

Метод автоматического совмещения разнородных цифровых изображений дистанционного зондирования Земли

А.Н. Борисов¹, В.В. Мясников¹, В.В. Сергеев¹

¹ Самарский национальный исследовательский университет имени академика С.П. Королёва, 443086, Россия, г. Самара, Московское шоссе, д. 34

Аннотация

В работе предлагается метод автоматического совмещения разнородных цифровых изображений дистанционного зондирования Земли с использованием данных о съемке. Метод предназначен для совмещения цветных, полутонных, мультиспектральных и радарных изображений и их комбинаций с возможным отличием в пространственном разрешении до четырех (опционально – шестнадцати) раз. Основными этапами предлагаемого метода являются: опциональный этап повышения разрешения (до четырех раз); опциональный этап снижения числа каналов цифровых снимков до предустановленных трех или одного; этап поиска особых точек и этап их описания и совмещения. Для получения универсального и устойчивого решения на последних этапах в работе сопоставлялись лучшие известные алгоритмы: SIFT, SAR-SIFT, RIFT и обучаемый RoMa. Экспериментальные исследования с использованием указанных типов космических изображений показали однозначное преимущество обучаемой нейросетевой модели RoMa, которая и была настроена/обучена на множестве разнородных снимков. Для дополнительного повышения точности совмещения мы использовали априорные данные о снимках в виде данных их геопривязки.

Ключевые слова: цифровые изображения дистанционного зондирования Земли, автоматическое совмещение изображений, мультиспектральные изображения, радарные изображения.

Цитирование: Борисов, А.Н. Метод автоматического совмещения разнородных цифровых изображений дистанционного зондирования Земли / А.Н. Борисов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. – 2024. – Т. 48, № 6. – С. 932-943. – DOI: 10.18287/2412-6179-CO-1604.

Citation: Borisov AN, Myasnikov VV, Sergeev VV. Method of automatic coregistration of digital remote sensing images from different sources. Computer Optics 2024; 48(6): 932-943. DOI: 10.18287/2412-6179-CO-1604.

Введение

По мере роста количества спутников дистанционного зондирования Земли (ДЗЗ) задача совмещения цифровых изображений, полученных от сенсоров, использующих различные физические принципы регистрации информации, снимков разного пространственного и/или спектрального разрешения, снятых в разные моменты времени, сегодня становится все более актуальной. Исторически задача совмещения цифровых оптических изображений, называемая также задачей *регистрации*, часто возникала в приложениях фотограмметрии и смежных направлениях технического зрения [1]: нахождение соответствующих точек на различных изображениях одного и того же объекта является ключевым шагом для построения/реконструкции 3D-модели регистрируемого объекта. Развитие цифровых средств регистрации изображений различной физической природы (радарные, мульти- и гипер-спектральные, видео), развитие персональных мобильных устройств и платформ размещения средств регистрации, высокопроизводительных средств обработки данных привело к взрывному росту решений в области компьютерного зрения в целом и методов регистрации цифровых изображений в частности.

Один из возможных и часто используемых способов сопоставления изображений состоит из нескольких этапов, включающих (для упрощения рассматриваем задачу для двух изображений): поиск особых точек на изображениях, выделение из этих двух множеств некоторого множества пар соответствующих точек (точка на одном изображении соответствует точке на другом изображении) и оценку параметров преобразования координат между исходными изображениями. Для выполнения каждого из этапов на настоящий момент существует целый ряд решений/алгоритмов, основанных как на классических подходах компьютерного зрения [2], так и на нейросетевых [3]. Возможные постановки задачи и способы решений также могут варьироваться от описанного типового (соответствие по небольшому набору точек) до вариантов “dense”, когда соответствие находится между всеми возможными точками изображений. То есть в целом можно говорить о наличии значительного разнообразия постановок задачи регистрации и ее возможных решений.

Применительно к задаче регистрации/сопоставления разнородных/разновременных цифровых изображений ДЗЗ выделяются две существенные особенности. Во-первых, наличие принципиально различных по физи-

ческим принципам средств регистрации, данных разного пространственного и спектрального разрешения, а также одновременности съемки. Это существенно усложняет процесс поиска пар особых точек и последующего совмещения. Большинство известных алгоритмов обнаружения и сопоставления особых точек для снимков, полученных с различных аппаратов ДЗЗ, оказываются неработоспособны. С другой стороны, в отличие от классических цифровых фотографий, цифровые снимки ДЗЗ имеют дополнительную информацию об условиях регистрации, включающую и координаты снимка, называемые его *геопривязкой*. Типовым вариантом является указание координат для угловых и центральной точек снимка. Эти координаты с определенной погрешностью, зависящей от условий съемки, рельефа Земли в месте съемки и пространственного разрешения цифрового снимка, обеспечивают т.н. «привязку» снимка к земной поверхности. Хотя точность геопривязки снимка для многих космических аппаратов (КА) относительно высокая, простое «наложение» двух снимков, полученных с различных аппаратов в различное время, демонстрирует довольно значительное их рассогласование. Подобное расхождение взаимного положения снимков значительно влияет на последующий анализ этих данных, в частности, на результаты их совместной обработки (англ. fusion – слияние). Указанная проблема показывает целесообразность взаимного совмещения цифровых изображений ДЗЗ, даже несмотря на наличие их координатной «привязки».

С учетом первой из указанных выше особенностей решения задачи совмещения цифровых изображений ДЗЗ, ее решение на практике производится специализированными (т.е. отличными от классических) методами и, как правило, адаптированными под конкретные источники данных. В частности, для описания особых точек используются специальным образом модернизированные алгоритмы: для РСА-изображений ДЗЗ (РСА – радар с синтезированной апертурой, англ.: SAR – synthetic aperture radar) используют SAR-SIFT [4] или PSO-SIFT[5], дескриптор/детектор OS-SIFT [6] используется для поиска точек соответствия между оптическими и РСА-данными одного разрешения и др. Также из-за существенных различий в данных разных сенсоров при работе с изображениями ДЗЗ все большее распространение приобретают детекторы особых точек, построенные на основе нейронных сетей: LoFTR [7], R2D2 и Superpoint [8] и др. Их основным преимуществом является возможность обучения под конкретный тип изображения, однако предсказуемость результата при замене сенсора и/или параметров его функционирования остается недостаточно очевидной. Недостатком является и необходимость значительных вычислительных затрат при обучении детектора.

Обобщая изложенное, можно заключить, что единый/универсальный подход, хорошо работающий для

разнородных цифровых изображений ДЗЗ (полученных с различных платформ, сенсоров, в разное время и с различным пространственным и/или спектральным разрешением), авторам настоящей работы не известен.

В представленной работе мы предлагаем единый метод, основанный на ряде существующих решений, их улучшений и модификаций и позволяющий сформировать объединенное согласованное по координатам покрытие территории разнородными цифровыми изображениями ДЗЗ. Метод основан на приведении всех входных данных к единому «формату», для которого настраивается/обучается одна из state-of-the-art (нейросетевых) моделей совмещения изображений. Новыми элементами метода является способ приведения к единому «формату», а также способ дополнительной фильтрации пар точек на основе априорной информации и снимка. Ключевым моментом также является выбор алгоритма совмещения из первоначального опорного множества доступных для сравнения (в коде) алгоритмов. Построение предлагаемого метода осуществляется в рамках подхода, включающего целый ряд этапов, детально изложенных во втором параграфе.

Работа организована следующим образом. В первом параграфе представлен обзор существующих методов и алгоритмов совмещения как обычных оптических снимков ДЗЗ, так и РСА-данных дистанционного зондирования. Во втором параграфе представлен метод совмещения разнородных цифровых изображений ДЗЗ и последовательное описание этапов его построения. В третьем параграфе дано экспериментальное сравнение предложенного метода и доступных реализаций иных современных решений. Завершает работу заключение и список использованной литературы.

1. Современное состояние исследований

1.1. Методы совмещения в компьютерном зрении

Как отмечалось выше, задача совмещения цифровых оптических изображений исторически возникала в приложениях фотограмметрии и смежных направлениях технического зрения [1] и рассматривалась вне контекста обработки цифровых изображений ДЗЗ. Первые методы совмещения изображений были основаны на максимизации некоторой метрики, например корреляции или взаимной информации [9]. В этом случае сопоставление изображений приобретает характер оптимизационной задачи, что с учетом изменения характеристик изображения на каждом шаге требует больших вычислительных ресурсов.

Более современным подходом к решению задачи совмещения изображений является совмещение на основе сопоставления особых точек. Особые точки обычно расположены на визуальных особенностях изображения: границах, углах, центрах замкнутых контуров и т.д. В этом случае задача сопоставления изображений состоит из этапов обнаружения особых

точек (выполняется т.н. *детекторами*) или т.н. точек интереса (англ.: RoI – Region of Interest), построения описания их локальной окрестности (выполняется т.н. *дескрипторами*), последующего сопоставления описаний особых точек (т.н. *дескрипторов* – N -мерных векторов признаков) и формирования их пар на основе ограничений, возникающих из специфики геометрического преобразования между изображениями. Последний этап обычно выполняется с использованием алгоритмов RANSAC [10] или PROSAC [11]. При сравнении дескрипторов особых точек для формирования их пар часто используется стандартное евклидово расстояние, хотя могут быть использованы и иные метрики/меры.

Наиболее известным алгоритмом *описания особых точек* является запатентованный алгоритм SIFT [12], породивший ряд модификаций [13, 14]. Более новые алгоритмы данного типа, например, ORB [15] и SURF [16], обещают лучшее качество сопоставления изображений, однако SIFT до настоящего времени остается наиболее распространенным алгоритмом.

Примером алгоритма *поиска особых точек* является детектор углов FAST [17], который использует дерево решений, построенное на обучающей выборке, для улучшения обнаружения углов на целевом изображении.

Минусом многих детекторов и дескрипторов является отсутствие адаптивности – алгоритмы расчета, используемые для поиска и/или описания особых точек, детерминированы и фиксированы. Это порождает проблемы при использовании их на изображениях, имеющих существенные различия – особенно в случае изображений из различных доменов (например, оптических изображений и карт глубины). Проблема адаптации алгоритмов обнаружения и описания особых точек на изображениях разных доменов решается с помощью методов машинного обучения, включающих и нейронные сети.

Многие современные алгоритмы сопоставления изображений, в том числе с использованием особых точек, основаны на искусственных нейронных сетях (ИНС). Примерами являются алгоритмы LIFT [18], R2D2 [19], Superpoint [20], DISK [21], ALIKED [22]. В зависимости от алгоритма для поиска особых точек и расчета дескрипторов может быть использована как одна ИНС, так и несколько отдельных ИНС. Во многих случаях (особенно в случае использования сверточных ИНС – СНС) результатом работы алгоритма являются тензор дескрипторов, в котором содержатся дескриптор для каждого пикселя или группы пикселей, и маска ключевых точек, отмечающая особые точки. Отсюда следует, что может быть выполнено попиксельное сопоставление изображений, что подводит нас к следующему (бездетекторному) типу алгоритмов сопоставления изображений.

Бездетекторные (англ.: detector-free) методы сопоставления изображений выполняют одновременно

обнаружение и сопоставление особых точек на изображении (т.е. результатом работы является набор пар точек). Обычно на одном из этапов работы метода производится расчет полей признаков для сопоставляемых изображений, за которым следует их сопоставление с последующим подавлением немаксимумов (англ.: non-maximum suppression). Примерами алгоритмов подобного типа являются LoFTR [23], NCNet [24].

Наиболее современным подходом является т.н. *попиксельное сопоставление* (англ.: dense matching), при котором результатом работы алгоритма является карта соответствий между пикселями первого и второго изображений. Такой подход потенциально позволяет строить соответствия между изображениями в условиях сильных искажений. Недостатком такого подхода является высокая вычислительная сложность. Однако, в отличие от классического оптимизационного подхода, алгоритмы этого типа отрабатывают за один шаг. Примером алгоритма из данного класса, не использующего машинное обучение, является SIFTFlow [25]. Примерами методов попиксельного сопоставления изображения являются DKM [26], RoMa [27]. Метод LoFTR из предыдущей категории на одном из этапов выполняет попиксельное сопоставление на большом масштабе, за которым следует определение точных координат точек.

Стоит отметить, что если поля двух изображений связаны между собой аффинным преобразованием, то использование попиксельного преобразования является избыточным и может приводить к обнаружению ложных соответствий. Подобная проблема может быть относительно просто решена путем отбора некоторого достаточного (для оценки параметров преобразования) количества наиболее качественных пар точек.

1.2. Методы совмещения космических снимков ДЗЗ

В рамках данного подпараграфа рассматриваются существующие методы совмещения снимков, прошедших радиометрическую коррекцию, геопривязку и ортокоррекцию.

Совмещение оптических снимков

Задача совмещения оптических снимков, полученных от одного источника, во многом аналогична задаче совмещения обычных изображений и при малом числе каналов снимка может решаться путем использования рассмотренных выше подходов.

Большую сложность представляет задача совмещения оптических снимков, полученных от разных источников в силу различий в чувствительности сенсоров, пространственном и/или спектральном разрешении. Среди методов сопоставления изображений, не использующих ключевые точки, можно отметить сопоставление на основе корреляции [28], взаимной информации [29] и нелинейного отображения тонов [30].

Помимо универсальных методов, в работах [31,32] рассматриваются методы сопоставления снимков от конкретных спутников. В данных работах для сопоставления снимков активно используются дополнительные данные (данные геопривязки, данные о положении детектора при съемке).

Наборы данных, используемые в задачах оценки методов совмещения оптических снимков, зачастую собираются авторами самостоятельно, что может затруднять последующее сравнение алгоритмов. Во многих случаях используются наборы данных, предназначенных для задач совместного анализа снимков. Примером могут быть наборы, предлагаемые в ежегодных соревнованиях IEEE GRSS Data Fusion Contest. В частности, набор данных, использовавшийся в 2021 IEEE GRSS Data Fusion Contest [33] включает данные от спутников:

- Sentinel-2A – мультиспектральные оптические снимки, 11 полос, исходное разрешение – 15 м;
- Landsat-8 – мультиспектральные оптические снимки, 12 спектральных полос, исходное разрешение – 30 м, для панхроматического канала – 15 м;
- Suomi NPP (полоса DNB датчика VIIRS, исходное разрешение – 500 м).

Другим примером может служить набор [34], включающий в себя пары снимков Landsat-8 и Terra MODIS (6 спектрально близких полос).

Совмещение оптических и радарных снимков

Совмещение радарных и оптических снимков представляет собой отдельную задачу в силу наличия спекл-шума на PCA-изображениях. Специфика шума не позволяет автоматически перенести методы и алгоритмы, используемые для работы с оптическими изображениями, на радарные изображения.

Так же как в рассмотренной выше задаче сопоставления изображений в компьютерном зрении, первые алгоритмы сопоставления пар оптических и PCA-изображений ДЗЗ опирались на *оптимизацию некоторой глобальной метрики*, например, взаимную информацию [35] или нормализованную корреляцию [36].

Отдельно стоит отметить методы совмещения указанных изображений, которые используют синтез псевдорадарного изображения из оптического для приведения обоих изображений к одному «типу» [37]. В этом случае задача сопоставления разнородных изображений ДЗЗ может рассматриваться как задача сопоставления однородных.

Для сопоставления пар оптических и PCA-изображений ДЗЗ также используются и *детекторы/дескрипторы особых точек*. Примерами являются SIFT-подобные детекторы PSO-SIFT [5], OS-SIFT [6].

В отдельную категорию следует выделить метод RIFT [38], поскольку он предназначен для *сопоставления разнотипных изображений* и потому представляет особый интерес. Другим интересным примером

является дескриптор CFOG [39], который также используется для сопоставления разнотипных изображений. Поскольку CFOG является только дескриптором, используемым для сопоставления особых точек, он должен использоваться в паре с алгоритмом поиска особых точек (в исходной работе используется детектор углов Харриса [40]).

Для сопоставления пар оптических и PCA-изображений ДЗЗ также используются и *ИНС*, например, в работе [8] приводятся результаты применения некоторых нейросетевых методов сопоставления изображений для задачи сопоставления радарных и оптических снимков.

Примерами *наборов данных*, которые могут быть использованы в задаче сопоставления пар оптических и радарных изображений ДЗЗ, являются:

- SEN1-2 [41], где сопоставляются цветные оптические и одноканальные радарные снимки;
- SEN12MS [42], где сопоставляются мультиспектральные оптические и двухканальные радарные снимки,
- OSDataset [43], где сопоставляются оптические и радарные изображения в оттенках серого,
- набор 2023 IEEE GRSS Data Fusion Contest [33], где сопоставляются цветные оптические и одноканальные радарные снимки.

Среди указанных наборов только OSDataset декларирует обучение сопоставлению оптических и радарных снимков в качестве основного назначения.

Совмещение радарных снимков

Задача совмещения радарных снимков от разных источников авторам не встречалась. Однако задача совмещения радарных снимков одного и того же космического аппарата и сенсора, вообще говоря, возникает как задача интерферометрического анализа PCA-снимков – InSAR [44], однако в данном случае совмещаются снимки одного и того же сенсора, что существенно упрощает задачу.

Метод SAR-SIFT [4] предназначен только для сопоставления радарных изображений. Методы OS-SIFT, PSO-SIFT и RIFT, появившиеся позже и используемые для сопоставления разнородных изображений, способны сопоставлять PCA-снимки.

Публичных наборов данных по задаче совмещения PCA-снимков от разных источников не найдено.

Разновременные радарные снимки могут быть найдены в ежегодных соревнованиях IEEE GRSS Data Fusion Contest. В частности, набор данных, использовавшийся в 2021. IEEE GRSS Data Fusion Contest [33] включает данные от спутника Sentinel-1 – радарные снимки с VV- и VH-поляризацией.

2. Метод совмещения разнородных данных ДЗЗ

2.1. Общее описание метода

Предлагаемый нами единый метод сопоставления пар разнородных изображений ДЗЗ представлен на

рис. 1. Входными данными для метода являются пары изображений ДЗЗ, отличающиеся по разрешению не более чем в 16 раз, и априорные данные о съемке в виде первичной геопривязки этих изображений. Для определенности будем считать, что менее детальное изображение на схеме – это второе изображение. При отличии разрешений изображений более чем в 4 раза ко второму изображению применяется операция искусственного повышения разрешения одним из предопределенных методов (для упрощения рассмотрения, мы данную операцию пока опускаем).

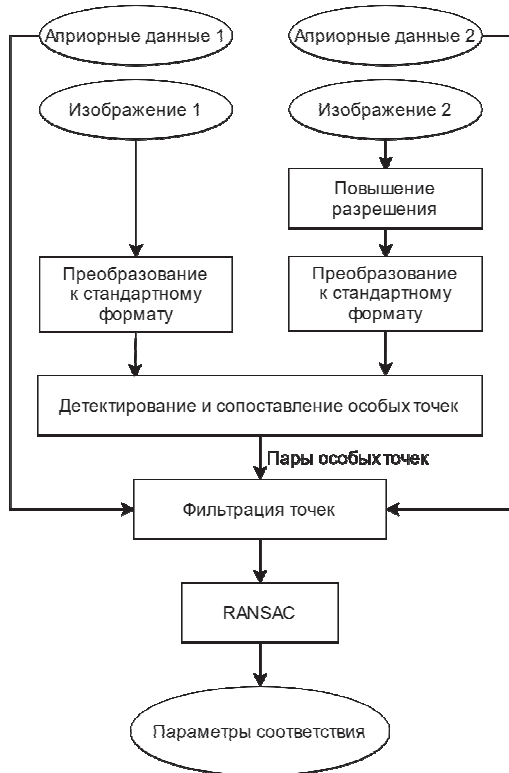


Рис. 1. Схема предлагаемого метода совмещения

Первым общим этапом преобразования изображения является приведение их к стандартному виду. Под этим в большинстве случаев понимается понижение числа каналов изображений до типового количества – трех (в случае одноканальных изображений канал просто дублируется еще два раза для получения трехканального итогового изображения).

На втором этапе по двум «стандартизированным» изображениям некоторым «оптимальным» алгоритмом находятся особые точки и делается их описание. Их множества передаются на третий этап обработки.

На третьем этапе метода происходит фильтрация точек и формирование сокращенных подмножеств на основе априорных данных о геопривязке изображений.

Наконец, на последнем шаге происходит окончательный отбор пар точек соответствия с использованием алгоритма RANSAC и определение параметров преобразования перспективы между парами изображений.

Более формальное описание можно выполнить следующим образом. Пусть даны пара изображений

$I_1, I_2 \in \mathbb{R}^{W \times H \times K}$, методы предобработки изображений L_1, L_2 , метод сопоставления изображений M и метод построения матрицы проективного преобразования R .

Результатом применения метода M к результатам предобработки I_1, I_2 является набор пар точек P_0 :

$$P_0 = \{(p_1, p_2)_i \mid p_1, p_2 \in \mathbb{R}^2\}_{i=1}^N = M(L_1(I_1), L_2(I_2)), \quad (1)$$

где p_0, p_1 – координаты особых точек 1 и 2 изображения соответственно.

Методы сопоставления изображений на основе ключевых точек могут выдавать ложные соответствия точек, при этом без дополнительной информации нельзя определить, какие из соответствий являются правильными. При наличии дополнительных данных можно заранее отсеивать ложные соответствия точек.

Пусть $P_1 \subseteq P_0$ – множество пар точек, отфильтрованных с учетом дополнительной информации (в случае, если фильтрация не применяется $P_1 = P_0$). Тогда итоговая матрица преобразования, связывающего изображения H , может быть получена применением R к P_1 .

$$H, P_2 = R(P_1), \quad (2)$$

где P_2 – множество пар точек, которые были признаны корректными и были учтены при построении матрицы H .

Новыми элементами предложенного метода являются:

- способ приведения изображений ДЗЗ к единому «формату»,
- способ дополнительной фильтрации пар точек на основе априорной информации о снимке;
- выбор и настройка алгоритма детектирования для поиска точек соответствия и их описания;
- способ совмещения пар точек на основе предварительной их фильтрации по данным геопривязки с учетом пространственного разрешения снимков конкретной модальности и априорной точности их геопозиционирования.

Каждый этап предложенного метода мы конкретизируем (то есть определяем/строим/обучаем соответствующий алгоритм) в результате решения некоторой подзадачи, как правило, оптимизационной. Более конкретно, отдельные составляющие метода получаются в результате последовательного выполнения следующих этапов.

- 1) Подготовка исходного (далее называем – *опорного*) множества алгоритмов совмещения цифровых изображений, включающего различные (специализированные, универсальные) решения для попарного совмещения цифровых снимков с сопоставимым разрешением (под сопоставимым разрешением будем понимать разрешения изображений, отличающиеся не более чем в 4 раза). Учитывая, что существует два принципиально различных типа снимков ДЗЗ – оптические (обо-

значим V : монохроматические, цветные, мультиспектральные, гиперспектральные) и радарные (обозначим R), опорное множество алгоритмов должны обеспечивать решение задачи сопоставления как минимум для трех вариантов данных: VV , RV/VR и RR ;

2) выбор или построение алгоритма уменьшения числа каналов изображений ДЗЗ;

3) сокращение опорного множества алгоритмов за счет выбора тех, которые допускают работу с малоканальными (трехканальными и одноканальными) изображениями;

4) определение достижимых параметров точности решения задачи совмещения с использованием выбранных алгоритмов. Задача рассматривается для случая использования априорной информации в виде данных о съемке, то есть координатной привязки космических снимков. Таким образом, на этом этапе происходит *доработка алгоритмов* из выбранного множества для учета ими указанной информации;

5) обучение модифицированных алгоритмов сопоставления на подготовленной базе данных пар изображений базы данных. *Наилучший алгоритм совмещения* (включая алгоритмы поиска особых точек, их описания и нахождения пар) определяется в результате предварительного экспериментального исследования как наилучший среди доступных на подготовленной нами базе данных (тестах);

6) сопоставление результатов совмещения снимков, достижимых с помощью предложенного решения, с существующими частными решениями, доступными в виде кода.

Ниже в этом параграфе даны комментарии относительно выполнения этапов 1, 2, 4, 5. Этап 3 тривиален и не описывается. Этап 6 представлен в третьем параграфе настоящей работы. Отдельным не рассмотренным выше этапом является формирование базы данных изображений, на которых будут сопоставляться алгоритмы. В настоящей работе конкретизация этого этапа представлена также в третьем параграфе (п.3.1).

2.2. Подготовка опорного множества алгоритмов сопоставления изображений

В качестве опорного множества алгоритмов сопоставления изображений, среди которых определяется наилучший, использовались следующие: SIFT, SAR-SIFT, RIFT и RoMa. Дадим их краткую характеристику.

SIFT [12] является широко известным запатентованным методом сопоставления изображений. Он работает с полутоновыми (одноканальными) изображениями, использование его для многих каналов требует сведения изображения к полутоновому.

SAR-SIFT [4] является модификацией метода SIFT, предназначенной для сопоставления PCA-

изображений. Аналогично SIFT, SAR-SIFT работает с полутоновыми (одноканальными) изображениями.

RIFT (Radiation-Invariant Feature Transform) [38] – метод сопоставления разнородных полноцветных (RGB, трехканальных) изображений. В оригинальной работе авторы сопоставляют цветные оптические изображения с радарными изображениями, картами глубины и даже просто картами местности, соответствующей снимку. Таким образом, на первоначальном этапе отбора метод RIFT выглядел достаточно перспективно в силу своей универсальности.

Три приведенных метода используют типовой подход к сопоставлению кадров – на основе поиска и описания особых точек – и не используют обучение или ИНС. Последний выбранный в опорное множество метод – RoMa – относится к иной категории.

RoMa (Robust Matcher) [27] – метод попиксельного сопоставления на основе ИНС – в настоящий момент является state-of-the-art методом сопоставления RGB-изображений. В оригинальной работе использовался унимодальный (не универсальный) метод, но представлял интерес в силу топовых качественных показателей на одной конкретной модальности снимков. Метод RoMa для пары сопоставляемых изображений $I_1, I_2 \in \mathbb{R}^{W \times H \times 3}$ возвращает два тензора. Первый тензор $P \in \mathbb{R}^{W \times H \times 4}$ задает попиксельное (dense) соответствие между изображениями, второй тензор $P \in \mathbb{R}^{W \times H}$ – меру качества соответствия отдельных пикселей. Поскольку попиксельное сопоставление в случае совмещения снимков, прошедших ортокоррекцию, нам в данной работе представлялось избыточным, мы ограничили результаты работы метода подмножеством пар точек, имеющих наибольшую/наилучшую меру качества соответствия. На основе полученного подмножества настраивалось итоговое геометрическое преобразование между кадрами.

Напомним, что поскольку в работе рассматриваются два типа модальностей снимков (цветные, радарные), то работоспособность алгоритмов сопоставления должна обеспечиваться в трех комбинациях входных данных: VV , VR , RR . Как следствие, обучаемые детекторы/дескрипторы должны быть обучены для конкретной комбинации входов. Процесс обучения и его результаты представлены в параграфе 3.2 экспериментальной части работы. Мы обучали модель ИНС RoMac для сопоставления мультиспектральных оптических снимков (VV), оптических и PCA-снимков (VR) и пар PCA-снимков (RR).

2.3. Уменьшение числа каналов изображений ДЗЗ

Алгоритмы опорного множества при сопоставлении изображений работают с одним (SIFT, SAR-SIFT) или тремя (RIFT, RoMa) каналами. Это означает невозможность их прямого использования для работы с мульти- и гиперспектральными снимками. Доработка самих алгоритмов для работы с многоканальными изображениями представляет собой отдельную слож-

ную задачу. Кроме того, подобная доработка не будет универсальной – менять алгоритмы придется для разного числа каналов.

Альтернативным подходом, используемым в некоторых работах по сопоставлению снимков ДЗЗ с различными спектральными каналами [32], является выбор каналов, попадающих в пересечение регистрируемых спектральных диапазонов. Однако данный подход также не универсален: для различных пар снимков число каналов для каждого снимка может оказаться индивидуальным, что возвращает нас к первой проблеме (для каждого числа каналов свой алгоритм). Дополнительно возникает проблема различной спектральной «ширины» каналов различных снимков (например, канал гиперспектрального изображения может иметь ширину 5 нм, а некоторого мультиспектрального – десятки нанометров). Поэтому такой «упрощенный» подход также представляется нам малоэффективным и неперспективным.

Наиболее удачным решением указанной проблемы, на наш взгляд, является совместное снижение размерности /канальности сопоставляемых цифровых изображений до заданного числа каналов. Это заданное число каналов может быть выбрано, исходя из возможностей метода поиска и описания особых точек. В нашем случае мы ориентируемся на снижение числа каналов до одного или трех каналов соответственно. Для этого мы используем известный *метод главных компонент* [45], или PCA (англ.: Principle Component Analysis), применение которого будет иметь специфику в зависимости от сопоставляемых изображений. Кратко опишем используемый метод.

Пусть $X = \{x_i | x_i \in \mathbb{R}^n\}_{i=1}^N$ – набор точек в n -мерном пространстве, Σ – ковариационная матрица X . В задаче понижения размерности снимков ДЗЗ в роли X выступает множество векторов – (спектральных) профилей «яркости» пикселей изображения, полученного от конкретного источника.

Далее пусть $\lambda_i, a_i, i = \overline{1, n}$ – собственные значения и соответствующие им собственные вектора матрицы Σ , расположенные в порядке убывания собственных значений ($\lambda_i > \lambda_{i+1}$). Матрица $A = (a_1, \dots, a_n)^T$ задает переход к ортогональному базису $\{a_i\}_{i=1}^n$. С учетом того, что при вычислении Σ данные центрируются, вычисление представления исходного вектора x в новом базисе вычисляется по формуле:

$$PCA(x) = A(x - \bar{X}). \tag{3}$$

На множестве $Y = \{y_i = PCA(x_i) | x_i \in X\}$ дисперсия компонент $Y_k = \{y^k | y \in Y\}$ убывает с ростом k ($D(Y_k) \geq D(Y_{k+1})$). Таким образом, оставляя только K первых компонент, можно уменьшить размерность данных с наименьшими потерями. В случае многоканальных снимков данных подход позволяет свести изображения к трем или одному каналу.

Специфика применения метода для ситуаций со снимками одинаковых и разных доменов различается. Принципиально различных ситуаций три:

- сопоставляются снимки одного и того же домена (то есть полноцветные с одного сенсора или сенсоров с одинаковым спектральным диапазоном, мультиспектральные с одинаковым спектральным диапазоном и числом каналов и т.п.). В этом случае базис PCA рассчитывается по объединенному набору сопоставляемых снимков;
- сопоставляются снимки (спектральные), которые отличаются регистрируемым спектральным диапазоном. В этом случае число каналов для каждого типа снимка предварительно ограничивается до диапазона пересечения их спектральных диапазонов. При нехватке каналов используются дополнительно смежные с выбранными каналами. Получаемые снимки с «усеченным» числом каналов используются в методе PCA, который применяется по отдельности для каждого типа снимков;
- сопоставляются снимки разных доменов (например, радарные и оптические либо спектральные с непересекающимися спектральными диапазонами). В этом случае метод PCA применяется независимо к каждому из доменов.

Обратим внимание: базисы PCA, полученные для снимков разных модальностей, оказываются неизменяемыми. В результате, для каждого типа данных ДЗЗ (а иногда для пары типов) необходимо формировать собственный базис главных компонент.

2.4. Фильтрация особых точек на основе априорных данных о геопривязке снимков

Как уже было отмечено выше, для изображений ДЗЗ доступны данные геопривязки, которые с некоторой погрешностью позволяют определить реальные координаты расположения изображения в заданной геодезической системе координат. Мы предлагаем при формировании множеств пар особых точек использовать данные геопривязки снимков для предварительного отсева пар, физически не допускающих такого сопоставления.

Более формально, пусть P_1 – множество пар соответствующих особых точек, $T_1, T_2 \in \mathbb{R}^{33}$ – матрицы преобразования между пиксельными и мировыми координатами соответственно. Тогда процесс фильтрации пар точек на основе геопривязки может быть описан следующим образом.

Пусть $h(p)$ – оператор приведения пиксельных координат к гомогенным координатам, а $h^{-1}(p)$ – оператор приведения гомогенных координат к пиксельным:

$$h(p) = (x, y, 1 | p = (x, y))^T, \tag{4}$$

$$h^{-1}(p) = \left(\frac{x}{w}, \frac{y}{w} | p = (x, y, w) \right)^T. \tag{5}$$

Оператор проецирования $\text{Pr}(T, p)$ задает проективное преобразование:

$$\text{Pr}(T, p) = h^{-1}(Th(p)), \quad (6)$$

где $T \in \mathbb{R}^{33}$ – матрица, задающая проективное преобразование, $p \in \mathbb{R}^2$ – координаты точки.

В данных обозначениях набор точек $P_1 \subseteq P_0$, отфильтрованных с учетом данных геопривязки, может быть определен следующим образом:

$$d(p_1, p_2) = \text{Pr}(T_1, p_1) - \text{Pr}(T_2, p_2)_2, \\ P_1 = \{(p_1, p_2) | (p_1, p_2) \in P_0, d(p_1, p_2) \leq \delta\}, \quad (7)$$

где δ – параметр алгоритма, задающий максимально отклонение между точками в реальных координатах. Данный параметр задается априорно, исходя из данных о точности геопривязки конкретных типов/доменов снимков. А именно, если δ_1 – погрешность координат первого типа снимков, а δ_2 – второго, то полагаем $\delta = \delta_1 + \delta_2$.

Необходимо заметить, что учет данных геопривязки при решении задачи сопоставления снимков ДЗЗ ранее уже рассматривался в работах [32, 46] других авторов. Но в этих работах совмещение выполнялось для конкретных пар сенсоров. В предлагаемом варианте представлен алгоритм общего вида для снимков произвольных типов/доменов.

3. Экспериментальные исследования

3.1. Наборы данных

Для обучения и сравнения предложенного метода совмещения разнородных данных ДЗЗ с существующими лучшими решениями использовался набор данных 2021 IEEE GRSS Data Fusion Contest [33]. Набор данных состоит из 98 локаций, каждой из которых соответствуют по 4 снимка Sentinel-1 (PCA-изображения), 4 снимка Sentinel-2 (мультиспектральные изображения), 3 снимка Landsat-8 (мультиспектральные изображения) и 9 снимков VIIRS (мультиспектральные/ИК-изображения), снятые в июле-августе 2020 г. Все снимки совмещены и приведены к единому пиксельному (800×800) и пространственному (10 м/пиксель) разрешению.

Из набора данных были исключены снимки со значительной облачностью. Итоговый обучающий набор состоял из 720 пар снимков Sentinel-2 и Landsat-8, 960 пар снимков Sentinel-2 и Sentinel-1 и 720 пар снимков Sentinel-1, базовый тестовый набор – из 171 пары снимков Sentinel-2 и Landsat-8, 288 пар снимков Sentinel-2 и Sentinel-1 и 228 пар снимков Sentinel-1.

Расчет матриц понижения размерности для снимков Sentinel-2 и Landsat-8 производится по всем пикселям снимков, полученных от соответствующего сенсора.

Поскольку модель RoMa требовала, чтобы яркость пикселей, подаваемых на вход снимков, находилась в диапазоне $[0, 1]$ (что неверно ни для оптических, ни для радарных снимков данного набора), по обучающему набору данных рассчитывались 0,1- и 99,1-процентили значений яркости, которые принимались за минимум и максимум и затем использовались для приведения изображения к итоговому диапазону значений яркости. Примеры итоговых изображений показаны на рис. 2.

Поскольку размеры изображений 800×800 не являлись подходящими для модели RoMa, снимки из базового тестового набора разбивались на пять фрагментов по 448×448 пикселей (четыре фрагмента брались по углам, один – по центру снимка). Итоговый тестовый набор состоял из 855 пар фрагментов снимков Sentinel-2 и Landsat-8, 1440 пар фрагментов снимков Sentinel-2 и Sentinel-1 и 1140 пар фрагментов снимков Sentinel-1.

3.2. Настройка и обучение метода

В предложенном методе два настраиваемых компонента: матрицы понижения размерности (см. п.2.3) и параметры алгоритма RoMa (см. п.2.2). Учитывая тривиальность первой задачи, мы не приводим здесь ее численных результатов.

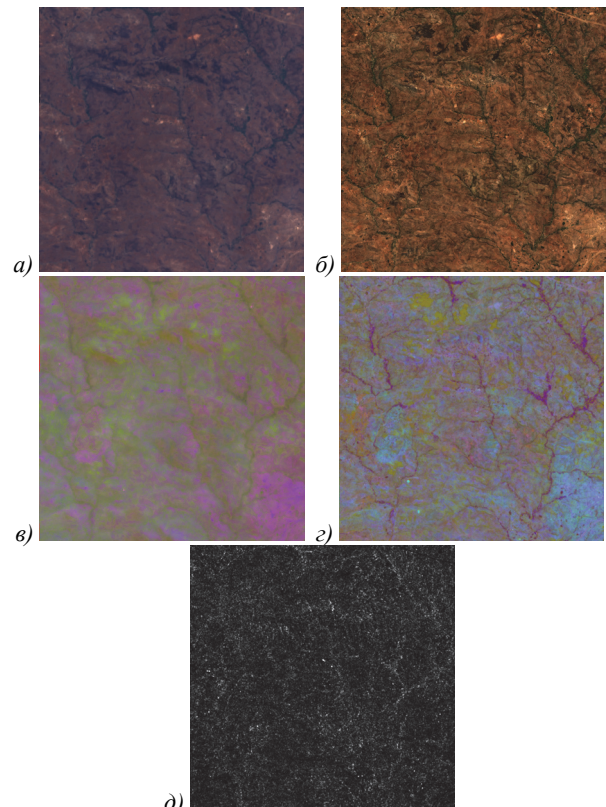


Рис. 2. Примеры изображений: а) RGB-каналы Landsat-8, б) RGB-каналы Sentinel-2, в) 3 главные компоненты Landsat-8, г) 3 главные компоненты Sentinel-2, д) VV-канал Sentinel-1

Обучение метода RoMa производилось на наборах данных из параграфа 3.1. Потенциально RoMa позволяет использовать при обучении любые размеры изображения по горизонтали/вертикали, кратные числу «14». Из соображений баланса потребления вычислительных ресурсов и размера изображения в рамках обучения и последующих экспериментов мы использовали размер входного изображения 448 × 448 пикселей. Поскольку этот размер отличался от размера реальных снимков, при обучении входные данные обучения случайным образом вырезались из исходных снимков с сохранением их пересечения (центр одного изображения должен быть виден на другом изображении). Аугментации, использованные при обучении, включали:

- случайный сдвиг на 64 пикселя;
- случайное масштабирование с коэффициентом от 0,9 до 1,1;
- поворот на угол до 30°.

Обучение проводилось на 150 000 пар изображений, использовался ПК с ЦП AMD Ryzen 5600X и 32 ГБ ОЗУ, снабженный графическим ускорителем Nvidia GeForce RTX 3090. Результаты обучения показаны в п.3.4 ниже.

3.3. Показатели качества

В качестве показателей качества задачи совмещения изображений, с учетом сделанных в параграфе 2.1 обозначений, выступали следующие величины.

Средняя ошибка отображения углов – ACE (англ.: average corner error). ACE рассчитывалась следующим образом (изображения из тестовых наборов имеют одинаковое разрешение и ориентацию):

$$C = \{(0,0), (0,H), (W,0), (W,H)\},$$

$$ACE = \frac{1}{4} \sum_{c \in C} c - \Pr(H, p_2)_2, \tag{8}$$

где W, H – размеры изображений по горизонтали и вертикали соответственно.

Совмещенными считаются изображения, для которых $ACE < 100$. Относительно общего количества изображений рассчитывается *доля совмещенных изображений*.

Доля корректных пар:

$$r = \frac{|P_2|}{|P_0|}. \tag{9}$$

Средняя ошибка локализации особых точек (в пиксельных координатах):

$$\varepsilon = \sum_{(p_1, p_2) \in P_2} \frac{p_1 - \Pr(H, p_2)}{|P_2|}. \tag{10}$$

3.4. Результаты

Параметры эксперимента:

- максимальное количество ключевых точек на изображении – 4096;
- максимальное отклонение физических координат, δ – 20 м.

Ниже представлены результаты экспериментов по совмещению космических снимков для двух ситуаций: без использования данных геопривязки (т.н. *базовый вариант*, табл. 1–4) и с их использованием (табл. 5–8). Заметим, что для космических изображений различного типа (в частности, сопоставления VR) возможна потеря работоспособности некоторых алгоритмов – в этом случае в ячейках указаны прочерки и/или нули. Лучшие числовые результаты выделены в таблицах жирным шрифтом.

На основании представленных данных можно сделать следующие выводы:

- 1 Широко известный алгоритм SIFT и специализированный SAR-SIFT оказались неработоспособными при сопоставлении разнородных данных. RIFT оказывается почти неработоспособным (см. табл. 1). Алгоритм RoMa – единственный, поддерживающий работу с разнотипными изображениями.
- 2 Метод RoMa обладает лучшими качественными показателями в базовом варианте исследования (без данных о геопривязке).
- 3 Использование данных геопривязки позволяет существенным образом повысить точность сопоставления снимков для всех рассмотренных алгоритмов (табл. таблицы 2–6, 3–7). Снижение средней доли корректных пар вызвано снижением общего количества сопоставляемых точек в результате операции фильтрации точек (снижение знаменателя).

Общим выводом по результатам экспериментальных исследований является преимущество модели RoMa. Фактические показатели предложенного в данной работе метода соответствуют строке «RoMa» в табл. 5–8.

Табл. 1. Доля совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	0,91	0	0,73
SAR-SIFT	0,795	0	1,0
RIFT	0,989	0,053	0,89
RoMA	1,0	0,997	1,0

Табл. 2. Среднее ACE совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	2,78	-	8,06
SAR-SIFT	5,91	-	4,15
RIFT	3,52	49,89	8,91
RoMA	2,56	12,65	0,24

Табл. 3. Средняя доля корректных пар для совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	0,048	-	0,004
SAR-SIFT	0,096	-	0,037
RIFT	0,081	0,003	0,039
RoMA	0,757	0,351	0,923

Табл. 4. Средняя ошибка локализации точек для совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	2,39	-	1,13
SAR-SIFT	2,80	-	1,74
RIFT	2,78	8,68	1,62
RoMA	1,89	6,37	0,567

Табл. 5. Доля совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	0,87	-	0,78
SAR-SIFT	0,74	-	0,992
RIFT	1,0	0,520	0,999
RoMA	1,0	0,992	1,0

Табл. 6. Среднее АСЕ совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	$6,1 \times 10^{-10}$	-	$7,6 \times 10^{-7}$
SAR-SIFT	$2,1 \times 10^{-9}$	-	$4,7 \times 10^{-13}$
RIFT	$3,6 \times 10^{-13}$	$8,6 \times 10^{-7}$	$5,7 \times 10^{-13}$
RoMA	$8,5 \times 10^{-13}$	$2,3 \times 10^{-11}$	$5,0 \times 10^{-13}$

Табл. 7. Средняя доля корректных пар для совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	0,019	-	0,004
SAR-SIFT	0,054	-	0,041
RIFT	0,042	0,002	0,045
RoMA	0,575	0,089	0,986

Табл. 8. Средняя ошибка локализации точек для совмещенных изображений

Метод совмещения	Тип совмещаемых данных		
	VV	VR	RR
SIFT	$1,3 \times 10^{-13}$	-	$1,5 \times 10^{-12}$
SAR-SIFT	$2,1 \times 10^{-13}$	-	$1,5 \times 10^{-13}$
RIFT	$9,7 \times 10^{-14}$	$2,6 \times 10^{-12}$	$1,0 \times 10^{-13}$
RoMA	$1,7 \times 10^{-13}$	$1,2 \times 10^{-13}$	$2,0 \times 10^{-13}$

Заключение

В работе предложен универсальный метод автоматического совмещения разнородных цифровых изображений ДЗЗ: цветных, полутонных, мульти-

спектральных и радарных изображений и их комбинаций. Метод оказывается работоспособен при существенных различиях в пространственном разрешении сопоставляемых изображений – допускаются различия до четырех (опционально – шестнадцати) раз. Ключевым элементом метода является нейросетевая state-of-the art модель RoMa, которая до настоящей работы для совмещения космических снимков не использовалась. Модель RoMa была настроена/обучена на множестве разнородных космических снимков для решения задачи определения пар особых точек. Дополнительное использование априорных данных о снимках (их геопривязки) позволило добиться повышения точности совмещения цифровых космических снимков, превзойдя по качеству существующие решения, в том числе узкоспециализированные (SAR-SIFT).

Возможными направлениями развития настоящей работы могут являться следующие:

- снижение числа каналов разнородных изображений ДЗЗ как решение единой оптимизационной задачи;
- развитие метода для работы с изображениями с отличием в пространственном разрешении более чем в четыре (опционально – шестнадцать) раз.

Благодарности

Результаты исследования были получены при финансовой поддержке Российской Федерации в лице Минобрнауки России (Соглашение № 075-15-2024-558).

References

- [1] Horn BKP. Robot vision. Cambridge: MIT Press; 1986. ISBN: 978-0-262-08159-7.
- [2] Forsyth DA, Ponce J. Computer vision: A modern approach. New Jersey: Prentice Hall; 2002. ISBN: 978-0-13-085198-7.
- [3] Davies ER, Turk MA, eds. Advanced methods and deep learning in computer vision. San Diego, CA: Academic Press; 2022. ISBN: 978-0-12-822109-9.
- [4] Dellinger F, Delon J, Gousseau Y, Michel J, Tupin F. SAR-SIFT: A SIFT-like algorithm for SAR images. IEEE Trans Geosci Remote Sens 2015; 53(1): 453-466. DOI: 10.1109/TGRS.2014.2323552.
- [5] Ma W, Wen Z, Wu Y, Jiao L, Gong M, Zheng Y, Liu L. Remote sensing image registration with modified SIFT and enhanced feature matching. IEEE Geosci Remote Sens Lett 2017; 53(1): 3-7. DOI: 10.1109/LGRS.2016.2600858.
- [6] Xiang Y, Wang F, You H. OS-SIFT: A robust SIFT-Like Algorithm for High-Resolution Optical-to-SAR image registration in Suburban areas. IEEE Trans Geosci Remote Sens 2018; 56(6): 3078-3090. DOI: 10.1109/TGRS.2018.2790483.
- [7] Zhao J, Yang D, Li Y, Xiao P, Yang J. Intelligent matching method for heterogeneous remote sensing images based on style transfer. IEEE J Sel Top Appl Earth Obs Remote Sens 2022; 15: 6723-6731. DOI: 10.1109/JSTARS.2022.3197748.
- [8] Chen J, Xie H, Zhang L, Hu J, Jiang H, Wang G. SAR and optical image registration based on deep learning with co-attention matching module. Remote Sensing 2023; 15(15): 3879. DOI: 10.3390/rs15153879.

- [9] Brunelli R. *Template matching techniques in computer vision: Theory and practice*. Chichester, UK: John Wiley & Sons Ltd; 2009. ISBN: 978-0-470-51706-2.
- [10] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981; 24(6): 381-395. DOI: 10.1145/358669.358692.
- [11] Chum O, Matas J. Matching with PROSAC – progressive sample consensus. *Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition 2005*: 220-226. DOI: 10.1109/CVPR.2005.221.
- [12] Lowe DG. Object recognition from local scale-invariant features. *Proc Seventh IEEE Int Conf on Computer Vision 1999*: 1150-1157. DOI: 10.1109/ICCV.1999.790410.
- [13] P Yan K, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. *Proc 2004 IEEE Computer Society Conf on Computer Vision and Pattern Recognition 2004*: II-II. DOI: 10.1109/CVPR.2004.1315206.
- [14] Ghassabi Z, Shanbehzadeh J, Sedaghat A, Fatemizadeh E. An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. *EURASIP J Image Video Process* 2013; 2013: 25. DOI: 10.1186/1687-5281-2013-25.
- [15] Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. *Proc 2011 Int Conf on Computer Vision 2011*: 2564-2571. DOI: 10.1109/ICCV.2011.6126544.
- [16] Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features. In Book: Leonardis A, Bischof H, Pinz A, eds. *Computer Vision – ECCV 2006*. 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part I. Berlin, Heidelberg: Springer-Verlag; 2006: 404-417. DOI: 10.1007/11744023_32.
- [17] Rosten E, Drummond T. Machine learning for high-speed corner detection. In Book: Leonardis A, Bischof H, Pinz A, eds. *Computer Vision – ECCV 2006*. 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part I. Berlin, Heidelberg: Springer-Verlag; 2006: 430-443. DOI: 10.1007/11744023_34.
- [18] Yi KM, Trulls E, Lepetit V, Fua P. LIFT: learned invariant feature transform. In Book: Leibe B, Matas J, Sebe N, Welling M, eds. *Computer Vision – ECCV 2016*. 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI. Cham, Switzerland: Springer International Publishing AG; 2016: 467-483. DOI: 10.1007/978-3-319-46466-4_28.
- [19] Revaud J, De Souza C, Humenberger M, Weinzaepfel P. R2D2: Reliable and repeatable detector and descriptor. *arXiv Preprint*. 2019. Source: <<https://arxiv.org/abs/1906.06195>>. DOI: 10.48550/arXiv.1906.06195.
- [20] DeTone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detection and description. *2018 IEEE/CVF Conf on Computer Vision and Pattern Recognition Workshops (CVPRW) 2018*: 224-236. DOI: 10.1109/CVPRW.2018.00060.
- [21] Tyszkiewicz M, Fua P, Trulls E. DISK: Learning local features with policy gradient. *NIPS'20: Proc 34th Int Conf on Neural Information Processing Systems 2020*; 33: 14254-14265.
- [22] Zhao X, Wu X, Chen W, Chen PC, Xu Q, Li Z. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Trans Instrum Meas* 2023; 72: 5014016. DOI: 10.1109/TIM.2023.3271000.
- [23] Sun J, Shen Z, Wang Y, Bao H, Zhou X. LoFTR: Detector-free local feature matching with transformers. *2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2021*: 8918-8927. DOI: 10.1109/CVPR46437.2021.00881.
- [24] Rocco I, Arandjelović R, Sivic J. Efficient neighbourhood consensus networks via submanifold sparse convolutions. In Book: Vedaldi A, Bischof H, Brox T, Frahm JM, eds. *Computer Vision – ECCV 2020*. 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX. Cham, Switzerland: Springer Nature Switzerland AG; 2020: 605-621. DOI: 10.1007/978-3-030-58545-7_35.
- [25] Liu C, Yuen J, Torralba A. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Trans Pattern Anal Mach Intell* 2011; 33(5): 978-994. DOI: 10.1109/TPAMI.2010.147.
- [26] Edstedt J, Athanasiadis I, Wadenbäck M, Felsberg M. DKM: Dense kernelized feature matching for geometry estimation. *2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2023*: 17765-17775. DOI: 10.1109/CVPR52729.2023.01704.
- [27] Edstedt J, Sun Q, Bökman G, Wadenbäck M, Felsberg M. RoMa: Robust dense feature matching. *Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition 2024*. p. 19790-19800.
- [28] Dong Y, Jiao W, Long T, He G, Gong C. An extension of phase correlation-based image registration to estimate similarity transform using multiple polar fourier transform. *Remote Sens* 2018; 10(11): 1719. DOI: 10.3390/rs10111719.
- [29] Chen HM, Arora MK, Varshney PK. Mutual information-based image registration for remote sensing data. *Int J Remote Sens* 2003; 24(18): 3701-3706. DOI: 10.1080/0143116031000117047.
- [30] Hel-Or Y, Hel-Or H, David E. Matching by tone mapping: Photometric invariant template matching. *IEEE Trans Pattern Anal Mach Intell* 2014; 36(2): 317-330. DOI: 10.1109/TPAMI.2013.138.
- [31] Stumpf A, Michéa D, Malet J-P. Improved Co-Registration of Sentinel-2 and Landsat-8 Imagery for Earth surface motion measurements. *Remote Sens* 2018; 10(2): 160. DOI: 10.3390/rs10020160.
- [32] Rengarajan R, Choate M, Hasan MN, Denevan A. Co-registration accuracy between Landsat-8 and Sentinel-2 orthorectified products. *Remote Sens Environ* 2024; 301: 113947. DOI: 10.1016/j.rse.2023.113947.
- [33] 2021 IEEE GRSS Data Fusion Contest Track DSE. 2024. Source: <<https://www.grss-ieee.org/community/technical-committees/2021-ieee-grss-data-fusion-contest-track-dse/>>.
- [34] Li J, Li Y, He L, Chen J, Plaza A. Spatio-temporal fusion for remote sensing data: An overview and new benchmark. *Sci China Inf Sci* 2020; 63: 140301. DOI: 10.1007/s11432-019-2785-y.
- [35] Xie H, Pierce LE, Ulaby FT. Mutual information based registration of SAR images. *2003 IEEE Int Geoscience and Remote Sensing Symposium (IGARSS 2003) 2003*; 6: 4028-4031. DOI: 10.1109/IGARSS.2003.1295351.
- [36] Wang Y, Yu Q, Yu W. An improved Normalized Cross Correlation algorithm for SAR image registration. *2012 IEEE Int Geoscience and Remote Sensing Symposium 2012*: 2086-2089. DOI: 10.1109/IGARSS.2012.6350961
- [37] Hu C, Zhu R, Sun X, Li X, Xiang D. Optical and SAR image registration based on pseudo-SAR image generation strategy. *Remote Sens* 2023; 15(14): 3528. DOI: 10.3390/rs15143528.

- [38] Li J, Hu Q, Ai M. RIFT: Multi-modal image matching based on radiation-invariant feature transform. arXiv Preprint. 2018. Source: <<https://arxiv.org/abs/1804.09493>>. DOI: 10.48550/arXiv.1804.09493.
- [39] Ye Y, Bruzzone L, Shan J, Bovolo F, Zhu Q. Fast and robust matching for multimodal remote sensing image registration. IEEE Trans Geosci Remote Sens 2019; 57(11): 9059-9070. DOI: 10.1109/TGRS.2019.2924684.
- [40] Harris C, Stephens M. A combined corner and edge detector. Proc 4th Alvey Vision Conf 1988: 147-151. DOI: 10.5244/c.2.23.
- [41] Schmitt M, Hughes LH, Zhu XX. The SEN1-2 dataset for deep learning in SAR-optical data fusion. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 2018; IV-1: 141-146. DOI: 10.5194/isprs-annals-IV-1-141-2018.
- [42] Schmitt M, Hughes LH, Qiu C, Zhu XX. SEN12MS – A curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion. arXiv Preprint. 2019. Source: <<https://arxiv.org/abs/1906.07789>>. DOI: 10.48550/arXiv.1906.07789.
- [43] Xiang Y, Tao R, Wang F, You H, Han B. Automatic registration of optical and SAR images via improved phase congruency model. IEEE J Sel Top Appl Earth Obs Remote Sens 2020; 13: 5847-5861. DOI: 10.1109/JSTARS.2020.3026162.
- [44] Chen Z, Zhang L, Zhang G. An improved InSAR image co-registration method for pairs with relatively big distortions or large incoherent areas. Sensors 2016; 16(9): 1519. DOI: 10.3390/s16091519.
- [45] Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag 1901; 2(11): 559-572. DOI: 10.1080/14786440109462720.
- [46] Ye Y, Yang C, Zhu B, Zhou L, He Y, Jia H. Improving co-registration for Sentinel-1 SAR and Sentinel-2 optical images. Remote Sens 2021; 13(5): 928. DOI: 10.3390/rs13050928.

Сведения об авторах

Борисов Алексей Николаевич, 1996 года рождения, в 2019 г. окончил с отличием Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет). В настоящее время работает ассистентом кафедры геоинформатики и информационной безопасности Самарского университета. Круг научных интересов включает высокопроизводительные вычисления, вычисления общего назначения на графических процессорах, обработку изображений. E-mail: borisovalexey1996@gmail.com

Мясников Владислав Валерьевич, 1971 года рождения, в 1994 г. окончил с отличием Самарский национальный исследовательский университет имени академика С.П. Королева (Самарский университет), в 2008 защитил диссертацию на соискание степени доктора физико-математических наук. Работает профессором кафедры геоинформатики и информационной безопасности Самарского национального исследовательского университета имени академика С.П. Королева. Круг научных интересов: компьютерное зрение, распознавание образов, искусственный интеллект, машинное обучение и геоинформатика. Имеет более 200 публикаций, в том числе более 100 статей и три монографии (в соавторстве). E-mail: vmvas@geosamara.ru

Сергеев Владислав Викторович, 1951 года рождения. В 1974 году окончил Куйбышевский авиационный институт (ныне Самарский национальный исследовательский университет имени академика С. П. Королёва). В 1978 г. защитил кандидатскую, а в 1993 году – докторскую диссертацию. В настоящее время является заведующим кафедрой геоинформатики и информационной безопасности Самарского университета. Область научных интересов: цифровая обработка сигналов, анализ изображений, распознавание образов, геоинформатика. E-mail: vserg@geosamara.ru

ГРНТИ: 89.57.35

Поступила в редакцию 26 августа 2024 г. Окончательный вариант – 9 сентября 2024 г.

Method of automatic coregistration of digital remote sensing images from different sources

A.N. Borisov¹, V.V. Myasnikov¹, V.V. Sergeev¹

¹ Samara National Research University, 443086, Samara, Russia, Moskovskoye Shosse 34

Abstract

In this paper, a method for the automatic alignment of diverse digital Earth remote sensing images using survey data is proposed. The method is designed to align color, grayscale, multispectral, and radar images, as well as their combinations, with potential differences in spatial resolution of up to four times (optionally – sixteen times). The main stages of the proposed method include: an optional upscaling (up to four times); an optional number of image channels reduction to three or one; keypoint detection, their description and alignment. To achieve a universal and robust solution in the latter stages, the best-known algorithms were compared: SIFT, SAR-SIFT, RIFT, and the trainable RoMa. Experimental studies using the indicated types of space images demonstrate a clear advantage of the trainable neural network model RoMa trained on a variety of heterogeneous images. For additional improvement of the alignment accuracy, we utilized a priori data about the images in the form of their georeferencing information.

Keywords: digital remote sensing images, automatic image coregistration, multispectral images, radar images.

Citation: Borisov AN, Myasnikov VV, Sergeev VV. Method of automatic coregistration of digital remote sensing images from different sources. *Computer Optics* 2024; 48(6): 932-943. DOI: 10.18287/2412-6179-CO-1604.

Acknowledgements: This work was supported by the Ministry of Science and Higher Education, Russia (Agreement No.075-15-2024-558).

Authors' information

Alexey Nikolayevich Borisov, (b. 1996), graduated with honors from Samara National Research University (Samara University) in 2019. Currently works as an assistant at Geoinformatics and Information Security department at Samara University. Research interests: high-performance computing, GPGPU, image processing.
E-mail: borisovalexey1996@gmail.com

Vladislav Valerievich Myasnikov, received his DrSc degree in Physics & Maths (2008). Currently he works as a professor at the Geoinformatics and Information Security department in Samara National Research University. The range of scientific interests: computer vision, pattern recognition and artificial intelligence, machine learning and geoinformatics. He has about 200 publications, including more than 100 articles and three monographs.
E-mail: vmyas@geosamara.ru

Vladislav Viktorovich Sergeev, (b. 1951), graduated from Samara National Research University (Samara University) in 1974. In 1978 he received the Candidates of Technical Sciences degree and in 1993 – the Doctor of Technical Sciences degree. Currently he works as the head of Geoinformatics and Information Security department at Samara University. Research interests are digital signals and image processing, GIS and pattern recognition.
E-mail: vserg@geosamara.ru

Received August 26, 2024. The final version – September 9, 2024.
