

УДК 004.932

## МОДИФИЦИРОВАНИЕ МЕТОДА ПОИСКА И ДЕСКРИБИРОВАНИЯ УСТОЙЧИВЫХ ТОЧЕК SIFT ДЛЯ СОПОСТАВЛЕНИЯ ОПТИЧЕСКИХ И РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ

© 2022 г. В. В. Волков<sup>1,2,\*</sup>

<sup>1</sup> *Московский физико-технический институт (национальный исследовательский университет)  
141701 г. Долгопрудный, Институтский пер., д. 9, Россия*

<sup>2</sup> *Институт проблем передачи информации им. А.А. Харкевича РАН  
127051 Москва, Большой Каретный пер., д. 19, стр. 1, Россия*

\*E-mail: volkov-vl-v@yandex.ru

Поступила в редакцию 28.04.2022 г.

После доработки 17.05.2022 г.

Принята к публикации 08.07.2022 г.

Сопоставление изображений — это задача выравнивания (т.е. нахождения общей системы координат) двух или более изображений одной сцены. Достаточно сложным частным случаем данной задачи является сопоставление разнородных изображений — например, оптических и радиолокационных (optical-to-SAR). Необходимость подобного сопоставления возникает в задачах комплексирования изображений и детектирования объектов. Одним из способов сопоставления изображений являются поиск устойчивых точек и их сопоставление по дескрипторам-описаниям, вычисляемым для каждой точки с последующим нахождением преобразования при помощи геометрической модели. В данной работе исследовано влияние модификаций поиска устойчивых точек SIFT на точность детектора и точность всего алгоритма сопоставления оптических и радиолокационных изображений. Дополнительно в работе предлагается модификация SIFT дескриптора и приведено его сравнение с SIFT дескриптором. При сопоставлении изображений рассматривались искажения только в классе сдвигов. Все исследования производились на опубликованном датасете из 100 выравненных пар optical-SAR изображений.

*Ключевые слова:* сопоставление изображений, повторяемость устойчивых точек, дескрибирование устойчивых точек

DOI: 10.31857/S0235009222040060

### ВВЕДЕНИЕ

Сопоставление изображений — это выравнивание изображений одной сцены, полученных в разное время, под разным ракурсом и/или при помощи разных сенсоров. Задача сопоставления изображений, полученных от разных сенсоров, называется сопоставлением разнородных изображений (“multimodal image registration”). Сопоставление оптических и радиолокационных SAR (Synthetic Aperture Radar) изображений — частный случай такого сопоставления. Сопоставление optical-to-SAR широко используется в задачах дистанционного зондирования Земли (ДЗЗ), таких как комплексирование изображений (Сидорчук, Волков, 2018), детектирование объектов (Errico et al., 2015), обнаружение изменений (Plank et al., 2016).

Оптические изображения хорошо интерпретируются человеком и не содержат спекл-помех, в то время как SAR изображения не подвержены

влиянию атмосферы и освещенности (Ye et al., 2020; Hamdi et al., 2021). Также на SAR изображениях легче различать застроенные области, благодаря характерной высокой интенсивности их пикселей, вызванной множественным переотражением (Сидорчук, Волков, 2018; Errico et al., 2015).

Сложность сопоставления оптических и SAR изображений обусловлена тем, что (i) изображения подвергаются разным типам и силе шума. Например, зернистый шум присутствует как на оптических, так и на SAR изображениях, но на SAR изображениях шум проявляется сильнее. Более того (ii) значения интенсивности оптических и SAR изображений в некоторых участках изображений могут быть не скоррелированы даже в отсутствие шума. Кроме того, (iii) геометрическое положение (например, высота над землей) сенсора в пространстве может различаться в момент съемки, вследствие чего некоторые линии или

формы объемных объектов могут не совпадать (особенно это заметно на снимках гористой местности).

Методы сопоставления optical-SAR изображений разделяют на две категории: плотные методы (intensity-based methods) и разреженные методы (feature-based methods). Плотные методы основаны на поиске двух соответствующих (т.е. “похожих” по той или иной метрике) подызображений. Для сопоставления используются нормализованная кросс-корреляция (normalized cross-correlation) (Shi et al., 2012) и функция взаимной информации (mutual information) (Suri, Reinartz, 2009). Такие модели могут обладать высокой (субпиксельной) точностью, однако, имеют большую вычислительную сложность и зачастую требуют предварительного приближенного выравнивания изображений (Gong et al., 2013); также они могут быть неэффективны для пар изображений с большими геометрическими различиями (Suri, Reinartz, 2009).

Разреженные методы (feature-based methods) основаны на (i) нахождении некоторых “выделяющихся” точек или элементов изображений (features), легко отличимых как на картинке оптического диапазона, так и радиолокационного; (ii) последующего построения геометрического преобразования, которое сопоставляет найденные элементы изображений. Например, такие методы могут искать точки (Fan et al., 2012; Ma et al., 2016; Paul, Pati, 2018), углы (Xiang et al., 2018; Paul, Pati, 2019; Xiong et al., 2019), линии (Wang et al., 2012) и объекты (например, дороги) (Kunina et al., 2019). Разреженные методы вычислительно проще и зачастую эффективнее (Gong et al., 2013) плотных методов в случаях, когда присутствуют значительные геометрические искажения между изображениями. Однако они могут проигрывать плотным методам в максимальной точности сопоставления (Gong et al., 2013).

#### **Общий алгоритм сопоставления изображений**

В данной статье уделяется основное внимание группе разреженных методов, использующих в качестве “выделяющихся элементов” устойчивые точки. В общем виде разреженные алгоритмы сопоставления на основе устойчивых точек состоят из следующих шагов:

- Нахождение устойчивых точек.
- Вычисление дескрипторов для каждой устойчивой точки.
- Нахождение сопоставлений между устойчивыми точками изображений (например, путем сравнения дескрипторов).
- Нахождение геометрического преобразования между изображениями (например, методом RANSAC (Fischler, Bolles, 1981)), которое воспро-

изводит наибольшую часть найденных на предыдущем шаге сопоставлений точек.

#### **Геометрическая модель**

Методы робастного оценивания, такие как RANSAC, позволяют восстановить правильное геометрическое преобразование даже в случаях, когда значительная часть сопоставлений между устойчивыми точками (найденная на 2-м шаге алгоритма) сформирована неправильно (Чеканов и др., 2020; Tropin et al., 2019; Шемякина, 2017). Тем не менее высокая доля правильных сопоставлений *точек* позволяет повысить долю и точность правильных сопоставлений *изображений* (Skoquukina et al., 2020; Xu et al., 2015). Также точность сопоставления изображений может быть выше, если точки распределены по изображению более равномерно (Тропин и др., 2019), поскольку плотные скопления точек могут создавать менее точные пары (из-за взаимного “перемешивания”) и иметь низкую обусловленность.

#### **Нахождение устойчивых точек и дескрипторов**

Примером способов для более равномерного распределения точек по изображению являются NMS (non-max-suppression) (Yu et al., 2018) и биннинг (Xiong et al., 2019). NMS среди всех пар точек на одном изображении, расположенных ближе некоторого порога, оставляет только одну точку, обладающую наибольшей “важностью”. “Важность” точки определяется детектором устойчивых точек. Таким образом, алгоритм избавляется от скоплений точек в одном месте. Биннинг – это предварительное разделение изображения на непересекающиеся блоки. Для каждого блока происходит независимое вычисление устойчивых точек, после чего найденные множества точек объединяются. Данный метод позволяет избежать скопления точек в одной области изображения, обеспечивая наличие устойчивых точек по всему изображению.

Стоит отметить, что формирование “правильного” сопоставления с участием устойчивой точки невозможно, если на другом изображении не была найдена устойчивая точка в месте, которое соответствует той же позиции в реальной сцене. Поэтому эффективный метод должен находить такие точки, которые с высокой вероятностью будут найдены и на другом изображении той же сцены. Для численной оценки этого качества в литературе используется метрика повторяемости (repeatability) (Xiang et al., 2018). Повторяемость – это доля или среднее количество (определения разнятся) устойчивых точек таких, что для устойчивой точки на одном изображении сцены существует достаточно близкая устойчивая точка на втором изображении сцены. В ходе исследования

было обнаружено, что в разных работах встречаются разные способы вычисления повторяемости. В данной статье приводится пример работ с указанием применяемых в них формул вычисления повторяемости (п. 1).

Для сопоставления точек возникает необходимость вычисления дескрипторов (“описаний”), по которым можно определить, соответствуют ли две точки одному и тому же географическому месту. Для этого алгоритм вычисления дескрипторов должен для точек, соответствующих одному географическому месту, привязывать похожие по той или иной метрике (например, L2) дескрипторы, и, наоборот, для точек из разных мест – разные дескрипторы.

### Постановка задачи

В работе рассматривается метод нахождения устойчивых точек SIFT с точки зрения его эффективности для сопоставления optical-SAR изображений. Разбираются его несколько простых модификаций и оценивается их влияние на результат сопоставления. Дополнительно предлагается метод дескрибирования устойчивых точек, являющийся модификацией SIFT дескриптора. В работе можно выделить несколько смысловых частей.

В п. 1 уделяется внимание метрике повторяемости, используемой для оценки эффективности нахождения устойчивых точек и развернуто показывается, что в литературе нет общепринятого определения повторяемости. В п. 2 описываются используемые данные. В п. 3 представлен общий вид используемого алгоритма сопоставления изображений с описанием его отдельных шагов. Здесь описаны используемые модификации для алгоритма нахождения устойчивых точек SIFT и предлагаемый алгоритм вычисления дескрипторов. В п. 4 показан алгоритм определения оптимальных параметров и результаты алгоритмов, описанных в п. 3. Проведена оценка влияния модификаций для алгоритма нахождения точек на повторяемость и точность сопоставления optical-SAR изображений. Дополнительно проведено сравнение предложенного дескриптора с SIFT дескриптором.

Отдельно отметим, что в литературе встречаются модификации SIFT алгоритма для сопоставления optical-SAR изображений (Fan et al., 2012; Gong et al., 2013; Paul, Pati, 2018), однако, в работах не предоставлено открытого кода, что не позволяет провести объективное сравнение.

## 1. ПОВТОРЯЕМОСТЬ

В литературе встречаются различные способы оценки точности алгоритмов сопоставления оптических и радиолокационных изображений. Ча-

ще всего для оценки всего алгоритма находят долю правильно сопоставленных изображений. В качестве критерия правильно сопоставленной пары изображений можно привести метрику RMSE (Paul, Pati, 2019). Дополнительно к оценке всего алгоритма зачастую в работах приводится оценка его первой части, а именно, детектирование устойчивых точек. Метрики для оценки точности детектирования точек называют метриками повторяемости.

В литературе встречаются разные варианты метрик повторяемости. Отсутствие единой метрики повторяемости не позволяет объективно сравнивать результаты различных работ. Рассмотрим примеры работ и применяемые в них метрики повторяемости.

В работе (Li et al., 2019) авторы оценивают повторяемость своего алгоритма на датасете, состоящем из шести мини датасетов по десять пар изображений в каждом: optical-optical, infrared-optical, optical-SAR, depth-optical, map-optical и day-night. Авторы не выложили датасет в общий доступ. В качестве метрики повторяемости они использовали отношение количества пар повторяемых устойчивых точек к среднему количеству устойчивых точек на обоих изображениях (формула (1)).

$$Rep = \frac{2N_{Rep}}{(n_1 + n_2)}, \quad (1)$$

где  $n_1, n_2$  – количество точек на первом и втором изображении соответственно,  $N_{Rep}$  – количество пар повторяемых точек.

Повторяемой парой точек считается пара, удовлетворяющая

$$\|x_i^1 + Hx_i^2\| < \theta,$$

где  $H$  – правильное преобразование между двумя изображениями,  $x_i^1, x_i^2$  – координаты устойчивых точек на первом и втором изображении соответственно,  $\theta$  – порог повторяемости ( $\theta = 3$  в (Li et al., 2019)).

Авторский метод получил 22.9% по метрике повторяемости и превзошел такие детекторы, как FAST (Rosten et al., 2008), Harris, BRISK (Leutenegger et al., 2011), SIFT (Lowe, 2004), SURF (Bay et al., 2008), MSER (Matas et al., 2004).

В работе (Fan et al., 2018) авторы оценили свой алгоритм на десяти парах optical-SAR изображений (из которых две пары – смоделированные изображения), где также отметили, что использовали метрику повторяемости, описанную в работе (Schmid et al., 2000). Данная метрика, как и предыдущая, считает количество пар повторяемых точек, но только на общих (т.е. пересе-

кающихся) участках двух изображений (формула (2)).

$$Rep = \frac{N_{Rep}}{\min(n_1, n_2)}, \quad (2)$$

где  $n_1, n_2$  — количество точек на общем участке первого и второго изображения соответственно,  $N_{Rep}$  — количество пар повторяемых точек на общем участке двух изображений.

Заметим, что метрика по формуле (2) будет в диапазоне  $[0,1]$  тогда и только тогда, когда каждая точка не может иметь более одной пары. Данного уточнения в работе (Schmid et al., 2000) замечено не было, как и описания принципа, по которому разрешаются такие “конфликтные” ситуации (например, для одной точки подходит несколько точек, какую взять в пару?). В работе (Fan et al., 2018) порог расстояния между повторяемыми точками равен 1.5 пикселям. Результаты показали повторяемость более 50% для всех пар, кроме одной.

Был найден ряд работ, в которых вычисление повторяемости производилось на частично или полностью смоделированных данных (Xiang et al., 2018; Wang et al., 2020; Yu et al., 2021). В работе (Xiang et al., 2018) моделирование optical-SAR пары изображений происходит путем добавления к оптическому изображению с высоким разрешением гауссового и зернистого шума соответственно. В работе исследуется робастность повторяемости алгоритмов детектирования устойчивых точек в зависимости от степени шума. В качестве метрики повторяемости используется отношение количества повторяемых точек к общему количеству устойчивых точек на обоих изображениях, что эквивалентно метрике по формуле (1). Порог для определения повторяемости двух точек равен двум пикселям. Авторы оценили повторяемость на двух смоделированных изображениях и получили значения от 50 до 59% в зависимости от уровня шума.

В работе (Wang et al., 2020) для оценки робастности алгоритмов детектирования также вычислялась повторяемость на смоделированных данных. Моделирование оптического и SAR изображения происходило при помощи добавления гауссового и зернистого шума соответственно. В качестве метрики повторяемости используется метрика, отмеченная в формуле (1), где порог повторяемости равен двум пикселям. Результаты вычислялись для трех пар optical-SAR изображений, к каждой из которых добавили свой шум с несколькими уровнями силы. Повторяемость от 40 до 65% в зависимости от пары картинок и уровня шума.

В работе (Yu et al., 2021) оценка повторяемости проводилась на искусственных двух парах изображений, не использующих спутниковые снимки

в качестве основы. Оценка точности алгоритма сопоставления изображений проводилась на optical-SAR спутниковых снимках, но без оценки повторяемости. Метрика повторяемости совпадает с указанной в формуле (1), где порог повторяемости равен 1.5 пикселям. Значения повторяемости указаны примерно от 41 до 75% в зависимости от пары изображений и уровня шума.

В ряде других работ повторяемость вычисляется на настоящих парах оптических и SAR изображений, но без описания используемого алгоритма повторяемости (Paul, Pati, 2019; Chen et al., 2017; Xie et al., 2021). Поэтому воспроизвести аналогичный метод оценки повторяемости невозможно. В работе (Paul, Pati, 2019) приводятся значения повторяемости, полученные на реальных шести парах optical-SAR изображений. Повторяемость для каждой пары optical-SAR изображений указана около 30–40%. В работе (Chen et al., 2017) приводится среднее значение повторяемости на других шести парах optical-SAR изображений без описания метрики повторяемости. Метод, предложенный авторами, показал повторяемость 57.8%. В другой работе оценка метода детектирования точек производится на двух парах optical-SAR изображений (Xie et al., 2021). Порог расстояния между повторяемыми точками брался равным двум пикселям. Повторяемость равна 14.5 и 16%.

Таким образом, из девяти работ четыре используют метрику повторяемости по формуле (1), две работы — по формуле (2) и в трех работах метрика повторяемости не указана.

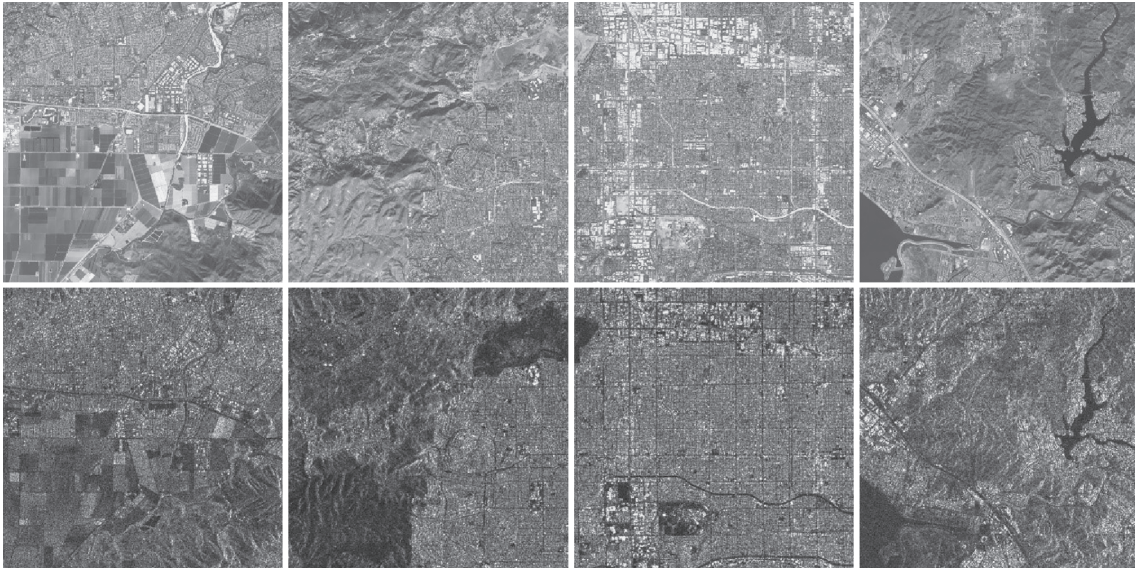
## 2. ОПИСАНИЕ ДАННЫХ

Используемый в данной работе датасет состоит из 100 пар optical-SAR изображений размером  $1024 \times 1024$  и опубликован в работе (Волков, Швец, 2021). Изображения в датасете собраны из открытых источников, а лицензии к ним позволяют их распространять и изменять<sup>1</sup>. В качестве источника оптических и SAR данных задействован сайт Copernicus Open Access Hub<sup>2</sup>. В качестве SAR изображений использовались изображения со спутника Sentinel-1A со следующими параметрами: тип продукта Level-1 Ground Range Detected (GRD), режим сенсора Interferometric Wave (IW) swath mode, поляризация VH. В качестве оптических изображений рассматривались трехканальные RGB спутниковые снимки со спутника Sentinel-2A (тип продукта: S2MSI1C). Изображения обоих типов были приведены к пространственному разрешению 10 метров/пиксель. Каждая пара optical-SAR изображений вы-

<sup>1</sup> <https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions>

<sup>2</sup> <https://scihub.copernicus.eu/>





**Рис. 1.** Примеры пар изображений из датасета. Верхний ряд – оптические изображения, нижний – соответствующие им SAR изображения.

равнена друг относительно друга с применением геопривязки. Для увеличения точности выравнивания было проведено ручное сопоставление изображений с помощью проективного преобразования. Полученные изображения выравнены преимущественно с субпиксельной точностью, но иногда встречаются небольшие области, где точность ниже (погрешность до двух пикселей). В датасете преимущественно присутствуют изображения сцен городов и полей. Примеры пар изображений представлены на рис. 1.

Готовый к использованию датасет и сопровождающие его метаданные доступны для скачивания и опубликованы в (Волков, Швеиц, 2021). Также данные могут быть скачаны в ручном режиме с ресурсов, указанных выше.

### 3. ОБЩИЙ ВИД АЛГОРИТМА СОПОСТАВЛЕНИЯ

В данной статье рассматривается алгоритм сопоставления оптических и радиолокационных изображений с использованием устойчивых точек. Общий вид рассматриваемого алгоритма сопоставления следующий:

- Нахождение устойчивых точек.
- Вычисление дескриптора для каждой устойчивой точки.
- Нахождение сопоставлений между точками на основе их дескрипторов.
- Нахождение геометрического преобразования при помощи геометрической модели.

В рамках работы находятся оптимальные параметры подалгоритмов, используемых на каждом

из шагов, а также предлагается новый метод дескрибирования устойчивых точек специально для случая оптических и радиолокационных спутниковых изображений. Рассмотрим детальнее каждый из шагов.

#### 3.1. Детектор устойчивых точек

В качестве детектора устойчивых точек использовался SIFT-детектор (из библиотеки OpenCV<sup>3</sup>). Скопление найденных устойчивых точек в одном месте изображения может ухудшать итоговый результат сопоставлений (набор близко расположенных точек может быть плохо обусловлен, кроме того близко расположенные точки могут иметь похожие дескрипторы и путаться друг с другом на этапе сопоставления точек). Для улучшения распределения устойчивых точек на изображении широко используются такие методы, как биннинг (Xiong et al., 2019) и NMS (Yu et al., 2018). В данной работе мы рассматриваем оба метода и оцениваем их эффективность для случая оптических и радиолокационных изображений.

Биннинг представляет собой разбиение изображения на непересекающиеся блоки (в нашем случае  $256 \times 256$  пикселей) перед применением детектора устойчивых точек. При биннинге детектор применяется к каждому блоку независимо, после чего множества найденных точек объединяются, таким образом, итоговый набор точек будет более равномерно распределен по всему изображению (рис. 2).

<sup>3</sup> <https://docs.opencv.org/master/index.html>

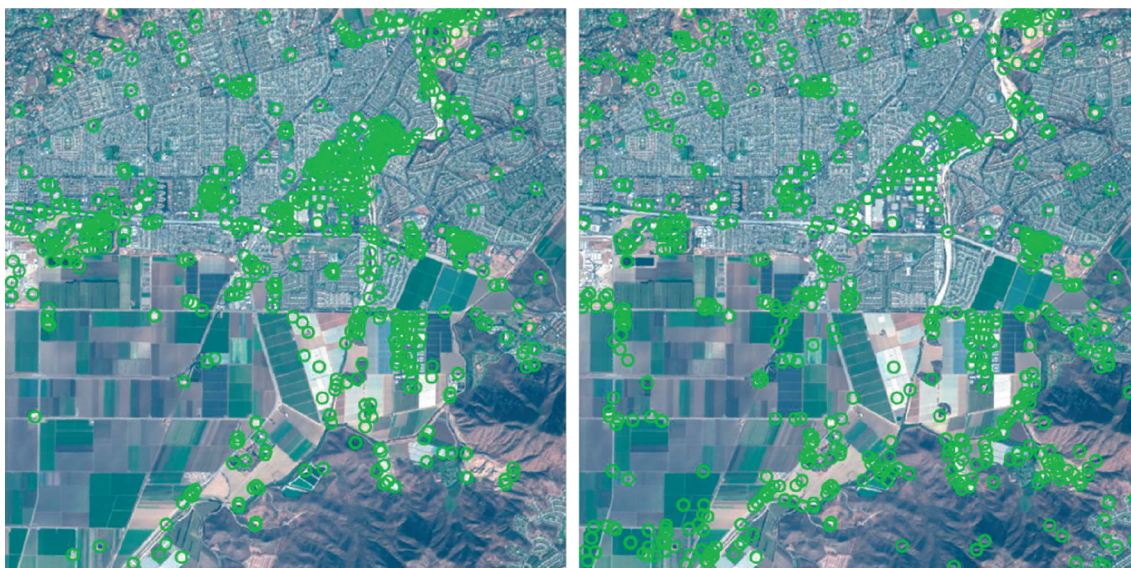


Рис. 2. Пример работы SIFT детектора. Слева без применения биннинга, справа – с применением биннинга.

В наших экспериментах максимальное количество точек, которые могут быть найдены на одном блоке, ограничивалось 50 точками (т.е. суммарно до 800 точек на изображении). Если было найдено больше точек, то отбирались те точки, которые были оценены SIFT-детектором как более важные (наибольший).

Несмотря на биннинг, скопления точек все равно возможны внутри каждого из блоков – для противодействия такому эффекту мы использовали алгоритм фильтрации NMS (non-max-suppression) (Yu et al., 2018), который из любых двух точек, расположенных достаточно близко (определяется параметром), выбирает одну с наибольшим. Влияние биннинга и NMS на получаемый набор устойчивых точек для optical-SAR случая описан в п. 4.1.

### 3.2. Вычисление дескрипторов

После нахождения устойчивых точек на каждом из изображений, необходимо найти соответствия между ними. Поскольку каждая пара точек (точка на оптическом изображении – точка на SAR изображении) должна быть “проверена” на совпадение, то общая сложность операции составляет  $N \cdot M$ , где  $N$  и  $M$  – число точек на соответствующих изображениях. Таким образом, квадратичная сложность алгоритма делает желательным наличие быстрого способа проверки – соответствуют ли две точки друг другу или нет.

Классическим методом для решения этой задачи является нахождение дескрипторов для каждой из точек – векторов рациональных чисел таких, что для похожих точек дескрипторы близки по L2 или другой быстро считающейся метрике, а

для отличающихся – далеки. Тогда после однократного подсчета дескрипторов для каждой из точек (всего  $N + M$  точек) для проведения  $N \cdot M$  операций можно использовать быструю векторизованную операцию L2.

В данной работе мы используем как классический метод дескрибирования алгоритма SIFT, так и предлагаем свой алгоритм, разработанный специально для сопоставления optical-SAR изображений.

### 3.3. Нахождение сопоставлений между точками

После нахождения дескриптора для каждой из устойчивых точек и нахождения расстояний между дескрипторами для каждой пары точек, необходимо на основе этих расстояний сформировать сопоставления между ними. Чем выше доля и количество корректно сопоставленных пар точек, тем, как правило, выше точность сопоставления изображений.

В данной работе использовался метод ближайшего соседа. Для любой пары дескрипторов на optical-SAR изображениях вычислялось дескрипторное расстояние (L2 норма). Две точки считаются сопоставленными, если соответствующие им дескрипторы взаимно ближайшие (т.е. не существует другого дескриптора такого, что дескрипторное расстояние с ним меньше найденного) и дескрипторное расстояние меньше настроиваемого порога. Пороги настраивались индивидуально для каждого используемого алгоритма. Подробное описание процедуры сопоставления и использованных порогов приведено в п. 4.



### 3.4. Нахождение геометрического преобразования

На этом шаге происходит вычисление геометрического преобразования между изображениями по найденным сопоставлениям между точками. Методы для нахождения такого преобразования, например, RANSAC (Fischler, Bolles, 1981), зачастую не требуют, чтобы все сопоставления точек были найдены верно. При целевой точности сопоставления изображений, условно в 10 пикселей, для большинства точек допустим шум, например, в 3 пикселя (некоторая часть точек может иметь больший шум). Кроме того, при достаточном высоком количестве точек с более сильным ограниченным равномерным шумом (т.е. большинство точек равномерно шумят в определенной окрестности вокруг точного их положения) можно достигнуть сопоставления с субпиксельной точностью.

В работе мы рассматриваем только случай, когда сопоставляемые изображения искажены друг относительно друга только сдвигом (что относится и к обучающим, и к тестовым парам изображений). Это позволяет использовать простую голосящую схему для нахождения геометрического преобразования между изображениями.

На вход алгоритм получает координаты пар сопоставленных устойчивых точек  $(x_i^{opt}, y_i^{opt})$ ,  $(x_i^{sar}, y_i^{sar})$ ,  $i \in [1, N]$ , где  $N$  – количество пар. Точки с одинаковым индексом формируют пару. Остальными параметрами алгоритма являются размер бина двумерной гистограммы  $size_{bin}$ , размер ядра размытия  $\sigma_m$  и размер изображения (для определения границ двумерной гистограммы). На выходе выдаются координаты сдвига  $(x_s, y_s)$ .

Алгоритм:

- Для каждой сопоставленной пары точек  $(x_i^{opt}, y_i^{opt})$ ,  $(x_i^{sar}, y_i^{sar})$  вычисляем  $\Delta x_i = x_i^{sar} - x_i^{opt}$ , и  $\Delta y_i = y_i^{sar} - y_i^{opt}$ .

- По этим значениям вычисляем двумерную гистограмму распределения сдвигов между сопоставленными точками. Размер бина гистограммы  $(size_{bin}, size_{bin})$ . Затем нормируем гистограмму, поделив значения в каждом бине на максимальное значение в гистограмме.

$$H_i = \frac{H_i}{\max_i H_i}, \quad (3)$$

где  $2H_i$  – значение  $i$ -го бина, а  $\max_i H_i$  – максимальное значение бина в гистограмме.

- К полученной двумерной матрице применяем двумерный гауссовый фильтр с сигмой  $\sigma_m$ .

- Искомым сдвигом  $(x_s, y_s)$  являются координаты центра бина с максимальным значением.

### 3.5. Критерии оценки

В качестве критерия оценки точности всего алгоритма использовалось количество успешно сопоставленных изображений. Пара изображений считалась успешно сопоставленной, если величина найденного сдвига составила меньше  $T$  (т.е.  $\sqrt{x_s^2 + y_s^2} \leq T$ ), где для изображений с даунскейлингом в 2 раза  $T = 2.5$  и без даунскейлинга  $T = 5$  при размере изображений  $512 \times 512$  и  $1024 \times 1024$  соответственно.

Для настройки параметров и для тестирования получившегося алгоритма датасет был разделен на две части – обучающую (76 пар) и тестовую (24 пары). Подбор оптимальных параметров происходил только на обучающей части датасета, а метрики считались на тестовой.

### 3.6. Алгоритм вычисления дескрипторов

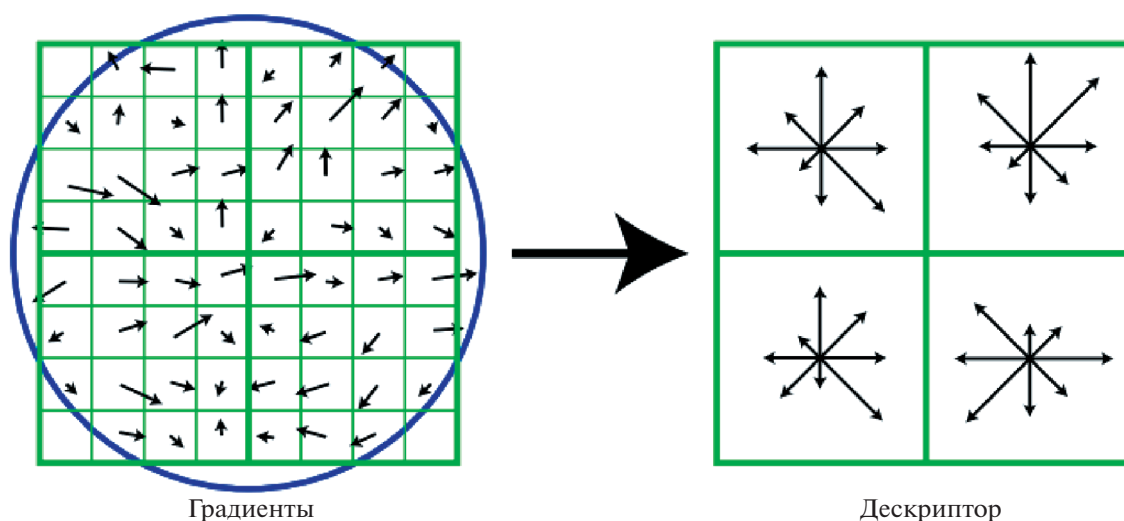
В этом параграфе мы представляем наш алгоритм вычисления дескрипторов, представляющий собой упрощенную модификацию дескриптора SIFT (Lowe, 2004). Для начала рассмотрим SIFT-дескриптор и его недостатки в рамках задачи optical-SAR сопоставления изображений.

#### SIFT-дескриптор

С подробным алгоритмом можно ознакомиться в статье-первоисточнике (Lowe, 2004). Здесь мы отметим основные интересующие нас моменты. Во-первых, SIFT имеет встроенную предобработку изображения (фильтр Гаусса). Во-вторых, основным шагом при подсчете SIFT-дескриптора является построение так называемой “гистограммы ориентаций”. Для ее формирования вокруг устойчивой точки вырезается патч, который делится на не пересекающиеся “кирпичики” в виде сетки  $4 \times 4$ . В каждом пикселе патча считается градиент и вычисляются его длина и угол наклона. “Гистограмма ориентаций” вычисляется для каждого “кирпичика” по этим градиентам – 8 бинов по углу наклона. Значение в каждом бине равно сумме длин градиентов с соответствующим углом наклона, умноженных на гауссовую функцию с окном, равным размеру патча (рис. 3, синий круг).

Таким образом, чем дальше градиент от центра патча, тем меньшее влияние он оказывает на значение бина гистограммы. На рис. 3 приведен пример гистограмм ориентаций для случая патча  $8 \times 8$  пикселей с разбиением на  $2 \times 2$  “кирпичика” (Lowe, 2004).

По “гистограмме ориентаций” формируется дескриптор путем объединения всех значений бинов. Каждый “кирпичик” содержит восемь чисел, конкатенацией векторов всех 16 “кирпичиков” ( $4 \times 4$ ), алгоритм получает вектор-дескрип-

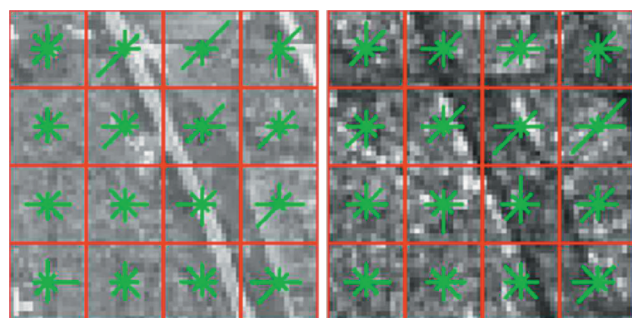


**Рис. 3.** Слева нарисованы градиенты на патче размера  $8 \times 8$ , которые умножаются на вес, определенный гауссовым окном (круг). Эти градиенты образуют гистограммы градиентов справа. Каждый вектор справа отображает сумму длин градиентов с близким направлением в соответствующей области патча.

тор из 128 чисел. Далее вектор нормируется, после чего значения выше некоторого порога “обрезаются” до этого порога и вектор снова нормируется. Более подробное описание алгоритма дескрибирования описано в статье-первоисточнике (Lowe, 2004).

Задача сопоставления оптических и радиолокационных изображений имеет, как минимум, две особенности, ограничивающие эффективность SIFT дескриптора:

– *Инвертирование градиентов.* На оптических и SAR изображениях некоторые объекты отображаются похожим образом, но встречаются и объекты, где интенсивности инвертированы. Например, дорогам на оптическом снимке часто соот-



**Рис. 4.** Пример инвертированности градиентов. Слева патч оптического изображения, справа – SAR. В каждом из 16 квадратиков нарисованы вектора градиентов, соответствующих сумме всех градиентов в этом квадратике с соответствующим направлением вектора (сектор в  $360^\circ$  разбивается на восемь одинаковых секторов).

ветствует высокая интенсивность, в то время как на SAR снимке зачастую, наоборот, низкая (рис. 4).

Внутри квадратиков отмечена “гистограмма ориентаций” с единичным весом. У квадратиков около дороги хорошо видно изменение диагональной составляющей градиента между оптическим и SAR патчем.

– *Одинаковые параметры предобработки оптических и радиолокационных изображений.* SIFT дескриптор имеет встроенную предобработку изображения в виде гауссового фильтра, но его параметры настраиваются самим алгоритмом без участия человека. Оптические и SAR изображения сильно различаются по уровню шума и детализации, потому что оптимальные параметры предобработки для них отличаются (например, наши эксперименты показали, что для SAR изображений нужно использовать большую сигму, чем для оптических), а также могут потребовать дополнительных индивидуальных методов предобработки для одного из типов изображений.

Далее рассмотрим наш алгоритм вычисления дескрипторов.

### *Предобработка изображений*

Перед применением нашего алгоритма вычисления дескрипторов применялась предобработка изображений. SAR изображения подвержены сильному зернистому шуму, поэтому было решено использовать гауссовский фильтр для борьбы с шумом с  $\sigma_{sar}^{pre}$  (рис. 5, вторая и четвертая колонки).

Оптические изображения тоже подвержены шуму, но намного меньше, поэтому к ним также применяется гауссовский фильтр, но с меньшей

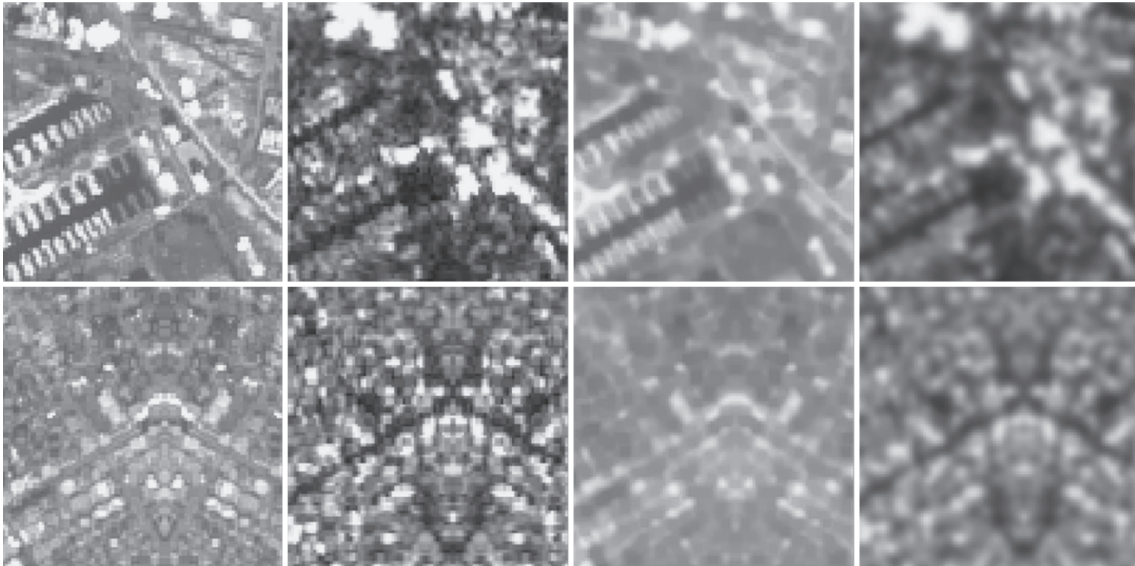


Рис. 5. Пример предобработки патчей. Слева направо: оптический патч, SAR патч, оптический патч с предобработкой ( $\sigma_{opt} = 1$ , окно раздутия и эрозии 1 пиксель), SAR патч с предобработкой ( $\sigma_{sar} = 1.5$ ).

силой размытия. Дополнительно на некоторых фрагментах оптических изображений (например, пристань, рис. 5, верхний ряд) присутствуют мелкие детали, где градиент сильно различается с SAR картинкой, на которой эти детали не различимы. Для ослабления этого эффекта к оптическим картинкам, помимо гауссовского фильтра, применяется последовательность раздутия и эрозии с окном в 1 пиксель. Пример патчей с примененной предобработкой представлен на рис. 5. На верхнем патче видно, что на оптическом патче видна структура пристани, в то время, как на SAR патче структура слабо видна, что может влиять на вычисление градиентов. После применения гауссового фильтра и морфологии структура пристани на оптическом патче уже не такая явная.

#### Предлагаемый алгоритм вычисления дескрипторов

Данный алгоритм основан на алгоритме SIFT-дескриптора.

Пусть имеются координаты устойчивой точки  $(x, y)$ . Вокруг этой точки вырезается патч размера  $64 \times 64$  с центром в точке  $(x, y)$ . В случае, если патч выходит за пределы изображения, то недостающая часть заполняется путем отзеркаливания изображения (рис. 5, нижний ряд – пример частично отзеркаленного патча). Далее выполняется следующий алгоритм:

- Вычисляются градиент в каждом пикселе патча, его длина и угол наклона.
- Патч равномерно делится на 16 непересекающихся “кирпичиков” в виде сетки  $4 \times 4$ .

- Для каждого “кирпичика” вычисляется “гистограмма градиентов по направлению”, состоящая из восьми бинов. Значение в бине равно сумме длин градиентов с соответствующим бину направлением. Визуально будем представлять эти гистограммы в виде восьми векторов, отложенных из одной точки.

- Так как некоторые области на оптических и SAR изображениях, соответствующие одному географическому месту, могут иметь противоположно направленные градиенты (например, дороги на оптических изображениях, как правило, имеют большую интенсивность, чем у окружающей области, а на SAR, наоборот (рис. 4)), то направления, повернутые на  $180^\circ$ , будем считать, как одно направление. Таким образом, получаем четыре вектора, каждый из которых равен сумме двух коллинеарных векторов.

- Нормируем вектора так, чтобы сумма длин векторов равнялась единице.

- Объединяем четыре вектора в 16 “кирпичиках” в единый вектор из 64 значений (по рядам от верхнего левого к нижнему правому).

На рис. 6 представлены примеры гистограмм градиентов нашего метода с четырьмя и восемью направлениями и гистограммы градиентов SIFT дескриптора.

Гистограммы градиентов для нашего метода строились по алгоритму, описанному выше. Гистограммы градиентов для SIFT дескриптора были построены по значениям SIFT дескриптора. Вектор из 128 чисел разбивался на 16 блоков по восемь чисел в каждом. Каждый блок соответствовал одной клетке на патче.



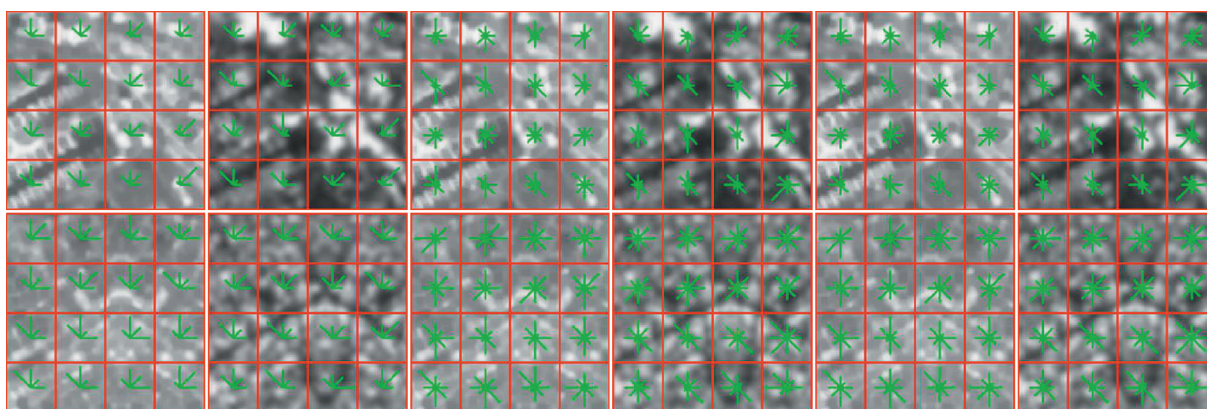


Рис. 6. Гистограммы градиентов для optical-SAR патчей. Слева направо: наш метод с четырьмя направлениями (первая и вторая колонки), наш метод с восемью направлениями (3- и 4-я колонки), SIFT дескриптор (5- и 6-я колонки).

Предложенный алгоритм вычисления дескрипторов показал результат лучше, чем SIFT дескриптор (п. 4.2), а алгоритм заметно проще. Замеры производительности не проводились, так как используемый алгоритм не оптимизирован.

#### 4. НАХОЖДЕНИЕ ОПТИМАЛЬНЫХ ПАРАМЕТРОВ АЛГОРИТМОВ

В этом параграфе мы представим ряд модификаций, предлагаемых нами для повышения точности нашего алгоритма дескрибирования, оцениваем влияние этих модификаций на качество сопоставления и сравниваем результат с SIFT-дескриптором. Устойчивые точки для всех экспериментов были одинаковыми, алгоритм их нахождения описан в п. 3.1. Также приведем сравнение методов с и без использования NMS и биннинга на примере SIFT детектора и дескриптора (п. 4.1). Отметим, что нахождение оптимальных параметров алгоритмов всегда происходит только на обучающей части датасета (76 пар).

##### 4.1. Применение NMS алгоритма и биннинга

В этом разделе представлено два эксперимента. В первом эксперименте рассматривалось влияние NMS (non-max-suppression) алгоритма, примененного на шаге детектирования точек, на точность сопоставления optical-SAR изображений. В качестве детектора устойчивых точек и дескриптора использовались SIFT детектор и дескриптор соответственно. Для подбора параметров детектора использовалась метрика повторяемости (п. 1, формула (1)). Параметры SIFT-детектора (из библиотеки OpenCV<sup>4</sup>) приведены ниже:

- *nOctaveLayers*: 3
- *contrastThreshold*: 0.04

<sup>4</sup> <https://docs.opencv.org/master/index.html>

- *edgeThreshold*: 10
- $\sigma$ : optical 2.5, SAR 3.0

Первые три параметра взяты по умолчанию и не менялись, так как при их изменении повторяемость или оставалась такой же, или ухудшалась. Подбор параметра  $\sigma$  происходил индивидуально для оптических и SAR изображений. Порог повторяемости  $\theta$  фиксировался равным 5 пикселей. Оценка повторяемости для нескольких значений  $\sigma$  без применения NMS представлена в табл. 1, а. Из таблицы видно, что оптимальное значение достигается при  $\sigma_{opt} = 2.5$  и  $\sigma_{opt} = 3.0$ .

Результаты для ( $\sigma_{opt} = 2.5, \sigma_{opt} = 3.0$ ) и ( $\sigma_{opt} = 3.0, \sigma_{opt} = 3.0$ ) похожи, поэтому дополнительно отметим, что для  $\theta = \{1, 3, 4\}$  в первом случае доля повторяемости выше на 0.1%, а при  $\theta = 2$  доля повторяемости одинакова, что делает первый случай лучше.

В табл. 1, б приведены результаты повторяемости с теми же параметрами детекторов, но с применением NMS с окном 5 пикселей.

Оптимальные значения  $\sigma_{opt}$  и  $\sigma_{sar}$  остались как и в случае без NMS. Из табл. 1, а, б видно, что в случае с NMS повторяемость хуже, особенно если сравнивать количество повторяемых точек (оно сократилось примерно на треть). Это обусловливается тем, что среди близко расположенных точек удаляются все близко расположенные точки (ближе 5 пикселей), кроме одной. Доля повторяемых точек также уменьшилась, но не значительно — относительная разница составила около 3%. Отсюда можно сделать вывод, что точки детектируются скоплениями, но при этом уменьшение скопившихся рядом точек почти не влияет на долю повторяемых точек. При этом на этапе сопоставления точек будет уменьшена проблема сопоставления близкорасположенных точек, которые могут иметь похожие дескрипторы. Таким обра-

**Таблица 1.** Оценка повторяемости для нескольких значений  $\sigma$  для порога повторяемости  $\theta = 5$  с различными комбинациями применения биннинга и NMS. Числом записано среднее количество повторяемых точек на одну пару изображений, а в скобках их доля среди всех точек

а) Без биннинга и без NMS				
SIFT	$\sigma_{sar} = 2.0$	$\sigma_{sar} = 2.5$	$\sigma_{sar} = 3.0$	$\sigma_{sar} = 3.5$
$\sigma_{opt} = 2.0$	188.9 (11.8%)	198.5 (12.4%)	194.6 (12.2%)	181.3 (11.3%)
$\sigma_{opt} = 2.5$	190.2 (11.9%)	204.6 (12.8%)	<b>206.5 (12.9%)</b>	194.2 (12.1%)
$\sigma_{opt} = 3.0$	179.3 (11.2%)	195.2 (12.2%)	206.0 (12.9%)	201.1 (12.6%)
$\sigma_{opt} = 3.5$	161.4 (10.1%)	177.9 (11.1%)	194.9 (12.2%)	198.8 (12.4%)
б) Без биннинга и с NMS				
SIFT	$\sigma_{sar} = 2.0$	$\sigma_{sar} = 2.5$	$\sigma_{sar} = 3.0$	$\sigma_{sar} = 3.5$
$\sigma_{opt} = 2.0$	119.8 (11.5%)	126.1 (11.9%)	126.3 (11.8%)	119.4 (10.9%)
$\sigma_{opt} = 2.5$	123.9 (11.6%)	133.6 (12.4%)	<b>137.6 (12.5%)</b>	131.0 (11.7%)
$\sigma_{opt} = 3.0$	119.4 (10.9%)	129.5 (11.7%)	139.6 (12.4%)	138.8 (12.2%)
$\sigma_{opt} = 3.5$	108.2 (9.8%)	119.5 (10.7%)	133.6 (11.8%)	138.9 (12.1%)
в) С биннингом и без NMS				
SIFT	$\sigma_{sar} = 2.0$	$\sigma_{sar} = 2.5$	$\sigma_{sar} = 3.0$	$\sigma_{sar} = 3.5$
$\sigma_{opt} = 2.0$	164.1 (10.3%)	170.9 (10.8%)	169.2 (10.7%)	159.1 (10.1%)
$\sigma_{opt} = 2.5$	165.9 (10.4%)	175.7 (11.1%)	<b>180.7 (11.5%)</b>	173.7 (11.0%)
$\sigma_{opt} = 3.0$	153.8 (9.7%)	167.2 (10.6%)	178.9 (11.4%)	176.4 (11.2%)
$\sigma_{opt} = 3.5$	140.1 (8.8%)	153.6 (9.7%)	172.6 (11.0%)	174.1 (11.1%)
г) С биннингом и с NMS				
SIFT	$\sigma_{sar} = 2.0$	$\sigma_{sar} = 2.5$	$\sigma_{sar} = 3.0$	$\sigma_{sar} = 3.5$
$\sigma_{opt} = 2.0$	104.8 (10.0%)	110.2 (10.4%)	111.6 (10.4%)	105.8 (9.7%)
$\sigma_{opt} = 2.5$	108.6 (10.1%)	116.7 (10.8%)	<b>122.0 (11.1%)</b>	118.4 (10.6%)
$\sigma_{opt} = 3.0$	103.3 (9.5%)	112.6 (10.3%)	122.4 (11.0%)	123.0 (10.9%)
$\sigma_{opt} = 3.5$	94.0 (8.5%)	104.4 (9.4%)	118.9 (10.6%)	122.9 (10.8%)

зом, в дальнейшем будем использовать устойчивые точки, найденные при помощи NMS.

В табл. 1, в, г представлены результаты аналогичных экспериментов, только с применением биннинга с окном  $256 \times 256$  (без NMS в табл. 1, в и с NMS в табл. 1, г). Количество и доля повторяемых точек с применением биннинга уменьшились примерно на 4–10% относительно значений без биннинга. Оптимальные значения  $\sigma_{opt}$  и  $\sigma_{sar}$  сохранились. Напомним, что биннинг позволяет более равномерно распределить точки по изображению, что может улучшить точность сопоставления изображений на этапе нахождения геометрического преобразования в зависимости от используемой геометрической модели. Поэтому в следующем эксперименте мы оценим влияние биннинга и NMS на точность сопоставления изображений.

Для этого по найденным устойчивым точкам вычислялась точность сопоставления изображений (п. 3) для четырех комбинаций точек из про-

шлых экспериментов: без биннинга и без NMS, только с NMS, только с биннингом и с биннингом и с NMS. В качестве дескриптора использовался SIFT-дескриптор, а в качестве геометрической модели – голосующая схема (п. 3.4). Напомним, что параметры настраивались только на обучающей части датасета. Порядок подбора оптимальных параметров был следующим:

- Размер патча для вычисления дескриптора брался  $64 \times 64$  пикселя.

- Начальным дескрипторным порогом брались значение 175.

- Далее настраивались параметры голосующей схемы. Для этого фиксировался размер бина гистограммы  $size_{bin} = \{1, 2, 3\}$  и для каждого  $size_{bin}$  находилось оптимальное значение  $\sigma_m$ . Получаем три числа точности сопоставления изображений, из которых выбираем наибольшее.

**Таблица 2.** Оптимальные параметры и точность сопоставления изображений для четырех комбинаций точек с использованием биннинга и NMS

SIFT	Без модификаций	NMS	Биннинг	Биннинг + NMS
Параметры	<ul style="list-style-type: none"> <li>• <math>des\_thr: 175</math></li> <li>• <math>size_{bin}: 2</math></li> <li>• <math>\sigma_m: 5</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>des\_thr: 175</math></li> <li>• <math>size_{bin}: 1</math></li> <li>• <math>\sigma_m: 5</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>des\_thr: 175</math></li> <li>• <math>size_{bin}: 2</math></li> <li>• <math>\sigma_m: 4</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>des\_thr: 160</math></li> <li>• <math>size_{bin}: 2</math></li> <li>• <math>\sigma_m: 3</math></li> </ul>
Обучающая	53/76	53/76	54/76	55/76
Тестовая	12/24	12/24	13/24	17/24
Итого	65/100	65/100	67/100	72/100

**Таблица 3.** Точность сопоставления изображений для двух методов дескрибирования: SIFT и предлагаемый в статье метод на полноразмерных изображениях (первые две колонки) и на даунскейленных в два раза изображениях. В первой строке приведены параметры для преобработки, голосующей схемы и дескрипторный порог

Дескриптор	SIFT	Наш	SIFT + даунскейл	Наш + даунскейл
Параметры	<ul style="list-style-type: none"> <li>• <math>des\_thr: 160</math></li> <li>• <math>size_{bin}: 2</math></li> <li>• <math>\sigma_m: 3</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>\sigma_{opt}^{pre}: 1</math></li> <li>• <math>\sigma_{sar}^{pre}: 2.5</math></li> <li>• <math>w: 1</math></li> <li>• <math>size_{bin}: 1</math></li> <li>• <math>\sigma_m: 10</math></li> <li>• <math>des\_thr: 2</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>des\_thr: 160</math></li> <li>• <math>size_{bin}: 0.5</math></li> <li>• <math>\sigma_m: 3</math></li> </ul>	<ul style="list-style-type: none"> <li>• <math>\sigma_{opt}^{pre}: 0.75</math></li> <li>• <math>\sigma_{sar}^{pre}: 1.5</math></li> <li>• <math>w: 0</math></li> <li>• <math>size_{bin}: 0.75</math></li> <li>• <math>\sigma_m: 3</math></li> <li>• <math>des\_thr: 1</math></li> </ul>
Обучающая	55/76	57/76	41/76	57/76
Тестовая	17/24	16/24	13/24	17/24
Итого	72/100	73/100	54/100	74/100

– В завершение донастраивался дескрипторный порог  $des\_thr$  при фиксированных найденных параметрах голосующей схемы.

Используемые параметры применяются ко всем парам изображений из датасета. С одной стороны, высокий размер  $size_{bin}$  приводит к меньшей точности сопоставления из-за уменьшенной дискретности (сдвиг вдоль каждой оси кратен  $size_{bin}$ ). С другой стороны, слишком низкий размер бина приводит к большому влиянию шума (слишком маленькие бины будут содержать несколько (единицы) точек). Применение гауссового фильтра позволяет учитывать сдвиги в соседних бинах.

Значения параметров и точность сопоставления изображений показаны в табл. 2.

Из таблицы видно, что наилучший результат наблюдается при применении биннинга и NMS вместе, подтверждая гипотезу, что более равномерное распределение точек увеличивает точность сопоставления изображений. В дальнейших экспериментах будем использовать набор точек, полученных с применением биннинга и NMS. Отметим, что без биннинга применение NMS не изменило результат. Это можно объяснить тем, что без биннинга точки более сконцентрированы в определенных местах изображения и в результате фильтрации многие из них удаляют-

ся. В случае с биннингом точки более распределены и “скопления” точек на изображении больше.

#### 4.2. Предлагаемый метод дескрибирования

В этом разделе представлены результаты нашего метода дескрибирования, описанного в п. 3.6, и проведено их сравнение с SIFT дескриптором. В качестве устойчивых точек были использованы точки, найденные в описанном выше эксперименте с применением биннинга и NMS. Параметры и результаты нашего метода дескрибирования представлены в табл. 3.

Параметры настраивались в порядке упоминания в списке по максимизации точности сопоставления изображений на обучающей выборке. Окно для морфологии  $w = 0$  означает отсутствие морфологии. Первые два столбца отражают точность сопоставления изображений алгоритмов, примененных к полноразмерным изображениям (как и в предыдущих экспериментах). Согласно табл. 3, наш метод дескрибирования немного выигрывает у SIFT дескриптора на полноразмерных картинках.

Поскольку изображения подвержены сильному зернистому шуму, было предложено произвести даунскейлинг изображений в 2 раза для уменьшения влияния шума. В этом случае размер



изображений меняется с  $1024 \times 1024$  до  $512 \times 512$ . Окно биннинга было изменено с  $256 \times 256$  до  $128 \times 128$ , а окно фильтрации NMS с 5 пикселей до 2.5. Размер патча для вычисления дескриптора остался прежним  $64 \times 64$  пикселя. Порог для сопоставления изображений был уменьшен с 5 пикселей до 2.5. Оптимальные параметры детектора подбирались аналогично описанному в п. 4.1. Параметры детекторов получились следующие:  $\sigma_{opt} = 1.5$  и  $\sigma_{sar} = 1.5$ .

Сравнение методов дескрибирования на даунскейленных картинках представлены в табл. 3 (последние два столбца). Алгоритм с предлагаемым методом дескрибирования немного улучшился, в то время как алгоритм с SIFT дескриптором сильно ухудшился.

Согласно результатам табл. 3 можно наблюдать, что алгоритм с предлагаемым методом дескрибирования показывает более высокую точность сопоставления, чем с SIFT дескриптором при оценке с использованием голосующей схемы.

## ЗАКЛЮЧЕНИЕ

В данной работе было продемонстрировано отсутствие общепринятого подхода в вычислении повторяемости устойчивых точек, что не позволяет объективно оценить точность алгоритмов детектирования. В статье приведены некоторые работы и используемые в них метрики повторяемости.

На опубликованном датасете (Волков, Швец, 2021) оптических и радиолокационных изображений было продемонстрировано влияние алгоритмов NMS (non-max-suppression) и биннинга на повторяемость и точность сопоставления изображений на примере SIFT детектора и дескриптора. Для этого датасет из 100 выравненных пар изображений был разделен на обучающую часть (76 пар) и тестовую часть (24 пары). Подбор оптимальных параметров алгоритма происходил только на обучающей части. Было показано, что применение алгоритмов NMS и биннинга уменьшает повторяемость устойчивых точек, но точность сопоставления изображений выше с применением этих алгоритмов. Наилучшая точность сопоставления наблюдалась при совместном использовании NMS и биннинга. В исследовании рассматривались только сдвиговые искажения изображений.

Дополнительно в данной работе предлагается новый метод дескрибирования устойчивых точек, являющийся модификацией SIFT дескриптора. Данный метод сравнивался с SIFT дескриптором по точности сопоставления изображений. Эксперименты показали, что наш метод показал более лучший результат по сравнению с SIFT дескриптором.

## ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено за счет гранта Российского научного фонда (проект № 20-61-47089).

## СПИСОК ЛИТЕРАТУРЫ

- Волков В.В., Швец Е.А. Набор данных и метод оценки алгоритмов сопоставления оптических и радиолокационных изображений на основе устойчивых точек. *Информационные технологии и вычислительные системы*. 2021. № 2. С. 44–57. <https://doi.org/10.14357/20718632210205>
- Сидорчук Д.С., Волков В.В. Комплексирование радиолокационных изображений и оптических снимков в видимом и тепловом диапазонах с учетом различий в восприятии яркости и цветности. *Сенсорные системы*. 2018. Т. 32. № 1. С. 14–18. <https://doi.org/10.7868/S0235009218010031>
- Тропин Д.В., Шемякина Ю.А., Коноваленко И.А., Фараджев И.А. О локализации плоских объектов на изображениях со сложной структурой проективных искажений. *Информационные процессы*. 2019. Т. 19. № 2. С. 208–229.
- Чеканов М.О., Шипитько О.С., Ершов Е.И. Одноточечный RANSAC для оценки величины осевого вращения объекта по томографическим проекциям. *Сенсорные системы*. 2020. Т. 34. № 1. С. 72–86. <https://doi.org/10.31857/S0235009220010060>
- Шемякина Ю.А. Использование точек и прямых для вычисления проективного преобразования по двум изображениям плоского объекта. *Информационные технологии и вычислительные системы*. 2017. Т. 3. С. 79–91.
- Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-up robust features (SURF). *Computer vision and image understanding*. 2008. V. 110. № 3. P. 346–359. <https://doi.org/10.1016/j.cviu.2007.09.014>
- Chen M., Habib A., He H., Zhu Q., Zhang W. Robust feature matching method for SAR and optical images by using Gaussian-gamma-shaped bi-windows-based descriptor and geometric constraint. *Remote Sensing*. 2017. V. 9. № 9. P. 882. <https://doi.org/10.3390/rs9090882>
- Copernicus Open Access Hub, Terms and Conditions. URL: <https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions> (дата обращения 18.04.2022).
- Copernicus Open Access Hub. URL: <https://scihub.copernicus.eu/> (дата обращения 18.04.2022).
- Errico A., Angelino C.V., Cicala L., Persechino G., Ferrara C., Lega M., Vallario A., Parente C., Masi G., Gaetano R., Scarpa G. Detection of environmental hazards through the feature-based fusion of optical and SAR data: A case study in southern Italy. *International Journal of Remote Sensing*. 2015. V. 36. № 13. P. 3345–3367. <https://doi.org/10.1080/01431161.2015.1054960>
- Fan B., Huo C., Pan C., Kong Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geoscience and Remote Sensing Letters*. 2012. V. 10. № 4. P. 657–661. <https://doi.org/10.1109/LGRS.2012.2216500>

- Fan J., Wu Y., Li M., Liang W., Cao Y. SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor. *IEEE Transactions on Geoscience and Remote Sensing*. 2018. V. 56. № 9. P. 5368–5379. <https://doi.org/10.1109/TGRS.2018.2815523>
- Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981. V. 24. № 6. P. 381–395. <https://doi.org/10.1145/358669.358692>
- Gong M., Zhao S., Jiao L., Tian D., Wang S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Transactions on Geoscience and Remote Sensing*. 2013. V. 52. № 7. P. 4328–4338. <https://doi.org/10.1109/TGRS.2013.2281391>
- Hamdi I., Tounsi Y., Benjelloun M., Nassim A. Evaluation of the change in synthetic aperture radar imaging using transfer learning and residual network. *Computer Optics*. 2021. V. 45 (4). P. 600–607. <https://doi.org/10.18287/2412-6179-CO-814>
- Kunina I., Panfilova E., Gladkov A. Matching of SAR and optical images by independent referencing to vector map. In *Eleventh International Conference on Machine Vision (ICMV 2018)*. 2019. V. 11041. <https://doi.org/10.1117/12.2523132>
- Leutenegger S., Chli M., Siegwart R.Y. BRISK: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*. 2011. P. 2548–2555. <https://doi.org/10.1109/ICCV.2011.6126542>
- Li J., Hu Q., Ai M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Transactions on Image Processing*. 2019. V. 29. P. 3296–3310. <https://doi.org/10.1109/TIP.2019.2959244>
- Lowe D.G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*. 2004. V. 60. № 2. P. 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- 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 Geoscience and Remote Sensing Letters*. 2016. V. 14. № 1. P. 3–7. <https://doi.org/10.1109/LGRS.2016.2600858>
- Matas J., Chum O., Urban M., Pajdla T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*. 2004. V. 22. № 10. P. 761–767. <https://doi.org/10.1016/j.imavis.2004.02.006>
- OpenCV: Main page. URL: <https://docs.opencv.org/master/index.html> (дата обращения 18.04.2022).
- Paul S., Pati U.C. Automatic optical-to-SAR image registration using a structural descriptor. *IET Image Processing*. 2019. V. 14. № 1. P. 62–73. <https://doi.org/10.1049/iet-ipr.2019.0389>
- Paul S., Pati U.C. Optical-to-SAR image registration using modified distinctive order based self-similarity operator. In *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. 2018. P. 1–5. <https://doi.org/10.1109/SCEECS.2018.8546950>
- Plank S., Twele A., Martinis S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. *Remote Sensing*. 2016. V. 8. № 4. P. 307. <https://doi.org/10.3390/rs8040307>
- Rosten E., Porter R., Drummond T. Faster and better: A machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence*. 2008. V. 32. № 1. P. 105–119. <https://doi.org/10.1109/TPAMI.2008.275>
- Schmid C., Mohr R., Bauckhage C. Evaluation of interest point detectors. *International Journal of computer vision*. 2000. V. 37. № 2. P. 151–172. <https://doi.org/10.1023/A:1008199403446>
- Shi W., Su F., Wang R., Fan J. A visual circle based image registration algorithm for optical and SAR imagery. In *2012 IEEE International Geoscience and Remote Sensing Symposium*. 2012. P. 2109–2112. <https://doi.org/10.1109/IGARSS.2012.6351089>
- Skoryukina N., Faradjev I., Bulatov K., Arlazarov V.V. Impact of geometrical restrictions in RANSAC sampling on the ID document classification. In *Twelfth International Conference on Machine Vision (ICMV 2019)*. 2020. V. 11433. <https://doi.org/10.1117/12.2559306>
- Suri S., Reinartz P. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*. 2009. V. 48. № 2. P. 939–949. <https://doi.org/10.1109/TGRS.2009.2034842>
- Tropin D.V., Nikolaev D.P., Slugin D.G. The method of image alignment based on sharpness maximization. In *Eleventh International Conference on Machine Vision (ICMV 2018)*. 2019. V. 11041. <https://doi.org/10.1117/12.2522903>
- Wang H., Wang C., Li P., Chen Z., Cheng M., Luo L., Liu Y. Optical-to-SAR Image Registration Based On Gaussian Mixture Model. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2012. V. 39. P. 179–183.
- Wang L., Sun M., Liu J., Cao L., Ma G. A Robust Algorithm Based on Phase Congruency for Optical and SAR Image Registration in Suburban Areas. *Remote Sensing*. 2020. V. 12. № 20. P. 3339. <https://doi.org/10.3390/rs12203339>
- 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 Transactions on Geoscience and Remote Sensing*. 2018. V. 56. № 6. P. 3078–3090. <https://doi.org/10.1109/TGRS.2018.2790483>
- Xie Z., Liu J., Liu C., Zuo Y., Chen X. Optical and SAR Image Registration Using Complexity Analysis and Binary Descriptor in Suburban Areas. *IEEE Geoscience and Remote Sensing Letters*. 2021. V. 19. P. 1–5. <https://doi.org/10.1109/LGRS.2021.3071870>
- Xiong X., Xu Q., Jin G., Zhang H., Gao X. Rank-Based Local Self-Similarity Descriptor for Optical-to-SAR Image Matching. *IEEE Geoscience and Remote Sensing*

- Letters*. 2019. V. 17. № 10. P. 1742–1746.  
<https://doi.org/10.1109/LGRS.2019.2955153>
- Xu C., Sui H., Li H., Liu J. An automatic optical and SAR image registration method with iterative level set segmentation and SIFT. *International Journal of Remote Sensing*. 2015. V. 36. № 15. P. 3997–4017.  
<https://doi.org/10.1080/01431161.2015.1070321>
- Ye S.P., Chen C.X., Nedzved A., Jiang J. Building detection by local region features in SAR images. *Computer Optics*. 2020. V. 44. № 6. P. 944–950.  
<https://doi.org/10.18287/2412-6179-CO-703>
- Yu H., Yang W., Liu Y. Coarse-to-fine accurate registration for airborne Sar images using SAR-FAST and DSP-LATCH. *Progress In Electromagnetics Research*. 2018. V. 163. P. 89–106.  
<https://doi.org/10.2528/PIER18070801>
- Yu Q., Ni D., Jiang Y., Yan Y., An J., Sun T. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021. V. 171. P. 1–17.  
<https://doi.org/10.1016/j.isprsjprs.2020.10.019>

## Modification of the method of detecting and describing keypoints SIFT for optical-to-SAR image registration

V. V. Volkov<sup>a,b</sup> #

<sup>a</sup> *Moscow Institute of Physics and Technology,  
141701 Dolgoprudny, Institutskiy per., 9, Russia*

<sup>b</sup> *Institute for Information Transmission Problems (Kharkevich Institute) RAS,  
127051 Moscow, Bolshoy Karetny per., 19, Russia*

# *E-mail: volkov-vl-v@yandex.ru*

Image registration is the alignment (i.e. finding a general coordinate system) of two or more images of the same scene. Complex case of this task is the multimodal image registration, for example, optical-to-SAR. The need for such registration appeared in image fusion and object detection. One of the ways for image registration is to detect keypoints, calculate descriptor for each keypoint and form keypoints pairs by comparing descriptors. Further, the geometric transformation between two images is calculated by geometric model. In this paper modifications for the algorithm of detection SIFT keypoints was evaluated for optical-to-SAR image registration. Additionally, in this paper a modification of the SIFT descriptor was proposed and compared with SIFT descriptor. Image distortions is considered to be only shifts. All studies were carried out on a published dataset of 100 aligned pairs of optical-SAR images.

*Key words:* image registration, repeatability of keypoints, description of keypoints

### REFERENCES

- Volkov V.V., Shvets E.A. Dataset and Method for Evaluating Optical-to-Sar Image Registration Algorithms Based on Keypoints. *Informacionnye tekhnologii i vychislitel'nye sistemy* [Journal of Information Technologies and Computing Systems]. 2021. V. 2. P. 44–57.  
<https://doi.org/10.14357/20718632210205> (in Russian).
- Sidorchuk D.S., Volkov V.V. Fusion of radar, visible and thermal imagery with account for differences in brightness and chromaticity perception. *Sensornye sistemy* [Sensory systems]. 2018. V. 32 (1). P. 14–18.  
<https://doi.org/10.7868/S0235009218010031> (in Russian).
- Tropin D.V., Shemiakina J.A., Konovalenko I.A., Faradjev I.A. Localization of planar objects on the images with complex structure of projective distortion. *Informacionnye processy* [Information processes]. 2019. V. 19 (2). P. 208–229 (in Russian).
- Chekanov M.O., Shipitko O.S., Ershov E.I. 1-point RANSAC for axial rotation angle estimation by tomographic projections. *Sensornye sistemy* [Sensory systems]. 2020. V. 34 (1). P. 72–86.  
<https://doi.org/10.31857/S0235009220010060> (in Russian).
- Shemiakina J.A. Ispol'zovanie tochek i pryamyh dlya vychisleniya proektivnogo preobrazovaniya po dvum izobrazheniyam ploskogo ob'ekta [Using points and straight lines to calculate a projective transformation from two images of a flat object]. *Informacionnye tekhnologii i vychislitel'nye sistemy* [Journal of Information Technologies and Computing Systems]. 2017. V. 3. P. 79–91 (in Russian).
- Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-up robust features (SURF). *Computer vision and image understanding*. 2008. V. 110. № 3. P. 346–359.  
<https://doi.org/10.1016/j.cviu.2007.09.014>
- Chen M., Habib A., He H., Zhu Q., Zhang W. Robust feature matching method for SAR and optical images by using Gaussian-gamma-shaped bi-windows-based descriptor and geometric constraint. *Remote Sensing*. 2017. V. 9. № 9. P. 882.  
<https://doi.org/10.3390/rs9090882>
- Copernicus Open Access Hub, Terms and Conditions. URL: <https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/TermsConditions> (дата обращения 18.04.2022).
- Copernicus Open Access Hub. URL: <https://scihub.copernicus.eu/> (дата обращения 18.04.2022).

- Errico A., Angelino C.V., Cicala L., Persechino G., Ferrara C., Lega M., Vallario A., Parente C., Masi G., Gaetano R., Scarpa G. Detection of environmental hazards through the feature-based fusion of optical and SAR data: A case study in southern Italy. *International Journal of Remote Sensing*. 2015. V. 36. № 13. P. 3345–3367.  
<https://doi.org/10.1080/01431161.2015.1054960>.
- Fan B., Huo C., Pan C., Kong Q. Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT. *IEEE Geoscience and Remote Sensing Letters*. 2012. V. 10. № 4. P. 657–661.  
<https://doi.org/10.1109/LGRS.2012.2216500>.
- Fan J., Wu Y., Li M., Liang W., Cao Y. SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor. *IEEE Transactions on Geoscience and Remote Sensing*. 2018. V. 56. № 9. P. 5368–5379.  
<https://doi.org/10.1109/TGRS.2018.2815523>.
- Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981. V. 24. № 6. P. 381–395.  
<https://doi.org/10.1145/358669.358692>.
- Gong M., Zhao S., Jiao L., Tian D., Wang S. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Transactions on Geoscience and Remote Sensing*. 2013. V. 52. № 7. P. 4328–4338.  
<https://doi.org/10.1109/TGRS.2013.2281391>.
- Hamdi I., Tounsi Y., Benjelloun M., Nassim A. Evaluation of the change in synthetic aperture radar imaging using transfer learning and residual network. *Computer Optics*. 2021. V. 45 (4). P. 600–607.  
<https://doi.org/10.18287/2412-6179-CO-814>.
- Kunina I., Panfilova E., Gladkov A. Matching of SAR and optical images by independent referencing to vector map. In *Eleventh International Conference on Machine Vision (ICMV 2018)*. 2019. V. 11041.  
<https://doi.org/10.1117/12.2523132>.
- Leutenegger S., Chli M., Siegwart R.Y. BRISK: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*. 2011. P. 2548–2555.  
<https://doi.org/10.1109/ICCV.2011.6126542>.
- Li J., Hu Q., Ai M. RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Transactions on Image Processing*. 2019. V. 29. P. 3296–3310.  
<https://doi.org/10.1109/TIP.2019.2959244>.
- Lowe D.G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*. 2004. V. 60. № 2. P. 91–110.  
<https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- 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 Geoscience and Remote Sensing Letters*. 2016. V. 14. № 1. P. 3–7.  
<https://doi.org/10.1109/LGRS.2016.2600858>.
- Matas J., Chum O., Urban M., Pajdla T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*. 2004. V. 22. № 10. P. 761–767.  
<https://doi.org/10.1016/j.imavis.2004.02.006>.
- OpenCV: Main page. URL: <https://docs.opencv.org/master/index.html> (дата обращения 18.04.2022).
- Paul S., Pati U.C. Automatic optical-to-SAR image registration using a structural descriptor. *IET Image Processing*. 2019. V. 14. № 1. P. 62–73.  
<https://doi.org/10.1049/iet-ipr.2019.0389>.
- Paul S., Pati U.C. Optical-to-SAR image registration using modified distinctive order based self-similarity operator. In *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. 2018. P. 1–5.  
<https://doi.org/10.1109/SCEECS.2018.8546950>.
- Plank S., Twele A., Martinis S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. *Remote Sensing*. 2016. V. 8. № 4. P. 307.  
<https://doi.org/10.3390/rs8040307>.
- Rosten E., Porter R., Drummond T. Faster and better: A machine learning approach to corner detection. *IEEE transactions on pattern analysis and machine intelligence*. 2008. V. 32. № 1. P. 105–119.  
<https://doi.org/10.1109/TPAMI.2008.275>.
- Schmid C., Mohr R., Bauckhage C. Evaluation of interest point detectors. *International Journal of computer vision*. 2000. V. 37. № 2. P. 151–172.  
<https://doi.org/10.1023/A:1008199403446>.
- Shi W., Su F., Wang R., Fan J. A visual circle based image registration algorithm for optical and SAR imagery. In *2012 IEEE International Geoscience and Remote Sensing Symposium*. 2012. P. 2109–2112.  
<https://doi.org/10.1109/IGARSS.2012.6351089>.
- Skoryukina N., Faradjev I., Bulatov K., Arlazarov V.V. Impact of geometrical restrictions in RANSAC sampling on the ID document classification. In *Twelfth International Conference on Machine Vision (ICMV 2019)*. 2020. V. 11433.  
<https://doi.org/10.1117/12.2559306>.
- Suri S., Reinartz P. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*. 2009. V. 48. № 2. P. 939–949.  
<https://doi.org/10.1109/TGRS.2009.2034842>.
- Tropin D.V., Nikolaev D.P., Slugin D.G. The method of image alignment based on sharpness maximization. In *Eleventh International Conference on Machine Vision (ICMV 2018)*. 2019. V. 11041.  
<https://doi.org/10.1117/12.2522903>.
- Wang H., Wang C., Li P., Chen Z., Cheng M., Luo L., Liu Y. Optical-to-SAR Image Registration Based On Gaussian Mixture Model. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2012. V. 39. P. 179–183.
- Wang L., Sun M., Liu J., Cao L., Ma G. A Robust Algorithm Based on Phase Congruency for Optical and SAR Image Registration in Suburban Areas. *Remote Sensing*. 2020. V. 12. № 20. P. 3339.  
<https://doi.org/https://doi.org/10.3390/rs12203339>.

- 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 Transactions on Geoscience and Remote Sensing*. 2018. V. 56. № 6. P. 3078–3090.  
<https://doi.org/10.1109/TGRS.2018.2790483>.
- Xie Z., Liu J., Liu C., Zuo Y., Chen X. Optical and SAR Image Registration Using Complexity Analysis and Binary Descriptor in Suburban Areas. *IEEE Geoscience and Remote Sensing Letters*. 2021. V. 19. P. 1–5.  
<https://doi.org/10.1109/LGRS.2021.3071870>.
- Xiong X., Xu Q., Jin G., Zhang H., Gao X. Rank-Based Local Self-Similarity Descriptor for Optical-to-SAR Image Matching. *IEEE Geoscience and Remote Sensing Letters*. 2019. V. 17. № 10. P. 1742–1746.  
<https://doi.org/10.1109/LGRS.2019.2955153>.
- Xu C., Sui H., Li H., Liu J. An automatic optical and SAR image registration method with iterative level set segmentation and SIFT. *International Journal of Remote Sensing*. 2015. V. 36. № 15. P. 3997–4017.  
<https://doi.org/10.1080/01431161.2015.1070321>.
- Ye S.P., Chen C.X., Nedzved A., Jiang J. Building detection by local region features in SAR images. *Computer Optics*. 2020. V. 44. № 6. P. 944–950.  
<https://doi.org/10.18287/2412-6179-CO-703>.
- Yu H., Yang W., Liu, Y. Coarse-to-fine accurate registration for airborne Sar images using SAR-FAST and DSP-LATCH. *Progress In Electromagnetics Research*. 2018. V. 163. P. 89–106.  
<https://doi.org/10.2528/PIER18070801>.
- Yu Q., Ni D., Jiang Y., Yan Y., An J., Sun T. Universal SAR and optical image registration via a novel SIFT framework based on nonlinear diffusion and a polar spatial-frequency descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2021. V. 171. P. 1–17.  
<https://doi.org/10.1016/j.isprsjprs.2020.10.019>.