Оптимизация архивов ЦКП «ИКИ-Мониторинг» с использованием методики адаптивного сжатия спутниковых изображений на основе алгоритма LERC с разделением по классам подстилающей поверхности

А. А. Прошин, Е. А. Лупян, М. А. Бурцев

Институт космических исследований РАН, Москва, 117997, Россия
E-mail: andry@smis.iki.rssi.ru

В настоящее время многолетние архивы различных типов спутниковых данных ДЗЗ широко используются для решения самых разных исследовательских и прикладных задач. Стремительное развитие систем ДЗЗ приводит к резкому возрастанию объёмов таких архивов, что делает всё более актуальной задачу поиска новых решений, позволяющих уменьшить объёмы хранения данных и не приводящих к потере качества их использования при решении разных задач. Для решения данной задачи в Институте космических исследований РАН (ИКИ РАН) была разработана новая методика адаптивного сжатия спутниковых изображений, использующая алгоритм с регулируемыми потерями LERC (*англ.* Limited Error Raster Compression). Она позволяет гарантированно сохранять текстуру изображений за счёт учёта статистики для выбора ошибки сжатия по разным классам подстилающей поверхности на каждом конкретном экземпляре данных и одновременно достигать достаточно высоких показателей сжатия — порядка 0,6 относительно данных, сжатых без потерь. Данная методика легла в основу работ по оптимизации архивов ЦКП «ИКИ-Мониторинг», описанию которых посвящена настоящая работа. Проведение такой оптимизации только для текущих архивов данных прибора MSI (англ. Multispectral Instrument) позволит освободить более 1,2 Пбайт мощностей хранения, имеющихся в ЦКП «ИКИ-Мониторинг», что соответствует примерно пятилетнему временному ряду таких данных.

Ключевые слова: спутниковые данные, ДЗЗ, архивы спутниковых данных, эффективное сжатие изображений, системы доступа к спутниковым данным

Одобрена к печати: 07.10.2026
DOI: 10.21046/2070-7401-2026-23-5-435-441

Введение

Данные дистанционного зондирования Земли (ДЗЗ) в настоящее время активно используются для решения широкого спектра научных и прикладных задач, связанных с мониторингом природной среды и антропогенных объектов. В последние десятилетия стремительно растёт как число космических аппаратов ДЗЗ, так и объёмы поступающей от них информации. Это приводит ко всё ускоряющемуся росту объёмов архивов спутниковых данных. В частности, архивы центра коллективного пользования (ЦКП) системами архивации, обработки и анализа данных спутниковых наблюдений для решения задач изучения и мониторинга окружающей среды (ЦКП «ИКИ-Мониторинг») ((Лупян и др., 2015, 2019), http://ckp.geosmis.ru/) в настоящее время содержат уже более 11 Пбайт спутниковых данных. Организация хранения и доступа к таким большим массивам данных очень ресурсоёмка, поэтому оптимизация для снижения объёмов хранения данных без потери их применимости для реальных задач и производительности систем доступа продолжает оставаться очень актуальной задачей.

В 2026 г. по результатам исследований в этой области была начата первая очередь оптимизации архивов данных высокого разрешения в составе ЦКП «ИКИ-Мониторинг» для снижения объёмов хранимой информации. Для этого осуществляется сжатие данных алгоритмом с контролируемой потерей LERC (*англ.* Limited Error Raster Compression) с использованием специализированной адаптивной методики сжатия, учитывающей распределение яркостей по отдельным классам подстилающей поверхности для подбора предельно допустимой ошибки сжатия.

Методика адаптивного LERC-сжатия по классам подстилающей поверхности

В 2024 г. в Институте космических исследований РАН (ИКИ РАН) была разработана методика сжатия спутниковых данных алгоритмом с потерями LERC ((Becker et al., 2016), https://esri.github.io/lerc/), которая достаточно подробно описана в работе (Прошин и др., 2024). Алгоритм позволяет задать максимальную вносимую ошибку в значение каждого пикселя, тем самым сохраняя радиометрические свойства изображения с заданной погрешностью, не внося при этом искажений в текстуру изображения. Сжатие спутниковых изображений с полученной по данной методике ошибкой позволяет уменьшить объём хранения примерно в два раза по сравнению с данными, сжатыми традиционными алгоритмами без потерь (DEFLATE, LZW (*англ.* Lempel-Ziv-Welch) и др.). В 2025 г. было проведено дополнительное тестирование применимости сжатых по этой методике данных для решения реальных задач, которое показало очень хорошие результаты (Прошин и др., 2025).

В основе выработанной методики лежит определение максимально допустимой ошибки посредством статистического анализа тестовой выборки соответствующего типа данных. Выбор ошибки при этом обусловлен такими параметрами, как номинальный диапазон или реальная гистограмма значений данных и радиометрическая точность прибора. Подобранная таким образом ошибка применяется ко всему изображению.

Вместе с тем этот подход, описанный в работах (Прошин и др., 2024, 2025), не учитывал возможное присутствие на изображениях зон (объектов) со значимо различными характеристиками (например, характерные диапазоны яркости и пространственной изменчивости). Так, для оптических данных к наиболее часто встречающимся подобным объектам можно отнести следующие:

- вода — неконтрастная поверхность с низким разбросом значений;
- снег и лёд — яркая, высококонтрастная поверхность;
- суша с растительностью — текстурированная поверхность со средним разбросом яркости;
- облачность — высококонтрастные, шумные объекты с максимальным разбросом.

При одновременном наличии на изображении крупных областей с такими разнородными объектами их гистограмма принимает вид кривой с несколькими ярко выраженными узкими максимумами, что делает подбор ошибки по ней затруднённым, а использование единой усреднённой ошибки при сжатии алгоритмом LERC может приводить к потенциальному искажению текстуры изображений на некоторых из них. Выбор же общей ошибки как ошибки для класса с минимальным разбросом значений может существенно ухудшить эффективность сжатия. Так, тесты показали, что такой подход к выбору ошибки позволяет избежать текстурных искажений на изображении, но обеспечивает выигрыш в сжатии порядка 25 % относительно данных, сжатых без потерь, что недостаточно.

Эти факторы привели к необходимости создания новой усовершенствованной методики сжатия, которая бы гарантировала отсутствие текстурных искажений для всех экземпляров сжатых данных, в том числе на однородных участках, и сохраняла необходимую для оптимизации архивов эффективность сжатия, близкую к полученной ранее, т. е. порядка двукратной. Для достижения этого результата была разработана адаптивная методика сжатия для оптических данных, в рамках которой используется максимальная ошибка, определённая для каждого отдельного изображения с учётом диапазонов спектральной яркости для разных типов подстилающей поверхности, встречающихся на этом изображении.

Созданная методика базируется на том, что на изображении выделяются основные классы подстилающей поверхности и ошибка по каждому из них определяется динамически на основе статистики значений яркости для него. Для динамического определения ошибки $E_c = \sigma_c L$ для каждого класса C используется стандартное отклонение значений внутри этого класса σ_c, умноженное на специально подобранный коэффициент L.

Для нахождения этого коэффициента применяется адаптированный вариант ранее разработанной методики сжатия при помощи алгоритма LERC. Выбирается такой коэффициент, при котором среднее по разнице между исходным и сжатым изображениями становится отрицательным, что наблюдается при последовательном уменьшении ошибки LERC. Обоснование такого подхода для детектирования структурных искажений при использовании алгоритма LERC приводится в ранее упомянутой работе (Прошин и др., 2024).

Хотя стандартное отклонение для неравномерных распределений значений яркости не является точной метрикой, в рамках отдельных классов изображения оно становится эффективным индикатором разброса значений.

В текущей реализации методики на изображении маскируется облачность, после чего отдельно определяются ошибки LERC для следующих типов подстилающей поверхности: вода, снег/лёд и суша — на основе стандартных отклонений соответствующих областей изображения. Облачные части изображения сжимаются с ошибкой, полученной для класса суши. В пользу именно такого решения был целый ряд доводов:

- неточность масок облачности, что может вызывать неаккуратное сжатие соседних пикселей других классов, особенно сильно могут пострадать области с полупрозрачной облачностью;
- пространственная переплетённость облачности разных типов с остальными классами, что может приводить к мозаичности изображения и наличию видимых границ;
- некритичность сохранения точной текстуры самой облачности для большинства реальных задач;
- относительная простота и быстродействие реализации алгоритма.

После вычисления допустимых ошибок для каждого из типов объектов производится независимое сжатие соответствующих частей изображения, а итоговое изображение получается их объединением.

Наиболее эффективен представленный алгоритм при наличии качественных масок различных классов на изображениях, но в случае их отсутствия грубая маска может быть получена на основе соответствующих индексов, рассчитываемых по спектральным каналам изображения.

Тестирование методики адаптивного сжатия

Для оценки качества сжатия алгоритмом LERC и его влияния на текстуру изображений с учётом адаптивной методики были сформированы тестовые наборы данных по каналам прибора MSI (*англ.* Multispectral Instrument) уровня обработки L1C (Sentinel-2). Выбор данного инструмента обусловлен двумя обстоятельствами. Во-первых, данные Sentinel-2 занимают значительную часть объёма архивов ЦКП «ИКИ-Мониторинг», и минимизация их объёма остаётся крайне актуальной. Во-вторых, для данных MSI доступны достаточно достоверные маски классов подстилающей поверхности — данные прибора MSI уровня L2A с атмосферной коррекцией, содержащие маску SCL (*англ.* Scene Classification Layer).

Наборы формировались так, чтобы в них входили максимально разные типы подстилающей поверхности в различных климатических зонах, а также наиболее сложные с точки зрения подбора ошибки участки, характеризующиеся однородностью в пределах одного типа поверхности и несколькими такими типами на одном изображении, т. е. гистограммой с несколькими ярко выраженными локальными максимумами. В качестве тестовых участков можно привести:

- юг Амурской области (пахотные земли, луга, лиственные леса);
- Красноярский край в среднем течении Енисея в районе Лесосибирска (хвойные, смешанные и лиственные леса);
- дельту Волги (луга, степи, открытые грунты);
- Черноморское побережье Краснодарского края (горы, лиственные леса, океан).

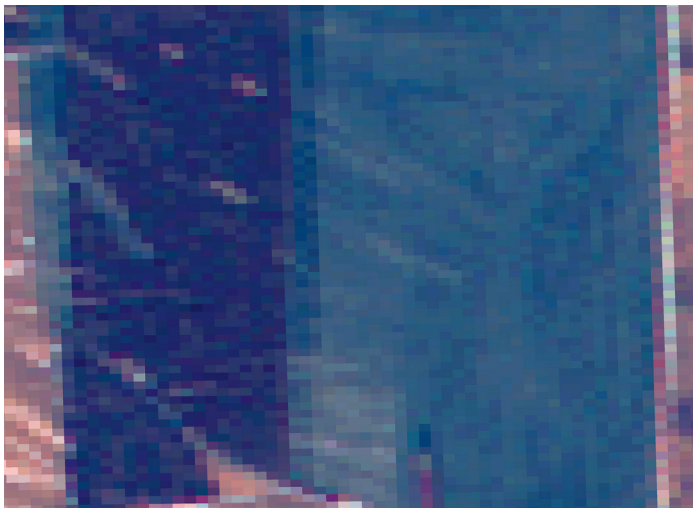


Рис. 1а. Пример сравнения одного и того же фрагмента, сжатого без потерь. Исходное изображение. Данные Sentinel-2 MSI, синтез R — канал 5 (705 нм), G — канал 3 (560 нм), B — канал 2 (490 нм)

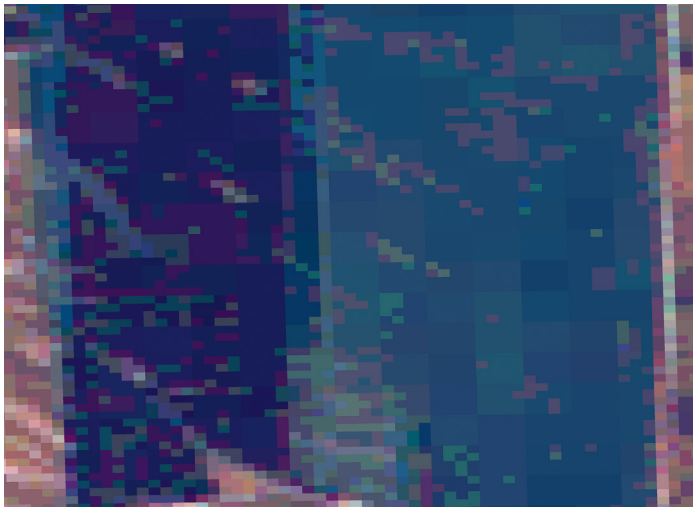


Рис. 1б. Пример сравнения одного и того же фрагмента, сжатого с потерями по базовой методике. Изображение, сжатое без учёта разных типов поверхности и реального диапазона. Отчётливо наблюдаются артефакты и искажения

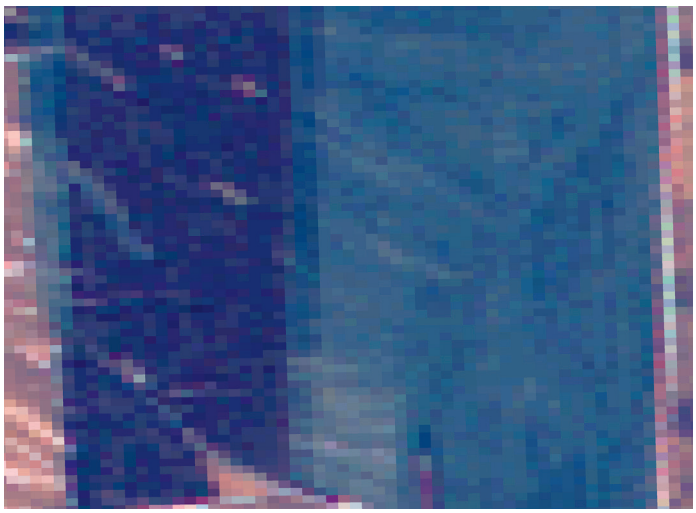


Рис. 1в. Пример сравнения одного и того же фрагмента, сжатого с потерями по адаптивной методике. Изображение, сжатое по адаптивной методике. Отличий от исходного изображения не наблюдается

Для тестовых данных по ранее упомянутой методике был определён коэффициент потерь L, составивший 0,05. С использованием этого коэффициента установлен сжатый набор данных по тестовым участкам. Статистика, полученная по тестовым фрагментам, показала, что средняя по фрагментам разность между исходными и сжатыми изображениями составила –0,39 при стандартном отклонении 0,40 (максимальный диапазон значений в самих изображениях более 10 000). С учётом одностороннего характера критерия это означает, что отрицательная средняя разность наблюдается в подавляющем большинстве фрагментов. При этом объём сжатых данных оказался на 39 % меньше по сравнению с данными, сжатыми без потерь, что достаточно близко к заявленной цели — к двукратному сжатию. Визуальное сравнение исходных и сжатых изображений показало, что в большинстве случаев они не различимы даже на пиксельном уровне. На *рис. 1* представлен пример сравнения оригинального изображения (*а*), изображения, сжатого с фиксированной ошибкой без адаптации к классам поверхности (*б*), и изображения, сжатого по адаптивной методике (*в*).

В целом, как показало тестирование, предложенная методика позволяет эффективно сжимать данные, допуская незначительные изменения (преимущественно занижение) в значениях яркости (фактически на уровне приборных ошибок), которые не отражаются на применимости данных в расчётных задачах и, что наиболее важно, не приводят к искажению текстуры изображения даже на однородных областях изображения с узкими динамическими диапазонами. Коэффициент сжатия для тестовых наборов относительно исходных, сжатых без потерь данных в среднем составил примерно 0,6.

Заключение

На основе предложенной методики в ИКИ РАН была разработана полностью автоматизированная процедура, обеспечивающая сжатие данных прибора MSI (стандартный продукт уровня обработки L1) (Baillarin et al., 2012), установленного на спутниках Sentinel-2. В настоящее время в рамках первой очереди оптимизации архивов ведутся работы по сжатию архивов таких данных, имеющихся в ЦКП «ИКИ-Мониторинг». Объём этих архивов на сентябрь 2026 г. составлял 3 Пбайт, т. е. при ожидаемом уровне сжатия (0,6) будет освобождено около 1,2 Пбайт дискового пространства. Такой выигрыш соответствует примерно годовому объёму всех данных, получаемых ЦКП, или же позволит разместить примерно пятилетний непрерывный ряд данных MSI по текущей зоне интереса ЦКП без привлечения дополнительных мощностей хранения. Также следует отметить, что проведённое сжатие данных потенциально позволит ускорить операции по доступу к ним и их обработке за счёт уменьшения их объёма и накладных сетевых расходов на передачу данных.

В дальнейшей перспективе планируется адаптировать и применить эту методику и к другим данным в архивах центра.

Исследование было выполнено с использованием возможностей ЦКП «ИКИ-Мониторинг» и при поддержке темы Министерства науки и высшего образования Российской Федерации «Большие данные в космических исследованиях: астрофизика, Солнечная система, геосфера» (№ 122042500019-6).

Литература

1. *Лупян Е. А.*, *Прошин А. А.*, *Бурцев М. А.*, *Балашов И. В.*, *Барталев С. А.*, *Ефремов В. Ю.*, *Кашницкий А. В.*, *Мазуров А. А.*, *Матвеев А. М.*, *Суднева О. А.*, *Сычугов И. Г.*, *Толпин В. А.*, *Уваров И. А.* Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263–284.
2. *Лупян Е. А.*, *Прошин А. А.*, *Бурцев М. А. и др.* Опыт эксплуатации и развития центра коллективного пользования системами архивации, обработки и анализа спутниковых данных (ЦКП «ИКИ-

Мониторинг») // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 3. С. 151–170. DOI: 10.21046/2070-7401-2019-16-3-151-170.
3. *Прошин А. А.*, *Лупян Е. А.*, *Бурцев М. А.* Особенности использования алгоритма сжатия изображений LERC для архивации данных ДЗЗ // Современные проблемы дистанционного зондирования Земли из космоса. 2024. Т. 21. № 6. С. 130–142. DOI: 10.21046/2070-7401-2024-21-6-130-142.
4. *Прошин А. А.*, *Лупян Е. А.*, *Бурцев М. А. и др.* Исследование влияния сжатия данных с потерями алгоритмом LERC на применимость данных ДЗЗ для решения задач мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2025. Т. 22. № 5. С. 63–72. DOI: 10.21046/2070-7401-2025-22-5-63-72.
5. *Baillarin S. J.*, *Meygret A.*, *Dechoz C. et al.* Sentinel-2 level 1 products and image processing performances // 2012 IEEE Intern. Geoscience and Remote Sensing Symp. IEEE, 2012. P. 7003–7006. DOI: 10.1109/IGARSS.2012.6351959.
6. *Becker P.*, *Plesea L.*, *Maurer T.* Optimizing cloud based image storage, dissemination and processing through use of MRF and LERC // The Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016. V. XLI-B4. P. 201–203. DOI: 10.5194/isprsarchives-XLI-B4-201-2016.

Optimization of IKI-Monitoring Shared Use Center archives using the adaptive satellite image compression method based on the LERC algorithm with classification by ground surface type

A. A. Proshin, E. A. Loupian, M. A. Burtsev

Space Research Institute RAS, Moscow 117997, Russia
E-mail: andry@smis.iki.rssi.ru

Currently, long-term archives of various types of Earth observation satellite data are widely used to address a variety of research and applied problems. Rapid development of Earth observation systems is leading to a radical increase in the volume of such archives, making it increasingly urgent to find new solutions that would reduce data storage requirements without compromising their applicability in research and practical use. The paper is devoted to the ongoing optimization of the archives of IKI-Monitoring Shared Use Center using a new method of satellite images adaptive compression based on the LERC (Limited Error Raster Compression) algorithm with adjustable loss. The method guarantees preservation of image texture by using statistical analysis to select the compression error for different classes of the underlying surface in each specific image, while simultaneously achieving sufficiently high compression ratios — approximately 0.6 relative to lossless-compressed data. Performing this optimization solely on the MSI (Multispectral Instrument) L1C data in IKI-Monitoring archives will free more than 1.2 PB of used storage space, which corresponds to approximately a five-year time series of such data.

Keywords: satellite imagery, Earth remote sensing, satellite data archives, imagery compression, satellite data access

Accepted: 07.10.2026
DOI: 10.21046/2070-7401-2026-23-5-435-441

References

1. Loupian E. A., Proshin A. A., Burtsev M. A., Balashov I. V., Bartalev S. A., Efremov V. Yu., Kashnitskii A. V., Mazurov A. A., Matveev A. M., Sudneva O. A., Sychugov I. G., Tolpin V. A., Uvarov I. A., IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, V. 12, No. 5, pp. 263–284 (in Russian).
2. Loupian E. A., Proshin A. A., Burtsev M. A. et al., Experience of development and operation of the IKI-Monitoring center for collective use of systems for archiving, processing and analyzing satellite data,

Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 2019, V. 16, No. 3, pp. 151–170 (in Russian), DOI: 10.21046/2070-7401-2019-16-3-151-170.
3. Proshin A. A., Loupian E. A., Burtsev M. A., Features of LERC image compression algorithm application in Earth remote sensing data archiving, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2024, V. 21, No. 6, pp. 130–142 (in Russian), DOI: 10.21046/2070-7401-2024-21-6-130-142.
4. Proshin A. A., Loupian E. A., Burtsev M. A. et al., Study of the impact of lossy data compression using the LERC algorithm on the applicability of remote sensing data for monitoring tasks, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2025, V. 22, No. 5, pp. 63–72 (in Russian), DOI: 10.21046/2070-7401-2025-22-5-63-72.
5. Baillarin S. J., Meygret A., Dechoz C. et al., Sentinel-2 level 1 products and image processing performances, *2012 IEEE Intern. Geoscience and Remote Sensing Symp.*, IEEE, 2012, pp. 7003–7006, DOI: 10.1109/IGARSS.2012.6351959.
6. Becker P., Plesea L., Maurer T., Optimizing cloud based image storage, dissemination and processing through use of MRF and LERC, *The Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, V. XLI-B4, pp. 201–203, DOI: 10.5194/isprsarchives-XLI-B4-201-2016.