

Определение способа хранения многоканальных данных при визуализации изображений на примере SPOT-4

Н.Ю. Севастьянова¹, Н.С. Виноградова¹

¹Уральский федеральный университет, Мира 19, Екатеринбург, Россия, 620002

Аннотация. Одной из особенностей хранения данных, получаемых системами дистанционного зондирования Земли (ДЗЗ), является широкое применение дисковых массивов большой вместимости. Аварийные ситуации, возникающие при использовании массивов, могут приводить к тому, что данные ДЗЗ, хранящиеся в несжатом виде, оказываются частично повреждены. Тем не менее, даже при неполном восстановлении такого рода данные могут быть в дальнейшем использованы при решении производственных задач. Однако такое восстановление иногда бывает затруднено неполным знанием о формате повреждённых данных. В статье рассмотрен подход к восстановлению данных ДЗЗ системы SPOT-4 в сегментном формате (SEG) без указания спецификации, повреждённых при аварии дискового массива.

1. Введение

SPOT – коммерческая оптическая спутниковая система наблюдения Земли высокого разрешения, функционировавшая в 1998–2013 и направленная на решение экологических и сельскохозяйственных задач, а также используемая в таких фундаментальных науках, как климатология и океанография [3,4,5]. Аппарат SPOT-4 являлся многозональной системой наблюдения, осуществлявшей съемку в четырех спектральных диапазонах (табл. 1).

Таблица 1. Каналы SPOT-4 [3].

Съемочный канал	Пространственное разрешение	Спектральный диапазон
панхроматический	10 м	0.61–0.68 мкм
V1: зелёный	20 м	0.50–0.59 мкм
V2: красный	20 м	0.61–0.68 мкм
V3: ближний инфракрасный	20 м	0.78–0.89 мкм

В 2011–2012 гг. Центр космического мониторинга Уральского федерального университета осуществлял прием сигнала космического аппарата SPOT-4, накопив более 1 Тб съемочных данных в исходном «сыром» формате (RAW) и в предобработанном формате SEG. После аварии дискового хранилища в 2016 году данные оказались временно утрачены, однако предварительный анализ результатов проведенных ремонтно-восстановительных работ показал, что окончательно потеряна только та часть данных, которая хранилась на конкретных вышедших из строя носителях (3 дисковых носителя из 20). При этом файловая система хранилища осталась слабо поврежденной, что позволило восстановить большую часть файлов

(в том числе и SEG), а цепочки потерянных фрагментов хотя и записывались нулевыми байтами, но сохраняли локализацию внутри файла. Поскольку данные SPOT-4 по настоящее время активно используются в образовательном процессе, крайне актуальным оказался вопрос о возможности визуализации и дальнейшего использования этих данных. В ходе решения данной задачи среди прочих возникла проблема опознавания способа хранения многоканальных данных. Существует три основных способа записи данных в многоканальных изображениях [6]:

- ВIL – каналы, разделенные по строкам;
- VIP – попиксельная запись спектральных каналов;
- BSQ – поканальная запись.

Представленные способы хранения имеют ассоциированный заголовок, который, как правило, представлен в кодировке ASCII. Этот заголовок содержит основную информацию об изображении, такую как формат, число строк и столбцов, наличие цветовой карты, также может содержать вспомогательную информацию, например, географическую привязку, авторство, время съемки.

VIL использует способ записи многоканальных изображений, при котором строки значений для каждого канала записываются последовательно. Например, в случае трехканального изображения, все три канала данных сначала записываются в первую строку, затем во вторую строку, и так далее, пока не будет записана последняя строка изображения.

VIP отличается от VIL тем, что данные для каждого пиксела записываются попиксельно. Например, при записи трехканального изображения данные для первого, второго и третьего каналов записываются для первого пиксела первого столбца; затем данные каналов записываются для первого пиксела второго столбца и так далее.

При поканальной записи (BSQ), информация для каждого канала записывается полностью. Другими словами, данные всех пикселов первого канала записываются в первую очередь, затем записываются все пиксели второго канала и так далее [6].

2. Алгоритм анализа данных SPOT-4

В качестве опорной информации использованы размеры изображений в байтах, которые могут быть совершенно различными, и количество каналов SPOT-4, равное четырем. Для анализа данных в среде MATLAB составлен алгоритм, позволяющий выбирать и анализировать способ хранения многозональных изображений. На первом этапе выполнена факторизация данных – разложение в произведение простых множителей, что позволило разбить размер изображения для определения столбцов и строк снимков. Для того, чтобы установить нужный способ записи данных, изображения открыты всеми тремя возможными вариантами. Визуальное сопоставление открытых снимков выявило существенные пространственные различия (рисунок 1).

На изображении, открытом в предположении, что данные сохранены способом VIL, отмечены полосы вдоль столбцов изображения, что связано с нарушением чередования пикселов. Изображение, открытое в предположении, что данные сохранены способом BSQ, так же несет в себе неправильное чередование пикселов, на рисунке видно произвольное чередование пикселов (больше похоже на полосы вдоль диагоналей). Изображение, открытое в предположении, что данные сохранены способом VIP, представляет собой правильное чередование. Таким образом, для данных ДЗЗ SPOT-4 корректным способом хранения многозональных данных является VIP.

2.1. Анализ особенностей VIL-формата

Как отмечено ранее из визуального сопоставления, изображение, открытое в предположении при помощи способа записи VIL, имеет выраженные полосы в направлении столбцов изображения. Для того, чтобы подчеркнуть полосы, использованы методы пространственной фильтрации. Опробованы следующие типы пространственных фильтров [1,2,7]:

- фильтры Лапласа;

- фильтры обнаружения линий;
- оператор Собеля;
- перекрестный фильтр Робертса.

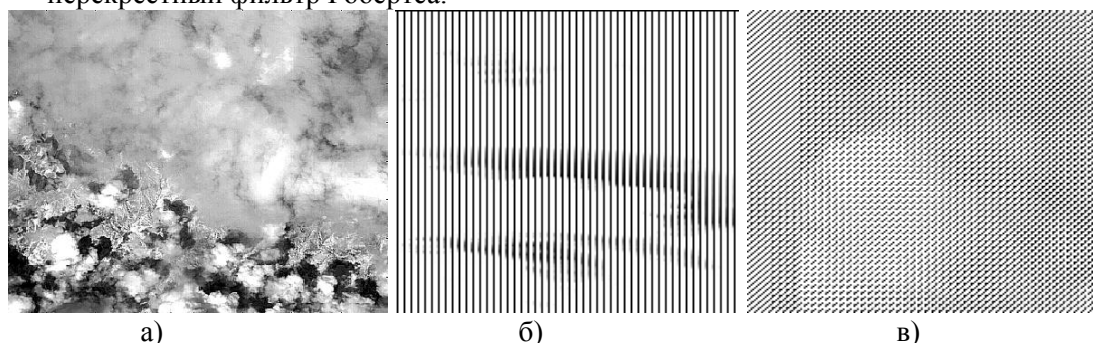


Рисунок 1. Фрагмент изображения, открытое форматами: а – VIP; б – VIL; в – BSQ.

Результаты представлены на рисунке 2.

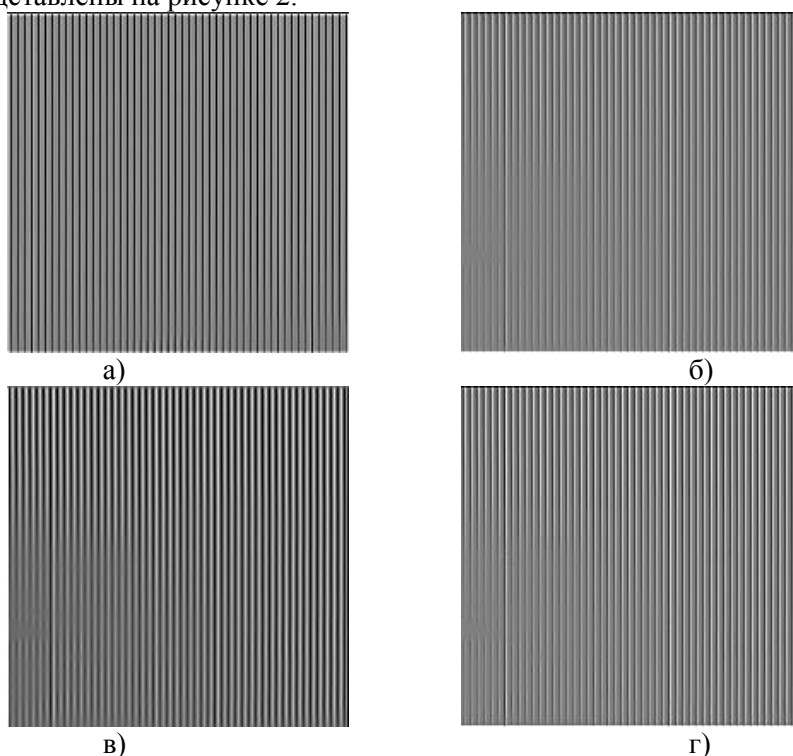


Рисунок 2. Результат фильтрации для изображения (VIL): а – типы Лапласа; б – типы обнаружений линий; в – оператор Собеля; г – перекрестный оператор Робертса.

Для всех результатов фильтрации (при разных каналах и разных фрагментах) посчитана дисперсия. Дисперсия вдоль строки характеризует разброс яркостей изображения, таким образом являясь количественным показателем наличия ориентированных вдоль столбцов изображения полос, возникающих при открытии изображения в предположении, что способом хранения цветных составляющих является VIL. Максимальное значение дисперсии получено при использовании фильтра Собеля. На рисунке 3 представлены результаты дисперсии для изображений, открытых в предположении способами VIP и BSQ, с помощью оператора Собеля. Максимальная дисперсия изображения, открытого в предположении верным способом хранения, не при каких фрагментах и каналах не превышает максимальную дисперсию другого изображения, открытого в предположении неправильным способом хранения. Так же можно заметить, что дисперсия снимка, открытого с правильным чередованием пикселей, не имеет резких перепадов и по сравнению со снимком, открытым при помощи VIL, является более

сглаженной. Таким образом, на основании визуального сопоставления и анализа дисперсии, можно сделать вывод, что оператор Собеля является наилучшим фильтром в задаче выделения линий в направлении столбцов изображения.

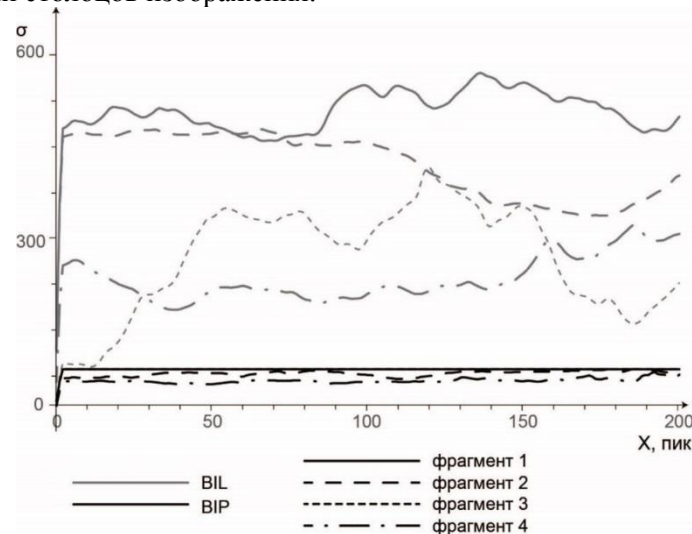


Рисунок 3. Дисперсия для разных фрагментов и каналов ВР и ВЛ.

На рисунке 4 представлены фрагменты отфильтрованных изображений помощи оператора Собеля. Изображение, открытое в предположении способом хранения ВР, не имеет резких изменений границ, в то время как на изображении, открытом в предположении ВЛ, проявляются резкие полосы в направлении столбцов изображения, что выражается количественно при расчете дисперсии изображения. Таким образом, полученный признак может быть использован как показатель корректности в определении формата хранения каналов данных SPOT-4

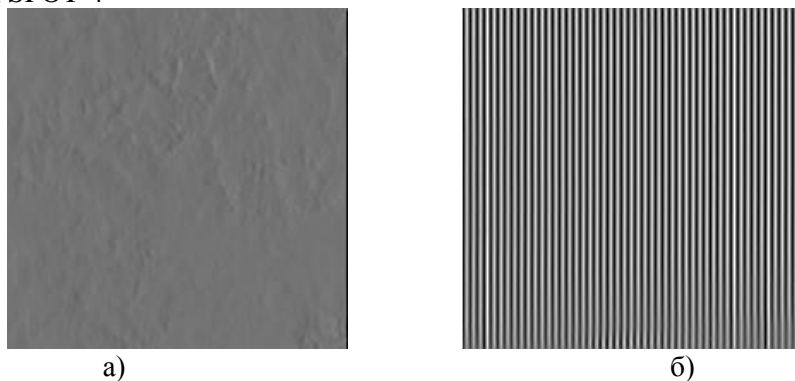


Рисунок 4. Применение оператора Собеля: а – изображение, открытое при помощи ВР; б – изображение, открытое при помощи ВЛ.

2.2. Анализ особенностей BSQ-формата

Как отмечено ранее, на изображении, которое открыто в предположении с помощью способа хранения BSQ, выражено произвольное чередование пикселей (больше похоже полосы вдоль диагоналей). Для того, чтобы подчеркнуть полосы, также использованы методы пространственной фильтрации. В ходе работы опробованные различные фильтры, которые были представлены ранее, а также и другие, такие как типы градиента и фильтрация при помощи гомогенности (однородности) (Рис. 5). Так как оператор Собеля и перекрестный оператор Робинсона используются только для выделения горизонтальных или вертикальных линий, изображение было повернуто на 45 градусов, после того сделана фильтрация. Угол составляет 45 градусов, поскольку SPOT-4 имеет четыре канала.

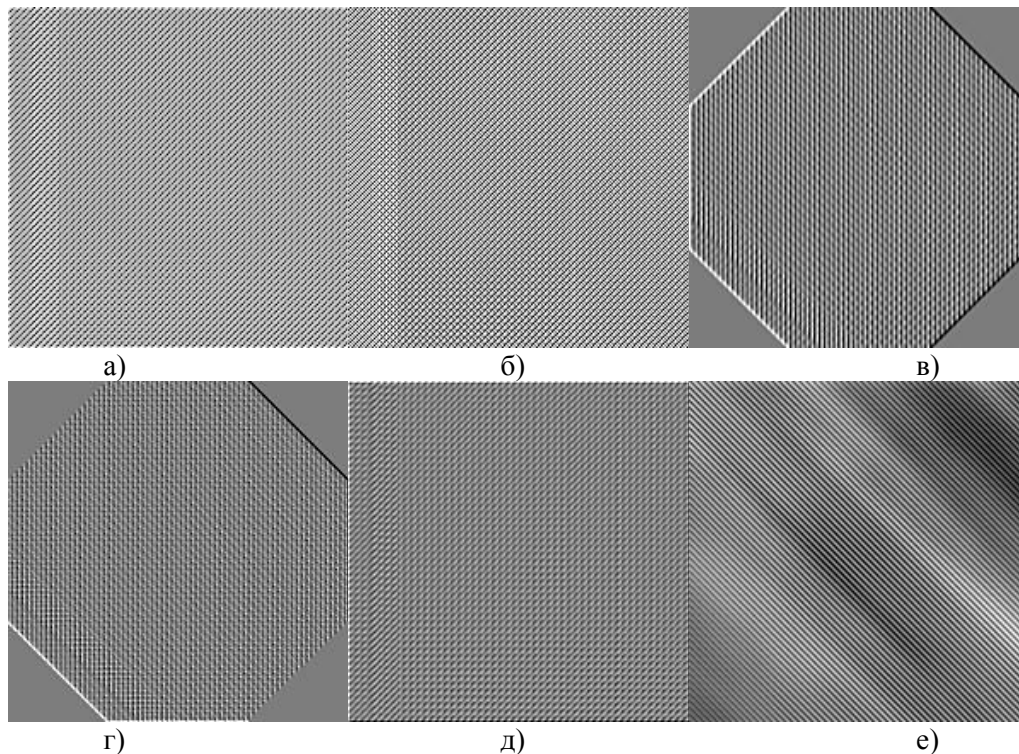


Рисунок 5. Виды фильтрации для изображения (BSQ): а – типы Лапласа; б – типы обнаружений линий; в – оператор Собеля; г – перекрестный оператор Робертса; д – типы градиента; е – гомогенный.

Для всех результатов фильтрации также посчитана дисперсия, которая показывает, что оператор Собеля предоставляет максимальный контраст между изображениями, открытыми в предположении VIP и BSQ. На рисунке 6 представлены результаты дисперсии для изображений, открытых в предположении VIP и BSQ, с помощью оператора Собеля. Дисперсия снимка, открытого в предположении способом VIP, является более сглаженной, чем дисперсия другого изображения. Таким образом, на основании визуального сопоставления и анализа дисперсии, можно сделать вывод, что оператор Собеля в данном случае тоже хорошо подчеркивает выделения линий в направлении столбцов изображения.

На рисунке 7 представлены фрагменты изображений, отфильтрованные при помощи оператора Собеля. Снимок, открытый в предположении при помощи BSQ, имеет вертикальные полосы, на другом снимке, который открыт в предположении при помощи VIP, такого не наблюдается.

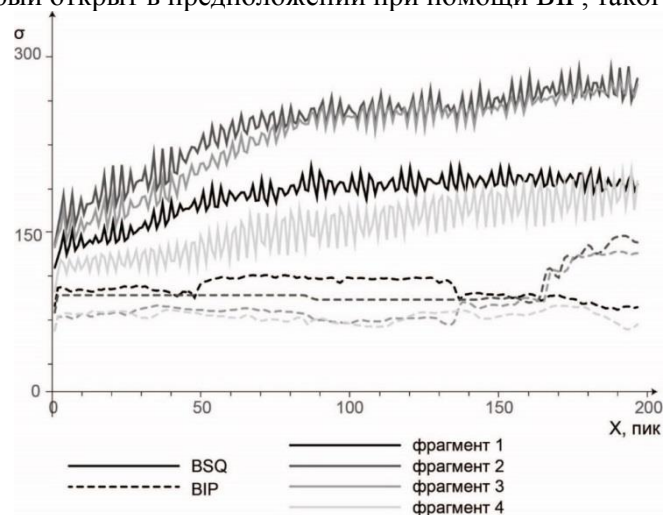


Рисунок 6. Дисперсия для разных фрагментов и каналов VIP и BSQ.

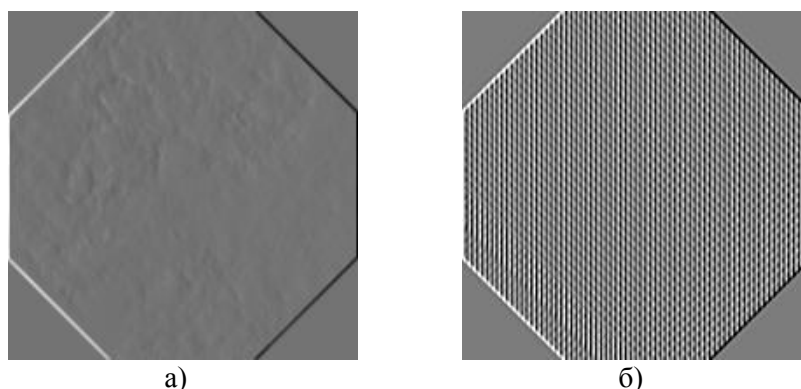


Рисунок 7. Применение оператора Собеля: а – для изображения, открытого при помощи ВР; б – для изображения, открытого при помощи BSQ.

3. Результаты

В среде MATLAB разработан алгоритм, позволяющий автоматически устанавливать корректный способ хранения многозональных изображений на примере данных ДЗЗ SPOT-4. В основе алгоритма лежит пространственная фильтрация Собеля, подчеркивающая геометрические особенности изображения исходного файла и файла, развернутого на угол 45 градусов. В случае некорректного способа открытия файла дисперсия вдоль строки изображения дает завышенные значения, таким образом становится возможным идентифицировать адекватный способ открытия данных, в данном случае это ВР. Алгоритм показывает надежные результаты на восьмидесяти семи различных фрагментах данных ДЗЗ SPOT-4 в четырех различных каналах. Предложенный алгоритм будет использован в дальнейшем для задач автоматизированного восстановления поврежденных участков файлов.

4. Литература

- [1] Сирота, А.А. Статистические алгоритмы обнаружения границ объектов на изображении / А.А. Сирота, А.И. Соломатин. – Вестник ВГУ. – 2008. – № 1. – С. 58-64.
- [2] Локшин, Г.Р. Дифракция. Пространственная фильтрация: учебное пособие по курсу «Общая физика» / Г.Р. Локшин. – Москва: МФТИ, 2016. – 156 с.
- [3] SPOT-4 (Франция) [Электронный ресурс]. – Режим доступа <http://www.geogr.msu.ru/science/projects/geoportal/data/satellites/spot4.php> (29.08.2018).
- [3] Основные технические характеристики [Электронный ресурс]. – Режим доступа <http://www.racurs.ru/?page=294> (29.08.2018).
- [5] SPOT-4 [Электронный ресурс]. – Режим доступа <https://directory.eoportal.org/web/eoportal/satellite-missions/s/spot-4> (02.09.2018).
- [6] Файлы растров ВР, ВР и BSQ [Электронный ресурс]. – Режим доступа <http://desktop.arcgis.com/ru/arcmap/10.3/manage-data/raster-and-images/bil-bip-and-bsq-raster-files.htm> (05.09.2018).
- [7] Методы нахождения границ изображения [Электронный ресурс]. – Режим доступа <https://habr.com/post/128753/> (10.09.2018).

Multi-channel data storage format definition for visualization tasks on the example of SPOT-4 images

N.Yu. Sevastianova¹, N.S. Vinogradova¹

¹Ural Federal University, Mira street 19, Ekaterinburg, Russia, 620002

Abstract. One of the features of data storage obtained by remote sensing systems is the widespread utilization of large-capacity disk arrays. Emergency situations arising from the use of arrays can lead to the fact that the remote sensing data, usually stored in uncompressed form, may become partially damaged. But even with incomplete recovery, this kind of data can be used in the future to solve production problems. However, this recovery is sometimes hampered by incomplete knowledge of the format of the corrupted data. The article describes an approach to recovery of remote sensing data of SPOT-4 system stored in the segment "SEG" format (without specification), damaged in case of disk array failure.