

敵対的生成ネットワークを用いた SAR 衛星画像と光学衛星画像の対応点探索法

鳥屋 剛毅^{*1} Ashraf Dewan^{*2} 北原 格^{*1}

Abstract --- 本稿では、合成開口レーダー(SAR)衛星画像と光学衛星画像といった観測モデルの異なる画像間の対応点探索法を提案する。画像の局所特徴ベースの対応点探索アルゴリズムを用いて SAR 画像と光学画像の画像レジストレーション(幾何学的なレジストレーション)を行う。対応点探索アルゴリズムでは、対応する位置で共通の画像特徴が取得されていることが前提とされるが、SAR 画像と光学画像では画像特徴が大きく異なる。本稿では、敵対的生成ネットワーク(GAN)を用いて、片方の画像の見え方をもう一方の画像の見え方に変換することにより、この問題を解決する手法を提案する。定量的評価を実施し、提案手法によって SAR 画像と光学画像の正確な対応点探索が実現できることを確認する。

Keywords: 画像レジストレーション, 合成開口レーダー(SAR), 敵対的生成ネットワーク(GAN)

1 はじめに

「百聞は一見に如かず」と言われるように、画像や動画など視覚情報を用いることにより短時間で状況を詳細に把握することが可能である。自然災害が発生した場合など緊急時において、衛星、航空機、ドローンなどリモートセンシングによって取得された空中画像であれば、広範囲を一括して観測し状況把握に有用である。こういった事情から、これらリモートセンシングデータは、災害後の対応および復旧作業に広く利用される[1]。観測の際、衛星による高高度からの観測であれば観測範囲が広がるが、可視光とその周辺帯域を観測する光学衛星画像による観測では、光源の不足や雲の影響により、情報が欠落するという問題が発生する。一方、合成開口レーダー(SAR: Synthetic Aperture Radar)による観測では、それら照明条件や天候による情報の欠落はなく、広範囲の情報を得ることが可能である。SAR 画像の各ピクセル値は、SAR 衛星に搭載されたレーダーアンテナより地表へ照射された、波長数 cm～数十 cm の電磁波の後方散乱(再帰性反射)の強度を表している。SAR で用いられる電磁波は可視光より長波長であるため、その反射特性の違いにより、同一領域を観測した場合でも SAR 画像と光学画像の見え目は大きく異なる。そのため、SAR 画像から地表上の情報を獲得する場合には、我々人間が見慣れた光学画像と SAR 画像との対応関係を求め、判読を支援する必要がある。この画像間の対応関係を求めるためには、両方の画像で共通して観測されている地点(ランドマークポイント)の情報が必要となるが、正確な対応関係を求めるためには、なるべく

広い範囲でなるべく多くのランドマークポイントを求めることが肝要である。そこで我々は、SAR 画像と光学画像の幾何学的レジストレーションを自動的かつ高精度に実現する手法を提案し、この問題の解決を試みる。

画像の幾何学的レジストレーションは、デジタル標高モデル(DEM: Digital Elevation Model)を使用する手法が一般的である[2]。しかし、レジストレーションの対象となる SAR 画像と光学画像の解像度が高い場合、DEM の解像度が不足することが懸念される。また、災害の発生後、地形が大幅に変化し、その結果 DEM がその状況に適さないものになってしまう場合がある。したがって、DEM に依存しない画像特徴ベースのレジストレーション手法が必要とされている。しかしながら、SAR 画像と光学画像の間に存在する画像特徴の差により、両者に共通する画像特徴を取得することは困難である。

そこで提案手法では、画像特徴のマッチング処理を実行する前に、敵対的生成ネットワーク(GAN: Generative Adversarial Networks)[3]により訓練された画像変換器を使用し、SAR 画像の外観を光学画像に変換する前処理を採用する。

図 1 に示すように、画像特徴ベースの特徴点検出、特徴量記述、および対応点探索アルゴリズムを使用し、マルチモーダル(SAR および光学)画像間の局所特徴対応を見つける方法を提案した[4]。本稿では、その手法の適用の有効性と性能限界を定量的に確認するため、実験的評価を実施した。

2 関連研究

DEM ベースではない、画像ベースのレジストレーションを行うには、観測周波数の違う二画像から共通の特徴量を検出する必要がある。テンプレートマッチングを

^{*1} 筑波大学

^{*2} Curtin University

用いた手法としては、二つの画像間の正規化相互相関(NCC)または相互情報量(MI: Mutual Information)を使用した手法が提案されている[5][6]。これらの方法は比較的広範囲の画像特徴を使用する必要がある。テンプレートのサイズが小さいとテンプレートのマッチング精度が低下し、テンプレートのサイズを大きくすると部分的な差異(災害前後など)やオクルージョンへのロバスト性が低下する。したがってこのようなテンプレートベースの手法は、災害前後に地形が部分的に変化する可能性があるため、災害現場に適用されることを想定すると、より局所の画像特徴量によってレジストレーションする必要がある。

画像ベースの幾何学的レジストレーションを実現するためには、画像の局所特徴量が利用される。特に、特徴点ベースの方法[7][8][9]は、テンプレートベースのマッチング方法と異なり、部分的な対応を使用して二画像間の対応関係を推定できるという利点がある。SIFT [7], SURF[10], ORB[11], AKAZE[12]などの特徴点検出器・特徴点記述子を使用する方法は、二つの画像間のマッチングにしばしば使用される。これらの方法は、幾何学的な変動(回転、スケールなど)および照明条件の変化にロバストな特徴量を記述する。これらは、対応する特徴点と同様の特徴を持つ特徴点のペア(対応点)を見つけ、最終的に画像間の対応関係の推定を実現する。この手法は局所特徴を使用するため、観測されたエリアの一部が災害によって変化した場合でも、その他の部分的な箇所から対応を取得できるという利点がある。

画像のモーダル変換は、深層ニューラルネットワーク(DNN)を用いた機械学習が広く利用されている[13][14]。本稿では、主にDNNを用いたGANに焦点を当てる。GANは、ランダムな入力からデータを生成する生成器(Generator)と、データの真贋を判断する識別器(Discriminator)を訓練し、両者の性能を競わせることによって本物のデータに類似したデータを生成する生成器を得るネットワークである。GANの一つの発展形として、生成器に変換元の画像を入力とするpix2pix[15]は、多目的に用いることの出来る画像変換手法として知られている。このpix2pixで得られる生成器は、画像を入力としてその画像を別モーダルの画像へ変換する「画像変換器」としての役割を有する。Pix2pixはその論文において[15]、線画像から実際の写真へのマルチモーダルな変換が可能であることを実証している。Pix2pixは、特徴点検出に必要なエッジとコーナー情報を保持するような学習を行うため、画像レジストレーションを目的としたSAR画像から光学画像への変換でも高いパフォーマンスが期待できる。

GANのもう1つの利点は、少ない訓練データセットでも効果が期待できる点である[15]。従来、機械学習には多くの訓練データセットが必要であったが、GANは生

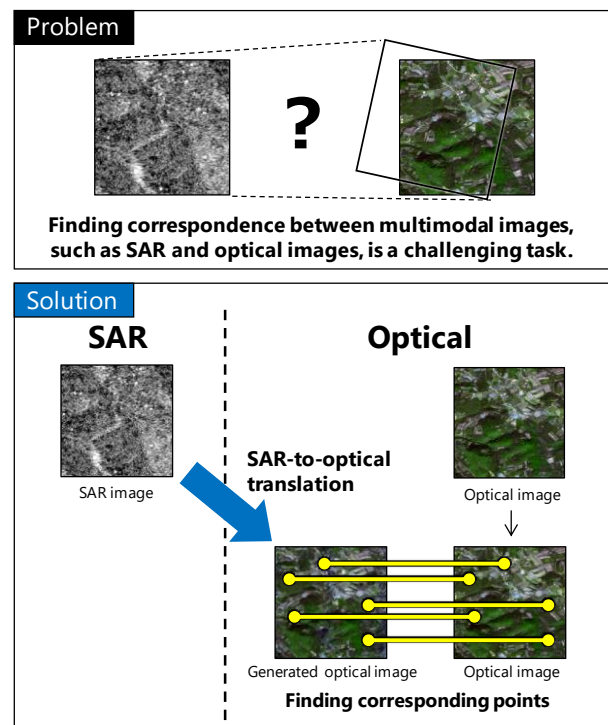


図 1 提案する SAR と光学の画像間の対応点探索手法の概要。対応点探索の前処理ステップとして GAN を使用し、マルチモーダル画像のレジストレーションを実現する。

Fig. 1 Outline of research on the proposed method of finding corresponding points between multimodal images.

Using GANs as a preprocessing step before keypoint matching, local correspondences are established for multimodal image registration.

成器モデルと識別器モデルにより、少ないデータセットで高いパフォーマンスが期待できる。我々の研究目的の場合、大量の訓練用 SAR・光学画像データセットを準備することは困難であるため、非常に重要な利点である。

本稿では、対応点探索アルゴリズムをマルチモーダルな画像間に適用する方法として、GAN によって SAR 画像を光学画像に変換しモーダルを揃える方法を提案する。GAN ベースの手法である pix2pix によって SAR から光学への画像変換器を訓練し、その画像変換器を SAR 画像に適用する。それによりマルチモーダルな画像間の幾何学的レジストレーションという課題が、モノモーダルな画像間の幾何学的レジストレーションという従来広く研究されてきた課題に変換され、特徴点ベースの対応点探索によって幾何学的なレジストレーションが実現される。

3 SAR から光学への画像変換のための GAN の訓練

本章では、SAR 画像から光学に似た画像(疑似光学画像)へ変換するための GAN の訓練プロセスについて説明する。入力として SAR 画像を、正解として光学画像を使用して GAN の訓練を行い、変換を学習した生成器

の推論により、疑似光学画像を得る。図 2 は、GAN ベースの画像変換を学習させる pix2pix モデルを示している。このモデルでは、生成器 G と識別器 D を互いに競わせ、より高品質の疑似光学画像を取得する。

すでに幾何学的なレジストレーションがされている SAR 画像と光学画像のペアを準備し、SAR 画像を pix2pix への入力 x_{train} として設定し、 y_{train} を生成器 G と識別器 D の訓練に対する正解として設定する。この生成器 G を用いて SAR から光学へ変換するプロセスにより、疑似光学画像は元の光学画像と同じモダリティを持ち、従来の対応点探索処理を適用することが可能となる。

4 特徴点検出器と特徴量記述子を用いた対応点探索

提案手法の対応点探索処理の概要を図 3 に示す。前章で説明したように、pix2pix モデルを使用して生成器 G を取得し、その生成器 G により SAR 画像を疑似光学画像へ変換する。対応点探索処理は、疑似光学画像と実際の光学画像の間で実行される。対応点を見つける過程で、結果に誤対応が含まれる場合があり、この誤対応によってレジストレーションの精度が低下するため、外れ値として削除する処理が必要となる。地図投影された画像間のレジストレーションの場合には、スケール、回転、及び並進の違いのみとみなすことができるため、誤対応を除去する処理は、特徴点のスケール値と勾配方向に従って行う事が可能である[16]。図 4 は、誤対応を減らした後、対応点をプロットした結果を示している。

5 評価実験

本章では、提案手法を検証するための実験について説明し、結果と考察を示す。

5.1 実験の目的と評価尺度

この実験の目的は、提案手法の有効性を検証するため、対応点探索の精度を定量的に評価することである。具体的には、対応点が検出された座標値の精度を評価する。この実験のデータセットとして、レジストレーション処理済みの SAR および光学衛星画像を使用した。評価尺度として、対応点の SAR 画像と光学画像それぞれにおける座標値のユークリッド距離を用いる。レジストレーションが完全であるならば、理想的には対応点はそれぞれ同じ座標値を持つため、評価尺度は 0 に近いほど高精度ということになる。

比較のため、SAR 画像と光学画像から直接対応点探索を行う手法である DLSC[17]と同条件で実験を行い、精度の比較を行った。

5.2 実験環境

本節では実験環境、主にデータセットとプログラムの実装およびその実行環境について述べる。

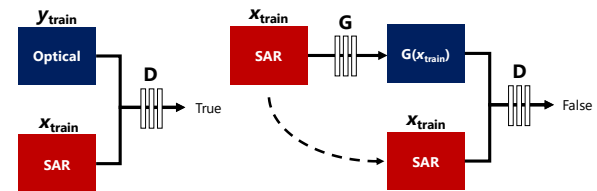


図 2 敵対的生成ネットワークの訓練。

Fig. 2 Training model structure of SAR-to-optical image translation.

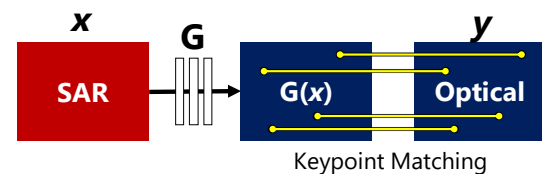


図 3 処理の流れ。3 章にて訓練された生成器は、入力の SAR 画像から疑似光学画像を生成する。

Fig. 3 Flow of the process. Trained G in Section 3 predicts optical-like images (generated optical image) from input SAR images.

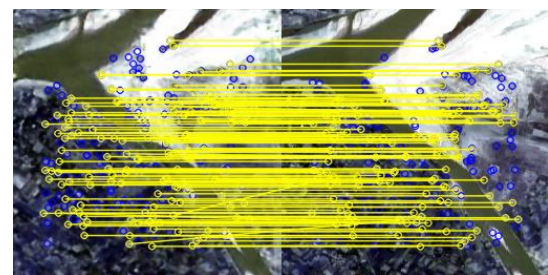


図 4 二画像の対応点探索の例。各黄色の線は対応点を、青い点は対応点がないことを意味する。

Fig. 4 Example of finding local correspondences of two images. Each yellow line connects corresponding points. Blue points mean that they do not have corresponding points.

5.2.1 データセット

この実験には、欧州宇宙局[18]の Sentinel-1 (SAR 衛星) および Sentinel-2 (光学衛星) の画像データから作成された SEN1-2 データセット[19]を使用した。データセットには、レジストレーション済みの Sentinel-1 および Sentinel-2 の画像パッチが含まれている。Sentinel-1 画像の各ピクセル深度は 8 ビットグレースケールであり、画像サイズは 256×256 ピクセルを有する。Sentinel-2 画像は、8 ビット RGB、 256×256 ピクセルである。Sentinel-2 画像はグレースケール画像に変換し GAN の正解データとした。空間分解能は、ピクセルあたり 10 m である。SEN1-2 データセットの「spring」データの 1128 セットの画像ペアを選定し、3 章で説明した pix2pix モデルを訓練するため 900 セットを訓練データ、100 セットを検証 (validation) データ、128 セットをテストデータに割り振る。

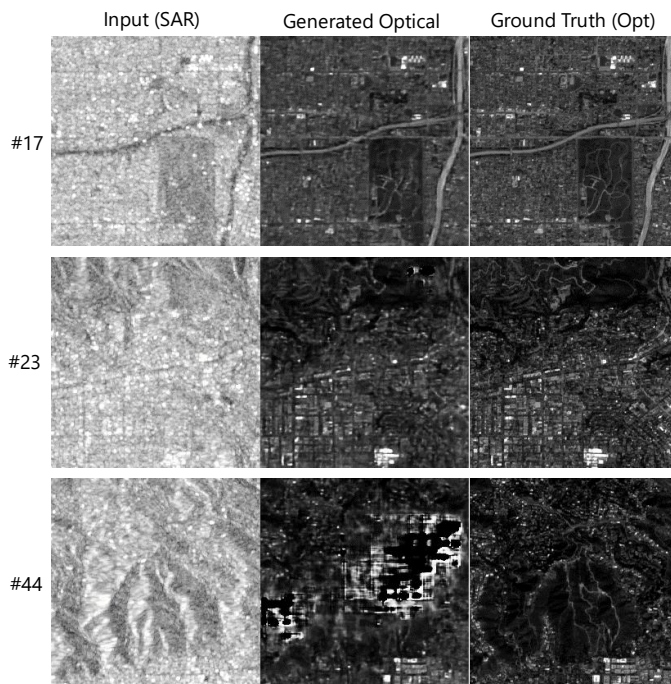


図 5 入力と結果の例. 左から, 入力 SAR 画像, 生成器を介して生成された疑似光学画像, およびグラウンドトゥルースの光学画像. PSNR 値は, 上から 26.198 dB(最良値), 23.878 dB(平均に最も近い), および 14.203 dB(最悪値)である.

Fig. 5 Example of the inputs and results. Input SAR images (left), output generated optical images through pix2pix trained model (center), and ground truth optical images (right). The PSNR values are 26.198 dB (top, the best case), 23.878 dB (middle, the closest one to the average) and 14.203 dB (bottom, the worst case).

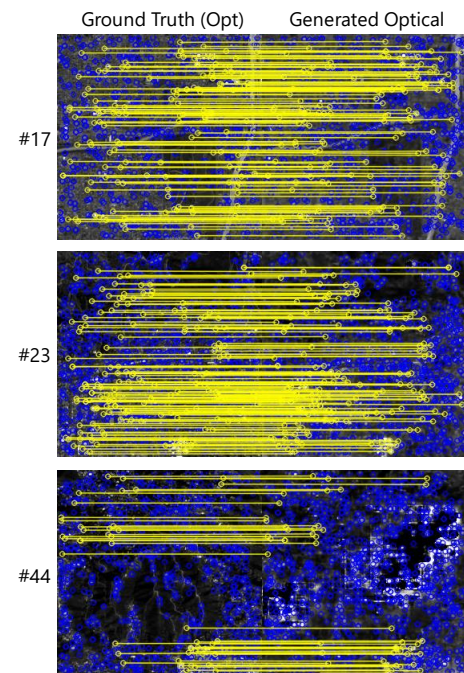


図 6 対応点探索の結果. PSNR 最悪値のケースにおいて, 崩壊した領域では局所的な特徴の対応は検出されないが, 残りの箇所に多くの対応点検出されている.

Fig. 6 Results of keypoint matching. Indeed, no local feature correspondences are extracted on a collapsed area in the worst PSNR case (bottom), but many corresponding points can be found in the rest.

た. 選定された画像の場所は, 災害時の重要性を鑑み, 主に市街地や住宅地など人工物の立地するエリアとした.

5.2.2 プログラムの実装および実行環境

GitHub Web サイトで公開されている pix2pix のソースコードをベースとして使用した[15]. PatchGAN[20]のパッチサイズは 128×128 ピクセルに設定し, バッチサイズは 4, エポック数は 2000 回とした. 並列処理に 6 GB の VRAM を搭載した NVIDIA GeForce GTX 1070 GPU を使用した. 訓練時間は 20 時間であった. 特徴点検出器・特徴量記述子として SIFT[7]を使用し, パラメータは OpenCV 3.4.3[21]のデフォルト値とした. また, 対応点の距離のしきい値を 3.0 ピクセルとした. これは 3.0 ピクセルを超える距離を持つ対応点は外れ値と見なされることを意味する.

DLSC のパラメータは, テンプレートウィンドウサイズは 64 ピクセル(テストイメージサイズの 1/4)に設定した. 提案手法で使用した, 距離のしきい値 3.0 ピクセルを DLSC にも適用した. 手法の実装には著者らが公開している MATLAB コードを使用した[17].

5.3 結果と考察

SAR から光学画像への変換の結果を図 5 に示す. pix2pix で得られた生成器によって, SAR 画像が適切に疑似光学画像へ変換されていることがわかる. 128 枚のテスト画像の PSNR(Peak Signal Noise Ratio)値は, 平均 23.90 dB, 標準偏差 2.142 dB, および中央値 24.35 dB であった. 図 5 は, それぞれ, PSNR の最良値・平均値に最も近い値・最悪値を持つ, 入力・推論結果・グラウンドトゥルースを示している. これらの画像に対応点探索を適用した場合の対応点検出の精度を評価した.

図 6 は, 光学画像と疑似光学画像の間の対応点探索結果をプロットしたものである. 十分な数の対応点検出され, pix2pix が SAR から光学への変換に効果的であることが確認された. 誤対応が残るケースもあるものの, なべて正対応に比べ低い割合であるため, RANSAC[22]などのロバスト推定方法によって除去することが可能である. PSNR 最悪値のケースである図 6 の 3 番目のデータセットを詳しく見ると, 疑似光学画像の中央部分は pix2pix による変換に失敗している. しかし, 中央付近の失敗した領域での対応点は検出されないもの

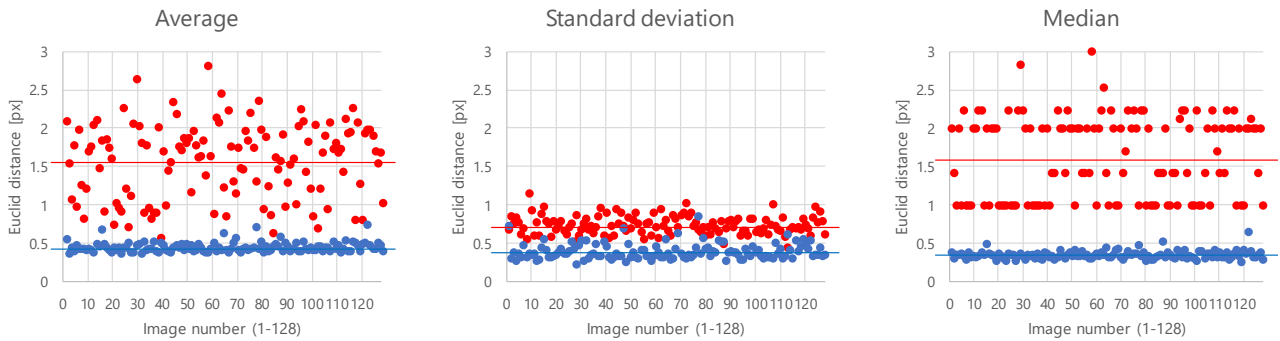


図 7 対応する各特徴点位置の差の平均、標準偏差、および中央値。赤と青の点は、それぞれ DLSC[17]と提案手法の結果を示す。赤と青の線はそれぞれの平均値を示している。DLSC では、平均値の平均は 1.58 ピクセル、平均標準偏差は 0.74 ピクセル、平均中央値は 1.63 ピクセル。提案手法では、平均は 0.45 ピクセル、平均標準偏差は 0.39 ピクセル、平均中央値は 0.35 ピクセルである。

Fig. 7 Average, standard deviation, and median of differences of each corresponding keypoint position. The red and blue points show the results of DLSC[17] and our proposed method, respectively. The red and blue lines show their mean values. In DLSC, the mean Average is 1.58 pixels, the mean Standard Deviation is 0.74 pixel, and the mean Median is 1.63 pixels. In the proposed method, the mean Average is 0.45 pixel, the mean Standard Deviation is 0.39 pixel, and the mean Median is 0.35 pixel.

の、残りの箇所では多くの対応点が検出されていることが分かる。画像の一部でしか対応点が見つからなかった場合でも、SIFT または他の特徴点検出/特徴量記述法は局所特徴を計算するため、画像全体を正しく変換できない場合であっても、正しく変換された部分の情報を使用して対応関係を推定することが可能である。

図 7 は、128 枚のテスト画像データセットの対応する各対応点の座標値差の平均、標準偏差、および中央値を示している。DLSC では、平均値の平均は 1.58 ピクセル、平均標準偏差は 0.74 ピクセル、平均中央値は 1.63 ピクセルであった。提案手法では、平均値の平均は 0.45 ピクセル、平均標準偏差 0.39 ピクセル、平均中央値 0.35 ピクセルであった。この結果は、 $\pm 1\sigma$ の範囲内であっても 1.0 ピクセル未満の平均精度を達成可能であることを示しており、したがって、この方法がレジストレーションに十分な精度を達成していると言える。

図 8 に、平面仮定を置いた系において、提案手法を用いて SAR 画像から光学画像への幾何学的レジストレーションを施した実例を示す。画像データは前出の SEN1-2[19]を用いた。二次元射影変換は、光学画像と疑似光学画像の間の対応点探索結果から推定され、最終的に SAR 画像を二次元射影変換によって光学画像上へ投影しレジストレーションを実現する。図 8 の結果より、提案手法によって高精度の幾何的レジストレーションが達成されていることを確認した。使用したデータセットはあまり広域のものではないこと、平らな表面の都市エリアを観察した画像であることから、平面仮定でも高精度な幾何的レジストレーションが可能であったと考えられる。あるいは、平面仮定がおけない場合であっても、必要に応じて、対応点の情報から 3D 投影を推定するこ

とも可能である[23]。

6 おわりに

モダルの異なる二画像(SAR 画像・光学画像)に対応点探索処理を適用する方法として、敵対的生成ネットワーク(GAN)をベースとした pix2pix を用い、SAR 画像を疑似光学画像に変換してモデルを揃える方法を提案した。この SAR から光学画像への変換手法によって疑似光学画像を取得した後、モノモデルとなった画像(疑似光学画像と光学画像)間で対応点探索処理を実行し、最終的に SAR 画像と光学画像の対応点を検出した。

対応点探索の精度を定量的に評価することにより、pix2pix で訓練された生成器によって生成された疑似光学画像と元の光学画像の間で、正確な対応点探索を実現できることが確認された。具体的には、 $\pm 1\sigma$ の範囲内であっても、1.0 ピクセル未満の平均精度を達成可能であることが示された。

本研究は JST CREST(JPMJCR16E3)の助成を受けたものである。

参考文献

- [1] V. Stefan et al., "Global trends in satellite-based emergency mapping," Science, vol. 353, issue 6296, pp. 247-252, Jul. 2016.
- [2] J. R. Jensen, "Geometric correction," in Introductory Digital Image Processing: A Remote Sensing Perspective, 4th ed., Prentice Hall Press Upper Saddle River, NJ, USA, 2016, pp. 235-272.
- [3] Goodfellow et al., "Generative Adversarial Nets," in Proc. Advances in neural information processing systems, Montreal, Canada, 2014, pp. 2672-2680.

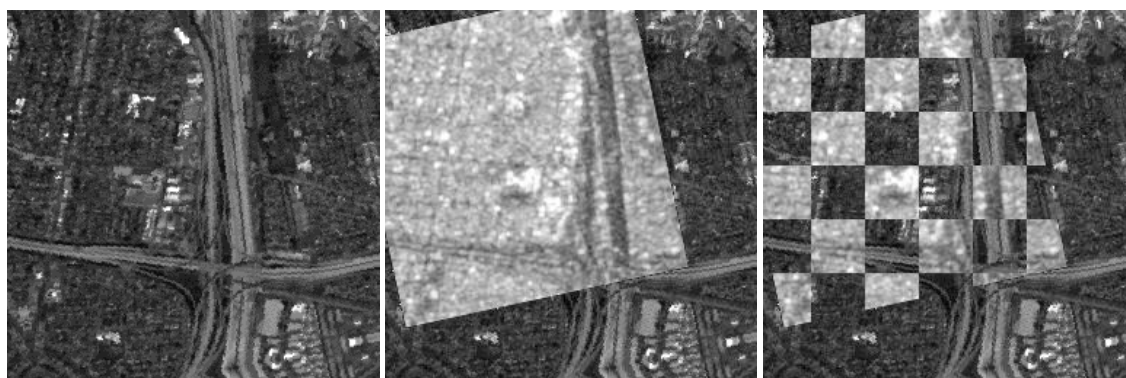


図 8 平面仮定における SAR および光学画像の幾何的レジストレーションの例。SAR 画像をターゲットの光学画像(左)にレジストレーションするために、疑似光学画像と光学画像の対応点探索の後、ホモグラフィ変換を施してレジストレーションした結果である(中央)。右は半分を透過し対応を見やすくしたものである。

Fig. 8 Example of SAR and optical image registration on the plane assumption. To overlay a SAR image to the target optical image (left), after finding the local feature correspondences between generated optical and optical images, the original SAR image was projected onto the optical image using estimated homography transform (center), and half of the area is transparent like a checkerboard pattern (right). It can be confirmed that image registration between SAR and optical images is performed with high accuracy.

- [4] H. Toriya et al., "SAR2Opt: Image alignment between multi-modal images using generative adversarial networks," IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019.
- [5] Dame et al., "Accurate real-time tracking using mutual information," in Proc. International Symposium on Mixed and Augmented Reality, Seoul, Korea, 2010, pp. 47-56.
- [6] Xiong et al., "Registration for SAR and optical images based on straight line features and mutual information," in Proc. IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 2016.
- [7] G. Lowe et al., "Object recognition from local scale-invariant features," In Proc. IEEE International Conf. on Computer Vision, vol. 99, no. 2, Kerkyra, Corfu, Greece, 1999, pp. 1150-1157.
- [8] R. Szeliski, "Feature detection and matching," in Computer Vision: Algorithms and Applications, 1st ed., Springer-Verlag London, 2011, pp. 200-207.
- [9] M. I. Patel et al., "Image registration of satellite images with varying illumination level using HOG descriptor based SURF," Procedia Comput. Sci., vol. 93, 2016, pp. 382-388.
- [10] H. Bay et al., "SURF: Speeded-Up Robust Features," in Proc. European conference on computer vision, Springer, Berlin, Heidelberg, 2006, pp. 404-417.
- [11] Rublee et al., "ORB: An efficient alternative to SIFT or SURF," in Proc. IEEE International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2564-2571.
- [12] P. F. Alcantarilla et al., "Fast explicit diffusion for accelerated features in nonlinear scale spaces," IEEE Trans. Patt. Anal. Mach. Intell., vol. 34, no. 7, 2011, pp. 1281-1298.
- [13] Simo-Serra et al., "Learning to simplify: fully convolutional networks for rough sketch cleanup," ACM T. Graphic, vol. 35, no. 4, issue 121, 2016.
- [14] K. Enomoto et al., "Image translation between SAR and optical imagery with generative adversarial nets," in Proc. IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 1752-1755.
- [15] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in Proc. Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017, pp. 1125-1134.
- [16] H. Toriya et al., "Mobile camera localization using aerial-view images," IPSJ Trans. CVA, vol. 9, no. 4, 2014, pp. 896-904.
- [17] Y. Ye et al., "Robust optical-to-SAR image matching based on shape properties," IEEE Geosci. Remote S., vol. 14, no. 4, 2017, pp. 564-568.
- [18] M. Berger et al., "ESA's Sentinel missions in support of earth system science," Remote Sens. Environ., vol. 120, 2012, pp. 84-90.
- [19] M. Schmitt et al., "The SEN1-2 dataset for deep learning in SAR-optical data fusion," ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., vol. IV-1, 2018, pp. 141-146.
- [20] Li, Chuan et al., "Precomputed real-time texture synthesis with markovian generative adversarial networks," In European Conference on Computer Vision. Springer, Cham, p. 702-716, 2016.
- [21] OpenCV, available: <https://opencv.org/>, accessed: 12 June 2019.
- [22] M. A. Fischler et al., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol. 24, no. 6, 1981, pp. 381-395.
- [23] R. Szeliski, "Feature-based alignment," in Computer Vision: Algorithms and Applications, 1st ed., Springer-Verlag London, 2011 pp. 275-283.

© 2020 by the Virtual Reality Society of Japan (VRSJ)