Greedy Cost Optimization Algorithm for Satellite Image Mosaic Problem

Manuel Combarro Simón[1][0000-0002-2699-1397], Jedrzej Musial[2][0000-0003-3018-5010], Grégoire Danoy[1] [0000-0001-9419-4210], Mohammed Alswaitti[1] [0000-0003-0580-6954], Andrei Tchernykh[3][0000-0001-5029-5212], Johnatan E. Pecero[1][0000-0002-2695-9960] and Pascal Bouvry[1][0000-0001-9338-2834]

[1] University of Luxembourg, avenue de l'Université. 2, 4365 Esch-sur-Alzette, Luxembourg
[2] Poznan University of Technology, Pl. Marii Skłodowskiej-Curie 5, 60-965 Poznan, Poland
[3] CICESE Research Center, Carretera Ensenada - Tijuana 3918, 22860 Ensenda, Mexico

Abstract. Satellite imagery solutions are widely used to study different regions of the earth. A single satellite image can cover only a limited area, in cases where a larger Area of Interest (AOI) is needed, merging several images into a mosaic is inevitable. Nowadays, with the large number of satellite images that are available for commercial use, choosing the set of images to build the mosaic might be challenging, especially if the user wants to keep the overall cost as low as possible. In this paper, we give a formal definition of this problem, and we prove that it is NP-hard. We also, propose a greedy heuristic to minimize the total cost of the images that are in a cover of the AOI and evaluate its performance against a random selection strategy for five AOIs.

Keywords: Satellite Image Mosaic, Combinatorial Optimization, Geometric Weighted Set Cover, Set Cover.

1 Introduction

The space industry is expanding and is no longer an exclusive market for governmental and military applications. According to the most recent European Union Agency for the Space Programme (EUSPA) report [1], the Global Navigation Satellite System and Earth Observation (EO) market had revenues of around €200 billion in 2022 and is expected to reach €500 billion by 2031. As the access to space has become cheaper, more private companies have entered the space business, increasing its popularity. There even exist companies that use space data without owning any space assets, thanks to services such as satellite-as-a-service [2].

Recently, the EO sector has witnessed significant developments, for instance, the number of satellite launches dedicated to EO in 2021 was higher than the launches between 2012 and 2016 [3].

There are several applications based on EO that analyze a vast AOI that can be only covered by combining several adjacent images into a bigger one, usually called a mosaic. Mosaics are crucial for applications like crop classification [4, 5], environmental monitoring [6, 7], and urban development analysis [8, 9], among others.

Mosaicking satellite images present several challenges such as geometric correction of the images [10, 11], color balancing [12, 13], image stitching [14, 15], etc. Also, another important problem is the presence of clouds in satellite images. Some research has been done to minimize the cloud covering percentage in the final mosaic [5, 16, 17]. As the number of available images is significantly increasing, a new challenge arises in choosing the optimal combination of images to be utilized to produce the mosaic by optimizing one or more criteria. This is equivalent to finding the best cover of the AOI. In theory, each of the n images that overlap the AOI could be part or not of a cove, therefore it could exist up to 2^n possible covers. To the best of our knowledge, there is no research about the former challenge. We named this problem as Satellite Image Mosaic Combination Problem (SIMCOP).

On some higher levels of abstraction, one kind finds some similarities between the SIMCOP and the Cloud Brokering Problem (CBP) [18], especially for the version with bundles [19]. The mentioned CBP describes the situation when a customer wants to buy cloud services. These are offered soloed or grouped in bundles (by providers). Another problem that can be somehow connected with the SIMCOP is the Internet Shopping Optimization Problem [20] in the different variations [21, 22], where a customer plans to buy products from Internet stores.

The main contribution of this paper is to model the SIMCOP problem and prove its NP-hardness. As a secondary contribution, a greedy heuristic is provided and compared against a randomized algorithm.

This paper is structured as follows: In Section 2, we give the mathematical definition of the problem. In Section 3, we prove that this problem is NP-hard. In Section 4, we explain the proposed algorithms. In Sections 5 and 6, we describe and show the experiments.

2 Problem definition

Consider an EO application that is interested in a specific AOI that is big enough to be covered by only one satellite image. Hence, a mosaic is needed. Depending on the application, the images that form the mosaic must meet certain requirements, such as minimum resolution, cloud coverage percentage, incident angle, and range of dates. The set of possible images that could be used to build the mosaics is formed by all the images that intersect the AOI and meet the requirements. Each image has an associated weight that represents its cost or one of the requirements (resolution, cloud coverage, etc.). The objective is to minimize the total weight of the mosaic.

Formally, the problem is defined as follows. Given a polygon A, requirements MR, where each value MR_j represents the maximum allowable value of requirement j, and a set I of n convex quadrilaterals (corresponding to all the satellite images in the range of date), with each quadrilateral i having a weight w_i, and requirement values r_i, where r_{ij} corresponds to the value of requirement j for quadrilateral i, we have the following objective:

$$min \sum_{i \in B} w_i \tag{1}$$

subject to the constraints

$$A \subseteq \bigcup_{i \in B} i \tag{2}$$

$$\forall i \in B,\ r_{ij} \leq MR_j \tag{3}$$

where B is a cover of the AOI.

This problem differs from the classical geometric weighted set cover [23, 24], where given a finite set of points P and a set I of weighted objects in the plane, the objective is to find a cover of P with minimum weight. The set to be covered here is not a set of points but a region in the plane with infinite cardinality. This makes it harder to verify that a certain configuration is a cover as we cannot check if all the points in the set are covered. Moreover, in this problem, each object (image) is characterized by a set of requirements.

3 NP-hardness

In this section, we analyze the computational complexity of SIMCOP. Moreover, we will prove its NP-hardness by proving the NP-completeness of its decision counterpart problem (let us call it SIMCOP-D). The latter has the same input as SIMCOP plus an additional parameter y. The question is whether or not there exists a selection of quadrilaterals covering the interested area (AOI) with a total weight of y or less.

Proposition 1 *Satellite Image Selection Problem is strongly NP-hard even if all weights of quadrilaterals are equal to one and there are no defined requirements.*

Proof Let us construct a polynomial transformation from the well-known, strongly NP-complete problem, Set Covering (Minimum Cover) Problem (SCP) [25, 26] to the SIMCOP-D problem.

The SCP can be described as follows. There is a given finite set of elements S, a collection C of subsets of S, and a positive integer $K \leq |C|$. The question is whether there is a possibility of covering the set S with K or fewer subsets $C' \subseteq C$ ($|C'| \leq K$) such that every element of S belongs to at least one member of C'?

It is understandable that if Y is a solution to SCP, then $|Y| \leq K$.

Given an instance of SCP, we construct the following instance of SIMCOP-D. Assume that there are no requirements (MR) for all of the quadrilaterals (or we can say that $MR_j \rightarrow \infty, \forall j$). There are i quadrilaterals ($i = C'$) which create a collection of quadrilaterals I ($I = C$). In the original definition of the SIMCOP, each quadrilateral covers a particular area. Let's assume that the size of the site is represented discretely - proportionally by the number of points (using some suitable measure, e.g., an area of 10x10 cm of the actual space is 1 point). Let's call these points specific points (elements). The customer would like to pick/order a number of quadrilaterals i to cover a desired area AOI (AOI $= S$), where AOI is a set of all required specific points (elements). Areas in $i, \forall i \in I$ that are outside the AOI can be omitted, therefore we can assume that they contain no specific points (elements). We assume that all elements of I are in the AOI (therefore, $I = C$ is a collection of subsets of $AOI = S$). All weights of quadrilaterals are equal to 1 ($w_i = 1, \forall i \in I$). The threshold value of the criteria is $y = K$.

Now we show that SCP has a solution if and only if there exists a solution X for the constructed instance of problem SIMCOP-D with $|X| \leq y$. Seeing that the transformation is polynomial, problem SIMCOP-D belongs to the class NP.

Let Y be a solution to SCP. Construct a solution for problem SIMCOP-D, in which the required specific points are realized with K quadrilaterals determined by $C' = i \in I$, i.e., i is a part of the solution if $i \in Y$ and i is not a part of the solution if $i \notin Y$ Since Y is a covering of S, all the required specific points are ordered. The cost of the corresponding solution X is $|X| \leq K$.

Now assume that there exists a solution X for the problem SIMCOP-D with the cost $|X| \leq y$. The number of chosen quadrilaterals for this solution should not exceed y because, otherwise, $|X| > y$ and corresponding solution Y will be $|Y| > y$. On the other hand, the number of quadrilaterals should not be less than y because otherwise at least one specific point will not be purchased. Therefore, there are exactly y quadrilaterals i. Since all purchased specific points cover the set AOI (S), the collections of chosen quadrilaterals ($i = C'$) represent a solution for SCP.

4 Proposed algorithm

In this section, we propose a greedy algorithm to tackle a specific scenario of the previously presented problem. Imagine a case in which an user is interested in getting the cheapest possible mosaic for certain AOI.

Let us assume that the satellite image cost is proportional to its area. This is true in several satellite image marketplaces when the images have the same resolution [27, 28]. In this case, a minimization of the cost is achieved by minimizing the area, so in equation (1) we can set the image's weight equal to its area.

Considering the previous, the proposed heuristic selects, in a greedy fashion, the image with the lowest ratio *(image's total area) / (image's area intersecting uncovered area of the AOI),* until the entire AOI is covered. If more than one image has the same ratio, the algorithm chooses the image that covers the biggest area inside the AOI. Algorithm 1 shows the pseudocode for this strategy.

We evaluate the proposed greedy algorithm against a random selection algorithm that randomly selects images until the AOI is fully covered. The random algorithm only selects an image if the image can cover an uncover area of the AOI.

Algorithm 1: Greedy Covering

INITIALIZE $B \leftarrow \{\}$
REPEAT UNTIL $Area(AOI) = 0$
  FOR all $i \in I$
    COMPUTE $AOI_i = i \cap AOI$
  Select $i_s \in I$ with $min(AOI_i \,/\, w_i)$ and $max(AOI_i)$
  $B = B \cup \{i_s\}$
  $I = I \setminus \{i_s\}$
  $AOI = AOI \setminus AOI_{i_s}$

9. RETURN *B*

5 Experimental setup

To perform the experiments, five AOIs in the world were selected: Mexico City (Mexico), Rio de Janeiro (Brazil), Paris (France), Lagos (Nigeria), and Tokyo Bay (Japan). For every AOI, there are 30 images available to the user to form a mosaic.

The constraints for the image selection were that the images should intersect the AOI, so every image can be part of a cover, and a cover of the AOI should be achieved. Considering those constraints, we randomly selected the images from the satellite image marketplace Up42 [29]. All the images used in the experiments were captured between 01-10-2022 and 01-01-2020 by the Pléiades Satellite Constellation[30].

As we are only interested in calculating the total area of the cover, it is not necessary to download the satellite images; it is sufficient to get the image corner's coordinates.

The value of a solution is the ratio between the area of all images forming the cover and the area of the AOI. The lower bound is 1, which means that both areas are equal, and it happens when the images of the cover don't overlap themselves and they are fully contained inside the AOI.

For each AOI and the corresponding 30 images, Greedy Covering is run one time generating one solution (as it is deterministic). On the other hand, Random Covering is run 100 times, generating 100 different solutions.

We compare Greedy Covering and Random Covering by counting how many random solutions (out of 100) were worse than the greedy solution and how much better was the greedy solution respect the average evaluation parameter of the random solutions.

6 Experimental analysis

In **Table 1**, it is presented a resume of the experiments. For each AOI it is shown the value of the Random Covering solution and the average of all the random solutions. Also, it is shown how well performed Greedy Covering compared to the average of the random solutions and how many random solutions were outperformed by the greedy solution.

Table 1. Experiments resume

AOI	Greedy Covering	Avg Random solutions	Greedy vs Avg Random %	Random solutions outperformed by Greedy Covering
Paris	3.53	4.33	18	84 / 100
Tokyo bay	3.18	5.22	39	99 / 100
Lagos	3.10	4.71	34	100 / 100
Mexico City	4.56	6.78	33	100 / 100
Rio de Janeiro	3.38	5.36	37	100 / 100

For all cases the proposed greedy heuristic was at least 33% better (cheaper) than the average random solution, except for Paris where it was 18%. Also, Paris was the AOI in which Greedy Covering outperformed the fewer random solutions but still was significant. For the other AOIs, just in Tokyo, one random solution was better than the greedy one, for the rest none were better.

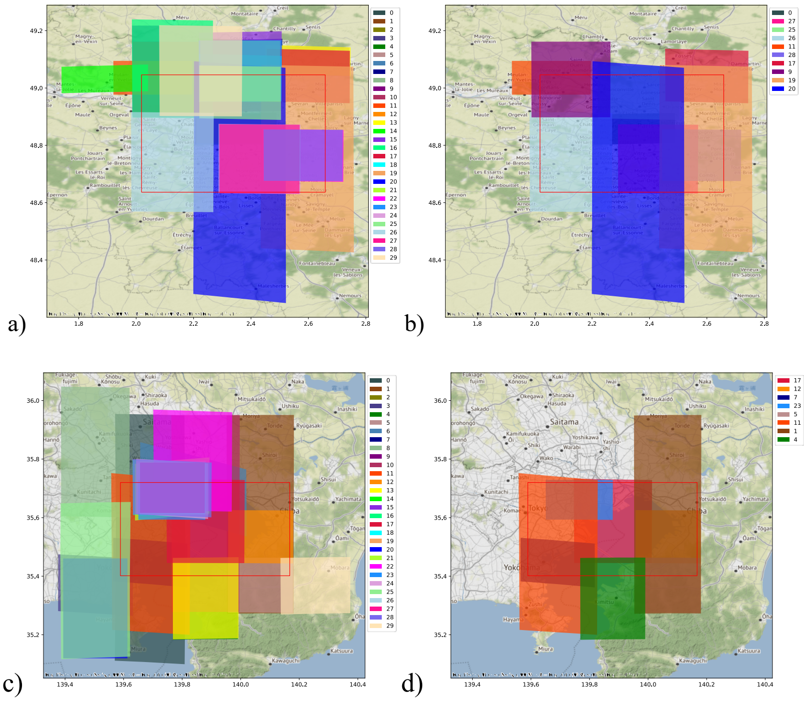

Fig. 1. AOIs corresponding to Paris and Tokyo Bay with the initial 30 satellite images and the greedy solution. The AOI is represented with a transparent square with red borders and each satellite image is represented with a unique color. For the greedy solution, it is shown in which order the satellite images were selected. (a) Paris with 30 initial images (b) Greedy solution for Paris. (c) Tokyo Bay with 30 initial images(d) Greedy solution for Tokyo Bay.

In **Fig. 1** a) and c), it is represented the AOIs corresponding to Paris and Tokyo Bay with the 30 images. In b) and d) is shown the greedy solution for those AOIs respectively. We can see that both solutions can be improved as there are some images that completely cover other ones. This happens because for the greedy strategy, at the moment of the pick, the images that are covered were a better choice than the image that covers it. In Paris, this situation is more evident because the last two images picked (19 and 20) covered almost 2/3 of the AOI, overlapping a bigger area, also these two images have to be in a cover as there some smalls area of the AOI that are only covered by them.

7 Conclusions

Satellite image mosaics are fundamental to monitoring and making predictions on vast regions of our planet. Selecting the images to build the mosaic is a challenging task, especially if the user wants to optimize one or several parameters, for example, the total cost of the selected images. With the purpose of modeling this situation, we presented the Satellite Image Mosaic Combination Problem (SIMCOP) and proved its NP-hardness. Furthermore, a greedy heuristic was proposed for this problem and was compared against a random selection strategy. Despite the greedy heuristic has room for improvement, it has, in general, a better performance than the random algorithm. For future work, we plan to propose heuristics with better performance and make experiments involving more AOIs and a larger number of images per AOI.

8 Acknowledgment

This work is partially funded by the Fonds National de la Recherche Luxembourg (FNR), Aides à la Formation-Recherche (AFR PhD) program under the ASTRAL Project (ref. 17043604), PolLux program under the SERENITY Project (ref. C22/IS/17395419), and by the joint research programme UL/SnT-ILNAS on Technical Standardization for Trustworthy ICT, Aerospace, and Construction.

References

1. EUSPA: EUSPA EO and GNSS Market Report. (2022).
2. ILNAS, ANEC: Aerospace Sector Luxembourg. (2021).
3. Union of Concerned Scientists: Satellite Database, https://www.ucsusa.org/resources/satellite-database, last accessed 2022/09/15.
4. Hall, O., Dahlin, S., Marstorp, H., Francisca Archila Bustos, M., Öborn, I., Jirström, M.: Classification of Maize in Complex Smallholder Farming Systems Using UAV Imagery.
5. Felegari, S., Sharifi, A., Moravej, K., Amin, M., Golchin, A., Muzirafuti, A., Tariq, A., Zhao, N.: Integration of Sentinel 1 and Sentinel 2 Satellite Images for Crop Mapping. Appl.
6. Nikolakopoulos, K.G., Lampropoulou, P., Fakiris, E., Sardelianos, D., Papatheodorou, G.: Synergistic Use of UAV and USV Data and Petrographic Analyses for the Investigation of Beachrock Formations: A Case Study from Syros Island, Aegean Sea, Greece. Miner. 2018, Vol. 8, Page 534. 8, 534 (2018).
7. Flood, N., Watson, F., Collett, L.: Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. Int. J. Appl. Earth Obs. Geoinf. 82, 101897 (2019).
8. Wang, H., Gong, X., Wang, B., Deng, C., Cao, Q.: Urban development analysis using built-up area maps based on multiple high-resolution satellite data. Int. J. Appl. Earth Obs. Geoinf. 103, 102500 (2021).
9. Piaggesi, S., Gauvin, L., Tizzoni, M., Cattuto, C., Adler, N., Verhulst, S., Young, A., Price, R., Ferres, L., Panisson, A.: Predicting City Poverty Using Satellite Imagery. (2019).
10. Sheng, Y., Alsdorf, D.E.: Automated georeferencing and orthorectification of amazon basin-wide SAR mosaics using SRTM DEM data. IEEE Trans. Geosci. Remote Sens. 43, 1929–1940 (2005).

11. Pradhan, B., Ahmed, A.A., Chakraborty, S., Alamri, A., Lee, C.W.: Orthorectification of WorldView-3 Satellite Image Using Airborne Laser Scanning Data. J. Sensors. 2021, (2021).
12. Cresson, R., Saint-Geours, N.: Natural Color Satellite Image Mosaicking Using Quadratic Programming in Decorrelated Color Space. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4151–4162 (2015).
13. Yu, L., Zhang, Y., Sun, M., Zhu, X.: Colour balancing of satellite imagery based on a colour reference library. Int. J. Remote Sens. 37, 5763–5785 (2016).
14. Megha, V., Rajkumar, K.K.: Automatic Satellite Image Stitching Based on Speeded Up Robust Feature. In: Proceedings - 2021 1st IEEE International Conference on Artificial Intelligence and Machine Vision, AIMV 2021. Institute of Electrical and Electronics Engineers Inc. (2021).
15. Sarl, H., Eken, S., Sayar, A.: An Approach for Stitching Satellite Images in a Bigdata Mapreduce Framework. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. pp. 351–355. Copernicus GmbH (2017).
16. Swaine, M., Smit, C., Tripodi, S., Fonteix, G., Tarabalka, Y., Laurore, L., Hyland, J.: Operational pipeline for a global cloud-free mosaic and classification of sentinel-2 images. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 43, 195–200 (2020).
17. Shen, C., Zhao, C., Yu, M., Peng, Y.: Cloud cover assessment in satellite images via deep ordinal classification. Int. Geosci. Remote Sens. Symp. 2018-July, 3509–3512 (2018).
18. Guzek, M., Gniewek, A., Bouvry, P., Musial, J., Blazewicz, J.: Cloud brokering: Current practices and upcoming challenges. IEEE Cloud Comput. 2, 40–47 (2015).
19. Musial, J., Kieffer, E., Guzek, M., Danoy, G., Wagle, S.S., Bouvry, P., Blazewicz, J.: Cloud Brokering with Bundles: Multi-objective Optimization of Services Selection. Found. Comput. Decis. Sci. 44, 407–426 (2019).
20. Blazewicz, J., Kovalyov, M., Musial, J., Urbanski, A., Wojciechowski, A.: Internet shopping optimization problem. Int. J. Appl. Math. Comput. Sci. 20, 385–390 (2010).
21. Musial, J., Lopez-Loces, M.C.: Trustworthy Online Shopping with Price Impact. Found. Comput. Decis. Sci. 42, 121–136 (2017).
22. Musial, J., Pecero, J.E., Lopez-Loces, M.C., Fraire-Huacuja, H.J., Bouvry, P., Blazewicz, J.: Algorithms solving the Internet shopping optimization problem with price discounts. Bull. Polish Acad. Sci. Tech. Sci. 64, 505–516 (2016).
23. Har-Peled, S., Lee, M.: Weighted geometric set cover problems revisited. J. Comput. Geom. 3, 65–85 (2012).
24. Chan, T.M., Grant, E., Könemann, J., Sharpe, M.: Weighted capacitated, priority, and geometric set cover via improved quasi-uniform sampling. In: Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1576–1585 (2012).
25. Garey, M.R., Johnson, D.S.: Computers and Intractability. W. H. Freeman and Company, New York (1979).
26. Karp, R.M.: Reducibility among Combinatorial Problems. In: Miller, R.E., Thatcher, J.W., and Bohlinger, J.D. (eds.) Complexity of Computer Computations. pp. 85–103. The IBM Research Symposia Series. Springer, Boston, MA (1972).
27. Skywatch Buy Satellite Imagery, https://skywatch.com/pricing/, last accessed 2022/11/28.
28. Satfotos Homepage, https://satfotos.com/, last accessed 2022/11/28.
29. UP42 Homepage, https://up42.com/, last accessed 2022/09/23.
30. Earth Online - Pleiades, https://earth.esa.int/eogateway/missions/pleiades, last accessed 2022/09/23.