

АСПЕКТЫ ИНФОРМАЦИОННОГО НАПОЛНЕНИЯ ИНФРАСТРУКТУР ПРОСТРАНСТВЕННЫХ ДАННЫХ НА ОСНОВЕ ЛАЗЕРНОГО СКАНИРОВАНИЯ

Рыльский И.А.¹, Парамонов Д.А.², Груздев Р.В.³

¹Рыльский Илья Аркадьевич – кандидат географических наук, старший научный сотрудник,
географический факультет,

Региональный центр Мировой системы данных, МГУ им. М.В. Ломоносова;

²Парамонов Дмитрий Андреевич – кандидат географических наук, заместитель генерального директора,
ООО «Проектстрой»;

г. Москва

³Груздев Роман Викторович - кандидат геолого-минералогических наук,
Институт природных ресурсов, экологии и криологии СО РАН,

г. Чита

Аннотация: за последние 25 лет лазерное сканирование из экспериментального метода превратилось в полностью автономное семейство методов дистанционного зондирования Земли. Сейчас эта группа методов обеспечивает получение наиболее точных и подробных наборов пространственных данных, при этом стоимость данных непрерывно падает, количество средств измерения (лазерных сканеров) непрерывно растет. Объемы данных, которые получают в ходе съемок, в ближайшие десятилетия уже позволяют создать первые субглобальные покрытия планеты. Тем не менее, оборотной стороной высокой точности и подробности становятся необходимость хранения фантастически больших объемов трехмерных данных без потери точности. При этом должна быть обеспечена возможность работы указанными данными как в 2Д, так и в 3Д режиме. Стандартные методы хранения (файловый метод, базы геоданных, архивация) решают проблему лишь частично. При этом существуют ряд других, альтернативных методик, позволяющих снять подобные ограничения и привести к появлению более гибких и функциональных инфраструктур пространственных данных.

Ключевые слова: лазерное сканирование, Большие данные, Цифровая земля, аэрофотосъемка, точки лазерных отражений, квадродерево, октодерево, беспилотный летательный аппарат, ГИС, дистанционное зондирование.

ASPECTS OF INFORMATION CONTENT OF SPATIAL DATA INFRASTRUCTURES BASED ON LASER SCANNING

Rylskiy I.A.¹, Paramonov D.A.², Gruzdev R.V.²

¹Rylskiy Ilya Arkadievitch – PhD in Geography, senior researcher,
WORLD DATA SYSTEM, GEOGRAPHICAL FACULTY, MOSCOW STATE UNIVERSITY,

²Paramonov Dmitriy Andreevitch – PhD in Geography, deputy general director,
OOO PROJECTSTROY,
MOSCOW

³Gruzdev Roman Viktorovitch – PhD in Geology,
INSTITUTE OF NATURAL RESOURCES, ECOLOGY AND CRYOLOGY SO RAN,
CHITA

Abstract: over the past 25 years, laser scanning has evolved from an experimental method into a fully autonomous family of Earth remote sensing techniques. Now this group of methods provides the most accurate and detailed sets of spatial data, while the cost of data is continuously falling, and the number of measuring instruments (laser scanners) is constantly growing. The volumes of data obtained during surveys will, in the coming decades, make it possible to create the first sub-global coverage of the planet. However, the downside to high precision and detail is the need to store fantastically large volumes of 3D data without losing precision. In this case, it must be possible to work with the specified data in both 2D and 3D modes. Standard storage methods (file method, geodatabases, archiving) solve the problem only partially. At the same time, there are a number of other, alternative techniques that can remove such restrictions and lead to the emergence of more flexible and functional spatial data infrastructures.

Keywords: laser scanning, Big data, Digital earth, aerial photography, laser reflection points, quadtree, octree, unmanned aerial vehicle, GIS, remote sensing.

Введение. Лазерное сканирование – один из самых молодых видов съемок. В общем случае лазерное съемочное устройство (лазерный сканер) представляет собой лазерный дальномер, осуществляющий единичное или групповое одномоментное измерение дальности с одновременным измерением углов отклонения луча (вертикальная и/или горизонтальная плоскость) [3]. При наличии у каждого из измеренных отражений точных меток времени данные могут быть синхронизированы с высокоточными инерциальными навигационными системами и средствами ГНСС для позиционирования в глобальных системах координат. Помимо координат, лазерный сканер может регистрировать амплитуду отраженного сигнала, альбедо отразившей сигнал поверхности, а также регистрировать более одного отраженного сигнала, выполнять классификацию сигналов по форме отраженного импульса, и много другое [9, 10, 11].

Метод существует с конца 1990-х годов. Количество выпущенных за это время лазерных сканирующих систем во всем мире составляет (по различным оценкам) несколько тысяч единиц, причем подавляющее большинство этих систем выпущено в последние годы (2016-2023). Наравне с другими методами дистанционного зондирования [2], лазерное сканирование является одним из основных источников, данных для ГИС, поскольку служит основой для производства наиболее точных ЦМР, ЦММ и векторных картографических продуктов [1].

Каждая из подобных систем характеризуется высочайшей производительностью данных. При этом большинство этих систем (в силу крайней дороговизны) постоянно находятся в эксплуатации. Все эти системы непрерывно производят огромное количество данных. Однако задача объединения этих разрозненных массивов, данных в сервисы, подобные Google Earth до сих пор не решена даже на концептуальном уровне. Хранение массивов, данных объемом в петабайты с сохранением возможности быстрого доступа требует иных подходов по сравнению с принятыми сейчас.

Актуальность данного исследования обусловлена непрерывным ростом производительности и численности систем лазерного сканирования, при этом проблематика хранения, управления, анализа и визуализации для подобного рода массивов данных разработана недостаточно. Также остро стоит проблема анализа данных лазерного сканирования с целью распознавания образов, объектов, фильтрации и чистки данных. Эти задачи не могут быть решены без упорядочения хранения изначально бесструктурных облаков трехмерных точек посредством создания топологических структур хранения и методов для этого вида данных.

Объектом исследования выступают пространственные данные лазерного сканирования. Предметом исследования выступают изучение организации структуры и хранения данных лазерного сканирования.

Целью данного исследования является оптимизация структуры хранения и доступа к данным лазерного сканирования, решение проблем нетопологичности массивов трехмерных точек отражений, решение задачи ускорения работы с данными и их анализа.

Задачи исследования:

- привести обзор известных типов и видов систем лазерного сканирования и их производительности;
- оценить существующие, возможные и оптимальные схемы хранения и существующие модели пространственных данных;
- проанализировать результаты и сделать выводы;

Методология данного исследования основана на теоретических и практических методах. В исследовании использованы результаты практического применения данных методов на крупных проектах съемок на территории РФ в период с 2004 по 2024 годы.

На сегодняшний день степень проработанности исследования указывает на ряд нерешенных вопросов. К таким относится то, что принятые в настоящее время методы хранения в открытых форматах (LAS, BIN, FBI, XYZ) устарели и требуют внедрение нового формата для BigData, который, по мнению авторов, следует основывать на использовании мало распространенной на сегодняшний день модели квадродерева (октодерева – для 3D данных).

Исследования на фактических данных, безусловно, добавляют актуальность проделанной работе и вносят определенный вклад в развитие и массовое внедрение использования и анализа данных лазерного сканирования.

Результаты исследования проанализированы, сделаны **выводы**.

Материалы и методы исследования.

Существующие системы лазерного сканирования. К середине 2024 года можно выделить несколько основных видов лазерного сканирования. Грубо все системы могут быть разделены на два больших класса, работающие на принципах измерения времени прохождения сигнала (Time-of-Flight), и на измерении фазы пришедшей волны излучения, сгенерированной ранее самим прибором на частотах существенно ниже частоты излучения самого сканера (Phase, или так называемые фазовые измерения).

Лазеры с фазовым измерением дальности генерируют плавно возрастающую и убывающую «волну излучения» с частотой максимального излучения от 2 до 30 мегагерц (длина волны от 10 до 150 метров). Измерения дальности при этом основаны на принципе определения дальности посредством умножения

доли принятой волны (фазы) на длину волны, что автоматически накладывает ограничение на скорость или дальность работы таких приборов. Так, при значительной дальности понижается частота сканирования и точность. Высокая производительность дается путем снижения дальности до 120, 50 и даже 15 метров – при этом точность измерений возрастает. Существуют и апробированы ряд алгоритмов для работы в условиях дальностей, превосходящих длину волны [8]. Точность при этом страдает: если на малых дистанциях она может составлять доли миллиметра, то на дистанции 150 м достижение точности измерения дальности в 20-30 мм весьма проблематично.

Лазеры, измеряющие время прохождения сигнала фактически не имеют вообще никаких разумных ограничений по дальности – она определяется фактически только мощностью лазерного импульса и точностью фокусировки луча. При этом они имеют ограничения по точности. На сегодняшний день наиболее совершенные образцы приборов обладают точностью определения дальности около 3 мм (на диапазоне дальностей от 5 до 500 м) и до 25 мм – на дальностях до 6 000 м и более [9, 11].

Каждый из классов, в свою очередь, может быть разделен на подклассы – сканеры для статических и для кинематических измерений. Системы для статической работы обычно называют наземными лазерными сканерами, системы для кинематических измерений (в движении) обычно еще подразделяют по классам носителей на авиационные лазерные сканеры (для БПЛА или пилотируемых носителей) и на мобильные лазерные сканеры (для работы с автомобилями, поездов, специальных носителей или будучи носимых человеком).

Наземные системы могут не иметь меток точного времени и не иметь возможности быть автоматически точно позиционированными на поверхности Земли. Кинематические системы в принципе не могут существовать без данной особенности, и всегда позиционируются сначала в глобальных системах координат (например, WGS84), и лишь потом подвергаются взаимному уравниванию и тому подобным процедурам.

Производительность всех современных кинематических сканирующих систем чрезвычайно велика. Наиболее совершенные системы воздушного сканирования – например, Riegl Q1560ii – обладают скоростью сканирования в 4 000 000 точек в секунду [9]. Наиболее совершенные мобильные системы – например, Riegl VMX RAIL – до 3 000 000 точек в секунду [10]. Системы других компаний (Leica, Ortech, ряд других) обладают более скромными характеристиками, но и у них производительность флагманских систем достигает 1 000 000 точек в секунду.

Системы статического сканирования также позволяют работать со скоростью более 1 000 000 измерений в секунду (Riegl VZ2000i). При этом, как правило, все системы параллельно оснащаются еще и фотокамерами, синхронизированными с лазерными сканерами. Это дает возможность (при качественной калибровке всех устройств) присвоить каждой точке сканирования еще и RGB-атрибуты яркости данного участка поверхности. То есть сделать облако лазерных точек цветным.

Таким образом, все описанные системы способны создавать пространственных данные бескомпромиссной точности с ранее не встречавшейся скоростью. Нетрудно подсчитать, что за 5 часов полноценной непрерывной работы система с производительностью 2 миллиона точек произведет около 36 миллиардов точек. В действительности – около 50 миллиардов, так как один импульс дает как правило более одного отражения при наличии растительности на территории [1,3]. Если округленно оценить объем хранения данных в 60 байт на точку, это дает около 3 терабайт данных только лазерного сканирования. Это – не считая обычно сопутствующих лазерному сканированию данных аэрофотосъемки.

В пересчете на 1 км² объем данных также выглядит достаточно впечатляюще. Сейчас даже при проведении аэрофотосъемки с пилотируемых носителей не является редкостью требование обеспечить точность в 10 и более точек на 1 м² и детальность снимков в 4-5 см/пиксел. Это требует 600 байт для данных лазерного сканирования [4] и примерно 200 байт для хранения готового ортофотоплана в формате JPG, всего – 800 байт на 1 м². Или 0.8 гигабайт для хранения данных на 1 км². Для хранения данных на площадь среднего по размерам субъекта РФ площадью в 100 000 км² потребуется уже 80 терабайт. Для хранения данных вышеуказанной детальности на территорию РФ потребуется до 14 петабайт пространства.

В настоящее время средний объем данных, производимых среднего размера аэрофотосъемочной компанией в течение года, без труда займет объем около 100 терабайт. За период с 2002 по 2024 годы в Российской Федерации данным методом реализованы несколько сотен крупных проектов съемки с площадью свыше 100 км², в части которых авторы исследования приняли участие [1, 3].

Приняв во внимание сказанное выше, можно отметить, что задача хранения и обработки данных лазерного сканирования в больших объемах в ближайшее время будет становиться все более сложной. Ее решение потребует все более специфических подходов, более характерных для работы с другими видами Big Data [3].

Возможности оптимизации работы с данными лазерного сканирования

Файловая система хранения данных – неудобная реальность

Несмотря на нарастающие сложности в использовании, наиболее распространенным методом хранения информации является использование файловой системы хранения данных с пространственной сегментацией. При реализации каждого проекта пользователь лазерной съемочной системы создает некоторую систему пространственных полигонов, каждому из которых соответствует некий файл с хранящейся в нем пространственной информацией (точки лазерных отражений, ортофотопланы, прочее). Это могут быть регулярные прямоугольники (планшеты), трапеции номенклатурных листов в государственных разграфках, куски буферных зон для линейно-протяженного объекта, перекрывающиеся полигоны сложной формы и тому подобные геометрические объекты. Связь между векторными объектами (как правило, это ГИС-данные с послойной системой хранения) и физическими файлами на диске обычно устанавливается с помощью гиперссылки или в виде текстовой строки с путем хранения данных.

Недостатки подобного метода вполне очевидны. В первую очередь, это необходимость хранения в одной или нескольких проекциях материалов без возможности быстрого перехода от одной проекции к другой. Проекты с подобным подходом обычно не могут быть просто так объединены между собой как по причине разных систем координат, так и разных видов «нарезки» и просто разных технических параметров съемки. Подобрать оптимальный размер «элемента хранения» также непросто – слишком малый полигон хранения требует работы постоянно с большим количеством кусков, слишком крупный не позволяет загрузить все данные в оперативную память или приводит к снижению производительности рабочей станции.

Однако наиболее неприятным аспектом подобной работы является сама форма хранения данных лазерного сканирования в виде точек [2]. Практически все современные форматы хранения данных – PLY, LAS, FBI, BIN и тому подобные – используют структуру данных в виде последовательных записей, каждая из которых содержит метку времени, координаты, цвет, амплитуду и ряд прочих параметров. Файл с данными лазерного сканирования таким образом представляет собой гигантскую таблицу.

Несмотря на то, что при данном виде записи, данные являются пространственными, они не являются топологическими. Даже элементарный запрос вроде «выбрать точки в радиусе X от точки Y » потребует полной переработки всего массива данных в текущем сегменте данных. Если точка находится ближе, чем на дистанции X от границы текущего сегмента, то обработка может быть остановлена, и потребуются подгрузка нового сегмента (зачастую вручную). Для избегания подобных ситуаций сегменты часто делают перекрывающимися [5], что не способствует облегчению работы алгоритмов и требует еще больше дискового пространства.

Это критическая проблема файлового сегментного подхода к хранению данных. По мере уменьшения размеров сегмента ее острота нарастает. При очень плотных данных (до 10000 точек на 1 м^2 и выше) размер сегмента может уменьшиться до 100 и даже 50 метров – при большем объеме будет наступать переполнение оперативной памяти. Это вполне возможная ситуация при реализации крупных проектов мобильного сканирования.

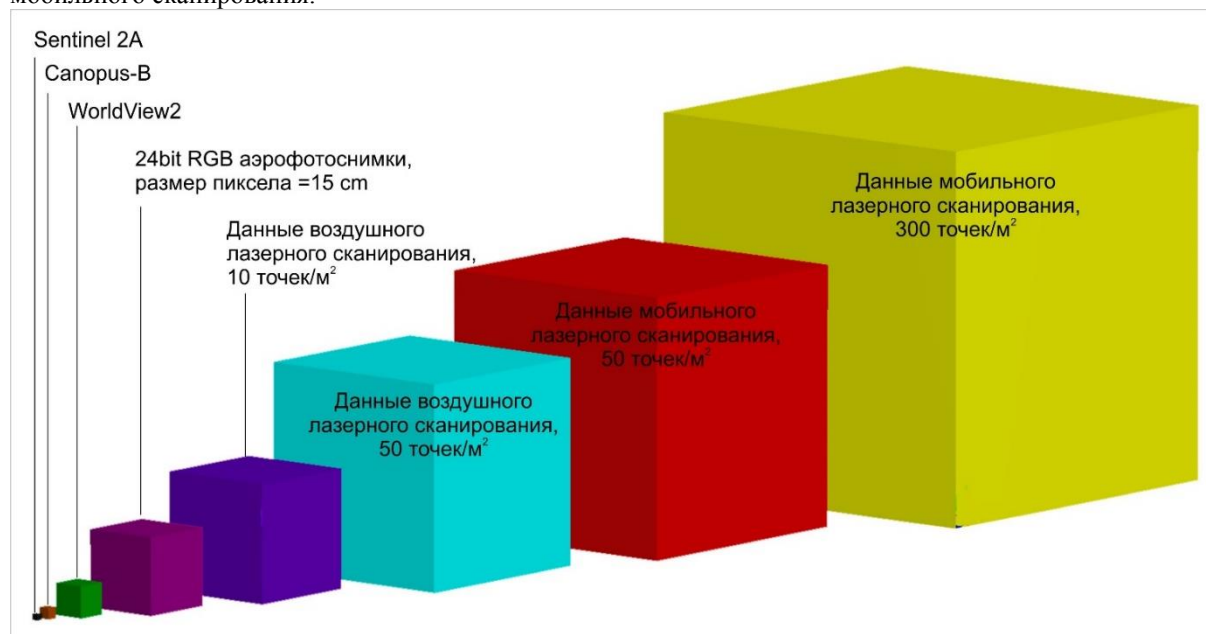


Рис. 1. Изменение объема пространственных данных различной детальности Архивация и пространственное загроуление.

Частично проблему маленького сегмента данных, а также проблему уменьшения общего объема хранимых данных можно решить за счет сочетания уменьшения пространственной точности данных и динамической архивации информации.

Итоговая точность значительной части систем воздушного лазерного сканирования редко превосходит 5 см, а систем мобильного сканирования – 1 см. Тем не менее, в большинстве случаев при хранении данных используется запись большого количества знаков после запятой – вплоть до десятых долей миллиметра. Конечно, при вышеописанных реальных точностях подобное является избыточно детальной записью координат.

Сознательно оценив реальную точность и необходимую плотность данных, пользователь может пойти на определенные «уступки». Например, ограничить детальность записи координат в 1 см. Кроме того, возможна фильтрация, при которой точки на дистанции в 1-2 см друг от друга считаются одной точкой (в том случае, если параметры съемки не дают оснований ожидать подобной дискретности данных). За счет этого возможно сжатие объема занимаемой исходной информации на 20-40 %. Подобный подход реализован в Bentley Point Tools, и частично может быть реализован и в TerraSolid.

Координаты точек изменяются достаточно медленно – у каждой последующей точки большинство цифр в координатах и метках времени совпадают с предыдущей и последующей точкой. Эта особенность создает неплохие возможности для реализации алгоритмов архивации данных. В сочетании с ранее описанным подходом, подобные методы позволяют сжать исходное облако данных в 2.2-3.5 раза по сравнению с исходным объемом данных.

Подобный подход применяется при хранении данных в формате LAZ. Естественно, это вызывает снижение скорости обработки, но позволяет снизить нагрузку на системы хранения. Проблему нетопологичности данных это не решает.

Базы геоданных

Одной из наиболее сильных сторон современных ГИС-пакетов (например, ArcGIS) является хранение материалов в виде базы геоданных. Подход решает проблему сегментации и перепроецирования материалов, в том числе – и точечных данных лазерного сканирования. Проблемой в этом случае является крайне медленная работа с точками лазерных отражений. Причина в том, что каждая точка трактуется как отдельный ГИС-объект, с присущими ей атрибутивными характеристиками, и обрабатывается со всей тщательностью. Платой за это является очень большой занимаемый объем хранения и медленная скорость работы. На практике работать в ArcGIS с большими объемами лазерных данных крайне неудобно. При этом принципиальных путей ускорения работы ArcGIS с базами геоданных на сегодняшний момент не просматривается.

Квадродерево

Квадродерево (дерево квадрантов, 4-дерево, Quadtree) – структура данных в информатике, в котором у каждого узла ровно 4 потомка [7]. Деревья квадрантов могут быть использованы для рекурсивного разбиения двумерного пространства по 4 области. Каждая область может быть квадратом, прямоугольником, или иметь произвольную форму. Термин quadtree был введен в 1974 году.

По своей сути, квадродерево позволяет разбивать пространство до адаптивного размера ячейки (adaptive cells) в зависимости от распределения информации. Аналогично квадродереву для двумерного пространства, для трехмерного пространства возможно построение аналогичной структуры – октодерева, или octree (делит трехмерное пространство на 8 октантов).

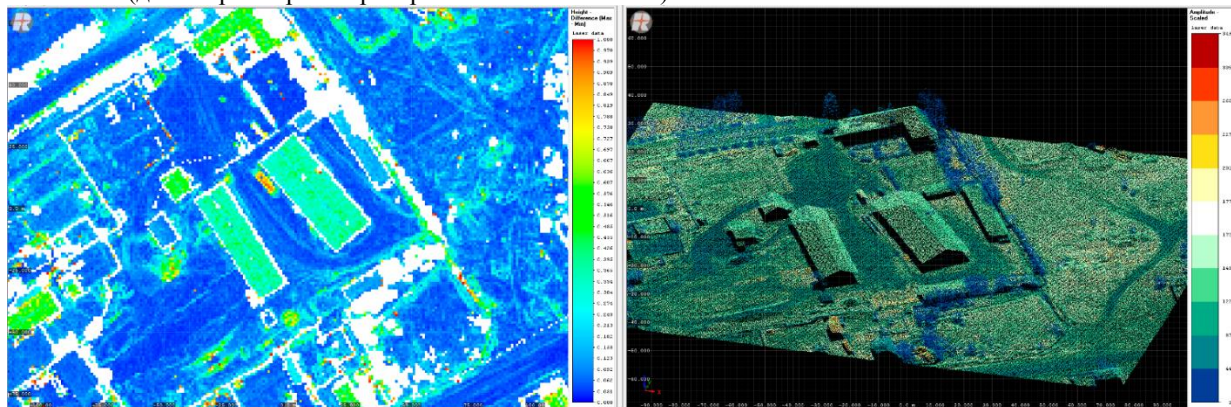


Рис. 2. Квадродерево с данными в программном обеспечении RiProcess (Riegl).

Идеология хранения и работы с данными в виде квадродерева используется в широко известном приложении Google Earth для иерархического упорядочивания и адресации элементов данных – небольших растровых изображений, формирующих итоговые мозаики космоснимков при работе в данном приложении.

Частично подход к аналогичной работе с данными используется в программном обеспечении RiProcess (Riegl), а также применяется для ускорения визуализации облаков точек в Bentley Point tools. Однако указанные программные среды в принципе не подходят для хранения крупных инфраструктур пространственных данных по целому набору параметров. Так, в RiProcess невозможен экспорт части данных по пространственному запросу – например, экспорт данных в пределах указанного полигона. Нет возможности неограниченно надстраивать массив данных без перестроения квадродерева по всему объему данных. И, самое главное, квадродерева в указанных программных продуктах являются именно двухмерными, а не трехмерными структурами. Тем не менее, в данном программном комплексе уже реализован подход к расчету статистических характеристик для каждой ячейки [8].

Подходы к решению проблем хранения лидарных данных

Принимая во внимание сказанное выше, считаем нужным отметить, что ни одна из программных сред (в том числе – современные ГИС-пакеты) не предлагает разумного решения проблемы хранения и работы с большими объемами данных лазерного сканирования в точечной форме.

Ранее подобные проблемы имели место и были успешно решены в геоинформатике путем постепенной унификации форм и форматов данных и алгоритмов их хранения (векторная графика, хранение в географических координатах, послойное представление данных, открытые форматы и алгоритмы). Этот процесс активно происходил в период с 1990 по 2010 год и успешно завершился появлением большого количества ГИС-пакетов, позволяющих легко управлять данными любых объемов и пространственного охвата, при этом обеспечивая возможность экспорта данных из одной программы в другую без потерь информации.

Аналогичный шаг необходимо сделать и в области хранения данных лазерного сканирования. Требования к ИПД на базе 3Д точек можно сформулировать следующим образом:

- возможность неограниченного увеличения охвата данных вплоть до глобального;
- возможность максимально быстрой визуализации (в любой форме) пространственных данных на любом уровне охвата (вплоть до рассмотрения отдельных точек и работы с ними);
- возможность слияния различных источников, данных без переработки ранее уложенных в ИПД материалов;
- возможность пространственных запросов и экспорта отдельных фрагментов данных;
- развитая система визуализации материалов в 2Д и 3Д форме.

Принимая во внимание все сказанное выше, можно сделать вывод о том, что именно форма хранения данных определяет возможности для достижения. Действительно, если взглянуть на историю форматов и технологий, то именно гибкость формата определяет приемлемость той или иной технологии обработки данных [6]. Из примеров можно отметить форматы JPG и MP3. Первый открыл дорогу для взрывного роста количества хранимых изображений и возможностей их передачи графики по сети Интернет, второй совершил революцию в области портативных плееров и передачи звуковых потоков онлайн. Аналогичный, хотя и менее громкий успех – у форматов DXF и AscINFO GRID.

Именно удачное решение в области создания формата создает предпосылки для быстреего развития технологий. Формат в этом отношении чем то напоминает патрон для огнестрельного оружия. Именно патрон является краеугольным камнем стрелкового комплекса. И именно под патрон проектируется оружие. В области пространственных данных дело обстоит аналогичным образом.

Примерные особенности формата для ИПД на базе 3Д точек.

Наилучшим способом хранения данных для вышеупомянутых целей является октодерево. Область моделируемого пространства – сфера с габаритами Земли. Работа с данными в градусах широты и долготы всегда сталкивается с проблемой перевода в какую-либо проекцию и порождает связанные с искажениями проекций сложности. Поэтому наиболее удобной – с точки зрения расчетов и последующего перехода к другим координатным системам – мы считаем картезианскую систему координат (начало отсчета – в центре Земли, координаты в метрах). В этом случае исходное пространство подразделяется на исходные 8 кубов с длиной стороны около 8388,608 км (нулевой уровень детальности), и далее на каждом уровне габариты куба уменьшаются в 2 раза. На 23-м уровне длина ребра куба достигнет 1 метра, на 30-м – 1 см, на 33-м – 1 мм [2].

В формате необходимо установить некий пространственный предел, менее которого размер куба быть не может [2, 7]. При более детальном приближении в 2Д режиме куб отображается в виде пиксела и более не детализируется, при приближении в 3Д режиме отображаются точки внутри куба (все). При отдалении от минимально возможного куба в 3Д режиме он отображается как одна точка, с дальнейшим иерархическим закруглением по мере отдаления [9]. Для ИПД с глобальным охватом разумным размером куба может быть 0.25-1.0 м, для локальных ИПД – 10 см. Назовем куб пространства подобного размера элементарным кубом.

Статистика и метаданные

Для каждого куба может быть посчитана (заранее, при создании файла) статистика 3Д точек. Это может быть «центр тяжести» куба (среднее положение из всех его точек), расчет средних координат

точек и их СКО для каждой из осей, осредненный цвет, амплитуда, отражающая способность, а также координаты осредненной плоскости (плоскость, сумма квадратов расстояний до которой от каждой из точек в кубе минимальна). Для каждого куба заранее просчитываются его географические координаты, а также координаты в какой-либо широко распространенной проекции (UTM). Аналогичным образом рассчитываются эллипсоидальная и ортометрическая высоты центра тяжести куба (с поправкой за модель геоида) и перепад высот в пределах куба.

Подобная структура данных при визуализации любого из ее участков в 2Д (вид сверху) или 3Д режиме (вид в изометрии) будет выглядеть как массив точек, где в каждом пикселе экрана визуализируется один куб, отображаемый одной точкой. Цвет точки определяется выбранным способом отображения (цветовая шкала) и может отображать высоту, разброс высот в пределах данного пиксела, амплитуду отраженного сигнала, и прочее [5].

Помимо пространственных свойств, для куба сохраняются метаданные. Для хранения даты и времени получения точек (при наличии в кубе разновременных съемок он отображается неким зарезервированным цветом, а при их отсутствии – в соответствии с выбранной цветовой шкалой, не включающей в себя зарезервированный цвет). Детальность хранения времени для куба – сутки, для отдельных точек вписывается точное GPS-время получения точек, допускающее их дополнительную обработку точек лазерных отражений. Также может быть предусмотрено хранение текстовых сведений об источнике данных, их точности и тому подобные сведения, которые могут быть сохранены в зарезервированных полях данных [7, 12].

Подходы к хранению на диске

Хранение данных в формате, идеология которого описана выше, непременно натолкнется на сложности хранения бесконечно больших массивов данных в одном файле. Технически это возможно. Тем не менее, в Google Earth схожая задача (хранение кэша снимков) решается путем хранения большого количества мелких файлов (куски растров), и при пополнении системы происходит простое добавление более детальных кусков растров в папку к уже имеющимся растрам. Схожий алгоритм используется и в других виртуальных моделях, основанных на растровых данных [3].

Похожий подход может быть реализован и в данном случае. Конечно, хранить данные в виде «один элементарный куб – один файл» неразумно. Но вполне возможно ввести хранение более крупных кубов (размером 100-1000 м) в виде отдельных файлов, а уже внутри них сохранять данные элементарных кубов. Подобный более крупный куб назовем «групповым кубом» [16].

В этом случае, при добавлении нового набора данных к старому потребуется часть новых групповых кубов просто переписать на новое место, а часть – перепросчитать, обновив статистику и метаданные каждого обновляемого группового куба. Однако весь остальной объем данных перепросчитывать не придется. Аналогичным образом возможно решение обратной задачи – вычленение части данных по пространственному запросу [15].

Выводы

Создание единого формата данных на базе алгоритмов работы с октодеревом позволит решить целую группу проблем BigData в области лазерного сканирования, такие как необходимость структурирования больших сегментов данных, быстрого доступа к произвольным участкам данных, а также минимизировать затраты на слияние и объединение перекрывающихся участков разновременных данных [15].

Важной особенностью структуры данных в виде октодерева при хранении точек лазерных отражений является возможность создания топологических структур данных, которые помимо статистической и пространственной информации содержат сведения о соседних сегментах данных [13, 15].

Несомненно, что подобная организация информации существенно более сложна, чем принятые в настоящее время методы хранения в открытых форматах (LAS, BIN, FBI, XYZ). Тем не менее, без решения данной проблемы дальнейшее развитие и массовое внедрение использования и анализа данных лазерного сканирования будет крайне затруднено [14].

Список литературы / References

1. Груздев Р.В., Рыльский И.А., Применение воздушных лидаров в высокоточной гравиразведке (на примере Восточного Забайкалья). Вестник Забайкальского Государственного Университета. Чита, 2022. №2, Т. 28, 6-18.
2. Капралов Е.Г., Кошкарёв А.В., Тикунов В.С. Основы геоинформатики. Учебное пособие для студентов вузов в 2-х книгах. М.: Академия, 2004. С. 480.
3. Рыльский И.А., Парамонов Д.А., Кожухарь А.Ю.1, Терская А.И. Создание виртуальной модели района Большой Сочи – Красная Поляна – плато Лаго-наки. Интеркарто, 2023. №29, С. 589-606.
4. Giuliani, G., Chatenoux B., De Bona A.: Building an Earth Observation Data Cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). Big Earth Data 1-2 (1), 100–117 (2017).

5. *Huang L. et al.* Octsqueeze: Octree-structured entropy model for lidar compression //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – C. 1313-1323.
6. *Janowski A., Szulwic J., Tysiac P.* Airborne and mobile laser scanning in measurements of sea cliffs on the southern Baltic. 15th International Multidisciplinary Scientific GeoConference SGEM, 2015. 114-124
7. *Octree homepage*, <https://en.wikipedia.org/wiki/Octree> Last accessed 16 Jan 2020.
8. *Rieger P.* Range ambiguity resolution technique applying pulse-position modulation in time-of-flight scanning lidar applications. *Optical engineering*, 2014. 53(6), 061614-061614.
9. [Electronic Resource]. URL: <http://www.riegl.com/nc/products/airborne-scanning/produktdetail/product/scanner/68/>. Last accessed 13 Jul 2020.
10. [Electronic Resource]. URL: <http://www.riegl.com/nc/products/mobile-scanning/produktdetail/product/scanner/67/> Last accessed 19 May 2020.
11. [Electronic Resource]. URL: <http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/33/> Last accessed 10 Jul 2020.
12. *Su Y.T., Bethel J., Hu S.* Octree-based segmentation for terrestrial LiDAR point cloud data in industrial applications //ISPRS Journal of Photogrammetry and Remote Sensing. – 2016. – T. 113. – C. 59-74.
13. *Truong-Hong L., Laefer D.F.* Octree-based, automatic building facade generation from LiDAR data //Computer-Aided Design. – 2014. – T. 53. – C. 46-61.
14. *Vo A.V. et al.* Octree-based region growing for point cloud segmentation //ISPRS Journal of Photogrammetry and Remote Sensing. – 2015. – T. 104. – C. 88-100.
15. *Wang M., Tseng Y.* Incremental segmentation of lidar point clouds with an octree-structured voxel space. *The Photogrammetric record* 26 (133), 32–57 (2011).
16. *Yang M.Z.X.L. et al.* Improved octree filtering algorithm of airborne LiDAR data in forest environment //Journal of Beijing Forestry University. – 2018. – T. 40. – №. 11. – C. 102-111.