(54) クレーン周辺の三次元地図作成における視差画像統合手法の検討

小林　知生[1]・須崎　純一[2]・繁森　央一[3]
大庭　哲治[4]・石井　順恵[2]・Marek OSOSINSKI[5]

[1]学生会員　京都大学大学院工学研究科（〒615-8540 京都市西京区京都大学桂）
E-mail: kobayashi.tomoki.28s@st.kyoto-u.ac.jp
[2]正会員　京都大学大学院工学研究科（〒615-8540 京都市西京区京都大学桂）
E-mail: {susaki.junichi.3r; ishii.yoshie.4k}@kyoto-u.ac.jp
[3]非会員　住友商事株式会社 （〒100-8601 東京都千代田区大手町 2-3-2）
E-mail: hisakazu.shigemori@sumitomocorp.com
[4]正会員　京都大学経営管理大学院（〒606-8501 京都市左京区吉田本町）
E-mail: oba.tetsuharu.5n@kyoto-u.ac.jp
[5]非会員　株式会社タダノ　技術研究所（〒761-0301 香川県高松市林町 2217-13）
E-mail: marek.ososinski@tadano.com

建設現場での人手不足が深刻化する昨今，クレーン操縦の自動化技術が注目されており，その実現のためにはクレーン周辺の三次元地図を精確かつ高速に作成することが必須ともいえる．本研究ではクレーンのブーム先端に取り付けた単眼カメラから取得した動画像に対し SGM を適用することで複数の視差画像を生成し，それらを統合することでクレーン周辺の三次元地図の生成を試みた．その際，画像の大部分を占める地物がフレームごとに異なる場合でも手法を適用できるよう，視差画像間で視差の基準合わせを行った．その結果， SLAM などを用いた先行研究では得ることが難しかった密な三次元地図を，上記の特殊な条件下においても準リアルタイムの処理時間で算出することが可能であることを示した．

Key Words: *photogrammetry, stereo matching, semi-global matching, disparity image, 3D mapping.*

1. はじめに

建設現場における人手不足が問題となっている昨今，クレーンの自動操縦技術が注目されている．この技術の実現にはクレーン周辺の高精度な三次元地図を作成することが必須ともいえる．また建設現場では状況が刻一刻と変化するため高速性も重要となる．これを踏まえ，クレーンブーム先端に設置した単眼カメラから撮影した動画像を基にクレーン周辺の三次元地図を高精度かつ準リアルタイムで作成する手法の開発を本研究の目的とした．

先行研究のうち旋回するブームの先端に取り付けた単眼カメラのみに頼るものとして王ほか [1)]や繁森ほか [2)]の研究がある．これらの研究では高速性に長けた三次元復元手法である Simultaneous Localization and Mapping (SLAM)[3)] を用いている．これらの手法の問題点として，SLAM はキーフレームの特徴点のみに対して深度推定を行うため，出力される三次元点群に空白部分が発生することがある．

そこで本研究ではフレーム間で視差画像を作成する Semi-Global Matching (SGM)[4)]の手法に注目した．視差画像とは，平行化した 2 枚の左右画像において左画像上の各画素に対応する右画像上の画素を検出し（ステレオマッチング），それら対応する画素の水平方向のずれ（視差）を基に，物体とカメラとの距離を可視化した画像である．視差画像においてもオクルージョンなどにより視差の計算に失敗し得るが，本研究では位置合わせした複数枚の視差画像を参照することで，従来の手法に比べ密な三次元地図が得られると考えた．また SGM を利用した先行研究 [5)]では，全フレームで同一高度の地物（地面など）が常に多くの面積を占めることが前提であり，それ以外の場合には視差画像同士を適切に統合できないという問題があった．そこで本研究では，隣接する視差画像において視差の基準合わせをすることで問題の解決を図った．

2. 使用データ

クレーンブームの運動には伸縮運動，起伏運動，旋回運動の 3 種類があるが，本研究では旋回運動のみを行うブームから撮影された動画像を対象とする．

今回用いる動画像はシミュレータ（Gazebo[6]）座標系において高度 15 m を水平に旋回するカメラから仮想的に撮影され，高度約 4 m の建物を模した地物やクレーン基部などを写している．長さは約 20.1 秒，フレームレートは 30 fps である．また画像の大部分を占める地物がフレームごとに異なる場合について検証するため，動画像の中盤で地物が画像の大部分を占めるよう調整した．

3. 本手法の概要

(1) 概要

本研究の提案手法では**図-1** の流れで処理を行った．それぞれの過程について簡単に説明する．

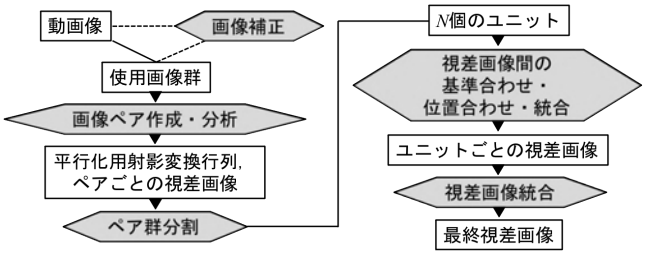


図-1 提案手法のフローチャート

(2) 使用画像およびユニットの生成

まず動画像から毎秒 f フレームの割合でフレームを抽出し元画像群とする．これに対し必要であればレンズ歪み補正などの画像補正を行ったのち使用画像群とする．

今回は全画像から直接 1 枚の視差画像を算出するのではなく，まず全画像から N 枚の全体視差画像を生成し，それらを重ね合わせて最終視差画像を算出することで，ノイズ除去を試みた．これの実装のため本手法では全画像を N 個の小さな画像群（以下ユニット）に分割した．具体的には，全使用画像の通し番号を n，ユニットの通し番号を m としたとき，$n \equiv m \pmod N$ となるように各ユニットに使用画像群を分配した．

(3) 画像ペアの分析

視差画像を算出する際の入力画像である 2 枚の画像（時系列順に左，右画像）は平行化されている必要がある．平行化とは，右画像上の任意の特徴点の y 座標が，それに対応する左画像上の特徴点の y 座標と等しくなるよう右画像を変換する処理である．この処理には左右画像の対応点情報を基に算出した射影変換行列を利用した．平行化後の左右画像の対応点を描画した例が**図-2** である．平行化の結果，各対応点の y 座標が同一となっている．

次に左右画像より視差画像を算出する．今回，視差画像を算出する手法としては，前述の SGM の改良版である Semi-Global Block Matching (SGBM)を用いた． SGBM ではブロックマッチング により隣接画素におけるノイズによる視差の不連続を軽減する．

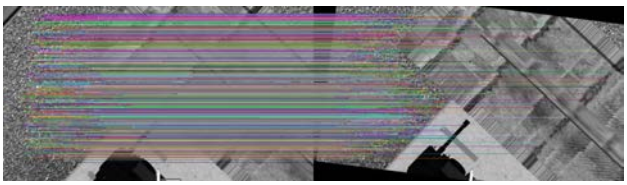


図-2 平行化後のペア画像対応点描画の例

(4) 視差画像の基準合わせ

平行化のための行列は先述のように画像内の対応点情報を基に算出するため，画像内に写る複数の地物の中でも対応点が多く検出される地物の像を一致させる（視差を 0 とする）行列が算出されやすい．ここで，検出される対応点数は地物のテクスチャなどのほか，地物が画像内を占める面積にも影響を受けるが，どの地物が画像内で多くの面積を占めるかはフレームによって異なる．そのためどの地物の（高度の）視差が 0 となるかが視差画像間で揃わない場合がある．このような状態のままでは視差画像同士を適切に統合できないため，隣接視差画像間で視差の基準を揃える必要がある．今回，隣接視差画像間の基準合わせ処理を以下の手順で実施した．

まず隣接する視差画像同士（時系列順に A，B）の重複部分における視差分布（d_a，d_b）を算出する．重複部分では同一地物が写されているはずであり，その結果 d_a，d_b も似た形状の分布となるはずである．しかし前述の理由で視差画像の基準が揃っていない場合には，分布形状自体は似ているもののずれが発生する場合がある．

今回は d_a に対しある量のずれ幅を与えて生成した新たな分布と d_b の類似度が最大となるずれ幅を基準合わせのためのオフセット値として利用した．

ここで視差分布同士の類似度指標としては Kullback-Leibler（KL）情報量[7]を用いた．KL 情報量は 2 つの分布に対し非負の値として計算され，分布同士の類似度が高いほど小さく（0 に近く）なるという性質がある．一般的に離散関数 $p(x)$，$q(x)$ に対する KL 情報量は式(1)で定義される．ただし，n，x はそれぞれ分布の階級数，階級値を表す．

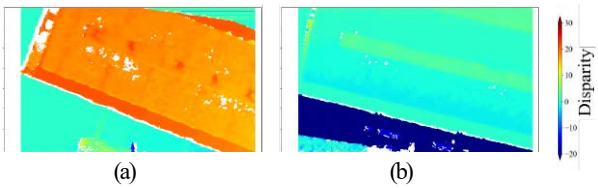
$$KL(p \parallel q) = \sum_{x=1}^{n} p(x) \log \frac{p(x)}{q(x)} \qquad (1)$$

これを踏まえ式(2)を満たすwをオフセット値とした．ただしZは整数全体の集合を表す．例えばwをAの各画素の視差値に足し算することでA，Bの視差の基準を合わせることができる．

$$w = \underset{x \in Z}{\operatorname{argmin}} KL(d_a + x \parallel d_b) \qquad (2)$$

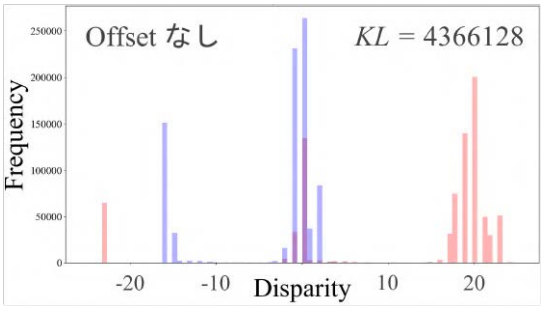
基準が揃っていない隣接視差画像の例を図-3に，また図-3で示す視差画像の重複部分における基準合わせ前後の視差分布を図-4にそれぞれ示す．

以上の処理により隣接視差画像間のオフセット値を求めることで，理論上全視差画像をユニット内の任意の視差画像に基準合わせできる．今回は全視差画像をユニット先頭の視差画像に基準合わせした．

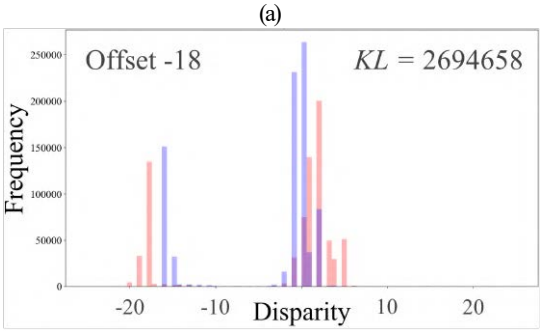


(a) (b)

図-3 基準が揃っていない隣接視差画像の例．
(a)では地面の，(b)では地物の視差が0となっている．



(a)



(b)

図-4 基準合わせ前(a)，後(b)の視差分布例．赤色は図-3の(a)，青色は(b)の視差画像の分布である．オフセット値により赤色の分布が横軸の負の方向に移動している．

⑸ 視差画像の投影変換

使用画像はある一点から撮影した中心投影の画像であり，それに伴い視差画像も中心投影の画像となる．しかし理想の最終視差画像は全ての地点を真上から見た正射投影の画像である．これを踏まえ，視差画像同士の統合前に各視差画像を中心投影から正射投影に変換した．投影変換時に視差値を画像中心に向けてずらす変換幅xは式(3)で求まる．

$$x = \frac{dl}{b' + d} \qquad (3)$$

ただしx: 変換幅 [px]，l: 画像主点からの距離 [px]，d: 視差 [px]，b': 主点間距離 [px]である．

⑹ 視差画像の統合

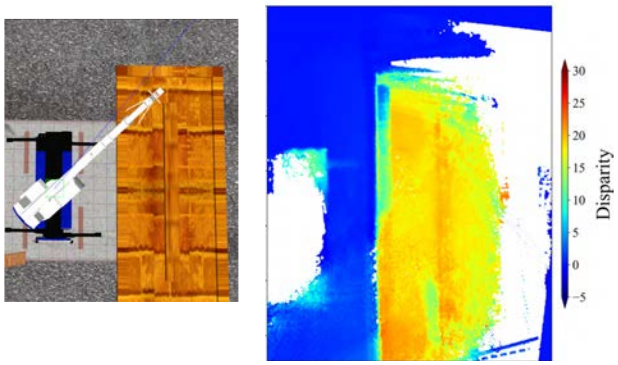
まず各ユニット内の視差画像を統合することで，ユニットごとに1枚の視差画像を算出する．この際，ペア画像の平行化時に算出した射影変換行列を用いて，各視差画像を自身が属するユニットの先頭の視差画像に位置合わせしたうえで重ね合わせた．また，視差画像が複数枚重なる画素では複数の視差の候補値が存在するが，閾値で外れ値を除外したうえでそれらの平均値を計算し，当該画素の視差値とした．

こうして得られたユニットごとの視差画像N枚を先ほどと同様に閾値，平均値を用いて統合し，1枚の最終視差画像を算出した．

上記の統合処理により，視差画像算出時のノイズを効果的に除去することができると考えた．

4. 結果

ここでは1秒あたり5フレーム（全609フレームのうち100フレーム）を使用し，それらを5つのユニットに分割した際の結果を示す．なお処理環境はCore i7，メモリ32GB，Ubuntu 20.04.6，python 3.9.16，OpenCV 4.5.3である．20.1秒の動画像に対しこの最終視差画像の作成に要した時間は94.4秒であった．提案手法にて算出した最終視差画像と，比較用のシミュレータ鳥瞰図を図-5に示す．



(a) (b)

図-5 シミュレータ鳥瞰図(a)と最終視差画像(b)．画像左にクレーン，右に建物を模した地物のモデルが写る．

また視差画像間基準合わせ処理の結果として，基準合わせ前後の隣接する 3 枚の視差画像を**図-6**，**図-7** に示す．

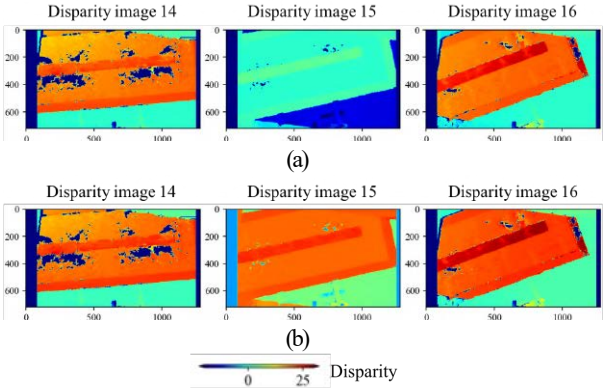


図-6 基準合わせ前(a)，後(b)の 3 枚の隣接視差画像例．ここではユニット 4 の 11~13 番目の視差画像を示す．

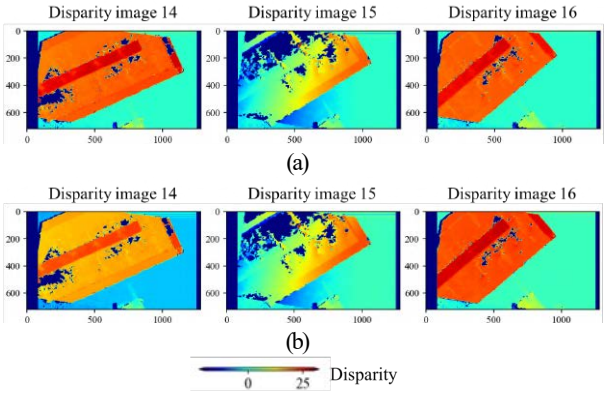


図-7 基準合わせ前(a)，後(b)の 3 枚の隣接視差画像例．ここではユニット 1 の 14~16 番目の視差画像を示す．

5. 考察

最終視差画像については，建物やクレーン基部などの地物の輪郭を捉えられている．また処理時間は 94.4 秒であり，準リアルタイムの処理時間と言えるものの，今後さらなる短縮を目指して検討を重ねる必要がある．

次に視差の基準合わせについての考察を述べる．まず**図-6** について，基準合わせ前において 11，13 番目の視差画像では地面の視差が，12 番目の視差画像では地物の視差がそれぞれ 0 であり，視差の基準が揃っていない．一方基準合わせ後においてはオフセット値の適用により 3 枚の視差画像を通して地面，地物の視差がほとんど同じであり，適切に基準合わせが行われていることが分かる．また**図-7** について，15 番目の視差画像が明らかに非現実的なものとなっている．平行化用行列の計算に使用された対応点が狭い範囲に偏在しており，行列が適切に計算されなかったことが原因だった．基準合わせ処理では，隣接視差画像の重複部分において視差分布の形状が似ていることが前提である．しかし**図-7** の 15 番目の視差画像のように算出に失敗した視差画像ではその前提が成立せず，基準合わせ処理に失敗し得る．実際，基準合わせ前では 14，16 番目の視差画像では視差の基準が揃っているものの，基準合わせ後では 14，16 番目の視差画像で地物，地面の視差がそれぞれ異なっており，処理によって逆に基準がずれる結果となった．今回，全 95 枚の視差画像のうち算出に失敗したのは**図-7** で示す 1 枚のみであり，最終視差画像に大きな影響は与えなかった．しかし実働画像における悪条件下では平行化の失敗が増え，結果に悪影響を与える可能性も十分にあると言える．

6. 結論

本研究では，視差画像作成および統合という手法でクレーン周辺環境の三次元地図作成に取り組んだ．

本研究では 1)SGM の手法によりクレーン周辺の視差画像の準リアルタイム作成が可能であること，2)視差画像間の基準合わせ処理により画像の大部分を占める地物がフレームごとに異なる場合でも手法の適用が可能になること，が判明した．

今後の課題として，何らかの指標を用いて画像ペアごとの平行化精度を定量的に評価し，精度が低いペアの除外処理を追加する必要があると考えられる．

参考文献

1) 王宇，鈴木宏正，大竹豊，小坂孝幸，運天弘樹，野口真児：吊荷カメラ画像によるクレーン作業空間の 3 次元再構成に関する研究，精密工学会大会学術講演会講演論文集，精密工学会，ROMBUNNO.C98，2020.

2) 繁森央一，須﨑純一：ORB-SLAM2 を用いたクレーン周辺の三次元地図のリアルタイム作成，第 63 回土木計画学研究発表会・講演集，2021.

3) MathWorks: SLAM, <https://jp.mathworks.com/discovery/slam.html>，（入手 2023.6.1）.

4) Hirschmuller, H.：Stereo processing by semi-global matching and mutual information, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No.2, pp. 328-342, 2008.

5) 小林知生，須﨑純一，繁森央一，米田瑞生，Marek Ososinski：クレーン周辺の三次元地図作成における頑健性の検討，写真測量とリモートセンシング，Vol. 61，No.6，pp. 391-394，2022.

6) Gazebo: 公式 HP, <https://gazebosim.org/home>，(入手 2023.6.1).

7) Kullback, S. and Leibler, R. A.: On information and sufficiency, *Ann. Math. Statist.*, Vol. 22, No.1, pp. 79-86, 1951.