

# 画像特徴点の数え上げによるマルチベースラインステレオ法

佐藤 智和<sup>†</sup> 横矢 直和<sup>†</sup>

<sup>†</sup> 奈良先端科学技術大学院大学 情報科学研究科 〒630-0192 奈良県生駒市高山町 8916-5

E-mail: †{tomoka-s,yokoya}@is.naist.jp

あらまし 本稿では、一般的なマルチベースラインステレオ法の奥行き推定で用いられる非類似度 SSD(Sum of Squared Differences) の代わりに、画像特徴点の座標からなる新しい評価尺度 TNIP(Total Number of Interest Points) を用いることで、照明条件の影響や射影歪みによる影響の少ない奥行き探索を実現する手法を提案する。提案手法では、画像上の輝度値情報を用いず、特徴点の座標から奥行き探索を行うため、動画像のように多数の画像から奥行きを推定する場合において計算量および必要メモリ容量を大幅に低減できる。また、画像の変形にロバストな特徴点を用いるため、カメラの移動に伴う画像パターンの変形に強い耐性を持つという特長がある。実験では、特徴点の数え上げに基づくマルチベースラインステレオ法によって、良好な奥行き画像が生成できることを示す。ただし本稿では、動画撮影時のカメラの位置・姿勢パラメータおよびカメラ内部パラメータは推定済みで既知とする。

キーワード 奥行き画像推定, マルチベースラインステレオ, 画像特徴点

## Multi-baseline Stereo by Counting Feature Points

Tomokazu SATO<sup>†</sup> and Naokazu YOKOYA<sup>†</sup>

<sup>†</sup> Nara Institute of Science and Technology 8916-5 Takayama, Ikoma, Nara 630-0192

E-mail: †{tomoka-s,yokoya}@is.naist.jp

**Abstract** This paper proposes a novel method for estimating depth from a long image sequence captured by a calibrated moving camera. Our idea for estimating a depth map is very simple; only counting interest points in images is integrated with the framework of multi-baseline stereo. Even by a simple algorithm, depth can be determined without computing similarity measures such as SSD and NCC that have been used for traditional stereo matching. The proposed method realizes robust depth estimation against image distortions and occlusions with lower computational cost than traditional multi-baseline stereo method. In experiments, validity and feasibility of our algorithm are shown for both a synthetic and real outdoor scene.

**Key words** depth map estimation, multi-baseline stereo, feature points

### 1. はじめに

ステレオ画像計測に基づく奥行き画像推定は、三次元形状復元、物体認識、拡張現実感等に用いられており、コンピュータビジョンの分野において最も重要な問題の一つである。このため現在においても、ステレオ画像計測に基づく奥行き画像推定法は広く研究されている。本稿では、これらステレオ画像に基づく奥行き画像推定手法の中でも、複数の画像を統合的に扱うことで、比較的精度の高い奥行き画像推定を実現できるマルチベースラインステレオ法について、新しい手法の提案を行う。

奥富らによって提案されたマルチベースラインステレオ法[1]は、複数の画像を一度に扱うことでステレオ法における対応点探索のあいまい性を低下させ、奥行き推定精度を向上させるという特長があり、現在広く用いられている[2]~[6]。また、近年

のカメラキャリブレーション手法の発展に伴って、自由な移動を伴って撮影された動画像や、広範囲に配置された複数のカメラによって撮影された画像群に対しても、マルチベースラインステレオ法の適用が試みられている[3]~[5]。

移動を伴って撮影した動画像を奥行き画像推定に用いれば、長いベースラインを容易に確保することができる。このため、動画像を入力として用いる手法[4],[5]は、カメラ間の距離に対して対象までの距離が相対的に長くなる屋外環境の三次元モデリングに適している。しかし、このような屋外環境を対象とする場合、従来手法で用いられてきたマルチベースラインステレオの評価尺度 SSSD(Sum of SSD)では、以下に示す複数の問題が発生する。

(1) 画像パターン変形による推定精度の低下: ビデオカメラの自由な移動による撮影では、対象とカメラの位置・姿勢の関係



図 1 画像特徴点の検出例

が大きく変化することから、対象物体上の同一の点の見え方が、動画の各フレームで大きく変化してしまう。従来から用いられてきたパターン類似度に基づく評価尺度 (SSD, NCC 等) は、このような画像上での幾何学的な変形に弱いために、奥行き推定精度が低下してしまうという問題がある。

(2) 遮蔽物の影響による誤推定の発生: あるフレームに撮像されている対象物体が、他のフレームにおいて遮蔽されている場合、SSSD に遮蔽物と対象物体の類似度評価値が加えられてしまうために、対象物体の正しい奥行き値を算出することが困難となる。

(3) 計算コストの増大: 動画を入力とした場合には、一度に多くの動画をを用いて奥行きを推定することで推定精度の向上を図ることが可能となるが、その反面計算コストや必要メモリ容量の増大といった問題が発生する。また、上記の問題 (1),(2) に対する改善策を推定手法に加えると、奥行き推定における計算コストの問題はさらに深刻化する。

このような問題を回避するため、本研究では、図 1 に示すような画像特徴点を用いる新しい奥行き推定手法について提案する。本稿で提案する奥行き推定手法の基本的な枠組みは、一般的なマルチベースラインステレオ法と同じであるが、従来用いられてきた SSSD の代わりに、特徴点の数え上げに基づく新しい評価尺度 TNIP (Total Number of Interest Points) を用いることで、画像の変形やオクルージョンにロバストな奥行き推定を実現する。本手法の基本的なアイデアは、“三次元空間中に特徴点が存在する場合には各画像への投影座標周辺にも特徴点が存在する”という仮定に基づいている。これにより、マルチベースラインステレオにおける奥行き探索時に、各画像上への投影座標に最も多く特徴点が存在する点を探索することで奥行き値を決定する。図 2 に、提案手法による奥行き画像推定の処理の流れを示す。本手法では、まず全ての入力画像上で特徴点を検出する (A)。次に、新しい評価尺度 TNIP によるマルチベースラインステレオ法によって特徴点の奥行き推定を行う (B)。また、信頼度を用いて奥行き推定結果を排除し (C)、最後に、粗な奥行きデータを補間することで密な奥行き画像を生成する (D)。

提案手法を用いることで、従来手法で問題となる三つの問題を解決することができる; (1) 本手法で用いる画像特徴点の検

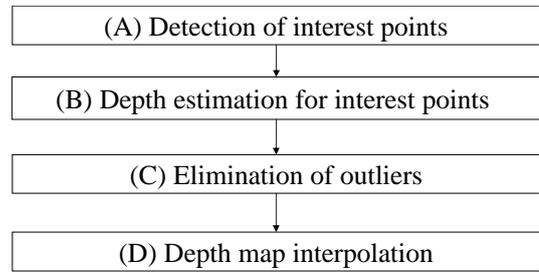


図 2 奥行き画像推定の処理の流れ

出位置は画像の幾何学的な変形に対して頑健である、(2) 新たな評価尺度 TNIP は遮蔽物による影響を受けにくい、(3) 特徴点の数を数え上げるだけで奥行きを決定できるため計算コストが小さい。ただし、本手法においては画像上で特徴点が存在する位置においてしか奥行きを算出できないという問題が新たに発生するが、一般的な屋外環境の三次元モデリングなどの分野においては、特徴点間の奥行き値は補間で十分な場合も多い。

以下、本稿では、まず 2 章において従来提案されている関連研究に対する提案手法の位置づけを述べる。次に 3 章において、本手法の基礎となる SSSD を用いる従来のマルチベースラインステレオ法について解説する。4 章では、マルチベースラインステレオに用いる新たな評価尺度 TNIP について提案する。また、5 章では、TNIP を用いた実際の奥行き推定処理であるステージ (A) からステージ (D) について詳述する。6 章では、シミュレーションおよび実環境を用いた実験結果を示し、提案手法の有用性を示す。最後に、7 章でまとめと今後の課題について述べる。

## 2. 関連研究

奥行き画像推定の分野においては、古くからエッジ線上のエッジを基準としてテンプレートマッチングを行う手法 [7] ~ [10] が用いられてきたが、本稿で提案する手法は、類似度評価尺度を用いないという点で一線を画するものである。画像上での類似度による評価尺度を用いない三次元形状の復元手法として、川戸 [11] が空間への投票に基づく手法を提案している。この手法では、キャリブレーション済みの画像系列に対して、エッジ情報をボクセル空間に投票し、投票数を基準として三次元形状を推定している。また、ボクセル空間をオクトツリーとして表現することで、計算量の削減を図っている。しかし、広域な屋外環境を対象とした場合には、ボクセル表現による三次元形状の復元は計算コストが膨大となるという問題があり、また、カメラキャリブレーションの誤差をどう扱うかといった課題も残されている。他方、structure from motion による手法 [12] では、複数画像上の画像特徴点を対応付けることで、カメラの位置関係やカメラに対する特徴点の三次元位置を得ることができる。本提案手法においても、カメラキャリブレーション手段の一つとして、structure from motion を想定しているが、画像上での対応関係を決定できない特徴点については三次元位置が推定されないために、これらを直接奥行き画像推定や屋外三次元モデリングに利用することは難しい。

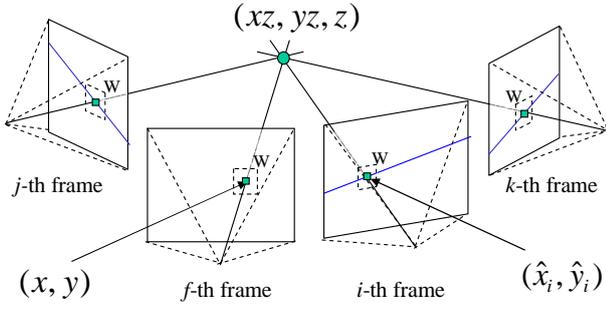


図3 画素  $(x, y)$  の三次元位置と各画像上への投影直線

### 3. 類似度による評価尺度 SSSD を用いたマルチベースラインステレオ

本章ではまず，マルチベースラインステレオ法で用いる座標系に関する定義を行う．次に，SSSD を用いたマルチベースラインステレオ法 [1] の基本原理について簡単に説明する．

#### 3.1 座標系の定義

移動を伴って撮影された動画像に対するマルチベースラインステレオ法では，図3に示すように，第  $f$  フレームにおける画素  $(x, y)$  の奥行き値  $z$  を第  $j$  フレームから第  $k$  フレームまでの画像を用いて推定する ( $j \leq f \leq k$ )．以下では，記述簡単化のため，カメラの焦点距離を1とし，レンズ歪み等はカメラ内部パラメータを用いて補正済みであるものとする．このとき，第  $f$  フレームにおける画素  $(x, y)$  の奥行き値  $z$  は，第  $f$  フレームのカメラ座標系において  $(xz, yz, z)$  と表現される．この三次元座標  $(xz, yz, z)$  は，以下の式によって第  $i$  フレーム上の画像座標  $(\hat{x}_i, \hat{y}_i)$  に投影される．

$$\begin{pmatrix} a\hat{x}_i \\ a\hat{y}_i \\ a \\ 1 \end{pmatrix} = \mathbf{M}_{fi} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix} \quad (1)$$

ただし， $a$  は媒介変数である．また， $\mathbf{M}_{fi}$  は第  $f$  フレームのカメラ座標系から第  $i$  フレームのカメラ座標系への  $4 \times 4$  の変換行列を表す．図3に示すように，マルチベースラインステレオにおいては，三次元座標  $(xz, yz, z)$  の第  $i$  フレームへの投影座標  $(\hat{x}_i, \hat{y}_i)$  は，各画像上においてエピポーラ線上に拘束される．

#### 3.2 SSSD を用いた奥行き推定

従来のマルチベースラインステレオ法では，SSD によって2枚の画像上における一定サイズのウィンドウ  $W$  内のパターン間の類似度を評価し，これらの総和 SSSD を最小化することで画素  $(x, y)$  の奥行き値  $z$  を決定する．第  $f$  フレームにおける画素  $(x, y)$  と，第  $i$  フレームにおける画素  $(\hat{x}_i, \hat{y}_i)$  周辺のパターンの非類似度 SSD は，以下のように定義される．

$$SSD_{fi}(x, y, z) = \sum_{(u,v) \subseteq W} \{I_f(x+u, y+v) - I_i(\hat{x}_i+u, \hat{y}_i+v)\}^2 \quad (2)$$

全ての入力画像に関する類似度を統合的に評価するために，以

下の SSD の総和 SSSD (Sum of SSD) を用いる．

$$SSSD_f(x, y, z) = \sum_{i=j}^k SSD_{fi}(x, y, z) \quad (3)$$

一般的なマルチベースラインステレオ法では，SSSD を最小にする奥行き値を探索することで，第  $f$  フレームにおける画素  $(x, y)$  の奥行き値  $z$  を決定する．ただし，SSSD の大域最小解を得るためには，あらかじめ設定した範囲内における全ての奥行き値を網羅的に探索する必要がある．

ここで，第  $f$  フレームにおける画素  $(x, y)$  に対応する物体が，第  $i$  フレームの画像上では他の物体に遮蔽されていると仮定すると，真の奥行き値  $z$  に対応する非類似度  $SSD_{fi}(x, y, z)$  が非常に大きな値となる．これによって， $(x, y)$  の真の奥行き値  $z$  に対応する SSSD 値も大きな値となり，結果として正しい奥行き値を推定することが困難となる．また，SSSD は画像上のパターンの類似度に基づいて評価を行うため，視点位置・視線方向の変化による見え方の変化に弱いという問題がある．

### 4. 特徴点の数え上げに基づくマルチベースラインステレオ

本章では，一般的なマルチベースラインステレオ法の枠組みを基礎として，画像上の特徴点の数え上げを行うことで頑健に奥行きを推定する手法について詳述する．一般に，三次元空間における物体の角や交点などの特徴的な点 (三次元特徴点) は，撮影画像上においてもエッジのコーナや交点として観測される．このような画像特徴点は，Harris オペレータ [13] や Moravec オペレータ [14] を用いることで容易に検出することが可能である．

本研究では，三次元特徴点の各画像への投影座標において，高い頻度で特徴点が発見されることに着目し，以下の新しい評価尺度 TNIP を最大化する奥行き値  $z$  を探索する．

$$TNIP_f(x, y, z) = \sum_{i=j}^k \sum_{(u,v) \subseteq W} H_i(\hat{x}_i+u, \hat{y}_i+v) \quad (4)$$

$$H_i(u, v) = \begin{cases} 1 & \text{interest point exists} \\ & \text{at } (u, v) \text{ in } i\text{-th frame} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

TNIP は，画素  $(x, y)$  の奥行き値  $z$  に対応する各画像上での探索位置  $(\hat{x}_i, \hat{y}_i)$  周辺に存在する特徴点の総数を表している．ただし，一般的に画像特徴点は三次元特徴点の投影座標に極めて近い範囲内に検出されるため，ここで用いるウィンドウ  $W$  のサイズは， $3 \times 3$  画素程度の小さいウィンドウを設定する．

SSSD の代わりに TNIP を用いて奥行き値  $z$  を探索することで，計算コストのかかる類似度評価を省略することができる．加えて，一般に特徴点の検出座標は画像パターンの変形や照明条件の変化に対してロバストであるため，頑健な奥行き推定が実現できる．また，遮蔽が起こった場合においても，SSSD の場合とは対照的に，それが直接 TNIP の値に対するペナルティとはならないために，正しく奥行きを推定することが可能とな

る．これらについては，後述するシミュレーション実験によっても有効性を示す．

## 5. 動画像からの密な奥行き画像推定

本章では，図 2 に示した密な奥行き画像推定の各ステージについて詳述する．本研究では，まず Harris オペレータを用いて全ての入力画像上で特徴点を検出する (A)．次に TNIP を用いた奥行き推定により，各画像の特徴点位置での奥行き値を算出する (B)．また，複数フレームでの奥行き推定結果を用い，奥行き推定の整合性によって定義される信頼度を用いることで，誤推定結果を排除する (C)．最後に粗な奥行き情報を補間することで密な奥行き画像を生成する (D)．

### 5.1 画像特徴点の検出

ステージ (A) では，画像の拡大・縮小や回転に対して頑健に特徴点を抽出可能な Harris オペレータ [13] を用いることで，各入力画像上の輝度エッジの交点やコーナなどの特徴点を検出する．Schmid ら [15] によれば，Harris オペレータは，画像の幾何学的な変形操作を行った場合に，他のインタレストオペレータに比べて，同じ位置に特徴点が検出される再現度が最も高いオペレータであるとされている．

Harris オペレータでは，入力画像上の座標  $\mathbf{x} = (x, y)$  の特徴量  $F(\mathbf{x})$  算出のために，まずガウシアンオペレータによる入力画像の平滑化処理を行う．次に一定の大きさの正方形窓  $W$  において，画像上の輝度  $I$  の勾配  $I_x, I_y$  を用いて以下に示す行列  $A$  を算出する．これにより特徴量  $F(\mathbf{x})$  は，行列  $A$  の固有値  $\lambda_1, \lambda_2$  の最小値として求めることができる．

$$A = \sum_{\mathbf{x} \in W} \begin{pmatrix} I_x(\mathbf{x})^2 & I_x(\mathbf{x})I_y(\mathbf{x}) \\ I_x(\mathbf{x})I_y(\mathbf{x}) & I_y(\mathbf{x})^2 \end{pmatrix} \quad (6)$$

$$F(\mathbf{x}) = \min(\lambda_1, \lambda_2) \quad (7)$$

本研究では，まず画像内の全ての画素の特徴量  $F(\mathbf{x})$  を算出し，次に一定サイズのウィンドウ内で特徴量  $F(\mathbf{x})$  が極大値となる点を画像特徴点として検出する．本ステージでは，動画像の全てのフレームの画像に対して特徴点の座標を求めておく．

### 5.2 画像特徴点による奥行き推定

ステージ (B) では，検出された全ての特徴点に対して 4 章で定義した TNIP 値を最大化する奥行きをそれぞれ探索することで，全ての画像上で粗な奥行き画像を算出する．ここで，奥行き値  $z$  は，あらかじめ設定した範囲内を網羅的に探索する．

本ステージにおける奥行きの探索では，画像の輝度情報は必要なく，代わりに特徴点の画像座標が必要となる．これにより，計算機に同時に保持すべきデータの必要メモリ容量は，例えば 8bit グレースケール画像に対して  $1/8$  となる．

### 5.3 誤推定結果の排除

図 3 に示したように，前節の手法において算出される画像特徴点  $p$  の奥行き値  $z$  は，特徴点  $p$  と他の複数のフレーム上の特徴点に投影されるため，これらの特徴点を関連付けるものであると考えられる．ステージ (C) では，この関連情報を用いることで，特徴点  $p$  の奥行き推定結果の信頼度  $C_p$  を以下

のように定義する．

$$C_p = \frac{\sum_{i \in L_p} \{0; p \notin L_i, 1; p \in L_i\}}{|L_p|} \quad (8)$$

$L_p$  は特徴点  $p$  に関連づけられた特徴点の集合である． $C_p$  は，特徴点  $p$  に関する奥行き推定値の整合性を表しており， $L_p$  に含まれる特徴点  $p$  が，逆に特徴点  $p$  を関連付ける割合を表している．関連づけられた全ての特徴点の奥行きが正しく推定されるときに限り， $C_p$  は最大値 1 をとる．本ステージでは，信頼度  $C_p$  が一定の閾値を下回る奥行き値は信頼性が低いと判断し，削除する．

### 5.4 奥行き画像の補間

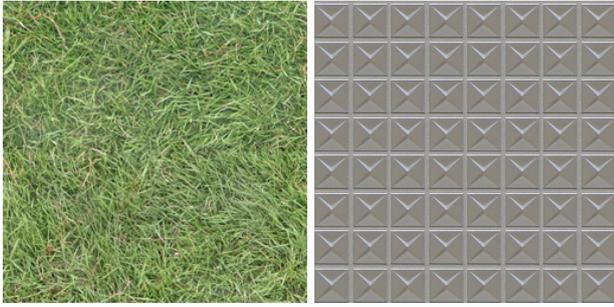
TNIP によるマルチベースラインステレオでは，特徴点以外の画素で奥行きを推定することができないため，密な奥行き画像を生成するためには何らかの補間処理が必要である．本稿では，特徴点が物体の角を表すことから，特徴点間の大部分で奥行きがなめらかに変化していることを仮定し，特徴点間の奥行き値に対する補間処理を行う．ステージ (D) ではまず，Delaunay の三角分割法 [16] を用いることで，入力画像を特徴点を頂点とする多数の三角形に分割する．次に，三角形の内部を画像特徴点の奥行き情報を用いて内挿することで，密な奥行き画像を生成する．

## 6. 実験

本章では，まず TNIP の有効性を示すために，計算機シミュレーションによる TNIP と SSSD の推定精度の比較を行う．次に，広域な屋外環境において奥行き画像を生成できることを示すために，屋外を撮影した実画像を用いた実験結果について示す．

### 6.1 計算機シミュレーションによる SSSD と TNIP の比較

ここでは，SSSD と TNIP の奥行き精度を計算機シミュレーションで比較することによって，提案手法の有効性を示す．  
実験条件：本実験では，2 枚の平面を仮想空間内に配置し，仮想ビデオカメラ (解像度:640 × 480 画素) でそれらを撮影することで，入力画像を作成した．2 枚の平面のテクスチャとしては，図 4 に示すように，自然物である草の模様 (plane 1) と人工物である繰り返しパターンのタイル模様 (plane 2) を用いた．実験に用いた平面の配置とカメラの動きを，図 5 に示す．仮想カメラは円弧を描くよう 1 度刻みで動き，91 枚の画像を撮影した．仮想カメラによって撮影された 91 枚の画像の一部を図 6 に示す．同図から分かるように，動画像後半の画像上では，plane 1 が plane 2 によって遮蔽されオクルージョンが発生している．本実験では，SSSD と TNIP の評価関数そのものの特性を比較するために，5.3 節および 5.4 節で述べた誤推定結果の排除および推定値の補間処理は行わなかった．また，カメラの内部および外部パラメータのキャリブレーション誤差を考慮するために，三次元位置の投影誤差に対して標準偏差  $\sigma$  を持つガウスノイズを付加し，これに加えて投影座標を画素単位に量子化した．本シミュレーション実験における奥行き値の探索範囲は，3,000mm から 35,000mm と設定した．



(a) plane 1 (b) plane 2  
 図 4 対象物体として用いた平面のテクスチャ

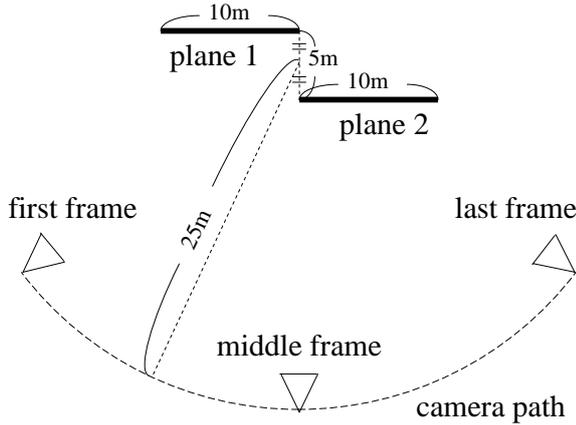
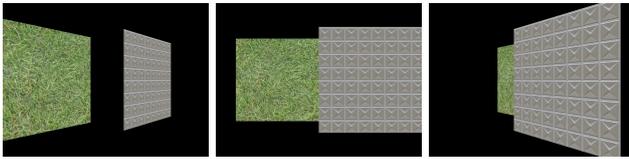


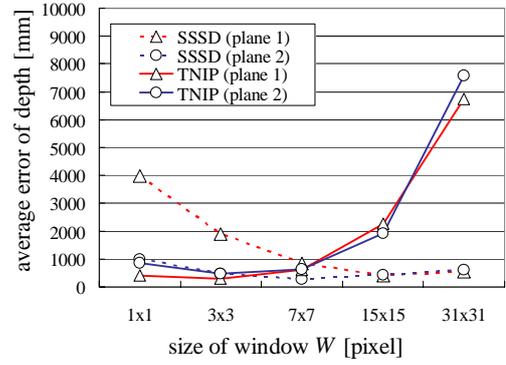
図 5 平面の配置とカメラの動き



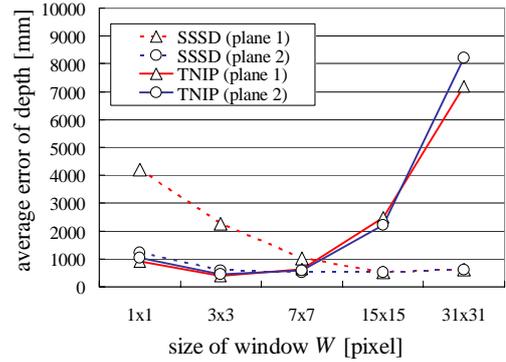
(a) first frame (b) middle frame (c) last frame

図 6 撮影された 91 枚の画像の一部

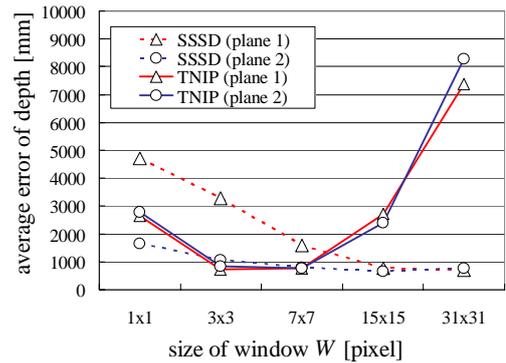
ウィンドウサイズとノイズレベルをパラメータとした実験: ここでは、式 (2) および式 (5) で用いるウィンドウ  $W$  のサイズと、三次元座標の投影位置の誤差レベル  $\sigma$  をパラメータとし、複数パターンのパラメータについて評価を行った。図 7 に本実験によって得られた TNIP および SSSD の推定精度について示す。同図から、TNIP においては  $3 \times 3$  画素の小さなウィンドウサイズを設定した場合に最も良い結果を得ることが分かる。これは、TNIP において大きなウィンドウサイズを設定すると、投影座標周辺に存在する他の画像特徴点との識別が困難となるためであると考えられる。一方、SSSD においては、遮蔽による影響から plane 1 の精度が plane 2 の精度に対して大きく劣っているのに対して、TNIP ではそのような傾向は見られない。これは、オクルージョンによって遮蔽が発生する箇所においても、TNIP が正しい奥行き値を算出できていることを意味する。ただし、SSSD においても  $15 \times 15$  画素程度の大きなウィンドウを用いることで、誤推定を回避し、精度の高い奥行き推定を実現できている。



(a)  $\sigma=0.0$  pixel



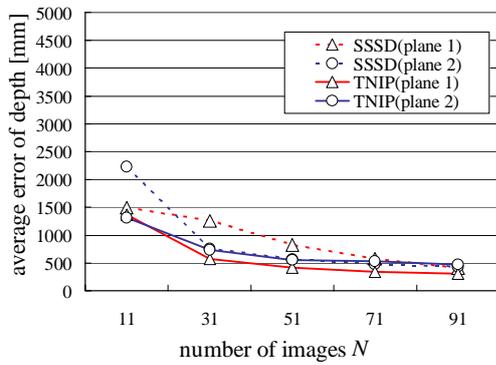
(b)  $\sigma=1.0$  pixel



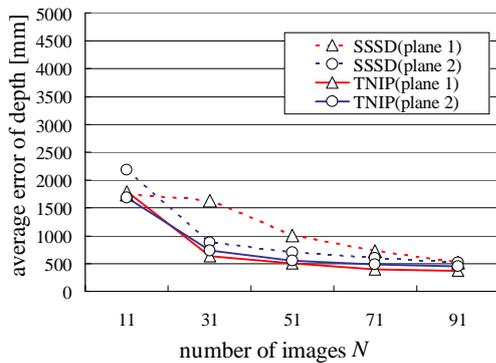
(c)  $\sigma=2.0$  pixels

図 7 奥行き推定における推定誤差の平均値 (パラメータ: ウィンドウサイズ, ノイズレベル)

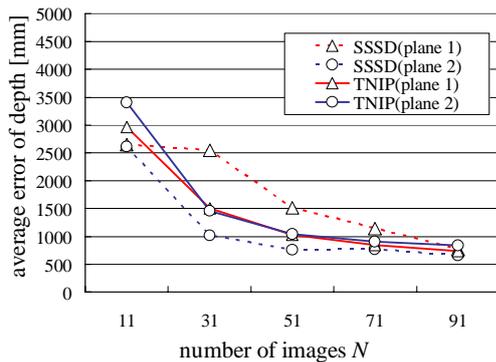
画像枚数とノイズレベルをパラメータとした実験: 次に、画像枚数の変化に対する双方の評価関数の特性を見るために、91枚の入力画像を  $N$  枚に間引いて入力画像とした。先の実験結果に基づき、ここでは TNIP と SSSD の評価関数算出に用いるウィンドウサイズをそれぞれ  $3 \times 3$ ,  $15 \times 15$  に固定した。図 8 に画像枚数とノイズレベルの変化に対するそれぞれの評価関数を用いた平均推定誤差を示す。同図から、どの結果においても、入力画像の枚数を増加させると、それに応じて推定精度が向上していることが確認できる。また、画像枚数が少ない場合には、 $15 \times 15$  画素のような大きなウィンドウを用いた場合でも、plane 1 に対する SSSD の推定精度が比較的悪いことが分かる。オクルージョンの起こらない plane 2 に関しては、TNIP と SSSD の間に有意な差は見られなかった。



(a)  $\sigma=0.0$  pixel



(b)  $\sigma=1.0$  pixel



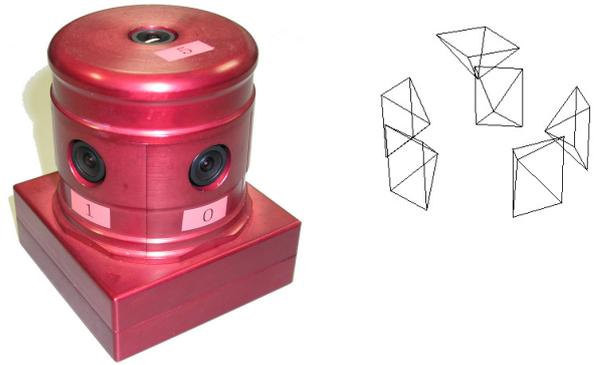
(c)  $\sigma=2.0$  pixels

図 8 奥行き推定における推定誤差の平均値 (パラメータ: 画像枚数, ノイズレベル)

計算時間: 表 1 に, 異なるウインドウサイズを用いた場合の, 1 画素の奥行き算出に要した平均時間を示す. ただし, ここでは入力として全 91 枚の画像を用いており, また計算時間は PC (CPU: Pentium-4 Xeon 3.20GHz dual, Memory: 2GB) を用いた場合の値である. 前述の結果と同表から,  $3 \times 3$  画素の TNIP を用いることで同程度の推定精度を達成できる  $15 \times 15$  画素の SSSD に対して約 40 倍程度高速に奥行きを推定できることが分かる.

表 1 1 画素の奥行き算出に要した平均時間 [ミリ秒]

window size $W$	$1 \times 1$	$3 \times 3$	$7 \times 7$	$15 \times 15$	$31 \times 31$
time for SSSD	10.9	24.2	80.3	335.6	1529
time for TNIP	7.2	8.2	9.2	10.7	19.0



(a) appearance

(b) view volume

図 9 全方位型マルチカメラシステム Ladybug



図 10 入力動画画像系列の 1 フレーム

## 6.2 屋外環境における奥行き画像推定

本実験では, 図 9 に示す, PointGreyResearch 社製の全方位型マルチカメラシステム Ladybug [17] を用い, 大学キャンパスを動きながら撮影した. Ladybug は合計 6 つのカメラユニットを持ち, 各カメラユニットはそれぞれ  $768 \times 1024$  画素の画像を 15fps の動画として撮影できる. 入力として用いた画像は, 図 10 に示す 6 枚を含め 3000 枚 (500 フレーム) である. 本実験において, 各カメラの内部パラメータおよびマルチカメラシステムの位置・姿勢パラメータは, それぞれトータルステーションとマーカボードによるキャリブレーション手法 [18] および基準マーカと自然特徴点の追跡による手法 [19] によってあらかじめ推定し, 利用した. 図 11 に, 奥行き画像推定に利用したカメラシステムの移動パラメータを示す. 図中の曲線および錘台はそれぞれ, 図 10 左上の画像に対応するカメラユニットの移動の軌跡および 20 フレーム毎の姿勢を表している. 事前に行った評価実験から, 得られたカメラパスの推定精度は, 位置に関して平均誤差 50mm, 姿勢に関して平均誤差  $0.07$  度であり, 利用したカメラパスの長さは 29m である [19].

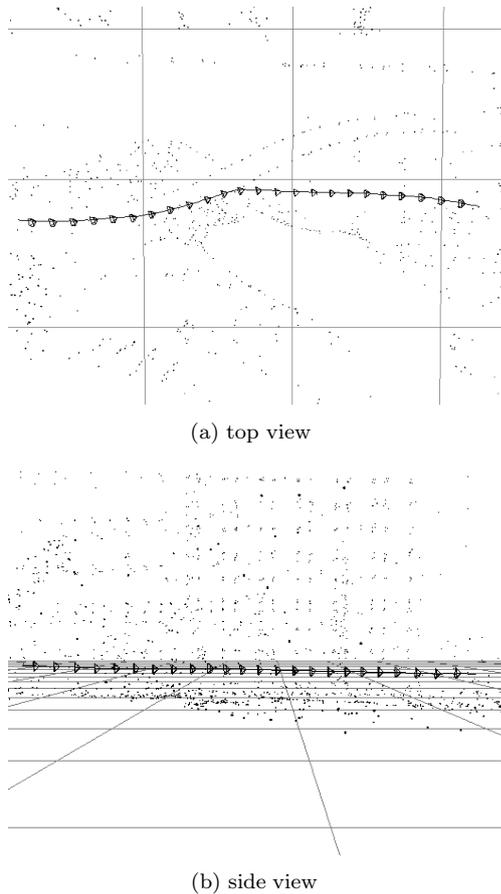


図 11 入力として用いたカメラパス (29m)

このような入力データを用い、5章に述べた手順で奥行き画像を生成した。まずステージ (A) では、各入力画像上において特徴点を検出した。検出された画像特徴点は、各入力画像において平均 1750 点 (1 フレームでの合計平均約 10500 点) であった。

次に、ステージ (B) では、TNIP によるマルチベースラインステレオ法によって、各画像特徴点の奥行き値を算出した。本実験では、第  $f$  フレームに存在する特徴点の奥行きを算出するために、第  $(f-100)$  フレームから第  $(f+100)$  フレームまで 2 フレームおき (606 枚, 101 フレーム) の画像内に存在する特徴点を利用した。また、特徴点探索ウィンドウのサイズは、前節の実験に基づき  $3 \times 3$  画素と設定し、奥行き探索範囲は 1,000mm から 80,000mm とした。

ステージ (C) では、奥行き推定結果の整合性に基づいて信頼度を算出し、信頼度の低い結果を削除した。本実験では、信頼度の閾値を 0.5 とした。図 12 に、図 10 に対応する画像特徴点の位置および、奥行き推定結果を輝度値に変換したものを示す。また、各画像中から無作為に選択された 6 つの特徴点 (図 12 参照) の奥行き値算出時の TNIP 値を、図 13 に示す。同図から、いずれの特徴点においても、正しいと思われる奥行き値付近で TNIP 値が最大値をとることが分かる。また、それ以外の奥行き値では TNIP 値に明白なピークが見られないために、これらの特徴点ではロバストな奥行き推定が実現されていることが確認できる。

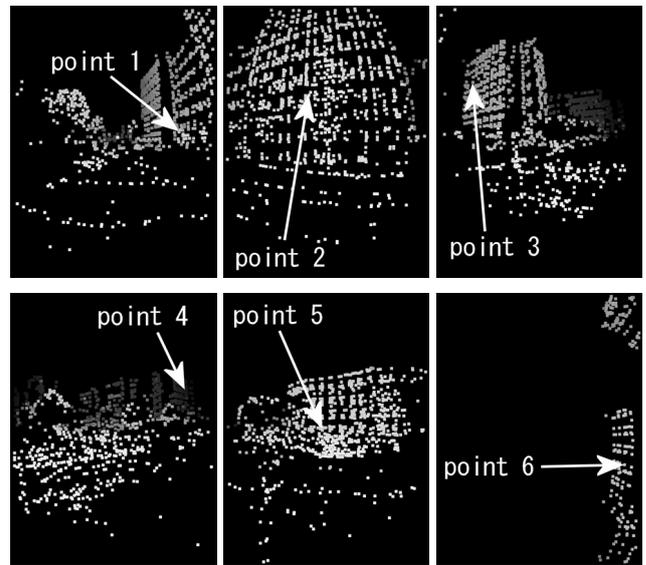


図 12 図 10 に対する奥行き推定結果

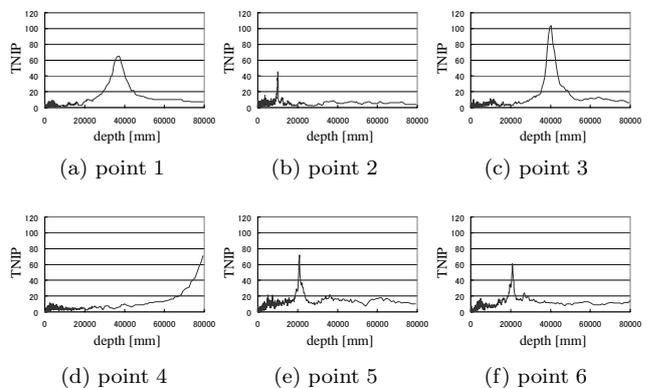


図 13 奥行き探索における TNIP 値

最後にステージ (D) において、密な奥行き画像を生成した。図 10 から得られた全方位画像を極座標展開したパノラマ画像を図 14 に、奥行き値の内挿処理によって生成された奥行き画像を極座標展開したものを図 15 に示す。図 14 と図 15 の比較により、良好に奥行き画像が推定されていることを確認できる。ただし、本研究では特徴点を頂点とする三角分割を行っているために、本来とは異なる奥行き値が算出されている箇所が見られる。今後、画像のエッジ情報等を用いた三角分割手法の導入が必要である。

## 7. まとめ

本稿では、画像特徴点の数え上げによるマルチベースラインステレオ法を提案した。本手法は、画像間の類似度計算を必要としないため、多数の画像を用いた場合にも比較的高速に奥行き画像を生成可能である。また、画像の輝度値を直接比較しないため、照明条件の変化や射影歪みの影響を受けにくく、ロバストな奥行き推定が可能である。今後は、推定された多数の奥行き画像を統合することで屋外環境の三次元モデル化を行う。

## 文献

- [1] M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence,



図 14 パノラマ展開した全方位動画の 1 フレーム

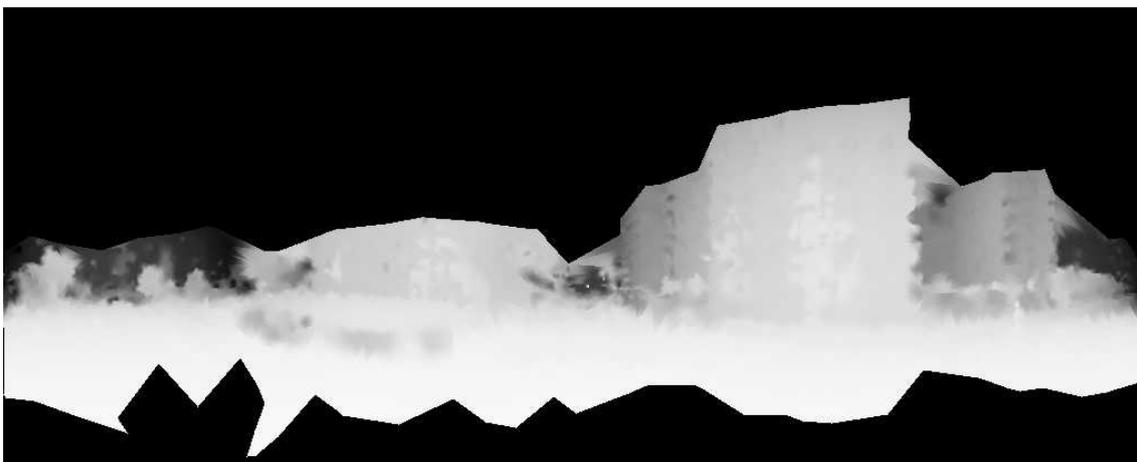


図 15 生成された全方位奥行き画像

- Vol. 15, No. 4, pp. 353–363, 1993.
- [2] S. B. Kang, J. A. Webb, C. Zitnick and T. Kanade: “A Multibaseline Stereo System with Active Illumination and Real-time Image Acquisition,” Proc. Int. Conf. on Computer Vision, pp. 88–93, 1995.
- [3] S. B. Kang and R. Szeliski: “3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo,” Int. Journal of Computer Vision, Vol. 25, No. 2, pp. 167–183, 1997.
- [4] W. Zheng, Y. Kanatsugu, Y. Shishikui and Y. Tanaka: “Robust Depth-map Estimation from Image Sequences with Precise Camera Operation Parameters,” Proc. Int. Conf. on Image Processing, Vol. II, pp. 764–767, 2000.
- [5] T. Sato, M. Kanbara, N. Yokoya and H. Takemura: “Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera,” Int. Journal of Computer Vision, Vol. 47, No. 1-3, pp. 119–129, 2002.
- [6] M. Okutomi, Y. Katayama and S. Oka: “A Simple Stereo Algorithm to Recover Precise Object Boundaries and Smooth Surface,” Int. Journal of Computer Vision, Vol. 47, No. 1-3, pp. 261–273, 2002.
- [7] H. H. Baker: “Edge Based Stereo Correlation,” Proc. Image Understanding Workshop, pp. 168–175, 1980.
- [8] W. E. L. Grimson: “Computational Experiments with a Feature-based Stereo Algorithm,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 7, No. 1, pp. 17–34, 1985.
- [9] S. Pollard, M. Pilu, S. Hayes and A. Lorusso: “View Synthesis by Trinocular Edge Matching and Transfer,” Image and Vision Computing, Vol. 18, No. 9, pp. 739–748, 2000.
- [10] M. Agrawal and L. S. Davis: “Trinocular Stereo Using Shortest Paths and the Ordering Constraint,” Int. Journal of Computer Vision, Vol. 47, No. 1-3, pp. 43–50, 2002.
- [11] 川戸: “空間への 2 段階投票による 3 次元情報の抽出”, 電子情報通信学会論文誌 (D-II), Vol. J77-D-II, No. 12, pp. 2334–2341, 1994.
- [12] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: “Bundle Adjustment a Modern Synthesis,” Proc. Int. Workshop on Vision Algorithms, pp. 298–372, 1999.
- [13] C. Harris and M. Stephens: “A Combined Corner and Edge Detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.
- [14] H. Moravec: “Towards Automatic Visual Obstacle Avoidance,” Proc. Int. Joint Conf. on Artificial Intelligence, p. 584, 1977.
- [15] C. Schmid, R. Mohr and C. Bauckhage: “Evaluation of Interest Point Detectors,” Int. Journal of Computer Vision, Vol. 37, No. 2, pp. 151–172, 2000.
- [16] P. Heckbert Ed.: Graphics Gems IV, pp. 47–59, Academic Press, 1994.
- [17] Point Grey Research Inc.: “Ladybug,” <http://www.ptgrey.com/>.
- [18] S. Ikeda, T. Sato and N. Yokoya: “High-resolution Panoramic Movie Generation from Video Streams Acquired by an Omnidirectional Multi-camera System,” Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System, pp. 155–160, 2003.
- [19] 佐藤, 池田, 横矢: “複数動画からの全方位型マルチカメラシステムの位置・姿勢パラメータの推定”, 電子情報通信学会論文誌 (D-II), Vol. J88-D-II, No. 2, pp. 347–357, 2005.