IIEEJ 論文 -- 年次大会 2022 Journal Track 論文特集 --

撮影者が写らない全方位画像生成のための撮影と合成

河 合 紀 彦†(正会員)　野 田 隆 成†

†大阪工業大学　情報科学部

Photographing and Compositing for Generating Omnidirectional Image Excluding Photographer

Norihiko KAWAI†(*Member*), Ryusei NODA†

†Faculty of Information Science and Technology, Osaka Institute of Technology

〈あらまし〉 全方位画像によるバーチャルリアリティ（VR）空間の作成では，カメラの撮影者が写り込んでいない画像を用いることが望まれる．そこで本論文では，1 枚の画像のみを用いて加工を行うのではなく，全方位カメラで撮影する複数枚の画像を用いることで撮影者が写り込んでいない全方位画像を生成する手法を提案する．提案手法では，まず，全方位カメラを中心として，撮影者が回転しながら全方位画像を複数回撮影する．次に，複数枚の全方位画像で特徴点マッチングを行い，そこから算出した平行移動量を用いて全ての全方位画像の見え方を統一する．最後に，それらの画像を補完し合うとともに色調補正を行うことで撮影者が写っていない全方位画像を生成する．

キーワード：全方位画像, パノラマ画像, 撮影者除去, 画像合成

<**Summary**> When creating a virtual reality space using omnidirectional images, it is desirable to use images in which the photographer does not appear in the image. This paper proposes a method to generate an omnidirectional image excluding the photographer by using multiple images taken by an omnidirectional camera, instead of processing only one image. In the proposed method, the photographer first takes multiple omnidirectional images while rotating around the omnidirectional camera. Next, feature point matching is performed on the multiple omnidirectional images, and the appearances of the images are unified using the amount of translation calculated from the feature point matching. Finally, the images are complemented and color-corrected to produce an omnidirectional image without the photographer.

Keywords: omnidirectional image, panoramic image, removal of photographer, image composition

1. はじめに

近年，全方位カメラの普及により，あらゆる場面で全方位カメラが使用されることが増えた．全方位カメラは年々小型化し，現在では片手に収まるサイズまで小さくなっている．よって手に持ちながら全方位カメラを撮影することで，手軽に全方位画像を撮影することができる．このような全方位カメラで撮影された全方位画像は，Google ストリートビューのような遠隔地を仮想体験できる VR 空間や，インターネットで不動産を内覧できる VR 空間などの構築で使用されている．しかし，手に持ちながら撮影すると，撮影者が大きく全方位画像に写りこんでしまい，背景が一部見えなくなってしまう．このため，上記のような使用目的では，撮影者などの動物体を除去することが好ましい．

この問題に対して，全方位カメラを三脚上に設置し，遠隔から撮影する方法がある．しかし，三脚を立てることができない水上や不安定な場所ではこの方法を用いることは難しい．また，いずれにしても三脚が画像に写り込む問題がある．写り込んだ三脚や撮影者をインペインティング[1)~5)]によって画像から除去する方法もあるが，手動で対象領域を指定する手間や，修復結果が実際の背景と異なるという問題もある．

これに対して，除去対象がいないときに，または異なる視点から実際の背景を撮影して，それを合成することで，不要なオブジェクトを除去し実際の背景を表示する方法が提案されている．例えば，Cohen[6)]は，固定カメラで撮影した動画を入力とし，グラフカットアルゴリズムにより画素ごとに移動物体のないフレームを選択してコピーすることで，背景のみの画像を生成する方法を提案している．ただし，固定カメ

ラを前提としており，フレーム間の位置合わせは必要ないが，対象物体が必ず動くことが前提となる．異なる方法として，オプティカルフローを用いてその背景の位置合わせを行い，他のフレームの画素値をコピーして背景を復元する方法がある[7),8)]．しかし，オプティカルフローを用いているため，フレーム間のカメラの動きが大きくないことが前提となる．

一方，全方位カメラで取得した全方位画像中の物体除去に関する研究もおこなわれている．新井ら[9)]は，地下街のパノラマビューアを生成するために，全方位カメラを三脚上に設置し，同一地点で数十秒の動画を撮影し，人などの動物体のいない全方位画像を生成している．ただし，固定撮影をしており，フレーム間での位置合わせは必要ないが，対象物体が必ず動くことが前提となる．一方，Flores ら[10)]は，異なる位置で撮影された全方位画像から得られる透視投影画像をホモグラフィによって射影変換し，合成する手法を提案している．この手法は，画像間の位置合わせに射影変換のみを用いているため，除去対象領域の背景が平面でない場合には，位置ずれが生じる問題がある．また，著者ら[11),12)]の手法では，動きながら撮影した全方位画像群に対して Structure-from-Motion[13),14)]を適用しカメラ位置姿勢を推定し，文献 11) では平面当てはめにより，文献 12) では Multi-view Stereo[15)] で取得した密な背景形状により，各画像を位置合わせし，動物体を除去しその背景を復元している．これらの従来手法は異なる位置で撮影された画像群を入力として用いることを前提としている．

本研究では，VR での使用を想定した撮影者の写らない全方位画像をできるだけ簡便に取得することを目指した，撮影および合成方法を提案する．提案手法では従来と異なり，一地点で撮影を行い，三脚などの機器を使用せず，全方位カメラを持った撮影者が全方位カメラを中心に回転しながら 3 枚以上撮影する．撮影した画像を特徴点マッチングにより位置合わせし，グラフカットおよびポアソン画像合成により画像選択および色調補正を伴う合成を行うことで，撮影者が写り込まない全方位画像を生成する．

提案手法の特徴として，一般的な透視投影画像とは異なる，画像の左端と右端が隣接関係にある正距円筒図法で展開された全方位画像を扱う．このため，画像の位置合わせ，グラフカット，ポアソン画像合成においても，その隣接関係を考慮した手法の開発および実装を行った．

2. 提 案 手 法

2.1 概要

提案手法ではまず，(1) 全方位カメラを中心として，撮影者がその周りを回転しながら撮影した複数枚のパノラマ画像を入力する．次に，(2) 各入力画像に対して特徴点検出[16)]を行う．検出された特徴点を中心とするパッチを設定し，2 枚の画像内のパッチ間の画素値の差の二乗和 SSD（Sum of Squared Differences）を計算することで最も類似したパッチを対応付け，RANSAC[17)]と最小二乗法によってパノラマ画像間の水平方向の平行移動量を算出する．次に，平行移動量を用いて各入力画像の見え方を統一する．次に，(3) 各入力画像の各画素で画素値の最頻値を算出し，最頻値を画素値とする最頻値画像を生成する．最後に，(4) グラフカット[18)]によって画素ごとの合成元画像を選択し，ポアソン画像合成[19)]により色調を補正し，撮影者の写らない全方位画像を出力する．以下では各手順について詳述する．

図 1 撮影方法
Fig. 1 Photographing method

図 2 取得画像例
Fig. 2 Examples of photographed images

2.2 撮影

全方位カメラを用いた全方位画像の撮影では，**図 1** に示すように全方位カメラを撮影者が横方向に突き出して持ち，カメラ位置をできるだけ動かさないように，撮影者がカメラを中心に回りながら 3 枚以上静止画像を撮影する．なお本研究では，RICOH THETA シリーズ[20)]の「天頂補正」や，Insta360 ONE X2[21)]の「自動水平補正」といった，カメラの重力方向が画像の下方向となるよう正距円筒図法でパノラマ展開する機能により，各全方位画像間で同一物体はおおよそ同じ高さにあるという前提を用い，**図 2** に示すような画像群が取得できるものとする．

2.3 平行移動画像の生成

1 枚の基準画像とそれ以外の画像間で特徴点マッチングを行い，全ての画像を基準画像の見え方に統一する．

2.3.1 平行移動量の算出

各入力画像をグレースケールに変換し，特徴点検出を行う．ここでは，Shi らの Good Features to Track[16)]を用いて画像間で対応付けを行うための特徴点を検出する．

次に各入力画像間の特徴点を対応付ける．今回想定する撮影方法および全方位画像の展開方法では，各入力画像の同一箇所でスケールや画素値の違いがほとんど生じないため，画素値の差の二乗和 SSD を類似度の指標として用いる．具体的には基準とする 1 枚の入力画像内の特徴点を中心とするパッチとそれ以外の各入力画像内の特徴点を中心とするパッチ間

の SSD を計算し，特徴点ごとに SSD が最も小さくなる特徴点を求め，対応付ける．なお本研究では，上述のとおり，各全方位画像は，カメラの重力方向を用いてパノラマ展開する機能により，同一物体はおおよそ同じ高さにあるという前提に基づき，特徴点間の Y 座標値の差がある一定範囲内の場合のみ SSD を計算することで探索範囲を限定する．これにより，計算時間を削減するとともに，誤対応の発生を抑制する．

次に，RANSAC[17]により誤対応を排除する．本研究での具体的な RANSAC の適用として，まず，(a)SSD で対応付けした対応点の組みの中からある 1 組をランダムに抽出し，その対応点間の水平方向の移動量（画素）を求める．なお，本研究では，画像の左端と右端がつながっている全方位画像を対象としているため，全ての移動量はプラス方向のみで計算する．次に，(b)SSD で対応付けした他の全ての対応点の組についても，同様に対応点間の水平方向の移動量を求め，(a) で求めた移動量との差の絶対値がある範囲内である対応点の組の数をカウントする．(c) これを一定回数繰り返し，最も多くの数をカウントした 1 組の対応点を決定する．(d) 最も多くの数をカウントした対応点と比較して，移動量の差がある範囲内に入っていない対応点の組をアウトライアとして除外する．

最後に，RANSAC で最も多くの数をカウントした対応点を含む，アウトライアを除外したインライアの対応点の移動量から最小二乗法によって最適な移動量を算出する．具体的には，インライアの各対応点の移動量を l_j（$j = 1, \cdots, J$，ただし，J はインライアの対応点の組数）とすると，以下のコスト C を最小化する移動量 l を求める．

$$C = \Sigma_{j=1}^{J}(l - l_j)^2 \tag{1}$$

ここでは，以下の式を満たす l を求めればよいため，全てのインライアの移動量の平均値が求めたい平行移動量となる．

$$\frac{\partial C}{\partial l} = 0 \tag{2}$$

2.3.2 見え方の統一

基準とした一枚の入力画像以外の入力画像を，上記で算出した平行移動量だけ水平方向に平行移動する．この処理により，全ての入力画像の背景の同一物体の X 座標値が等しくなり，見え方が統一される．図 2 左の画像を同図右の画像の見え方になるよう平行移動した例を**図 3** に示す．撮影者の位置のみが異なり，それ以外の物体や背景はおおよそ同じ X 座標を持っていることが確認できる．

2.4 最頻値画像の生成

後述するグラフカットに用いるため，全ての画素において全方位画像間の最頻値を算出する．撮影者が移動しながら複数枚を撮影しているため，上記の処理により見え方が統一された全ての全方位画像の同一画素を比較すると，撮影者よりその背景が写っている画像の枚数が多い可能性が高い．このため，画素ごとに最頻値を求めることで，撮影者が取り除かれた全方位画像を一時的に求める．

図 3 平行移動後の画像例
Fig. 3 Example of translated image

ここでは，Cheng らの Mean shift 法[22]により画素ごとの最頻値 $\mathbf{M}$ を算出する．具体的には，i 回目の反復処理により得られる最頻値 $\mathbf{M}_{i+1}$ は，N 枚の見え方が統一された入力画像の画素値の重み付き平均により以下のように計算される．

$$\mathbf{M}_{i+1} = \frac{\sum_{n=1}^{N} K(\mathbf{I}_n, \mathbf{M}_i)\mathbf{I}_n}{\sum_{n=1}^{N} K(\mathbf{I}_n, \mathbf{M}_i)} \tag{3}$$

ただし，K はガウスカーネルを表し，$i-1$ 回目の反復処理で得られた最頻値 $\mathbf{M}_i$，見え方が統一された n 枚目の入力画像の画素値 $\mathbf{I}_n$，定数 σ を用いて以下のように定義する．

$$K(\mathbf{I}_n, \mathbf{M}_i) = \exp\left(\frac{-||\mathbf{I}_n - \mathbf{M}_i||^2}{\sigma^2}\right) \tag{4}$$

ただし，$\mathbf{M}_1$ は N 枚の画像の単純平均とする．背景の同一箇所であっても複数の画像間で色に多少の違いが生じることが一般的であるが，反復処理により $\mathbf{M}$ が RGB 空間上で画素値が集まって分布する箇所に徐々に移動することから，$\mathbf{M}$ が最頻値であるとみなすことができる．

また，提案手法では，平行移動で位置合わせをしているものの，画像間で多少の位置ずれが生じている可能性もある．このため，最頻値画像に対して空間的な平滑化処理を行った画像を，次のグラフカットで用いる最頻値画像とする．

2.5 画像の選択と合成

2.5.1 画素ごとの画像選択

グラフカットを用いたエネルギー最小化[18]により，見え方が統一された入力画像の中から，各画素に対して適切な画像を選択し合成することで，撮影者を含まない全方位画像を得る．各画素の画像を選択するためのエネルギー関数を以下のように定義する．

$$E = \lambda \sum_{u \in A} E_1(f_u) + \frac{\kappa}{2} \sum_{(u,v) \in P} E_2(f_u, f_v), \tag{5}$$

ただし，f_u と f_v は，それぞれ画素 u と v の画像番号を表す．また，A は全方位画像における全画素の集合，P は全方位画像における隣り合う画素のペアの集合である．なお，全方位画像を対象としているため，最も左と最も右の画素も隣接画素とする．λ, k はデータ項 E_1 と平滑化項 E_2 のバランスをとるための重みである．

データ項 E_1 は以下のように定義する．

$$E_1(f_u) = ||\mathbf{I}_{f_u}(u) - \mathbf{M}(u)||, \tag{6}$$

ただし，$\mathbf{I}_{f_u}(u)$ は，画素 u における画素番号 f_u の RGB の画素値を表すベクトルであり，$\mathbf{M}(u)$ は見え方を統一した全入力画像群から生成した最頻値画像の画素 u の RGB の画素値である．E_1 が小さくなるような画像番号 f_u を選択することで，最頻値に近い画像を得ることができ，結果的に撮影者を画像から排除する．

平滑化項 E_2 は以下のように定義する．

$$E_2(f_u, f_v) = ||\mathbf{I}_{f_u}(u) - \mathbf{I}_{f_v}(u)|| + ||\mathbf{I}_{f_u}(v) - \mathbf{I}_{f_v}(v)|| \tag{7}$$

この項では，隣接する画素間で画像番号が頻繁に変化することを防ぎ，また画像番号が切り替わる箇所においても，切り替わる画像間で画素値の差が小さいところで切り替えることを促す．これにより，画像番号が切り替わる境界線を目立たなくする．

全体のエネルギー関数 E を最小化するにあたって，画像枚数が 2 枚より多いため，α-β 交換アルゴリズムを用いる．具体的には，2 枚の画像を抽出してグラフカットにより，その 2 枚の画像番号を入れ替える．これを全てのペアに対して行い，交換された画像番号の枚数がなくなるか，または一定回数に達するまで反復する．

2.5.2 ポアソン画像合成

ポアソン画像合成[19]とは，ポアソン方程式を解くことで，複数の画像の合成時に合成元画像の色調を合成先画像の色調に合わせる手法である．本研究では，ある 1 つの画像番号から得られた画素を合成先画像，それ以外の画像番号から得られた画素を合成元画像とする．ここでは，以下の定義される画像の勾配に関するエネルギー関数 L を最小化することで，合成元画像の合成後の画素値を決定する．

$$L = \sum_{(u,v)\in P} ((h_u - h_v) - (g_u - g_v))^2 \tag{8}$$

g_u はグラフカットによって選択された各画素に対応する画像を単純に貼り合わせた画像の，画素 u における RGB のいずれかの画素値である．h_u はポアソン画像合成後の求めたい画素値である．u，v は隣接する画素であり，全方位画像では画像の左端と右端の画素も隣接画素とする．エネルギー L が最小になる画素値 h を求めることで，画像番号が切り替わる境界において色の違いを目立たなくする．なお，エネルギー L の最小化処理を RGB 独立に処理を行う．

3. 実験と考察

提案手法の有効性を示すために，3 つの異なるシーンで実験を行った．実験では，具体的な内部の情報は公開されていないが，重力方向が画像の下方向となる全方位画像を出力する全方位カメラ RICOH THETA Z1 を用いて撮影を行い，全方位画像の解像度は 1024 × 512 ピクセルにリサイズして入力として使用した．SSD 算出におけるパッチのサイズは 31 × 31，最頻値算出における σ は 30，反復回数は 50 回，グラフカットにおける λ は 100 に固定し，α-β 交換アルゴリズムの最大反復回数を 8 回とした．また，**表 1** に示すスペックを持つ PC を用い，ポアソン画像合成での行列計算に Eigen ライブラリを用いた．以下では，まずシーンごとに 3 枚の入力画像を用いた場合の実験結果を示し，次に枚数を増やしたときの結果を示す．次に，各シーンや枚数での計算時間を示し，最後に本研究の限界と考察について述べる．

表 1 実験に用いた PC のスペック
Table 1 Specifications of PC used for experiments

OS	Windows 11
CPU	AMD Ryzen 7 3700X
メモリ	32GB
GPU	GeForce RTX 3080

3.1 3 枚の画像による実験

本節では，シーン A，B，C の 3 つのシーンの実験について順に示し，最後に考察を述べる．なお，各入力画像の図において，左上の入力画像の番号を 1，右上の入力画像の番号を 2，下の入力画像の番号を 3 とする．また，グラフカットの結果の選択画像番号の色について，青が 1，緑が 2，赤が 3 を示す．

3.1.1 シーン A：室内シーン

図 4 に室内で撮影した 3 枚の入力画像を示す．これらの画像では，撮影者の位置は同じだが背景が水平方向にずれていることがわかる．図 4 左上の画像を基準にして他の 2 枚を平行移動した画像を**図 5** に示す．図より，撮影者以外の背景の同物体はほぼ同じ X 座標になっていることがわかる．次に，図 4 左上の画像と図 5 の 2 枚の画像を合わせた計 3 枚の入力画像の最頻値を算出した画像およびその一部の拡大図を**図 6** に示す．最頻値画像では撮影者が排除されていることが確認できるが，全体的に縦方向にブレている画像が生成されている．元の入力画像は，加速度センサで重力方向を検出して生成されていると考えられるが，カメラを持つ手の動きによって方向に誤差が生じ，垂直方向の座標がずれてしまったと考える．そのため，単純な最頻値で画像生成すると，画像全体が縦方向にブレてしまう．

次に，グラフカットにおいて 3 種類のパラメータ κ で画像を生成した．各 κ の値で生成した結果の画像と，グラフカットにより生成した画像の各画素値に対応する入力画像番号を表した画像を**図 7** に示す．

$\kappa = 10$ の場合，生成画像は最頻値画像に近くなり，最頻値画像で人の肌色がうっすら残っているところについては，撮影者の手が合成されている．また，画像番号が頻繁に変更されており，蛍光灯やディスプレイでテクスチャのずれが生じていることが確認できる．$\kappa = 100$ の場合，画像番号の入れ

図 4 シーン A の入力画像
Fig. 4 Input images in Scene A

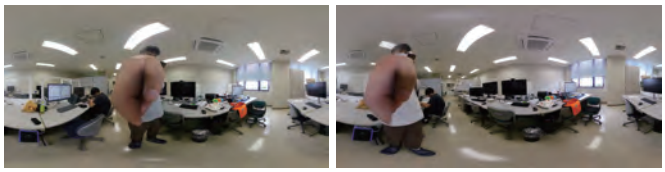


図 5 シーン A の平行移動後の画像
Fig. 5 Translated images in Scene A

図 6 シーン A の最頻値画像
Fig. 6 Mode image in Scene A

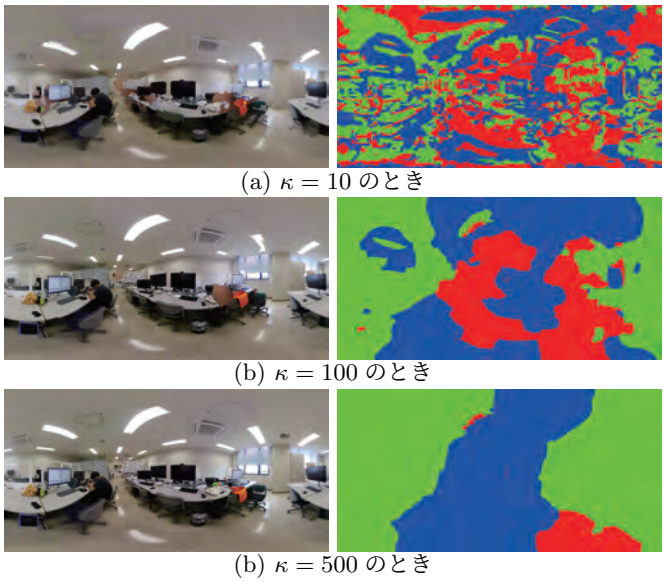
替わりが少なくなり，同じ画像番号が大きな領域に集まっている．その結果，テクスチャのずれが減少している．しかし，撮影者の手の肌色が残っている．$\kappa = 500$ の場合，どの領域でもほとんど違和感のない，撮影者の写っていない全方位画像が生成されている．しかし，カメラのホワイトバランスの自動調整などにより各入力画像間での色合いが多少異なっているため，画像の切り替わる境界が視認できる．

次に，$\kappa = 500$ の場合でポアソン画像合成を行った出力画像を**図 8** に示す．ポアソン画像合成を行った結果，画像番号が切り替わる境界で発生していたエッジが消え，色調が統一されていることがわかる．

3.1.2 シーン B：屋外シーン

図 9 に室外で撮影した入力画像を示す．このシーンでは高い建物と木が写っている．平行移動した 3 つの入力画像の最頻値画像およびその一部の拡大図を**図 10** に示す．このシーンでも最頻値画像が縦方向にブレていることがわかる．また，画像内の右の地面に撮影者の影が大きく残っていることも確認できる．

次に，$\kappa = 1000$ でグラフカットを行なって生成した画像およびその画像番号を**図 11** に示す．図より撮影者の写らない全方位画像を生成できていることが確認できる．しかし，生成画像の右側に撮影者の影が不自然に残っている．3 枚の入力画像でこの領域を見ると，1 枚目では撮影者の領域，2 枚目ではこの影の領域，3 枚目では撮影者の影ではない実際の背景となっている．この三者三様のため，最頻値画像においては影らしき色が残り，その結果，グラフカットで実際の背景が選択されなかった．また，このシーンでも，空や地面の領域の画像の切り替わる箇所でわずかではあるが色調の異なるエッジが発生している．



(a) $\kappa = 10$ のとき
(b) $\kappa = 100$ のとき
(b) $\kappa = 500$ のとき

図 7 シーン A のグラフカットによる合成結果画像と選択された画像
Fig. 7 Composite image and image number selected by graph cut in Scene A

図 8 シーン A のポアソン画像合成後の最終結果画像
Fig. 8 Final result image after poisson image editing in Scene A

図 9 シーン B の入力画像
Fig. 9 Input images in Scene B

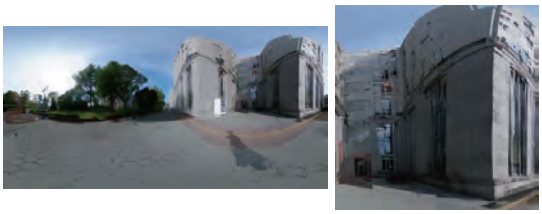


図 10 シーン B の最頻値画像
Fig. 10 Mode image in Scene B

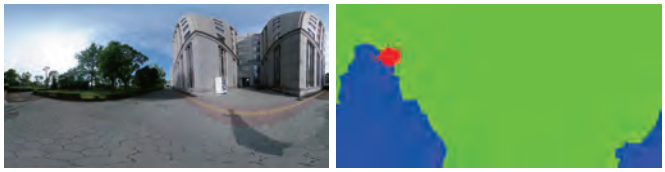


図 11 シーン B のグラフカットによる合成結果画像と選択された画像（$\kappa = 1000$）
Fig. 11 Composite image and image number selected by graph cut in Scene B ($\kappa = 1000$)

図 12 シーン B のポアソン画像合成後の最終結果画像
Fig. 12 Final result image after poisson image editing in Scene B

図 13 シーン C の入力画像
Fig. 13 Input images in Scene C

次にポアソン画像合成後の出力画像を**図 12** に示す．ポアソン画像合成後は空や地面の画像番号が切り替わる境界で発生していたエッジが消え，色調が統一されていることがわかる．

3.1.3 シーン C：狭い室内シーン

図 13 に狭い室内（エレベータ）で撮影した入力画像を示す．エレベーターの中には鏡が設置されている．平行移動した 3 つの入力画像の最頻値画像およびその一部の拡大図を**図 14** に示す．このシーンでも最頻値画像が縦方向にブレており，一部の領域で不自然な色調が現れている．

次に，$\kappa = 1000$ によるグラフカットで生成した画像およびその画像番号を**図 15** に示す．このシーンでも撮影者の写らない全方位画像を生成することができた．しかし，画像番号

図 14 シーン C の最頻値画像
Fig. 14 Mode image in Scene C

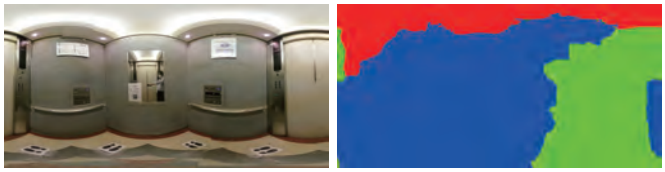


図 15 シーン C のグラフカットによる合成結果画像と選択された画像（$\kappa = 1000$）
Fig. 15 Composite image and image number selected by graph cut in Scene C ($\kappa = 1000$)

図 16 シーン C のポアソン画像合成後の最終結果画像
Fig. 16 Final result image after poisson image editing in Scene C

が切り替わる境界部分で縦方向にエッジのずれが随所に見られる．シーン A・B と異なり，カメラと被写体の距離が近くなる狭い空間では，重力方向のわずかな誤差が画像の垂直方向の大きなずれにつながる．この結果，エネルギーが最小となる最適な場所で切り替えたとしてもエッジのずれが生じたと考える．また，鏡の部分には撮影者写り込んでおり，シーン B と同様に，1 枚目の画像では撮影者が鏡に映っている領域，2 枚目の画像では撮影者の領域，3 枚目の画像では実際の背景の画像となっているため，実際の背景が適切に選択されていない．

次にポアソン画像合成を行った出力画像を**図 16** に示す．図から，このシーンでも画像番号が切り替わる境界で発生していた色調のずれ調整されている．エッジのずれもぼかされ目立たなくはなっているが，それでも大きなずれは解消されない．

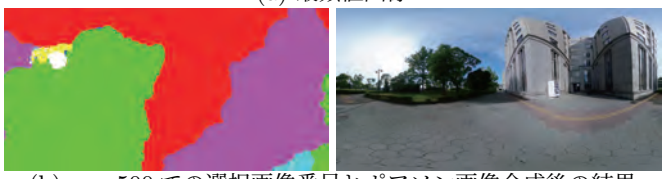
3.1.4 入力を 3 枚とした場合の考察

入力画像を 3 枚とした場合には，例え撮影者が適切に移動しその背景が観測できたとしても，強い光源による影や鏡のように撮影者が撮影者以外の領域のテクスチャに影響を及ぼすことがある．このとき，位置合わせをした後の同一画素が三者三様となり，グラフカットで適切な背景画像が選択されないという問題があることを確認した．このため，シーンによっては 4 枚以上の枚数が必要であると考える．

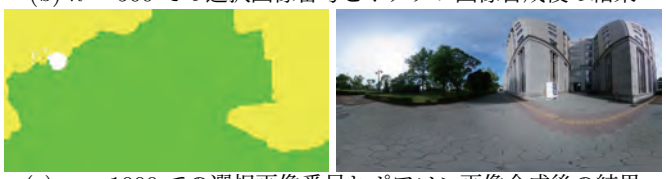
図 17 シーン B の追加の入力画像
Fig. 17 Additional input images in Scene B

図 19 シーン C の追加の入力画像
Fig. 19 Additional input images in Scene C

(a) 最頻値画像



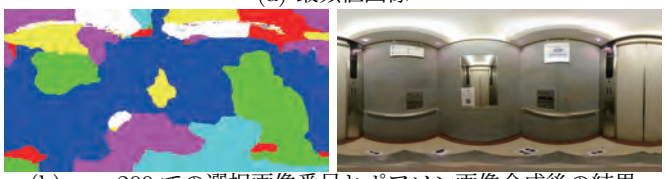
(b) $\kappa = 500$ での選択画像番号とポアソン画像合成後の結果



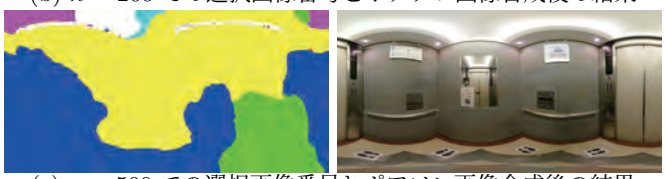
(c) $\kappa = 1000$ での選択画像番号とポアソン画像合成後の結果

図 18 シーン B の処理結果
Fig. 18 Processing results in Scene B

(a) 最頻値画像



(b) $\kappa = 200$ での選択画像番号とポアソン画像合成後の結果



(c) $\kappa = 500$ での選択画像番号とポアソン画像合成後の結果

図 20 シーン C の処理結果
Fig. 20 Processing results in Scene C

3.2 枚数による結果の変化

前節の結果を受け，シーン B・C について 7 枚の入力画像を用いて実験を行った．以下各シーンでの実験結果を示す．なお，各入力画像の番号については，図 9，13 の左上，右上，下の画像がそれぞれ 1，3，5 であり，新たに追加する入力画像の図の左上，右上，左下，右下の画像がそれぞれ 2，4，6，7 となる．また，グラフカットの結果の選択画像番号の色について，青が 1，緑が 2，赤が 3，シアンが 4，マゼンダが 5，黄が 6，白が 7 を示す．

3.2.1 7 枚の入力画像を用いたシーン B

実験では，図 9 の 3 枚に加えて，**図 17** に示す 4 枚を入力画像として使用した．計 7 枚の画像の位置合わせを行った後の，最頻値画像，$\kappa = 500$ および $\kappa = 1000$ でのグラフカットで得られた画像番号，ポアソン画像合成後の結果画像を**図 18** に示す．

図 18(a) より，最頻値画像においては，図 10 に示す入力が 3 枚のときの最頻値画像と比較して，影の領域がかなり小さくなっていることが確認できる．その結果，同図 (b)(c) に示す $\kappa = 500$，$\kappa = 1000$ のときのいずれの結果画像においても，図 12 に示す入力が 3 枚のときの結果では見られた影領域がなくなっている．黄色の誘導ブロックなど画像の切り替え位置において多少のずれは見られるが，画像全体として違和感が少なく撮影者のいない全方位画像が生成できている．

3.2.2 7 枚の入力画像を用いたシーン C

シーン C の実験では，図 13 の 3 枚に加えて，**図 19** の 4 枚を入力画像として用いた．計 7 枚の画像の位置合わせを行った後の，最頻値画像，$\kappa = 200$，$\kappa = 500$ でのグラフカットで得られた画像番号，ポアソン画像合成後の結果画像を**図 20** に示す．

図 20(a) に示す最頻値画像においては，図 14 に示す入力が 3 枚のときの最頻値画像と比較すると，鏡に映った撮影者の領域が多少小さくなっていることが確認できる．この最頻値画像を用いて，$\kappa = 200$，$\kappa = 500$ のパラメータでグラフカットを行った．同図 (c) に示す，$\kappa = 500$ の場合のポアソン画像合成による結果画像では，撮影者が鏡の中に残っている．一方，同図 (b) の $\kappa = 200$ の場合には，撮影者が鏡に映っていない．κ が小さい方が最頻値画像と似た画像が得られるため，鏡の中に撮影者が映っていない画像が選択された

表 2 各処理の計算時間（秒）
Table 2 Computation time for each process (sec.)

シーン-枚数-κ	見え方の統一	最頻値画像生成	グラフカット	ポアソン画像合成
A-3-10	0.65	0.85	4.38	44.97
A-3-100	0.63	0.83	18.91	47.93
A-3-500	0.66	0.87	81.55	43.91
B-3-500	0.68	0.83	64.56	62.45
B-3-1000	0.72	0.83	31.18	70.22
C-3-500	0.31	0.83	34.90	14.77
C-3-1000	0.31	0.83	57.53	19.02
B-7-500	2.06	1.79	179.25	116.03
B-7-1000	2.20	1.81	123.08	114.11
C-7-200	1.13	1.78	71.63	36.24
C-7-500	1.01	1.79	108.20	41.88

ためだと考える．一方，κ が大きい方の結果に比べて，画像の切り替わりが多いため，天井の枠のエッジのずれや床の足型のマークが正しく再現されていないといったテクスチャの違和感が生じている．

3.2.3 入力を 7 枚とした場合の考察

入力画像を 7 枚にした場合には，3 枚の場合と比較して，撮影者が影響を及ぼす影や鏡の写り込みに対しても，比較的適切な最頻値画像が生成され，撮影者が写らずまた影や映り込みもない全方位画像が生成できることを確認した．ただし，特に撮影対象が近くにある場合には，枚数を増やしたとしても，水平方向のみによる位置合わせでは，合成後にエッジのずれが生じる場合が多いことも確認した．

3.3 計算時間

これまでに示した実験結果を含む，異なるシーン・枚数・κ における処理時間を**表 2** に示す．基本的には入力画像の枚数が多い方が計算時間は増えるが，パラメータ κ の値によりグラフカットの時間が大きく変わることがわかる．グラフカットによりエネルギーが最小となる画像番号を得るにあたって，繰り返し処理で番号を交換するが，λ と κ のバランスによってその回数が大きく変化するためであると考える．また，ポアソン画像合成において，本実験では入力の 1 枚目の画像を合成先画像として画素値を固定し，それ以外の画像を合成元画像としてその画素値がパラメータとなっている．このため，その画素の割合によって計算時間が大きく変化したと考える．計算時間をできるだけ少なくするためには，グラフカットで最も多くの画素で選択された画像を合成先画像とすることが良いと考える．

3.4 本研究の限界と考察

提案手法では，各全方位画像間で同一物体はおおよそ同じ高さにあるという前提に基づき，水平方向の平行移動による見え方の統一を行ったが，実験結果からわかるように，パノラマ展開時の重力方向の検出の誤差による垂直方向の回転や，撮影時のカメラ位置のずれにより，特に撮影対象が近い領域では，合成後にずれが生じることを確認した．このため，結果を改善するためには，マッチングによる垂直方向の回転の補正や，自由視点画像生成で撮影位置の違いにより生じる位置ずれを補正する必要がある．

撮影者自身はいずれも消去できたが，鏡の中の撮影者の写り込みについては，除去できない場合が多かった．このため，除去する対象が人と限定できる場合には，グラフカットのコストに人物検出結果を導入することで，鏡の映り込みにも対応できるものと考える．

実験では全方位カメラから取得できるオリジナルな解像度より低い解像度にリサイズした画像を用いており，もし高解像度な画像に適用した場合には，処理時間の大幅な増加が予想される．このため，実用化を目指す場合には，グラフカットやポアソン画像合成の効率的な手法を考案する必要がある．

4. むすび

本論文では，全方位カメラで撮影する複数枚の画像を用いることで撮影者が写り込んでいない全方位画像を生成する手法を提案した．提案手法では，全方位カメラを中心に撮影者が回転しながらで複数枚撮影し，それらを特徴点マッチングにより算出した平行移動量を用いて全ての全方位画像の見え方を統一する．最後に，グラフカットおよびポアソン画像合成により画像を補完し合うことで撮影者が写っていない全方位画像の生成を実現した．

実験では，3 つの異なる特徴をもつシーンを用い，入力画像の枚数やパラメータを様々に変更して結果の出力および考察を行った．また，シーンやパラメータの違いによる計算時間の変化についても考察した．

今後の課題として，見え方の補正が合成結果に大きく影響することから，画像を統合する前に水平方向だけではなく垂直方向のズレを補正する必要がある．また，人物をより確実に取り除くためには，人物検出結果などをグラフカットのコストに導入することも考えられる．また，高解像度な画像にも適用するにあたって，計算時間を削減できるグラフカットやポアソン画像合成の効率的な手法を開発する必要がある．

謝辞 本研究の一部は，JSPS 科研費 JP18H03273, JP18H04116, JP21H03483 の助成による．

参考文献

1) A. Telea: “An Image Inpainting Technique Based on the Fast Marching Method”, Journal of Graphics Tools, Vol.9, No.1, pp.23–24 (2004).
2) A. Criminisi, P. Perez, K. Toyama: “Region Filling and Object Removal by Exemplar Based Image Inpainting”, IEEE Transactions on Image Processing, Vol.13, No.9, pp.1200–1212 (2004).
3) 河合紀彦, 佐藤智和, 横矢直和: “テクスチャの明度変化と局所性を考慮したパターン類似度を用いたエネルギー最小化による画像修復”, 電子情報通信学会論文誌 (D), Vol.J91-D, No.9, pp.2293–2304 (2008).
4) N. Kawai, N. Yokoya: “Image Inpainting Considering Symmetric Patterns”, Proc. of IAPR International Conference on Pattern Recognition, pp.2744–2747 (2012).

5) J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.-S. Huang, "Generative Image Inpainting with Contextual Attention", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.5505–5514 (2018).
6) S. Cohen: "Background Estimation as a Labeling Problem", Proc. of IEEE International Conference on Computer Vision, pp.1034–1041 (2005).
7) R. Xu, X. Li, B. Zhou, C.C. Loy: "Deep Flow-guided Video Inpainting", Proc. of IEEE Conference on Computer Vision and Pattern Recognition (2019).
8) T.T. Le, A. Almansa, Y. Gousseau, S. Masnou: "Object Removal from Complex Videos Using a Few Annotations," Computational Visual Media, Vol.22, No.5, pp.267–291 (2019).
9) 新井イスマイル, 堀磨伊也, 河合紀彦, 安部陽平, 市川昌宏, 里中裕輔, 新田竜規, 新田知之, 藤井陽光, 向井政貴, 堀見宗一郎, 牧田孝嗣, 神原誠之, 西尾信彦, 横矢直和: "Gooraffiti Umechika：人が消える地下街パノラマビューア", 情報処理学会論文誌, Vol.53, No.5, pp.1546–1557 (2012).
10) A. Flores, S. Belongie: "Removing Pedestrians from Google Street View Images", Proc. of International Workshop on Mobile Vision, pp.53–58 (2010).
11) N. Kawai, K. Machikita, T. Sato, N. Yokoya: "Video Completion for Generating Omnidirectional Video without Invisible Areas", IPSJ Transactions on Computer Vision and Applications, Vol.2, pp.200–213 (2010).
12) N. Kawai, N. Inoue, T. Sato, F. Okura, Y. Nakashima, N. Yokoya: "Background Estimation for a Single Omnidirectional Image Sequence Captured with a Moving Camera", IPSJ Transactions on Computer Vision and Applications, Vol.6, pp.68–72 (2014).
13) 佐藤智和, 池田聖, 横矢直和: "複数動画像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).
14) Visualsfm: A Visual Structure from Motion System, `http://ccwu.me/vsfm/` (2022).
15) M. Jancosek, T. Pajdla: "Multi-view Reconstruction Preserving Weakly-supported Surfaces", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.3121–3128 (2011).
16) J. Shi, C. Tomasi: "Good Features to Track", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.593–600 (1994).
17) M.A. Fischler, R.C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, Vol.24, No.6, pp.381–395 (1981).
18) Y. Boykov, V. Kolmogorov: "An Experimental Comparison of Min-cut/max-flow Algorithms for Energy Minimizaiton in Vision", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.26, No.9, pp.1124–1137 (2004).
19) P. Pérez, M. Gangnet, A. Blake: "Poisson Image Editing", ACM Transactions on Graphics, Vol.22, No.3, pp.313–318 (2003).
20) 360 度カメラ RICOH THETA, `https://theta360.com/ja/` (2022).
21) Insta360 ONE X2 - 全方位を思いのままに, `https://www.insta360.com/jp/product/insta360-onex2` (2022).
22) Y. Cheng: "Mean Shift, Mode Seeking, and Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17, No.8, pp.790–799 (1995).

(2022 年 6 月 8 日 受付)
(2022 年 7 月 11 日 再受付)

河 合 紀 彦 (正会員)
2010 年 奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．同年，日本学術振興会特別研究員（PD）兼，カリフォルニア大学バークレー校博士研究員．2011 年奈良先端科学技術大学院大学情報科学研究科助教．センスタイムジャパンを経て，2020 年大阪工業大学情報科学部情報メディア学科准教授，現在に至る．博士（工学）．画像処理，複合現実感，バーチャルリアリティに関する研究に従事．

野 田 隆 成
2022 年 大阪工業大学情報科学部情報メディア学科卒業．現在，三菱電機インフォメーションネットワーク株式会社に勤務．