Адаптивная методика сжатия изображений алгоритмом LERC с использованием информации о классах подстилающей поверхности

А.А. Прошин, Е.А. Лупян, М.А. Бурцев

Институт космических исследований Российской академии наук, Москва
Отдел «Технологии спутникового мониторинга»

Двадцать четвертая международная конференция
"СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА"
9 - 13 ноября 2026 г. в Москве

Введение

- Стремительное развитие систем ДЗЗ приводит к резкому возрастанию объемов многолетних архивов спутниковых данных, что делает все более актуальной задачу поиска новых решений, которые позволили бы уменьшить объемы хранения.
- В 2024 году в ИКИ РАН была разработана методика сжатия спутниковых данных алгоритмом с потерями LERC, который позволяет задать максимальную вносимую ошибку в значение пиксела, в рамках которой эта ошибка определяется на основе статистического анализа данных конкретного типа. Методика позволяет уменьшить объемы хранимой информации примерно вдвое по сравнению с использованием традиционных алгоритмов сжатия и при этом сохранить их применимость для разных задач, что было подтверждено работами, проведенными в 2025 году. В то же время она не учитывала должным образом разнородность изображений, что в некоторых достаточно редких случаях приводило к визуально заметным структурным искажениям текстуры.
- Доклад посвящен описанию новой адаптивной методики сжатия алгоритмом LERC, в рамках которой части изображения, относящиеся к разным типам подстилающей поверхности, сжимаются отдельно с использованием вычисляемой для них допустимой максимальной ошибки, что позволяет гарантировать отсутствие текстурных искажений и при этом сократить объемы данных примерно на 40 процентов.

Недостатки ранее реализованного подхода

Ранее реализованный алгоритм не учитывал возможное присутствие на изображениях зон (объектов) со значимо различными характеристиками (например, характерные диапазоны яркостей 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 яркостей;
- 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.

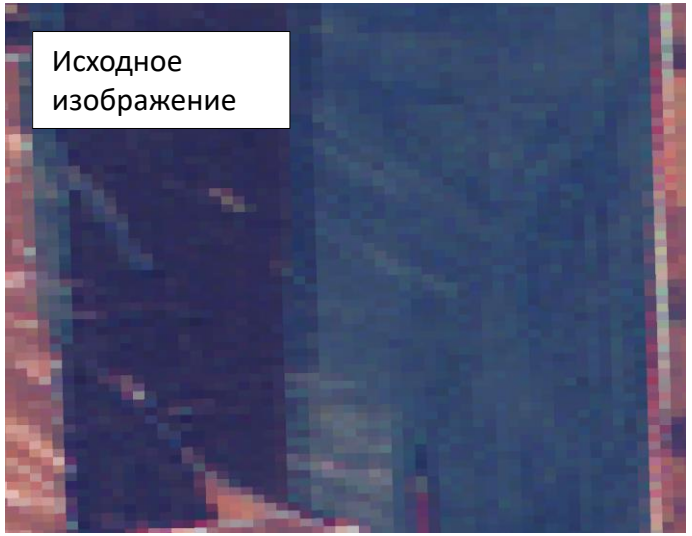
Адаптивная методика сжатия с учетом классов поверхности

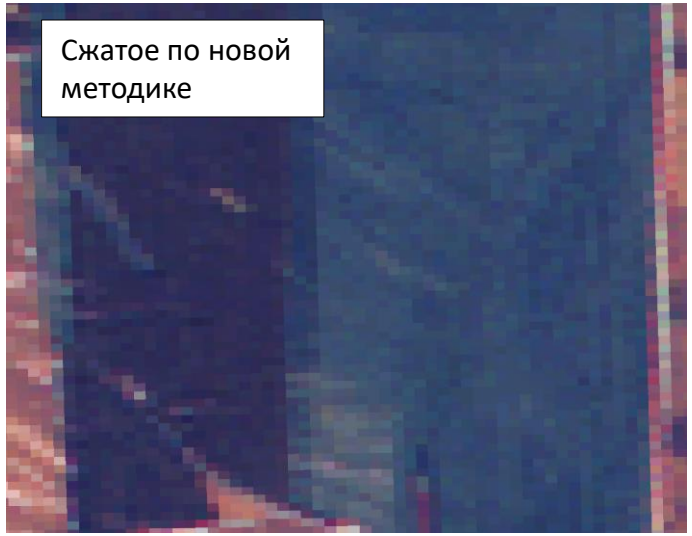
- Используется максимальная ошибка, определённая для каждого отдельного изображения с учётом диапазонов спектральной яркости для разных типов подстилающей поверхности, встречающихся на этом изображении.
- Созданная методика базируется на том, что на изображении выделяются основные классы подстилающей поверхности и ошибка по каждому из них определяется динамически на основе статистики значений яркостей 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 определения этого коэффициента используется 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.
- 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

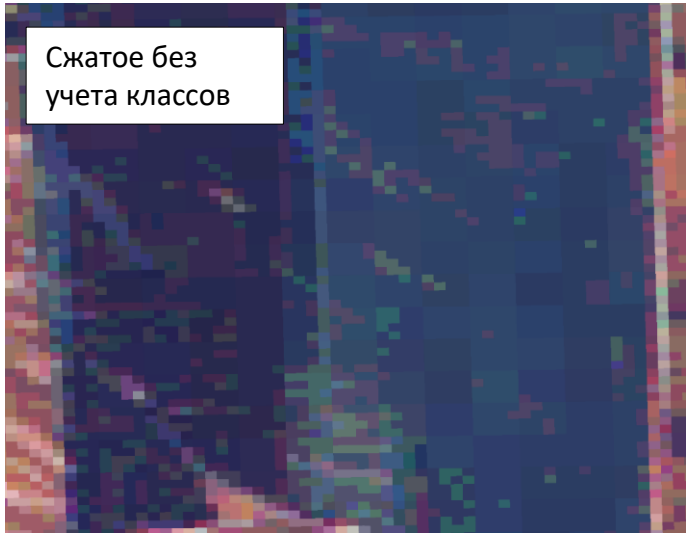
Тестирование методики адаптивного сжатия

- Использовались тестовые наборы данных по каналам прибора MSI уровня обработки L1C (Sentinel-2). Данные Sentinel-2 занимают значительную часть объёма архивов центра коллективного пользования «ИКИ-Мониторинг», и минимизация их объёма остаётся крайне актуальной. Также, для данных MSI доступны достаточно достоверные маски классов подстилающей поверхности — данные прибора MSI уровня L2A с атмосферной коррекцией, содержащие маску SCL (Scene Classification Layer).
- В наборы входили максимально разные типы подстилающей поверхности в разных 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, то есть гистограммой с несколькими ярко выраженными локальными максимумами.
- Для тестовых данных по методике был определён коэффициент потерь L, составивший 0.05, с использованием этого коэффициента был получен сжатый набор данных по тестовым участкам.
- Также визуальное сравнение исходных и сжатых изображений показало, что в большинстве случаев они не различимы даже на пиксельном уровне.

Примеры сжатия «сложного» изображения по разным методикам







Заключение

- С использованием реализованной автоматизированной процедуры в настоящее время ведутся работы по сжатию архивов данных прибора MSI, имеющихся в ЦКП «ИКИ-Мониторинг». Объем этих архивов на сентябрь 2026 года составлял 3 петабайта, т.е., при ожидаемом уровне сжатия (0,6), будет освобождено около 1,2 петабайт 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.
- В дальнейшей перспективе планируется адаптировать и применить эту методику и к другим данным в архивах центра, в том числе для тех, для которых требуемые маски необходимо будет рассчитывать «на лету».
- 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).