

LI-002

# 自然特徴点を利用したマルチベースラインステレオによる全方位奥行き画像推定

## Omni-directional depth-map estimation by multi-baseline stereo using interest points

佐藤 智和<sup>†</sup>  
Tomokazu SATO

池田 聖<sup>†</sup>  
Sei IKEDA

横矢 直和<sup>†</sup>  
Naokazu YOKOYA

### 1. はじめに

複数の画像を用いて画像の奥行き情報を推定するステレオ法は、コンピュータビジョンの分野で古くから研究され、三次元形状推定や、物体認識、拡張現実感などさまざまな分野で利用されている。これらステレオ法のなかでも、多数の画像を同時に利用することで各画素の奥行きを推定するマルチベースラインステレオ法 [1] は、一般的な二眼ステレオ法に比べて高い精度で奥行きを推定できることから注目され、広く用いられている [2, 3, 4, 5]。

一般的に、マルチベースラインステレオ法は、校正済みの複数台の固定カメラを用いて同時に撮影した画像からの奥行き推定問題に利用されている [1, 2, 5] が、近年、移動を伴って撮影された動画像を入力として用いることで、同時に扱う画像の枚数を増加させ、推定精度を向上させる試みが成されている [3, 4]。しかし、マルチベースラインステレオ法によって各画素の奥行き情報を推定するためには、入力画像間で多数の小領域の類似度を算出することが必要となるため、動画像のような多数の画像を入力とした場合には、推定精度が向上する反面、各入力画像への参照が頻発し、計算量や必要メモリ容量が膨大になるという問題がある。また、奥行き推定の評価尺度として一般に用いられている SSD (Sum of Squared Differences) は、照明条件の変化や画像の変形に対してロバストでないという問題があった。

そこで本稿では、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 の代わりに、自然特徴点の画像上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、自然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。また本研究では、全方位型マルチカメラシステムによって撮影される動画像系列を入力画像として用いることで、一度に全方位の奥行きを高精度に推定する。実験では、自然特徴点の座標を用いた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推定できることを示す。ただし本稿では、全方位動画像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。

以下、2章では、自然特徴点の画像上の座標を用いた奥行き画像の推定手法について述べる。3章では、カメラパラメータを推定済みの全方位動画像を入力とした、全方位奥行き画像の推定実験について報告する。最後に、4章でまとめと今後の課題について述べる。

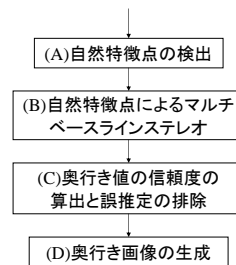


図 1: 処理の流れ

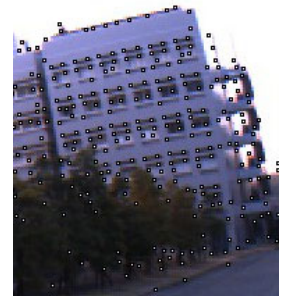


図 2: 自然特徴点の例

### 2. 自然特徴点による全方位奥行き画像推定

図 1 に、提案手法の処理の流れを示す。本手法では、まず図 2 に白点で示すような、画像上の輝度エッジの交点やコーナなどの自然特徴点を動画像の全てのフレームの画像上で検出する (A)。次に、検出された自然特徴点の画像上の座標を用いたマルチベースラインステレオ法によって画像上の自然特徴点の奥行き情報をフレーム毎に推定する (B)。また、ステップ (B) においてフレーム毎に推定された奥行き推定結果を統合的に用いることで、推定された自然特徴点の奥行き値の信頼度を算出し、誤推定を排除する (C)。最後に、入力画像上で奥行き値を内挿することで、全方位の奥行き画像を生成する (D)。以下では、処理の各ステップ (A) ~ (D) について順に述べる。

#### 2.1 自然特徴点の検出 (A)

本ステップでは、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 [6] を用いることで、図 2 に白点で示すような、各入力画像上の輝度エッジの交点やコーナなどを検出する。Schmid ら [7] によれば、Harris オペレータは、回転、拡大などの画像の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。

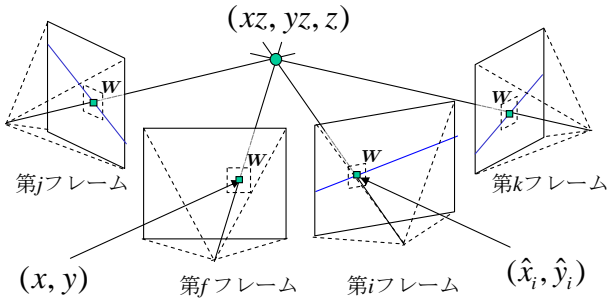
Harris オペレータによる入力画像上の座標  $\mathbf{x} = (x, y)$  の特徴量  $H(\mathbf{x})$  は、まずガウシアンオペレータによる入力画像の平滑化処理を行い、次に一定の大きさの正方形窓  $W$  において、画像上の輝度  $I$  の勾配  $I_x, I_y$  を用いて以下に示す行列  $\mathbf{A}$  を算出し、その行列の固有値  $\lambda_1, \lambda_2$  の最小値を求めることで得られる。

$$\mathbf{A} = \sum_{\mathbf{x} \in W} \begin{pmatrix} I_x(\mathbf{x})^2 & I_x(\mathbf{x})I_y(\mathbf{x}) \\ I_x(\mathbf{x})I_y(\mathbf{x}) & I_y(\mathbf{x})^2 \end{pmatrix} \quad (1)$$

$$H(\mathbf{x}) = \min(\lambda_1, \lambda_2) \quad (2)$$

本研究では、まず画像内の全ての画素の特徴量  $H(\mathbf{x})$  を算出し、次に一定サイズのウィンドウ内で特徴量  $H(\mathbf{x})$

<sup>†</sup> 奈良先端科学技術大学院大学 情報科学研究科


 図 3: 画素  $(x, y)$  の三次元位置と各画像上への投影直線

が極大値となる点を自然特徴点として検出する．本ステップでは，動画の全てのフレームの画像に対して自然特徴点の座標を求めておく．

## 2.2 自然特徴点によるマルチベースラインステレオ (B)

本ステップでは，ステップ (A) において動画の各フレームで検出された自然特徴点の座標を用い，マルチベースラインステレオ法を利用することで，自然特徴点の奥行き値を推定する．以下では，まず一般的なマルチベースラインステレオで用いる座標系について定義し，次に従来手法である小領域の類似度を評価尺度として用いるマルチベースラインステレオ法について述べる．最後に，提案手法である自然特徴点によるマルチベースラインステレオ法を用いて，自然特徴点の奥行き値を探索する手法について詳述する．

[マルチベースラインステレオにおける座標系の定義]  
 動画を入力と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  $f$  フレームにおいて撮影される画素  $(x, y)$  の奥行き値  $z$  を，その周辺に配置された第  $j$  フレームから第  $k$  フレームの画像を用いて推定する ( $j \leq f \leq k$ )．簡単のためにカメラの焦点距離を 1 とすれば，第  $f$  フレームにおける画素  $(x, y)$  に投影される三次元位置のカメラ座標系における座標は  $(xz, yz, z)$  となり，以下の式により，この三次元座標は第  $i$  カメラ ( $j \leq i \leq k$ ) の画像上  $(\hat{x}_i, \hat{y}_i)$  に投影される．

$$\begin{pmatrix} a\hat{x}_i \\ a\hat{y}_i \\ a \\ 1 \end{pmatrix} = \mathbf{M}_i \mathbf{M}_f^{-1} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix} \quad (3)$$

ただし， $a$  は媒介変数， $\mathbf{M}_f$  は第  $f$  フレームでの世界座標からカメラ座標への変換行列である．図 3 に示すように， $(\hat{x}_i, \hat{y}_i)$  は， $(xz, yz, z)$  と第  $f$  フレームの投影中心を結ぶ直線を各探索画像面上に投影した直線上に拘束される．

[類似度評価による従来のマルチベースラインステレオ]  
 小領域の類似度を評価尺度として用いる従来のマルチベースラインステレオ法 [1] では，図 3 に示すように，第  $f$  フレームにおける画素  $(x, y)$  を中心とするウィンドウ  $W$  と第  $i$  フレームにおける画素  $(\hat{x}_i, \hat{y}_i)$  を中心とするウィンドウ  $W$  の輝度値  $I$  の差の二乗和 SSD (Sum of

Squared Differences) を誤差の尺度として用いる．

$$SSD_{fi}(x, y, z) = \sum_{(u,v) \subseteq W} \{(I_f(x+u, y+v) - I_i(\hat{x}_i+u, \hat{y}_i+v))^2\} \quad (4)$$

SSD は，第  $f$  フレームと第  $i$  フレーム上のウィンドウ間のマッチング誤差であるが，類似度評価による従来のマルチベースラインステレオ法では，これらの和である SSSD (Sum of SSD) を最小とするような奥行き値  $z$  を探索する．

$$SSSD_f(x, y, z) = \sum_{i=j}^k SSD_{fi}(x, y, z) \quad (5)$$

[自然特徴点によるマルチベースラインステレオ]

ここでは，SSSD に代わる新しい評価尺度 TNIP について述べる．一般に，三次元空間内の物体の角など三次元空間中に存在する特徴点は，画像上に投影すると輝度エッジのコーナや交点のような自然特徴点として表れる．このような点は，Harris オペレータを用いて容易に検出できるため，“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点に自然特徴点を検出される”という仮定を用いて奥行き値を復元する．提案手法では，ステップ (A) で検出された自然特徴点の座標を特徴点の投影座標として用い，評価尺度 TNIP (Total Number of Interest Points) を以下のように定義する．

$$TNIP_f(x, y, z) = \sum_{i=j}^k \sum_{(u,v) \subseteq W} H_i(\hat{x}_i+u, \hat{y}_i+v) \quad (6)$$

$$H_i(u, v) = \begin{cases} 1; & \text{feature exists at } (u, v) \text{ in } i\text{-th frame} \\ 0; & \text{otherwise} \end{cases} \quad (7)$$

TNIP は，三次元座標  $(xz, yz, z)$  を動画の各フレームの画像上に投影した座標  $(\hat{x}_i, \hat{y}_i)$  周辺の小領域のウィンドウ  $W$  内に存在する自然特徴点数の総和を表しており (図 3 参照)，TNIP が大きければ，三次元座標  $(xz, yz, z)$  に実際の特徴点が存在する可能性が高いということになる．このため，現フレーム  $f$  内の座標  $(x, y)$  に存在する自然特徴点の奥行き値  $z$  は，特徴点と投影中心を結ぶ直線上の任意の座標  $(xz, yz, z)$  の中から，TNIP が最大となる奥行き値  $z$  を探索することで決定する．SSSD の代わりに TNIP を用いることで，奥行き探索において計算時間の大半を占める輝度値情報の比較処理が無くなり，計算量を大幅に削減できる．また，ステップ (A) において検出された自然特徴点の座標は，照明条件や射影歪みの影響を受けにくいいため，ロバストな奥行き探索を実現できる．本ステップでは，全ての入力画像上の全ての自然特徴点に対して，奥行き値  $z$  を算出する．

## 2.3 奥行き値の信頼度の算出と誤推定の排除 (C)

前節の手法において算出される自然特徴点  $p$  の奥行き値  $z$  は，自然特徴点  $p$  と他の複数のフレームの自然特徴点を関連付けるものである．ここでは，この関連情報を用いることで，自然特徴点  $p$  の奥行き推定結果の信頼度



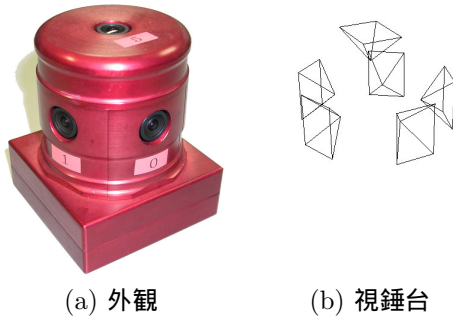


図 4: 全方位型マルチカメラシステム "Ladybug"

$C_p$  を以下のように定義する .

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

$L_p$  は自然特徴点  $p$  に関連づけられた特徴点の集合である .  $C_p$  は , 自然特徴点の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自然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 . 本ステップでは , 信頼度  $C_p$  が一定の閾値を下回る自然特徴点の奥行き値  $z$  は信頼性が低いと判断し , 削除する .

#### 2.4 奥行き画像の生成 (D)

先に述べた奥行き推定処理は , 各画像の自然特徴点の座標でのみ成立するため , 全ての画素の奥行き値を算出するためには何らかの補間処理が必要である . ここでは , ドローンの三角分割法 [8] を用いて , 入力画像を自然特徴点を頂点とする小領域に分割する . 次に内挿処理によって , 各画素の奥行き値を算出する . この処理をフレーム毎に繰り返すことで , 全てのフレームにおける奥行き画像を生成する .

### 3. 実験

本実験では , 図 4 に示す , PointGreyResearch 社製の全方位型マルチカメラシステム Ladybug[9] 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 の動画像として撮影できる . 入力として用いた画像は , 図 5 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 [10] および基準マーカと自然特徴点の追跡による手法 [11] によってあらかじめ推定し , 利用した . 図 6 に , 全方位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 5 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 , 本実験で用いたカメラパスの推定精度は , 位置に関して平均誤差 50mm , 姿勢に関して推定誤差  $0.07^\circ$  程度であり , 利用したカメラパスの長さは 29m である .



図 5: 全方位動画像の 1 フレームを構成する 6 枚の画像

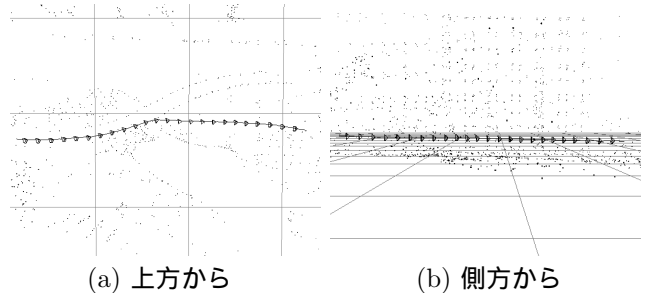


図 6: 入力として用いたカメラシステムの移動パラメータ

このような入力データを用いて , まず各入力画像上において 2.1 節で述べた手法によって , 自然特徴点を検出した . 本実験で検出された自然特徴点は , 各フレームの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 , 2.2 節において述べた手法によって , 自然特徴点の座標情報から奥行き値を算出し , 2.3 節で述べた手法によって信頼度の低い推定結果を削除した . 本実験では , 式 (6) における探索用ウィンドウサイズを  $3 \times 3$  , 信頼度の閾値を 0.5 に設定した . また , 各自然特徴点の奥行き推定には , その自然特徴点を含むフレームの前後それぞれ 100 フレーム分の画像を 2 フレームスキップで用いた .

図 7 に , 図 5 に対応する自然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 つの自然特徴点 (図 7 参照) の奥行き値算出時の TNIP 値を , 図 8 に示す . 同図から , いずれの自然特徴点においても , 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 値にピークが見られないために , これらの自然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 . 本実験において , 一枚の画像に含まれる自然特徴点の奥行き推定に要した時間は , PC(Pentium4-3.2GHz) を用いて平均 3 分程度であった .

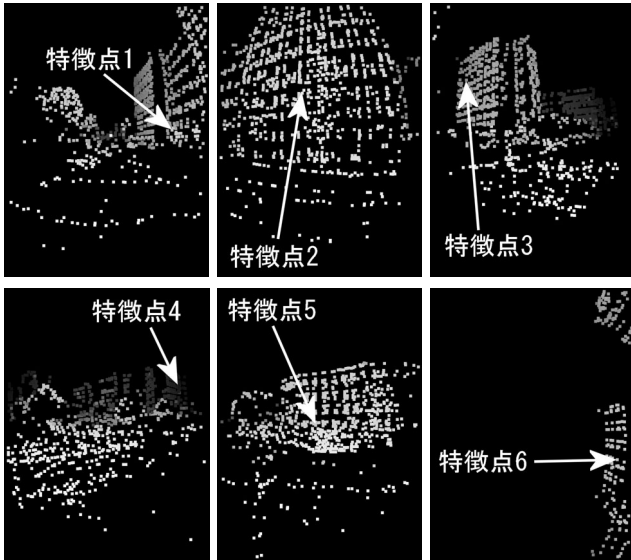


図 7: 自然特徴点の奥行き推定結果

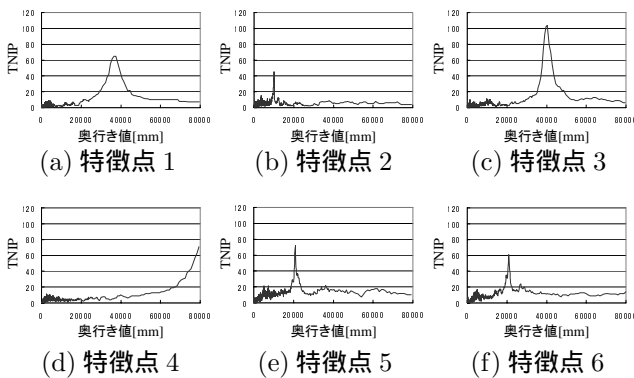


図 8: 奥行き探索における TNIP 値

最後に図 5 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 9 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 10 に示す。図 9 と図 10 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、本研究では自然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および、推定結果の評価が必要である。

#### 4. まとめ

本稿では、自然特徴点の座標を用いたマルチベースラインステレオ法による全方位奥行き画像の推定手法を提案した。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、推定された多数の全方位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。



図 9: 入力全方位画像の 1 フレームをパノラマ展開した画像

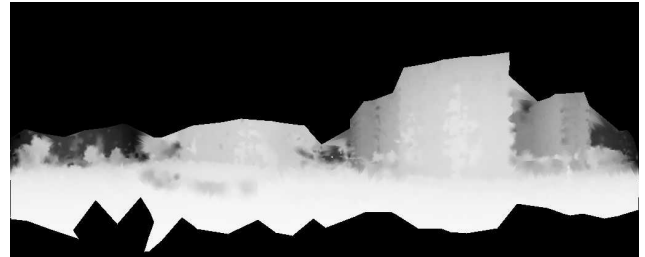


図 10: 生成された奥行き画像の 1 フレーム

#### 参考文献

- [1] M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353-363, 1993.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- [7] C. Schmid, R. Mohr and C. Bauckhage: "Comparing and Evaluating Interest Points," Proc. 6th Int. Conf. on Computer Vision, pp. 230-235, 1998.
- [8] P. Heckbert Ed.: Graphics Gems IV, pp. 47-59, Academic Press, 1994.
- [9] Point Grey Research Inc.: "Ladybug," <http://www.ptgrey.com/products/ladybug/index.html>.
- [10] 池田, 佐藤, 横矢: "全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画の生成とテレプレゼンスへの応用", 日本バーチャルリアリティ学会論文誌, Vol. 8, No. 4, pp. 443-450, 2003.
- [11] 佐藤, 池田, 横矢: "マルチカメラシステムによる全方位動画からのカメラ外部パラメータの復元", 情報処理学会研究報告, CVIM141-13, 2003.