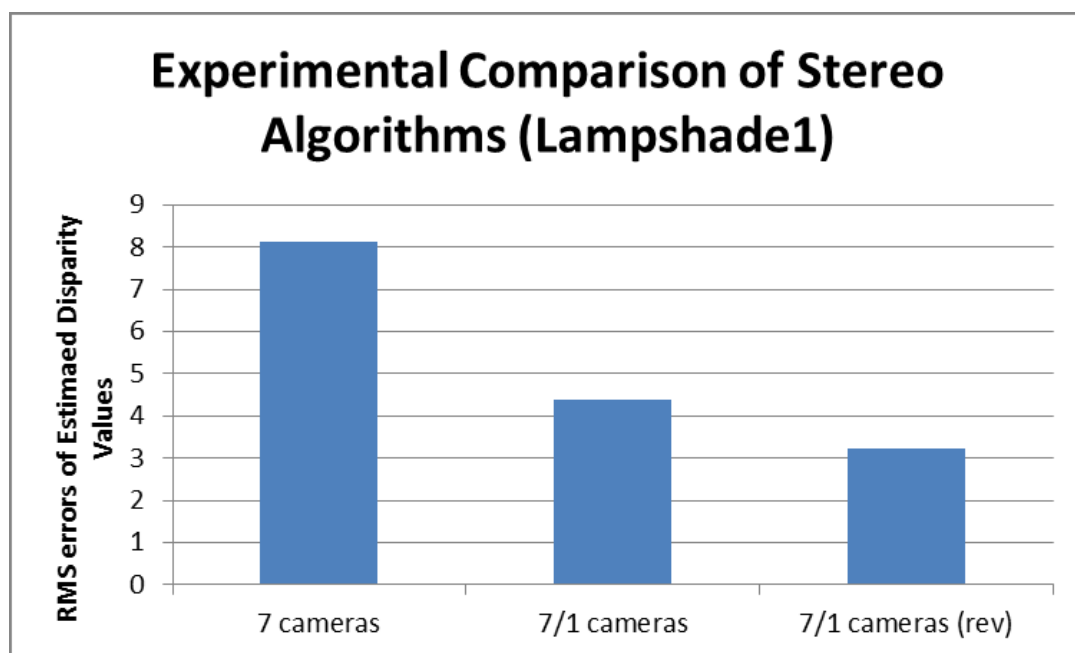


図 48 ベースライン6bにおける複数ステレオ法による推定視差の RMS 誤差

図 49 に7/1 cameras (rev)で推定した最良の推定視差結果を示す。このときの最適化手法はグラフカット α - β 交換、 $\lambda = 10$ である。

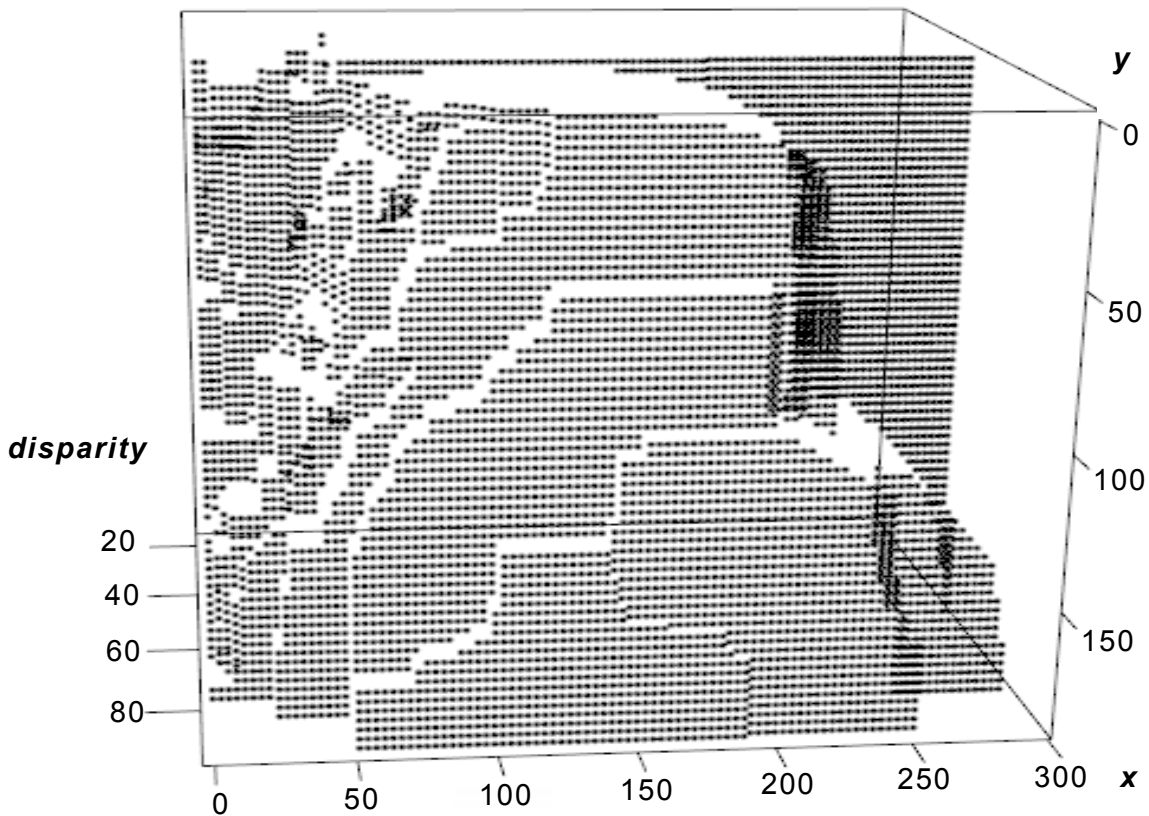


図 49 提案手法である 7/1 cameras (rev) で推定した最良の推定視差 ($\lambda = 10$ 、グラフカット α - β 交換)

さらに Middlebury ステレオデータセット 2001 の残りのデータに対する結果を示す。

図 52 は barn1 画像に対し、図 44 に類似度を差の絶対値(L1)に変更した提案アルゴリズムの結果を重ねて示したものである。センサに重畳するショットノイズやオクルージョンにより発生する、類似度が高い画素が存在しない場合に二次関数では推定結果が劣化することが知られていることから、追加実験として L1 も評価している。本実験では $L1 \cdot L2$ とともに三次元情報推定精度はほぼ変わらず、9/1 cameras (rev) 及び 9/2 cameras (rev) は、9 cameras の結果と比較してやや良好である。

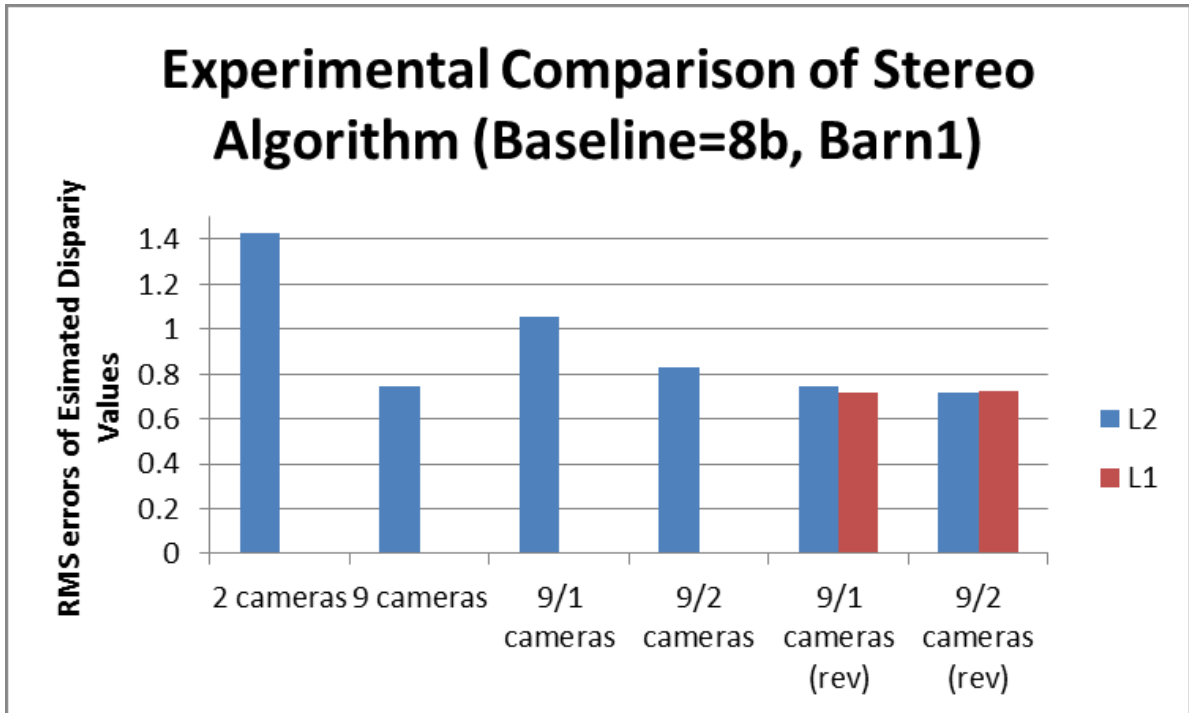


図 52 ベースライン8bにおける複数ステレオ法による barn1 画像[3]に対する推定視差の RMS 誤差、図 44 に9/1 cameras (rev)・9/2 cameras (rev)の L1 類似度の結果を追加

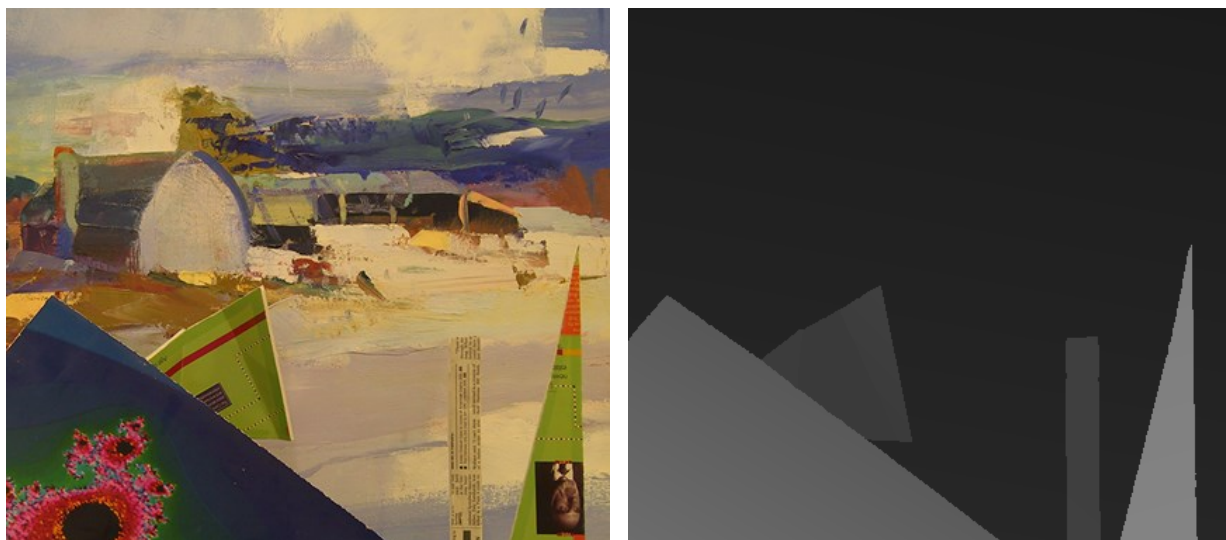


図 53 Middlebury データセット 2001 内の barn2 画像[3]。左: 左端位置から撮影した画像 / 右: 左から 3 番目の位置から撮影した画像から同 7 番目の画像に対する視差画像

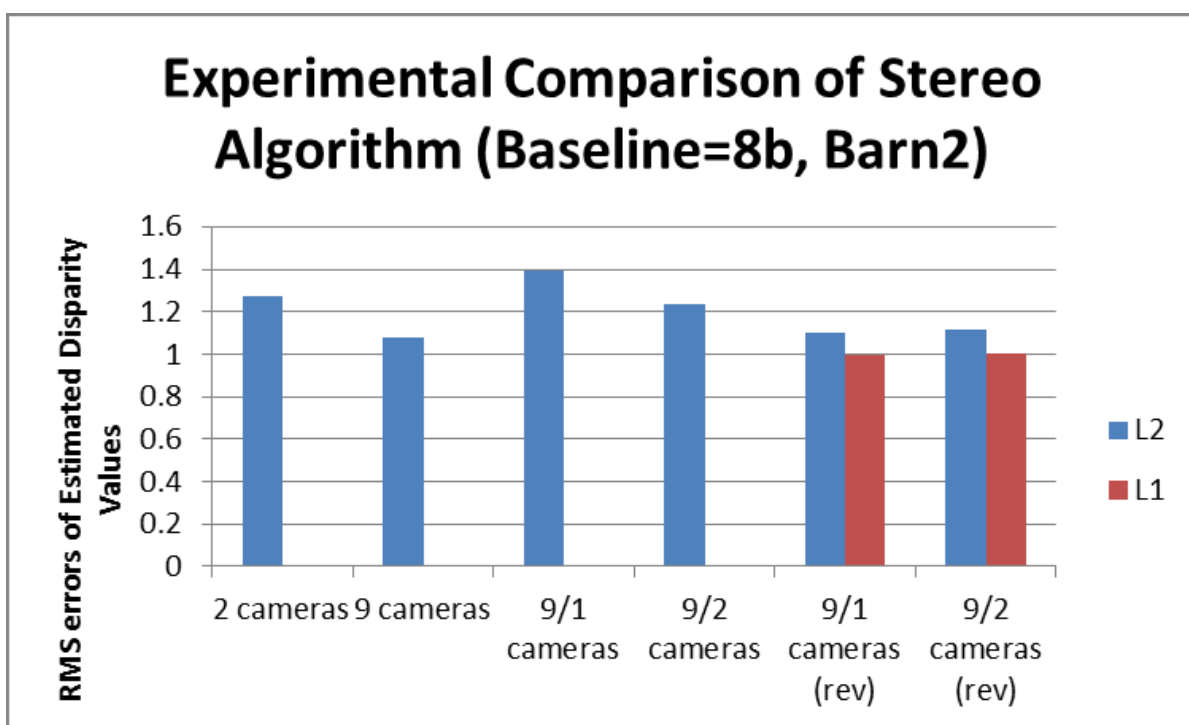


図 54 ベースライン8bにおける複数ステレオ法による barn2 画像[3]に対する推定視差の RMS 誤差

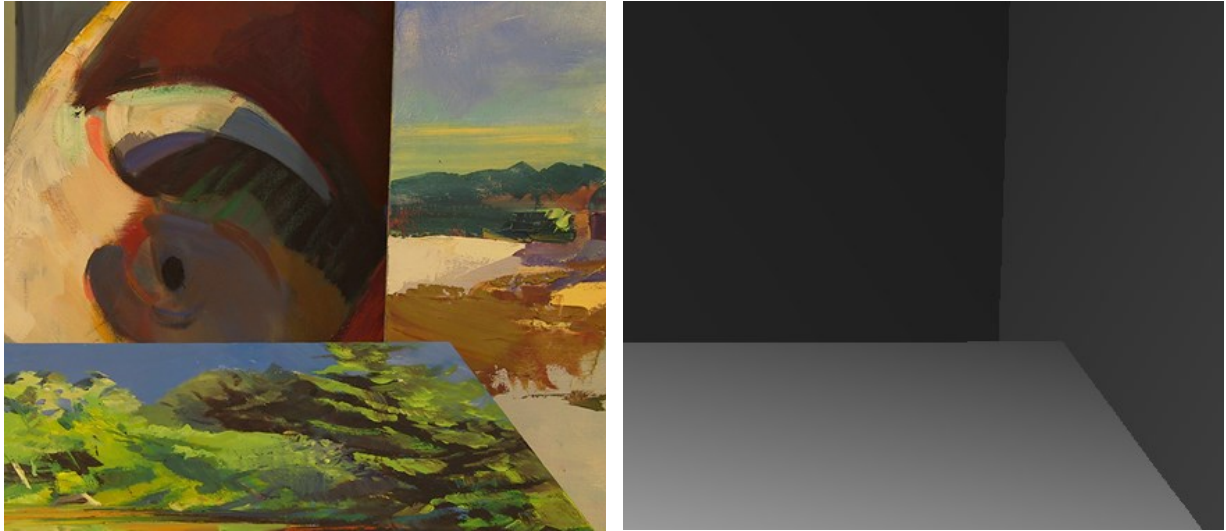


図 55 Middlebury データセット 2001 内の bull 画像[3]。左: 左端位置から撮影した画像 / 右: 左から 3 番目の位置から撮影した画像から同 7 番目の画像に対する視差画像

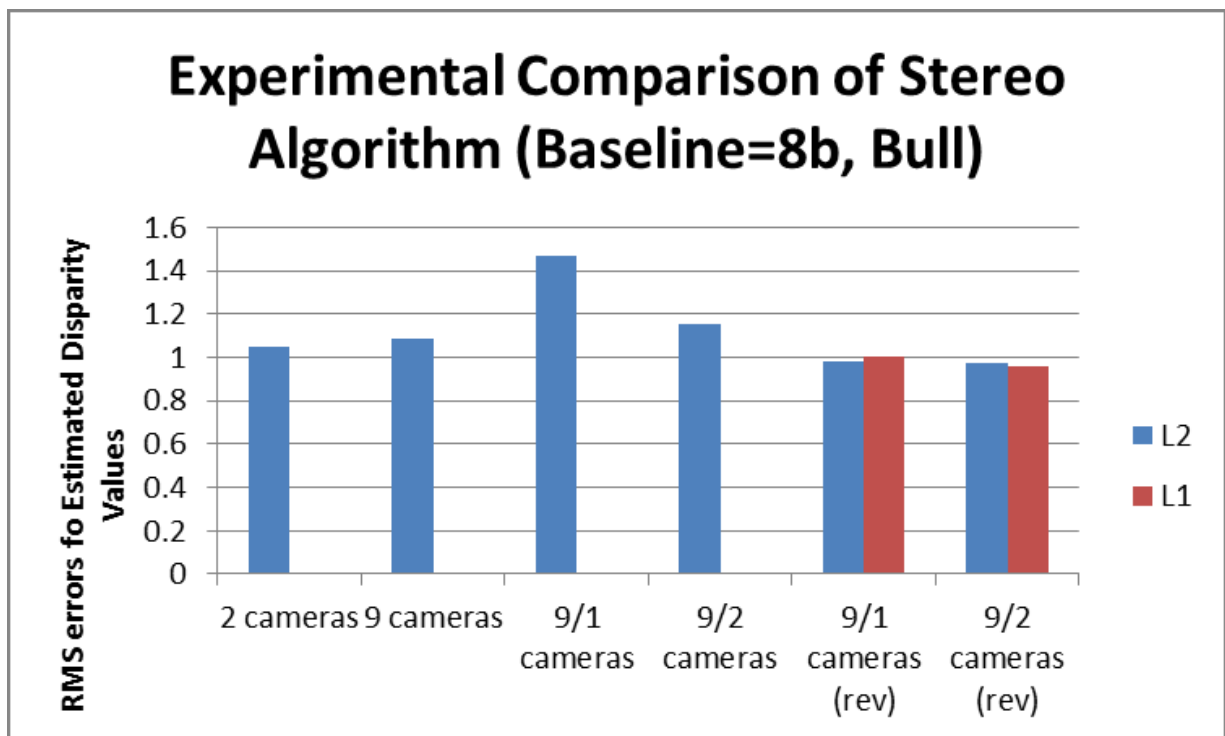


図 56 ベースライン8bにおける複数ステレオ法による bull 画像[3]に対する推定視差の RMS 誤差

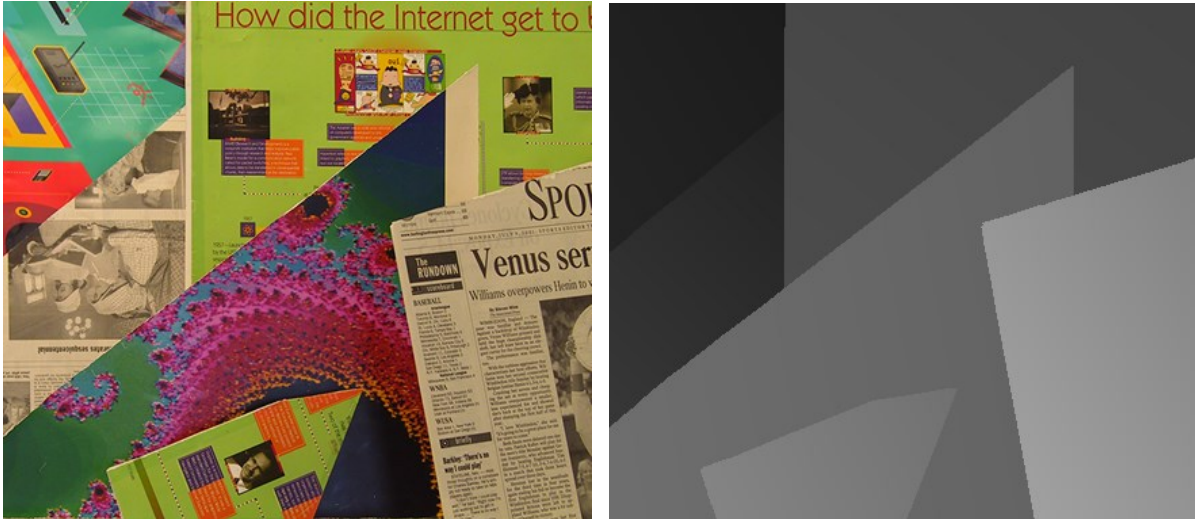


図 57 Middlebury データセット 2001 内の poster 画像[3]。左: 左端位置から撮影した画像 / 右: 左から 3 番目の位置から撮影した画像から同 7 番目の画像に対する視差画像

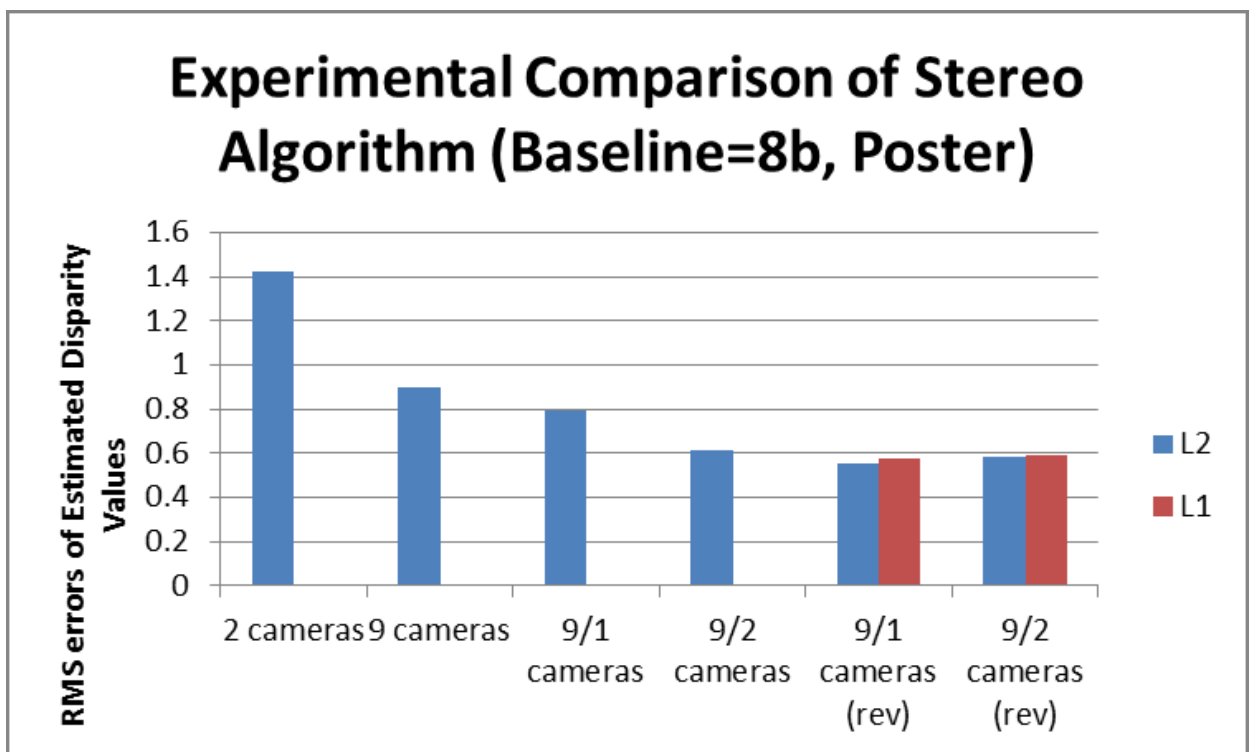


図 58 ベースライン8bにおける複数ステレオ法による poster 画像[3]に対する推定視差の RMS 誤差

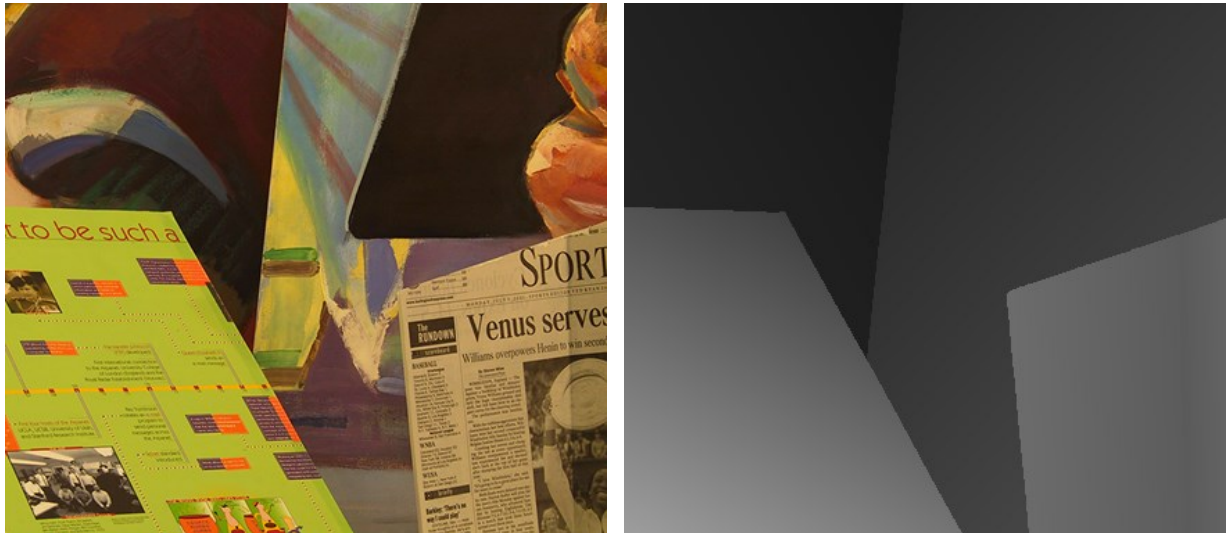


図 59 Middlebury データセット 2001 内の venus 画像[3]。左: 左端位置から撮影した画像 / 右: 左から 3 番目の位置から撮影した画像から同 7 番目の画像に対する視差画像

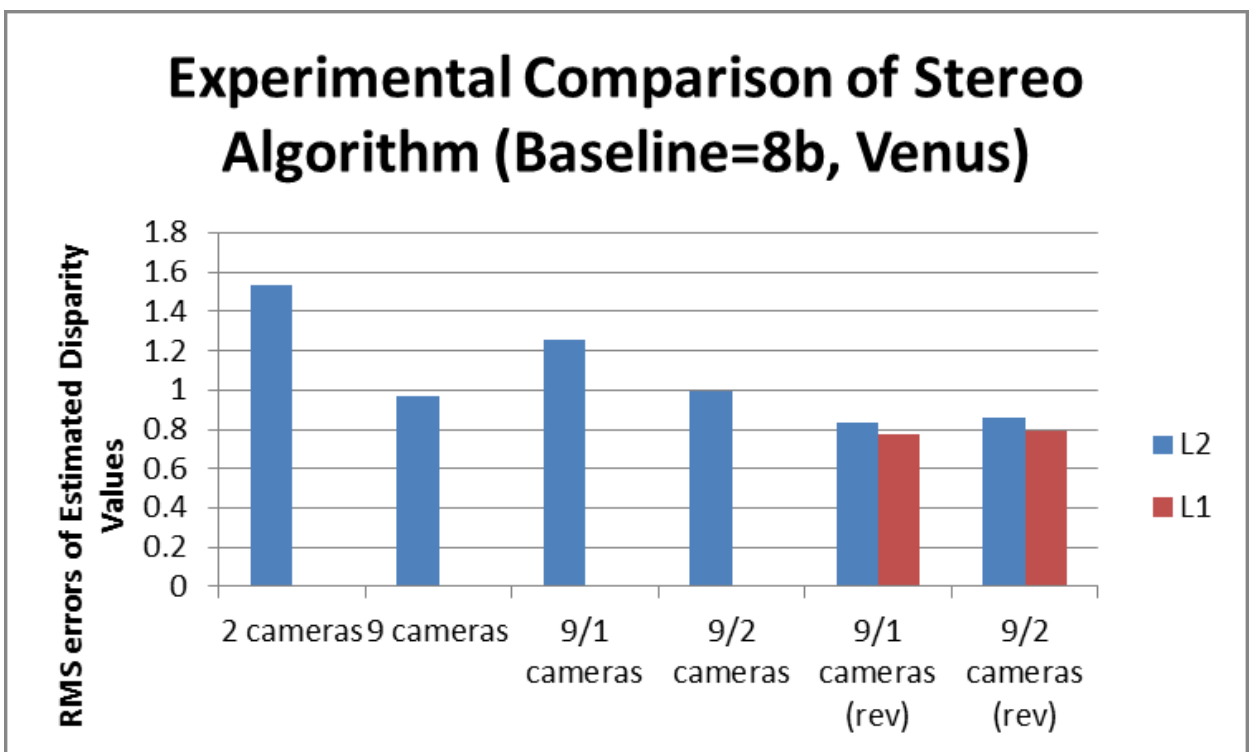


図 60 ベースライン8bにおける複数ステレオ法による venus 画像[3]に対する推定視差の RMS 誤差

図 53、図 55、図 57、図 59 はそれぞれ barn2 画像、bull 画像、poster 画像、venus 画像である。それぞれの図の左は左端位置から撮影した画像、同

右は左端から 3 番目の位置より撮影した画像を基準に、同 7 番目の画像に対する視差画像である。また図 54、図 56、図 58、図 60 は同じくそれぞれ barn2 画像、bull 画像、poster 画像、venus 画像に対する三次元情報推定精度をプロットした棒グラフである。Barn2 画像、bull 画像に対してはそれぞれ 9 cameras と提案手法が同等あるいはやや良好、poster 画像及び venus 画像では約 1 割・3 割の改善がそれぞれ観測されていることから、提案手法の有効性を確かめることができた。

図 61 に、実験評価結果をまとめて示す。

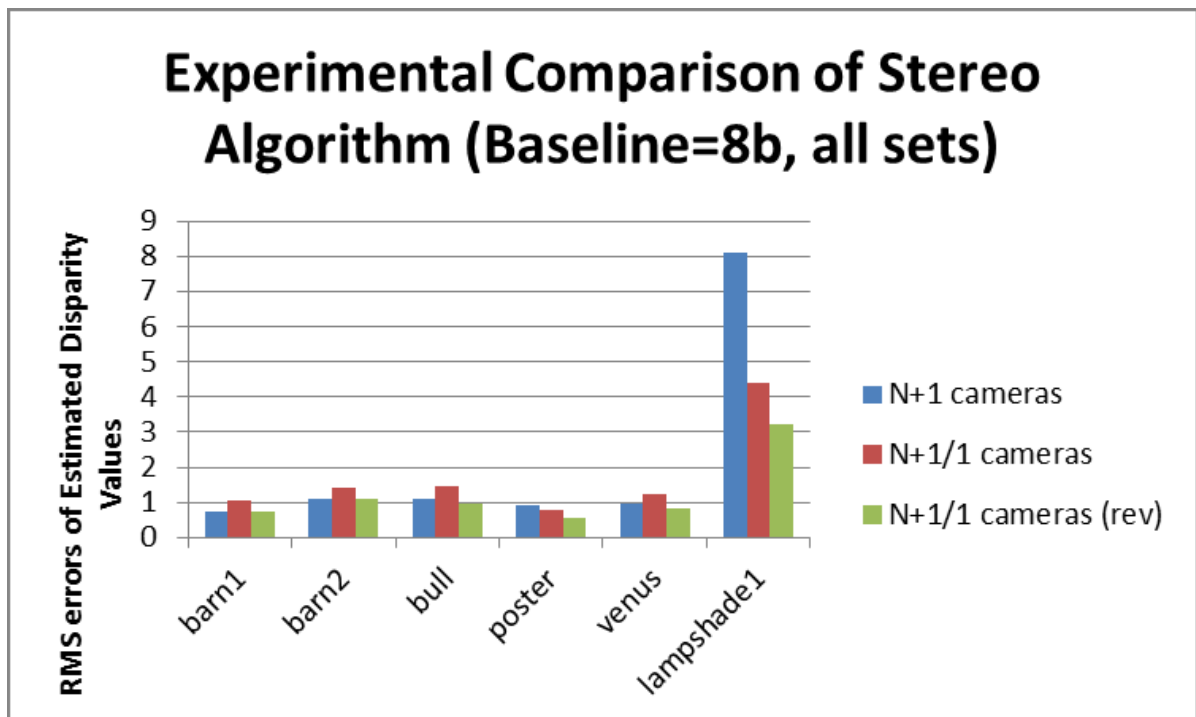


図 61 ベースライン8bにおける複数ステレオ法による推定視差の RMS 誤差 (サマリ)

第5章 まとめ

この論文では、異なる視点から撮影した画像から三次元情報を得ることを目的として、三角測量の原理に基づく手法について検討した。アクティブな手法として、著者らは多眼正弦波格子位相シフト法を提案した。その結果、一般的なマルチスリット光投影法よりも高精度かつ高速な計測が実現できることを示すとともに、本提案手法に基づく人間の顔及び頭部全周をそれぞれ計測するシステムについて述べた。

次にパッシブな手法で精度を向上させるにあたって課題となる視点依存の輝度変動を取り上げ、実際の画像セットでこの課題が観察できることを示した。視点依存の輝度変動に頑健な、近傍カメラ間比較の考えを導入し、新しいコスト関数をもつ二つのマルチベースラインステレオ法を提案した。まず視点位置に対して線形の輝度変動を許容するモデルを評価し、画像によっては従来のステレオ法を上回る精度をもつが一般的には劣化することを示した。次に輝度変動許容モデルの制約を強める手法を提案し、完全散乱仮定の下で同一のベースライン長をもつ二眼ステレオと同等な奥行き推定精度であることを理論的に証明した。さらに視点依存の輝度変動をもつ実際の画像セットを用いることにより、二眼ステレオだけでなく従来のマルチベースラインステレオと同等あるいは優れる精度をもつことを実験的に示した。

異なる視点から撮影した画像をもとに三次元情報を得ることは、コンピュータビジョンの重要なテーマのひとつであり、多様なアプリケーションへの応用が期待される。今後もより高速な撮影が可能でより高精度な計測が実現できるよう検討を進めてゆく。

参考文献

- [1] U. R. Dhond and J. K. Aggarwal: "Structure from Stereo – a Review," IEEE Transactions on Systems, Man, and Cybernetics, Vol.19, No.6, pp.1489-1510 (1989)
- [2] L. G. Brown: "A Survey of Image Registration Techniques," Computing Surveys, Vol.24, No.4, pp.325-376 (1992)
- [3] D. Scharstein and R. 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)
- [4] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski: "A Comparison and Evaluation of Multi-view Stereo Reconstruction Algorithms," Proceedings of Computer Vision and Pattern Recognition 2006, Vol.1, pp.519-526 (2006)
- [5] 井口征士, 佐藤宏介: 三次元画像計測, 昭晃堂 (1990)
- [6] Y. Shirai and M Suwa: "Recognition of Polyhedrons with a Range Finder," Proceedings of International Joint Conference on Artificial Intelligence, pp.80-87 (1971)
- [7] 青木則夫, 谷内田正彦, 辻三郎: "パターン投影法と2台のカメラを用いた3次元物体の計測," 電子情報通信学会論文誌, Vol.68-D, No.3, pp.384-391 (1985)
- [8] K. L. Boyer and A C. Kak: "Color-encoded structured Light for Rapid Active Ranging," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.9, No,1, pp.14-28 (1987)
- [9] J. L. Posdamer and M. D. Altschuler: "Surface Measurement by Space-Encoded Projected Beam Systems," Computer Graphics and Image Processing, Vol.18, No.1, pp.1-17 (1982)
- [10] J. Tajma and M. Iwakawa: "3-D Data Acquisition by Rainbow Range Finder," Proceedings of International Conference on Pattern Recognition, Vol.1, pp.16-21 (1990)
- [11] B. Carrihill: "The Intensity Ratio Depth Sensor," Computer Vision, Graphics, and Image Processing, Vol.2, No.3, pp.337-358 (1985)
- [12] D. M. Meadows, W. O. Johnson and J. B. Allen: "Generation of Surface Contours by Moiré Patterns," Applied Optics, Vol.9, No.4, pp.942-947 (1970)
- [13] J. H. Bruning: "Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses," Applied Optics, Vol.13, pp.2693-2703 (1974)

- [14] G. Indebetouw: "Profile Measurement Using Projection of Running Fringes," *Applied Optics*, Vol.17, No.18, pp.2930-2933 (1978)
- [15] G. T. Reid, R. C. Rixon and H. I. Messer: "Absolute and Comparative Measurements of Three-dimensional Shape by Phase Measuring Moiré Topography," *Optics and Laser Technology*, No.12, pp.315-319 (1984)
- [16] M. Halioua, R. S. Krishnamurthy, H. C. Liu and F. P. Chiang: "Automated 360 Profilometry of 3-D Diffuse Objects," *Applied Optics*, Vol.24, No.14, pp.2193-2196 (1985)
- [17] M. Halioua and H. C. Liu: "Optical Sensing Techniques for 3-D Machine Vision," *Optical Techniques for Industrial Inspection*, SPIE, Vol.665, pp.150-161 (1986)
- [18] G. Fornaro, G. Franceschetti, R. Lanari and E. Sansosti: "Robust Phase-Unwrapping Techniques: a Comparison," *Journal of Optical Society of America, A*, Vol.13, No.12, pp.2355-2366 (1996)
- [19] T. R. Judge and P. J. Bryanston-Cross: "A Review of Phase Unwrapping Techniques in Fringe Analysis," *Optics and Lasers in Engineering*, Vol.21, pp.199-239 (1994)
- [20] Microsoft Kinect <http://www.xbox.com/ja-JP/kinect> (2014年10月22日確認)
- [21] ASUS Xtion Pro Live, http://www.asus.com/jp/Multimedia/Xtion_PRO_LIVE/ (2014年10月22日確認)
- [22] K. Khoshelham and S. O. Elberink: "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications", *Sensors*, Vol.12, No.2, pp.1437-1454 (2012)
- [23] M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15, No.4, pp.353-363 (1992)
- [24] L. Matthies, T. Kanade and R. Szeliski: "Kalman Filter-Based Algorithms for Estimating Depth from Image Sequences," *International Journal of Computer vision*, Vol.3, No.3, pp.209-236 (1989)
- [25] M. J. Hannah: "Computer Matching of Areas in Stereo Images," Ph.D Thesis, Stanford University (1974)
- [26] H. Moravec: "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute (1980)

- [27] C. Harris and M. J. Stephens: "A Combined Corner and Edge Detector," Proceedings of Alvey Vision conference, pp.147-152 (1988)
- [28] D. G. Lowe: "Distinctive Image Features from Scale-Invariant keypoints," International Journal of Computer Vision, Vol.60, No.2, pp.91-110 (2004)
- [29] K. Mikolajczyk and C. Schmid: "A Performance Evaluation of Local Descriptors," IEEE transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.10, pp.1615-1630 (2005)
- [30] S. Winder and M. Brown: "Learning Local Image Descriptors," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1-8 (2007)
- [31] B. Trigg, P. F. McLauchlan, R. I. Hartley and A. W. Fitzgibbon: "Bundle Adjustment – A Modern Synthesis," Proceedings of International Workshop on Vision Algorithms, pp.297-372 (1999)
- [32] N. Snavely, S. M. Seitz and R. Szeliski: "Skeletal Graphs for Efficient Structure from Motion," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p.1-8 (2008)
- [33] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz and R. Szeliski: "Building Rome in a Day," Proceedings of IEEE International Conference on Computer Vision, pp.3057-3064 (2009)
- [34] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz and R. Szeliski: "Reconstructime Rome," Computer, Vol.43, No.6, pp.40-47 (2010)
- [35] N. Snavely, I. Simon, M. Goesele, R. Szeliski and S. M. Seitz: "Scene Reconstruction and Visualization from Community Photo Collections," Proceedings of the IEEE, Vol.98, No.8, pp.1370-1390 (2010)
- [36] S. Agarwal, N. Snavely, S. M. Seitz and R. Szeliski: "Bundle Adjustment in the Large," Proceedings of European Conference on Computer Vision, pp.29-42 (2010)
- [37] Q. Tian and M. N. Huhns: "Algorithms for Subpixel Registration," Computer Vision, Graphics and Image processing, Vol.35, pp.220-233 (1986)
- [38] T. Kanade: "Development of a Video-Rate Stereo Machine," Proceedings of Image Understanding Workshop, pp.549-557 (1994)
- [39] M. J. Black and A. Rangarajan: "On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision," International Journal of Computer Vision, Vol.19, No.1, pp.57-91 (1996)

- [40] R. D. Arnold: "Automated Stereo Perception," Technical Report AIM-351, Artificial Intelligence Laboratory, Stanford University (1983)
- [41] A. F. Bobick and S. S. Intille: "Large Occlusion Stereo," *International Journal of Computer Vision*, Vol.33, No.3, pp.181-200 (1999)
- [42] M. Okutomi and T. Kanade: "A Locally Adaptive Window for Signal Matching," *International Journal of Computer Vision*, Vol.7, No.2, pp.143-162 (1992)
- [43] E. P. Baltsavias and D. Stallmann: "Spot Stereo Matching for Digital Terrain Model Generation," *Proceedings of Swiss Symposium on Pattern Recognition and Computer Vision*, pp.61-72 (1993)
- [44] K. Konolige, "Small Vision Systems: Hardware and Implementation," *Proceedings of International Symposium on Robotics Research*, pp.203-212 (1997)
- [45] H. Hirschmüller, P. R. Innocent and J. M. Garibaldi: "Real-Time Correlation-Based Stereo Vision with Reduced Border Errors," *International Journal of Computer Vision*, Vol.47, No.1/2/3, pp.229-246 (2002)
- [46] D. G. Jones and J. Malik: "A Computational Framework for Determining Stereo Correspondence from a Set of Linear Spatial Filters," *Proceedings of European Conference on Computer Vision*, pp.397-410 (1992)
- [47] A. Ansar, A. Castano and L. Matthies: "Enhanced Real-Time Stereo Using Bilateral Filtering," *Proceedings of International Symposium on 3D Data Processing, Visualization and Transmission* (2004)
- [48] D. Bhat and S. Nayar: "Ordinal Measures for Image Correspondence," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.20, No.4, pp.415-423 (1998)
- [49] R. Zabih and J. Woodfill: "Non-Parametric Local Transforms for Computing Visual Correspondance," *Proceedings of the European Conference of Computer Vision*, pp.151-158 (1994)
- [50] I. Sarkar and M. Bansal: "A Wavelet-Based Multiresolution Approach to Solve the Stereo Correspondence Problem Using Mutual Information," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, Vol.37, No.4, pp.1009-1014 (2007)
- [51] H. Hirschmüller and D. Scharstein: "Evaluation of Stereo Matching Costs on Images with Radiometric Differences," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, Vol.31, No.9, pp.1582-1599 (2009)

- [52] J. Salvi, X. Armangué and J. Batlle: “A Comparative Review of Camera Calibrating Methods with Accuracy Evaluation,” *Pattern Recognition*, Vol.35, pp.1617-1635 (2002).
- [53] S. J. Kim and M. Pollefeys: “Robust Radiometric Calibration and Vignetting Correction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.30, No.4, pp.562-576 (2008).
- [54] J. Shi and C. Tomasi: “Good Features to Track,” *Proceedings of Computer Vision and Pattern Recognition*, pp.593-600 (1994)
- [55] I. J. Cox, S. Hingorani, B. M. Maggs and S. B. Rao: “A Maximum Likelihood Stereo Algorithm,” *Computer Vision and Image Understanding*, Vol.63, No.3, pp.542-567 (1996).
- [56] J. Kim, V. Kolmogorov and R. Zabih, “Visual Correspondence Using Energy Minimization and Mutual Information,” *Proceedings of International Conference on Computer Vision*, pp.1033-1040 (2003)
- [57] C. Strecha, T. Tuytelaars and L. Van Gool: “Dense Matching of Multiple Wide-Baseline Views,” *Proceedings of International Conference on Computer Vision*, pp.1194-1201 (2003)
- [58] J. Zhang, L. McMillan and J. Yu: “Robust Tracking and Stereo Matching under Variable Illumination,” *Proceedings of Computer Vision and Pattern Recognition*, Vol.1, pp.871-878 (2006)
- [59] S. Birchfield and C. Tomasi: “A Pixel Dissimilarity Measure that is Insensitive to Image Sampling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, No.4, pp.401-406 (1998)
- [60] V. Vaish, R. Szeliski, C. L. Zitnick, S. B. Kang and M. Levoy: “Reconstructing Occluded Surfaces Using Synthetic Apertures: Stereo, Focus and Robust Measures,” *Proceedings of Computer Vision and Pattern Recognition*, pp.2331-2338 (2006)
- [61] D. N. Bhat and S. K. Nayar: “Stereo and Specular Reflection,” *International Journal of Computer Vision*,” Vol.26, No.2, pp.91-106 (1998)
- [62] S. K. Nayar, X. S. Fang and T. Boult: “Removal of Specularities Using Color and Polarization,” *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp.583-590 (1993)
- [63] R. Klette, R. Kozera and K. Schlüns: “Shape from Shading and Photometric Stereo Methods,” *Handbook of Computer Vision and*

- Applications, Vol.2, Signal Processing and Pattern Recognition, pp.532-590 (1999)
- [64] V. Lohse, C. Heipke and R. L. Kirk: "Derivation of Planetary Topography using Multi-Image Shape-from-shading," Planetary and Space Science, Vol.54, pp.661-674 (2006)
- [65] C. Wöhler and P. d'Angelo: "Stereo Image Analysis of Non-Lambertian Surfaces," International Journal of Computer Vision, Vol.81, pp.172-190 (2007).
- [66] H. Jin, S. oatto and A. Yezzi: "Multi-view Stereo beyond Lambert," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.171-178 (2003)
- [67] L. Wang, R. Yang and J. E. Davis: "BRDF Invariant Stereo Using Light Transport Constancy," IEEE Transactions on Pattern Analysis and Macihne Intelligence," Vol.29, No.9, pp.1616-1626 (2007)
- [68] D. Terzopoulos: "Regularization of Inverse Visual Problems Involving Discontinuities," IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol.8, No.4, pp.413-424 (1986)
- [69] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen and C. Rother: "A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.30, No.6 (2008).
- [70] P. N. Belhumeur: "A Bayesian Approach to Bionocular Stereopsis," International Journal of Computer Vision, Vol.19, No.3, pp.237-260 (1996)
- [71] I. J. Cox: "A Maximum Likelihood N-camera Stereo Algorithm," International Conference on Computer Vision and Pattern Recognition, pp.733-739 (1994)
- [72] S. Geman and D. Geman: "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.6, No.6, pp.721-741 (1984)
- [73] D. Geiger and F. Girosi: "Parallel and Deterministic Algorithms for MRFs: Surface Reconstruction," IEEE Tansactions on Pattern Analysis and Machine Intelligence, Vol.13, No.5, pp.401-412 (1991)
- [74] J. Besag: "On the Statistical Analysis of Dirty Pictures," Journal of the Royal Statistical Society B, Vo.48, No.3, pp259-302 (1986)
- [75] J. Pearl: "Probablistic Reasoning in Intelligent Systems: Networks of

Plausible Inference,” Morgan Kaufmann Publishers (1988)

- [76] B. J. Frey and D. J. C. MacKay: “A Revolution: Belief Propagation in Graphs with Cycles,” *Advances in Neural Information Processing Systems* (1997)
- [77] W. T. Freeman, E. C. Pasztor and O. T. Carmichael: “Learning Low-Level Vision,” *International Journal of Computer Vision*,” Vol.40, No.1, pp.25-47 (2000)
- [78] M. Tappen and W. Freeman: “Comparison of Graph Cuts with Belief Propagation for Stereo, Using Identical MRF Parameters,” *IEEE International Conference on Computer Vision*, pp.900-907 (2003)
- [79] M. J. Wainwright, T. S. Jaakkola and A. S. Wilsky: “MAP Estimation via Agreement on Trees: Messagepassing and Linear Programming,” *IEEE Transactions on Information Theory*, Vol.51, No.11, pp.3697-3717 (2005)
- [80] V. Kolmogorov: “Convergent Tree-Reweighted Message Passing for Energy Minimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.28, No.10, pp.1568-1583 (2006)
- [81] B. Potez and T. S. Lee: “Efficient Belief Propagation for Higher-Order Cliques Using Linear Constraint Nodes,” *Computer Vision and Image Understanding*, Vol.112, No.1, pp.39-54 (2008)
- [82] D. Greig, B. Porteous and A. Seheult: “Exact Maximum a Posteriori Estimation for Binary Images,” *Journal of the Royal Statistical Society, Series B*, Vol.51, No.2, pp.271-279 (1989)
- [83] S. Roy and I. J. Cox: “A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem,” *Proceedings of International Conference on Computer Vision*, pp.492-499 (1998)
- [84] Y. Boykov, O. Veksler and R. Zabih: “Fast Approximate Energy Minimization via Graph Cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.11, pp.1222-1239 (2001)
- [85] Y. Boykov and V. Kolmogorov: “An Experimental Comparison of Min-cut / Max-flow Algorithms for Energy Minimization in Vision,” *IEEE Transactions of Pattern Analysis and Machine Intelligence*, Vol.26, No.9, pp.1124-1137 (2004)
- [86] V. Kolmogorov and R. Zabih: “What Energy Functions can be Minimized via Graph Cuts?,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.26, No.2, pp.147-159 (2004)
- [87] C. Rother, V. Kolmogorov, V. Lempitsky and M. Szummer: “Optimizing

- Binary MRFs via Extended Roof Duality,” Proceedings of Computer Vision and Pattern Recognition, pp.1-8 (2007)
- [88] N. Komodakis and N. Paragios: “Beyond Loose LP-Relaxations: Optimizing MRFs by Repairing Cycles,” Proceedings of European Conference on Computer Vision, pp.806-820 (2008)
- [89] R. Ishiyama, S. Sakamoto, J. Tajima, T. Okatani and K. Deguchi: “Absolute Phase Measurements using Geometric Constraints between Multiple Cameras and Projectors,” Applied Optics, Vol.46, Issue 17, pp.3528-3538 (2007)
- [90] A. Marugame, S. Sakamoto, T. Yoshinaga, K. Miyaoka, O. Fujimori, Y. Yamada, M. Kitano, H. Kabano, W. Tong and J. Tajima: “A Fast and Precise System for Taking High-density Human Head Measurements with Surrounding Range Finders,” Proceedings of International Conference on Pattern Recognition, Vol.3, pp.978-982 (2002)
- [91] R. Anderson and J. Parrish: “The Optics of Human Skin,” the Journal of Investigative Dermatology, Vol.77, No.1, pp.13-19 (1981)
- [92] M. van Gemert, S. Jacques, H. Sterenborg and W. Star: “Skin Optics,” IEEE Transactions on Biomedical Engineering, Vol.36, No.12, pp.1146-1154 (1989)
- [93] S. Sakamoto, I. J. Cox and J. Tajima: “A Multiple-baseline Stereo for Precise Human Face Acquisition,” Pattern Recognition Letters, Vol.18, No.9, pp.923-931 (1997)
- [94] 坂本静生, 甲藤二郎: ”視点依存の輝度変動に頑健なマルチベースラインステレオ法,” 画像電子学会誌, Vol.43, No.3, pp.280-291 (2014)
- [95] Middlebury Stereo Datasets 2006, available on <http://vision.middlebury.edu/stereo/data/scenes2006/>
- [96] D. Scharstein and R. Szeliski: “High-accuracy Stereo Depth Maps Using Structured Light,” Proceedings of Computer Vision and Pattern Recognition, Vol.1, pp.195-202 (2003)
- [97] 森口繁一, 一松信, 宇田川銈久: 岩波数学公式 II - 級数・フーリエ解析, 岩波書店 (1957)

業績リスト

査読付き論文

本論文の構成上重要な論文の行頭には◎を、申請者が第一著者である論文の行頭には○をそれぞれ付与して示す。

- ・ K. Abe 他 79 名: “Search for the Top Quark in e^+e^- Annihilation at $\sqrt{s} = 50\text{GeV}$: the First Result from the Venus Detector at Tristan,” *Journal of Physical Society of Japan*, Vol.56, pp.3763-3766 (1987)
- 坂本静生, 宮尾陽子, 田島譲二: “顔画像中からの目の特徴点抽出,” *電子情報通信学会論文誌 D-II*, Vol.76, No.8, pp.1796-1804 (1993)
- ◎S. Sakamoto, I. J. Cox and J. Tajima: “A Multiple-baseline Stereo for Precise Human Face Acquisition,” *Pattern Recognition Letters*, Vol.18, No.9, pp.923-931 (1997)
- ・ R. Ishiyama, M. Hamanaka and S. Sakamoto: “An Appearance Model Constructed on 3-D Surface for Robust Face Recognition against Pose and Illumination Variations,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol.35, No.3, pp.326-334 (2005)
- ・ 石山墨, 濱中雅彦, 坂本静生: “顔の三次元見えモデルを用いた任意姿勢・照明変動下の顔画像認識,” *電子情報通信学会論文誌 D-II*, Vol.88, No.10, pp.2069-2080 (2005)
- ・ R. Ishiyama, M. Hamanaka and S. Sakamoto: “Face Recognition under Variable Pose and Illumination Conditions Using 3D Facial Appearance Models,” *Systems and Computers in Japan*, Vol.38, No.2, pp.57-70 (2007)
- ・ R. Ishiyama, S. Sakamoto, J. Tajima, T. Okatani and K. Deguchi: “Absolute Phase Measurements using Geometric Constraints between Multiple Cameras and Projectors,” *Applied Optics*, Vol.46, Issue 17, pp.3528-3538 (2007)
- ◎坂本静生, 甲藤二郎: “視点依存の輝度変動に頑健なマルチベースラインステレオ法,” *画像電子学会誌*, Vol.43, No.3, pp.280-291 (2014)

国際発表

本論文の構成上重要な論文の行頭には◎を、申請者が第一著者である論文の行頭には○をそれぞれ付与して示す。

- S. Sakamoto: “Extracting Feature Points on Human Eye Photographs,” Proceedings of International Workshop on Machine Vision Applications, pp.461-464 (1992)
- ◎S. Sakamoto, I. J. Cox and J. Tajima: “A Multiple-baseline Stereo for Precise Human Face Acquisition,” Proceedings of the First International Conference, Audio- and Video-based Biometric Person Authentication, 1997, pp.419-428 (1997)
- G. House, S. Sakamoto and J. Tajima: “Multiple Object Labeling in Video Sequences without Skill,” Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, pp.101-105 (1998)
- R. Ishiyama, S. Sakamoto and J. Tajima: “A New Face Recognition System with Robustness against Illumination Changes,” Proceedings of IAPR Workshop on Machine Vision Applications, pp.28-30 (2000)
- R. Ishiyama and S. Sakamoto: “Geodesic Illumination Basis: Compensating for Illumination Variations in Any Pose for Face Recognition,” Proceedings of International Conference on Pattern Recognition, Vol.4, pp.297-301 (2002)
- A. Marugame, S. Sakamoto, T. Yoshinaga, K. Miyaoka, O. Fujimori, Y. Yamada, M. Kitano, H. Kabano, W. Tong and J. Tajima: “A Fast and Precise System for Taking High-density Human Head Measurements with Surrounding Range Finders,” Proceedings of 16th International Conference on Pattern Recognition, Vol.3, pp.978-982 (2002)
- R. Ishiyama and S. Sakamoto: “Fast and Accurate Facial Pose Estimation by Aligning a 3D Appearance Model,” Proceedings of International Conference on Pattern Recognition, pp.388-391 (2004)
- R. Ishiyama, H. Ikeda and S. Sakamoto: “A Compact Model of Human Postures Extracting Common Motion from Individual Samples,” Proceedings of International Conference on Pattern Recognition, Vol.1, pp.187-190 (2006)
- H. Suzuki, S. Sakamoto, T. Miyai, K. Nakano and M. Takeda: “A Novel Framework for Evaluation of ID Photo Quality,” International Biometric Performance Testing Conference 2012 (2012)

- S. Sakamoto: “Latest Trend of ISO/IEC JTC 1/SC 17 Standards regarding Biometrics,” Biometric Consortium Conference (2012)
- S. Sakamoto: “A Perspective from NEC regarding Vulnerability in Biometrics,” Biometrics Institute, Workshop on Biometrics Vulnerability Assessment Expert Group (2013)
- S. Sakamoto: “Advances in State-of-the-Art Face Recognition Solutions in Enabling Identity,” Biometrics 2013 (2013)
- S. Sakamoto: “Facial Biometrics and Surveillance,” Biometrics Institute, Developments of Facial Biometrics Seminar (2014)
- S. Sakamoto: “Facial Biometrics and Surveillance,” 1st Interpol Digital Security Research Seminar (2014)
- S. Sakamoto: “FR NIST Test Results,” Annual AFIS Internet Conference (2014)
- S. Sakamoto: “Facial Biometrics and its Deployment to Surveillance Applications,” Global Identity Summit (2014)

特許

以下では成立した日本・米国特許を、申請者が第一著者の行頭には○を付与して示す。

- 坂本静生: “上まぶた領域、目頭・目尻・上まぶた領域及び目の構造の検出方法及び装置,” 特許 2500726 号.
- 坂本静生: “特徴量選択方法及び装置と高速識別方法及び装置,” 特許 2551212 号.
- 坂本静生: 田島譲二, “距離画像取得方法及び装置,” 特許 2576243 号.
- 坂本静生: “距離画像取得方法及び装置,” 特許 2641945 号.
- 坂本静生: “人物動き領域検出装置,” 特許 2663878 号.
- 坂本静生: “口裂線検出方法及び装置,” 特許 2671556 号.
- 坂本静生: “人物顔画像照合装置,” 特許 2735028 号.
- 坂本静生: “目尻検出装置,” 特許 2757756 号.
- 坂本静生: “赤目検出補正装置,” 特許 2907120 号.
- 坂本静生: “三次元構造推定方法及び装置,” 特許 2910698 号.
- 坂本静生: “エッジ・領域検出方法及び装置,” 特許 2940254 号.
- 坂本静生: “物体検出・背景除去方法、装置およびプログラムを記録した記録媒体,” 特許 3114668 号.
- 坂本静生: “画像レイアウトシステム、方法及び記録媒体,” 特許 3134827 号.
- 坂本静生: “動画像アンカー作成システム,” 特許 3175760 号.
- 坂本静生: “3次元形状計測方法および装置ならびにプログラム,” 特許 4985123 号.
- ・内田薫, 坂本静生: “使用者認証方式、使用者認証装置、及びコンピュータプログラム,” 特許 5017905 号.
- S. Sakamoto: “Device for Obtaining Distance Information from an Object by Instantaneously Illuminating the Object by a Light Beam,” US5200792.
- S. Sakamoto: “Method for Detecting Eye Structure and Its Apparatus,” US5432866.
- S. Sakamoto: “Red-eye Detection / Retouch Apparatus,” US5990973.
- S. Sakamoto: “Method of and an Apparatus for 3-dimensional Structure Estimation,” US6049625.
- S. Sakamoto: “Method and Device of Object Detectable and Background Removal, and Storage Media for Storing Program Thereof,” US6333993.
- S. Sakamoto: “Method of and an Apparatus for 3-dimensional Structure

Estimation,” US6480620.

○ S. Sakamoto: “Method and Device of Object Detectable and Background Removal, and Storage Media for Storing Program Thereof,” US6603880.

○ S. Sakamoto: “Three-dimensional Shape Measurement Technique,” US7315643.

総説

以下では総説記事を、申請者が第一著者の行頭には○を付与して示す。

- 坂本静生: “顔画像から個人性と照明変動による見えの変化を見分ける,” 電子情報通信学会誌, Vol.87, No.3, pp.182-186 (2004).
- S. Sakamoto: “Development of Face Recognition Techniques at NEC Laboratories,” eNewsletter of IEEE Systems, Man and Cybernetics Society, Issue 12 (2005).
- 坂本静生: “顔画像の自動検索・照合技術～精度阻害要因抑制の立場から～,” 日本法科学技術学会誌, Vol.10, No.Supplement pp.9-10 (2005).
- 坂本静生: “顔認証の技術動向,” 電気学会誌, Vol.131, No.8, pp.550-553 (2011).
- IC 旅券調査委員会 WG2(バイオメトリクス): “IC 旅券と、顔認証技術導入へ向けての評価,” 電子情報通信学会誌, Vol.89, No.1, pp.52-55 (2006).

国内発表・講演・セミナー

以下では国内発表・講演・セミナーを、第一著者の行頭には○を付与して示す。

- 坂本静生, 幅淳二, 杉本章二郎, 長島順清: “O-STOS for Read-out of Anode Wires in BST,” 日本物理学会学術集会 (1986).
- 坂本静生, 田島譲二: “人物の顔照合のための特徴量解析,” 電子情報通信学会全国大会講演論文集, Vol.1990, No.Autumn, pp.6.411 (1990).
- 坂本静生, 宮尾陽子, 田島譲二: “顔画像中からの目の特徴点抽出,” 情報処理学会シンポジウム論文集, Vol.92, No.3, pp.II.143-II.150 (1992).
- 坂本静生, 田島譲二: “人物の顔照合のための特徴量解析(II),” 電子情報通信学会大会講演論文集, Vol.1993, No.Shunki, pp.7.400 (1993).
- 坂本静生, 田島譲二: “人物の顔照合のための特徴量解析,” 電子情報通信学会技術研究報告, Vol.94, No.341 (HC94 58-62), pp.7-14 (1994).
- 坂本静生, 田島譲二: “画像からの顔の検出,” 電子情報通信学会大会講演論文集, Vol.1995, No.Society D, pp.214 (1995).
- 坂本静生, 高梨伸彰, 田島譲二: “高速可変ベースライン-アクティブステレオカメラの試作と色に基づくリアルタイム物体検出 / 追跡,” 電子情報通信学会大会講演論文集, Vol.1996, No.Sogo, pp.175 (1996).
- 坂本静生, 佐藤洋一, 田島譲二: “自動赤目検出・補正処理,” 電子情報通信学会大会講演論文集, Vol.1996, No.Society D, pp.267 (1996).
- ・ Gregory House, 坂本静生, 高梨伸彰, 田島譲二: “アクティブビジョンを応用した人物監視システム,” 情報処理学会研究報告 CVIM, Vol.97, No.10, pp.123-130 (1997).
- 坂本静生, Ingemar J. Cox, 田島譲二: “視点に依存する画素値変動に強いマルチベースライン・ステレオ法,” 電子情報通信学会大会講演論文集, Vol.1997 No.Sogo, pp.156 (1997).
- 坂本静生, 新井雅年, 楠実奈子, 井上晃, 田島譲二: “新聞紙面向け顔画像処理: 精細な背景-人物間輪郭生成,” 電子情報通信学会大会講演論文集, Vol.1998 総合 7, pp.158 (1998).
- 坂本静生, Gregory House, 田島譲二: “競馬中継動画からのインデクシング,” 画像センシングシンポジウム講演論文集, Vol.4th, pp.111-116 (1998).
- 坂本静生, 新井雅年, 楠実奈子, 井上晃, 田島譲二: “新聞紙面向け顔画像処理,” 情報処理学会シンポジウム論文集, Vol.98, No.10, pp.II.397-II.403 (1998).
- ・ 石山塁, 坂本静生, 北野昌秀, 童衛青, 田島譲二: “顔用レンジファインダ,”

- 電子情報通信学会技術研究報告 PRMU, Vol.99, No.118, pp.35-42 (1999).
- ・今岡仁, 坂本静生: “顔の向きに依らない顔画像照合方法,” 電子情報通信学会技術研究報告 PRMU, Vol.99, No.118, pp.51-58 (1999).
 - 坂本静生: “顔画像による人物照合,” 日本機械学会情報・知能・精密機器部門講演会講演論文集, Vol.2000, pp.19-22 (2000).
 - 坂本静生, 石山壘, 北野昌秀, 童衛青, 田島譲二: “高精度・高速レンジファインダの開発,” 計測自動制御学会パターン計測部会研究会 (2000).
 - ・今岡仁, 坂本静生: “照明変動に強い顔照合: 顔画像からの表面反射率推定,” 電子情報通信学会技術研究報告 PRMU, Vol.100, No.312, pp.35-42 (2000).
 - ・石山壘, 坂本静生: “3次元形状と姿勢非依存な照明基底を用いた顔照合,” 電子情報通信学会技術研究報告 PRMU, Vol.101, No.58, pp.33-40 (2001).
 - ・坂野鋭, 赤松茂, 磯部義明, 坂本静生, 佐藤勝彦, 鷺見和彦, 長尾健司, 矢野博司: “顔画像認証アルゴリズムの精度評価方法: 標準化に向けて,” 電子情報通信学会技術研究報告 PRMU, Vol.101, No.524, pp.71-78 (2001).
 - ・石山壘, 坂本静生: “照明条件と姿勢の両者の自動補正による顔照合,” 電子情報通信学会技術研究報告 PRMU, Vol.101, No.524, pp.59-64 (2001).
 - ・丸亀敦, 坂本静生, 吉永朋弘, 宮岡一頼, 藤森修身, 山田芳信, 北野昌秀, 蒲野晴美, 童衛青, 田島譲二: “高速高密度頭部形状計測レンジファインダシステムの開発,” 電子情報通信学会技術研究報告 PRMU, Vol.101, No.524, pp.101-108 (2001).
 - ・濱中雅彦, 石山壘, 坂本静生: “測地照明基底を用いた顔照合における陰影部の影響について,” 情報科学技術フォーラム一般講演論文集, Vol.2002, No.3, pp.65-66 (2002).
 - ・丸亀敦, 坂本静生: “形状情報と色情報を併用した顔三次元データからの特徴部位の抽出,” 情報科学技術フォーラム一般講演論文集, Vol.2002, No.3, pp.199-200 (2002).
 - ・石山壘, 坂本静生: “測地照明基底を用いた照明変動に頑強な顔姿勢推定,” 電子情報通信学会総合大会講演論文集, Vol.2003, No.2, pp.330 (2003).
 - ・濱中雅彦, 石山壘, 坂本静生: “任意姿勢に対応した標準測地重みを用いた顔照合,” 電子情報通信学会技術研究報告 PRMU, Vol.103, No.296, pp.61-66 (2003).
 - ・石山壘, 濱中雅彦, 坂本静生: “3D物体表面上での見えモデル構築による姿勢と照明の変動に頑健な顔照合,” 情報処理学会研究報告 CVIM, Vol.2003, No.66, pp.61-68 (2003).
 - ・濱中雅彦, 石山壘, 坂本静生: “補正画像の鮮明度を用いた顔照合と大規模実験,” 電子情報通信学会総合大会講演論文集, Vol.2004, No.2, pp.233 (2004).

- ・石山 罌, 濱中雅彦, 坂本静生: “3D アピアランスモデルを用いた高速・高精度な顔姿勢推定,” 電子情報通信学会総合大会講演論文集, Vol.2004, No.2, pp.265 (2004).
- ・濱中雅彦, Ferreiro Jose, 石山 罌, 坂本静生: “3次元データをキーとした姿勢・照明変動に頑健な顔画像検索,” 情報処理学会研究報告 CVIM, Vol.2004, No.91, pp.41-48 (2004).
- ・丸亀敦, 坂本静生: “平均顔照明空間基底を用いた顔画像人種推定,” 電子情報通信学会総合大会講演論文集, Vol.2005, No.2, pp.221 (2005).
- ・池田浩雄, 石山 罌, 坂本静生: “行動に基づく関節自由度の制約を利用した人体姿勢推定,” 情報科学技術フォーラム一般講演論文集, Vol.4, No.3, pp.115-116 (2005).
- ・IC 旅券調査委員会 WG2: “IC 旅券と、顔認証技術導入へ向けての評価,” 電子情報通信学会技術研究報告 PRMU, Vol.105, No.374, pp.51-56 (2005).
- 坂本静生: “バイオメトリクスと本人認証 –旅券と顔認証を中心として–,” 国際協調問題研究会 (2005).
- 坂本静生: “顔画像認識技術の現状と将来 –変動要因抑制の立場から–,” 日本法科学技術学会学術集会 (2005).
- 坂本静生: “IC 旅券における顔認識技術,” 電気四学会関西支部専門講習会 (2007).
- 坂本静生: “IC カードとバイオメトリック指紋認証,” 自動認識セミナー (2007).
- 坂本静生: “SC17 におけるバイオメトリクス,” 日本自動認識システム協会, 国際標準化セミナー(SC37) (2008).
- 坂本静生: “IC カード国際標準化の動向: より使いやすく、より安全に ~アクセシビリティとバイオメトリクスの広がり~, ” 自動認識セミナー (2009).
- 坂本静生: “SC17 におけるバイオメトリクス,” 国際標準化セミナー(SC37) (2009).
- 坂本静生: “バイオメトリクスとその応用,” 立命館大学セミナー (2010).
- ・鈴木裕之, 坂本静生, 宮井貴志, 中野和也, 竹田賢史, 鈴木理道, 前田大輔: “ID用顔画像の品質評価用フレームワークの提案,” 情報科学技術フォーラム講演論文集, Vol.10, No.4, pp.489-490 (2011).
- 坂本静生: “バイオメトリック・ソリューションの最新動向,” 自動認識セミナー (2012).
- 坂本静生: “バイオメトリクス認証技術と防犯ソリューションへの応用,” 日本防犯設備協会 特別セミナー (2012).

- 坂本静生: “生体認証技術と映像シーン分析について,” 保安通信協会セミナー (2013).
- 坂本静生: “SC17 におけるバイオメトリクス,” 国際標準化セミナー(SC37) (2013).
- 坂本静生: “顔認証技術とその応用,” 日本鑑識学会秋季大会特別講演 (2013).
- 坂本静生: “企業・社会競争力強化へ向けたビッグデータ活用 特に顔画像利用について,” 日本自動認識システム協会バイオメトリクス部会セミナー (2013).
- 坂本静生: “顔認証技術とその多様な応用への広がり –USJ 様事例を題材として–,” 自動認識セミナー (2014).
- 坂本静生: “映像解析・顔認証ソリューションとラピッド DNA 分析,” 保安通信協会セミナー (2014).
- 坂本静生: “SC17 におけるバイオメトリクス,” 国際標準化セミナー(SC37) (2014).
- 坂本静生: “バイオメトリック認証システムとセキュリティ,” 電子情報通信学会技術報告, Vol.114, No.251, BioX2014-33, pp.29 (2014).
- ・山口理恵, 坂本静生, 鈴木宏哉, “スマートフォンを事例とする多要素認証,” SCIS 2015, 4C2-5 (2015)
- その他京都大学、つくば大学、横浜国立大学、電気通信大学講義各 1 件.

その他著作

以下ではその他の著作を、申請者が第一著者の行頭には○を付与して示す。

- 坂本静生: “VENUS/Barrel Streamer Tubes のアノードワイヤー信号の読み出しについて,” 物性研究, Vol.48, No.5, pp.658 (1987).
- 坂本静生: “顔画像中からの目の特徴点抽出,” 画像ラボ, Vol.5, No.5, pp.24-27 (1994).
 - ・井上晃, 坂本静生: “インテリジェント画像処理機能,” NEC 技報, Vol.49, No.9, pp.18-23 (1996).
- 坂本静生: “高精度・高速レンジファインダの開発,” 画像ラボ, Vol.10, No.12, pp.1-6 (1999).
- S. Sakamoto: “3D Model-based Face Recognition System with Robustness against Illumination Changes,” NEC Research and Development (2002).
- 坂本静生: “高精度・高速レンジファインダの開発 –複数カメラ・光源間の幾何学的拘束による絶対位相値計測,” 画像ラボ, Vol.10, No.12, pp.1-6 (1999).
- S. Sakamoto: “3D Face Recognition Technique Robust against Large Changes in Pose and Illumination Conditions,” New Breeze (2006).
- 坂本静生: “顔認証技術の進展と応用,” 機能材料, Vol.28, No.10, pp.19-24 (2008).
- 坂本静生: “ISO/IEC JTC 1/SC 17 (カード及び個人識別)におけるバイオメトリクス,” 自動認識, Vol.21, No.2, pp.73-75 (2008).
- 坂本静生: “顔認証技術の現状と将来,” 機能材料, Vol.28, No.10, pp.19-24 (2008).
- 坂本静生: “顔認証技術の現状と将来,” ニューメディア開発協会研究成果レポート, No.24, pp.16-19 (2009).
- 坂本静生: “バイオメトリクス製品とソリューションの現状と展望,” NEC 技報, Vol.63, No.3, pp.14-17 (2010).
 - ・蘇雷明, 坂本静生: “超大規模バイオメトリック認証システムとその実現,” NEC 技報, Vol.65, No.2, pp.52-56 (2012).
 - ・高屋正裕, 坂本静生: “生体認証の趨勢 先進国における動向,” 自動認識, Vol.34, No.5, pp.54-57 (2014).
 - ・浜田康志, 坂本静生: “顔認証製品と社会ソリューションでの活用,” NEC 技報, Vol.67, No.1, pp.52-55 (2014)
- 坂本静生: “標準活動トピックス: 羽田空港・成田空港における顔認証自動

化ゲート実験,” 情報処理学会 情報規格調査会, 情報技術標準
NEWSLETTER, No.104, pp.5-8 (2014)