

撮影者を含まない全方位画像の生成

野田 隆成† 河合 紀彦†

大阪工業大学 情報科学部 情報メディア学科†

1. はじめに

近年、小型化された全方位カメラの普及により、手に持ちながら手軽に全方位パノラマ画像を撮影する機会が増えている。このような全方位パノラマ画像は、ストリートビューや、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、撮影者が大きくパノラマ画像に写りこみ、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。

この問題の解決策として、全方位カメラを三脚に立てて、遠隔から撮影する方法がある。しかし、この方法では、水上や、傾斜のある場所では三脚を立てることができない。また、いずれにしても三脚が画像に写り込むという問題がある。写り込んだ三脚や撮影者をインペインティング[1]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。

そこで本研究では、三脚などの機材を使用せず、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複数枚撮影し、それらを統合することで、撮影者が写り込まない全方位パノラマ画像を生成する手法を提案する。

2. 提案手法

提案手法ではまず、カメラを中心として、撮影者が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。ただし本研究では、各全方位パノラマ画像は、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を用いる。次に、各入力画像に対して Harris コーナー検出[2]によって検出された特徴点を用いた対応付けによってパノラマ画像間の水平方向の平行移動を算出し、それを用いて各入力画像の見え方を統一する。次に、各入力画像の各画素で画素値の中央値を算出する。最後に、グラフカットによって画像を統合することで、撮影者の写

らない全方位パノラマ画像の出力する。以下では、平行移動量の算出による見え方の統一および、グラフカットによる画像の選択について詳述する。

2.1 平行移動の算出

各入力画像をグレースケールに変換し、Harris コーナー検出により特徴点抽出を行う。次に基準とする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と対応付け、特徴点間の水平方向の画素の距離を算出する。本研究では、上述の通り、各全方位パノラマ画像では、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、ある特徴点に対して SSD を計算する特徴点は Y 座標が一定範囲内のみとして探索範囲を限定することで、計算時間を削減し、また誤対応の発生を抑制する。また、対象画像が全方位画像であり、画像の左端と右端がつながっていることを考慮し、プラス方向のみの特徴点間の距離を計算する。

次に、RANSAC[3]により誤対応を排除し抽出した全てのサンプルの移動量の平均値を最小二乗法により求め、画像間の平行移動量を算出する。次に基準となる画像以外の入力画像は、上記で算出した移動量分だけ X 軸方向に移動させることで、見え方が統一される。

2.2 画像選択による画像合成

グラフカット [4]を用いたエネルギー最小化により、背景が統一された各画像の中から、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は以下のように定義する。

$$E = \lambda \sum_{u \in A} E_1(f_u) + \frac{k}{2} \sum_{(u,v) \in N} E_2(f_u, f_v) \quad (1)$$

データ項 E_1 は次のように定義される。

$$E(f_u) = \|I_{f_u}(u) - M(u)\| \quad (2)$$

$I_{f_u}(u)$ は、画素 u における画素番号 f_u の画素値であり、 $M(u)$ は全入力画像の中央値画像である。中央値に近い画像を得ることで撮影者を排除する。また、平滑化項 E_2 は次のように定義する。

$$E_2(f_u, f_v) = \|I_{f_u}(u) - I_{f_v}(u)\| + \|I_{f_u}(v) - I_{f_v}(v)\| \quad (3)$$

†Generation of omnidirectional image without photographer
†Ryusei Noda, Norihiko Kawai, Osaka Institute of Technology

u, v は隣接する画素であり、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、隣接する画素値の差が少ないところで画像番号を切り替えることを促す。これにより、画像番号が異なる境界線は隣接する画素間を目立たなくする。

3. 実験と結果

提案手法の有効性を示すために、室内外で撮影した各シーン3枚の全方位パノラマ画像を入力し、それらから撮影者が写らない画像を生成した。なお、画像の解像度は 1024×512 画素で、SSD 算出におけるパッチのサイズは 21×21 とした。図1に室内で撮影した3枚の入力画像を示す。この図では、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1 入力画像

次に、図1の左上の画像を基準に平行移動した2枚の画像と中央値を抽出した画像を図2に示す。



図2 平行移動画像と中央値画像

中央値画像では、撮影者は除去できているものの、画像間の垂直方向の位置ずれから、物体が二重に映っているところが多い。次に $\lambda = 100, \kappa = 1000$ でグラフカットを行なった出力画像と、グラフカットにより出力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3に示す。また、室外で撮影した入力画像およびそれに対して $\lambda = 100, \kappa = 1000$ で出力した画像とグラフカットにより出力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4に示す。

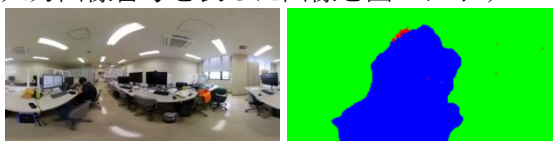


図3 グラフカットを行った結果

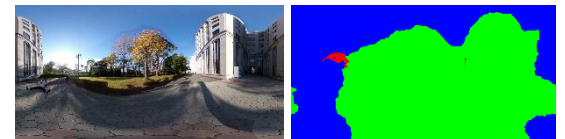


図4 屋外シーンでの入力画像と実験結果

実験の結果より、中央値画像と比較しても物体が二重に映ることなく撮影者を除去できている。ただし、図4では撮影者の影が残っていることがわかる。

4. おわりに

本研究では、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パノラマ画像を生成する手法を提案した。

実験では出力結果から撮影者を除去することができた。一方、今回は3枚に入力画像で実験を行なったが、影や反射がある環境では3枚の入力画像では不十分である。このため、今後はより多くの入力画像を用いて実験を行う必要がある。

謝辞 本研究の一部は、JSPS 科研費 (JP18H03273, JP18H04116, JP21H03483) の助成を受けて実施した。

参考文献

- [1] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. Proc. IEEE CVPR, pp. 5505–5514 (2018)
- [2] Chris, H., Mike, S., A Combined Corner and Edge Detector, Alvey Vision Club, pp. 23.1 – 23.6 (1988)
- [3] Martin, A., Robert, C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartograp, Communications of the ACM, 24(6), pp. 381-395 (1981)
- [4] Boykov, Y., Kolmogorov, V., An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE TPAMI 26(9), pp. 1124–1137 (2004)