

ОЦЕНИВАНИЕ КАЧЕСТВА РЕЗУЛЬТАТОВ ОБРАБОТКИ МАТЕРИАЛОВ АЭРОКОСМИЧЕСКОЙ СЪЕМКИ ПРИ МОНИТОРИНГЕ И УПРАВЛЕНИИ РАЗВИТИЕМ ТЕРРИТОРИЙ

В.А. Зеленцов

Санкт-Петербургский Федеральный исследовательский центр РАН
Россия, 199178, Санкт-Петербург, 14-я линия В.О., д. 39
E-mail: v.a.zelentsov@gmail.com

В.Ф. Мочалов

Военно-космическая академия имени А.Ф. Можайского
Россия, 197198, Санкт-Петербург, ул. Ждановская, д. 13
E-mail: vka@mil.ru

Ключевые слова: мультиспектральная аэрокосмическая съемка, автоматизированная обработка, показатели качества результатов обработки, наземные обследования.

Аннотация: Развитие технологии автоматизированной обработки материалов съемки должно сопровождаться совершенствованием методов оценивания качества выходных материалов. Показано, что в ряде случаев для расчета показателей качества обработки на основе объективных данных о реальном состоянии территории необходимо проведение дополнительных наземных обследований. Предлагается метод, позволяющий уточнить значения показателей качества результатов обработки материалов съемки. При этом предусматривается обоснованный выбор дополнительных точек наземных обследований.

1. Введение

В настоящее время важным источником исходных данных при мониторинге и управлении развитием территорий являются материалы мультиспектральной аэрокосмической съемки. Практически значимой становится задача регулярной оперативной и достоверной идентификации элементов ландшафта, оценивания их свойств в условиях сезонной изменчивости, выполняемых работ, или под влиянием антропогенных воздействий. Современные материалы съемки характеризуются расширенным числом спектральных каналов, улучшенным пространственным и энергетическим разрешением систематически регистрируемых исходных данных. Этот факт способствует успешному решению задач идентификации элементов ландшафта на основе применения алгоритмов автоматизированной обработки материалов съемки [1,3]. Как правило, результаты обработки сопровождаются сведениями о качестве представляемых выходных данных. Показатели качества обработки характеризуют степень совпадения выходных материалов с объективными данными из независимых источников. На основе объективных данных формируется контрольная выборка,

сведения которой считаются достоверными. Основными источниками объективных данных являются: материалы съемки более высокого пространственного разрешения; специализированные карты, экспертные знания; а также данные синхронных наземных обследований. При этом непосредственный выбор точек для наземных обследований может выполняться одним из следующих способов: полный охват территории, случайный выбор точек обследования или обоснованный выбор точек обследования.

Вместе с тем данные контрольной выборки могут устаревать. Проведенные исследования [2] показывают, что в ряде случаев более углубленные оценки качества результатов обработки могут быть получены на основе оперативного анализа результатов дополнительных обследований. При этом данные дополнительных обследований применяются для уточнения значений показателей, характеризующих качество результатов тематической обработки.

На рис. 1 приведена схема, отражающая порядок выполнения операций расчета предварительных и уточненных значений показателей качества результатов обработки материалов съемки.

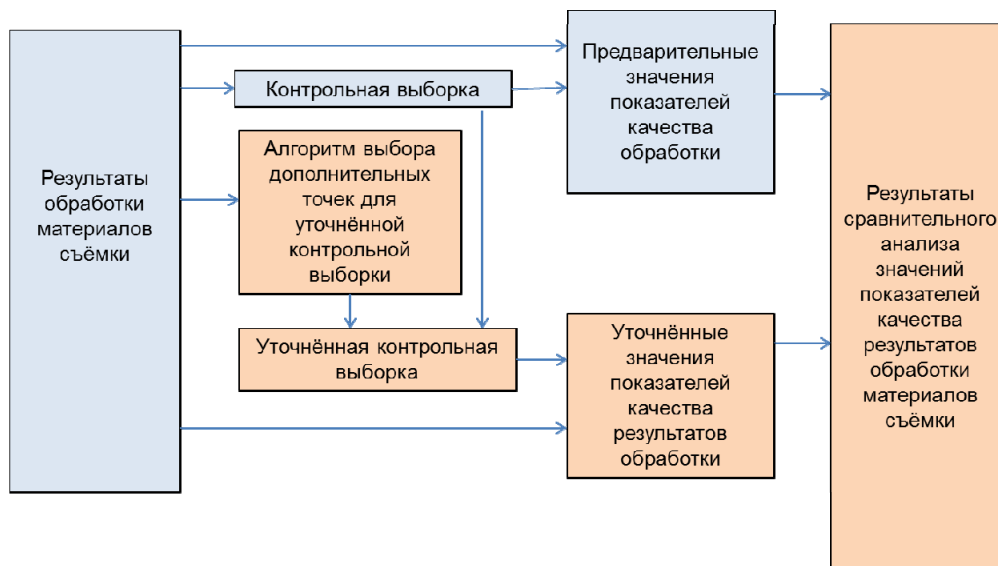


Рис. 1. Схема метода оценивания качества результатов обработки материалов съемки и роль алгоритмов выбора дополнительных точек для уточненной контрольной выборки.

Целью настоящего доклада является краткое описание разработанного метода оценивания качества результатов обработки материалов аэрокосмической съемки. Метод предусматривает расчет предварительных и уточненных значений показателей качества обработки. При этом уточненные значения показателей рассчитываются на основе данных, полученных в ходе дополнительных наземных обследований в обоснованно выбранных точках.

Приводимое описание включает:

- краткий анализ существующих методов и показателей оценивания качества результатов автоматизированной обработки материалов съемки;
- анализ возможностей алгоритмов кластеризации при выборе дополнительных точек для формирования уточненной контрольной выборки.

2. Методы и показатели оценивания качества результатов обработки материалов съемки

Рассмотрим задачу обработки материалов съемки, которая заключается в автоматизированной идентификации в пределах рассматриваемой сцены выбранных видов элементов ландшафта с известными характеристиками. Можно выделить следующие основные показатели, характеризующих качество идентификации: ошибки первого и второго рода; показатели, основанные на анализе значений матрицы соответствия или погрешности (Confusion Matrix); показатели, основанные на анализе значений «точности» P (Precision) и «чувствительности» R (Recall) идентификации.

Метод оценивания на основе анализа ошибок первого и второго рода [4] предполагает расчет ошибок: первого рода (α) – «пропуск цели», при котором пиксель, принадлежащий к рассматриваемой поверхности, не идентифицируется таковым; и ошибки второго рода (β) – «ложное срабатывание», когда пиксель, идентифицированный как принадлежащий поверхности, на самом деле таковым не является. Достоинством метода является простота логического объяснения качества результатов идентификации. Среди недостатков метода можно выделить ограниченный перечень оцениваемых показателей и ограниченное количество идентифицируемых поверхностей.

Более распространен метод оценивания качества идентификации, основанный на анализе содержания матрицы соответствия, или матрицы погрешности [2]. Содержание матрицы приведено в таблице 1.

Таблица 1. Матрица погрешностей и показатели точности идентификации.

		Данные контрольной выборки			Всего в строке	UA
		А	В	С		
Результаты идентификации	А	P_{AA}	P_{AB}	P_{AC}	P_{A+}	P_{AA}/P_{A+}
	В	P_{BA}	P_{BB}	P_{BC}	P_{B+}	P_{BB}/P_{B+}
	С	P_{CA}	P_{CB}	P_{CC}	P_{C+}	P_{CC}/P_{C+}
Всего в столбце		P_{+A}	P_{+B}	P_{+C}		
РА		P_{AA}/P_{+A}	P_{BB}/P_{+B}	P_{CC}/P_{+C}		

В таблице 1 приняты следующие обозначения: P_{ij} – количество пикселей, идентифицированных как принадлежащие элементу ландшафта i , на самом деле соответствующие элементу контрольной выборки j ; символ + обозначает результаты суммирования значений, приведенных в соответствующих строках и столбцах; UA (user's accuracy) – значения показателей, отражающих интересы должностного лица, принимающего решение при управлении развитием территории; РА (producer's accuracy) – значения показателей, отражающих интересы эксперта, представляющего результаты обработки материалов съемки.

Достоинством метода является возможность вычисления ряда дополнительных показателей для каждого элемента ландшафта, что может иметь важное значение при принятии управленческих решений.

Также известен метод, при котором осуществляется расчет и анализ значений показателей «точности» P и «чувствительности» R идентификации двух поверхностей в пределах сцены. [2, 5, 6]:

$$(1) \quad P = \frac{TP}{(TP+FP)}, R = \frac{TP}{(TP+FN)}.$$

В формулах (1) приняты следующие обозначения: TP - истинно положительный результат (количество пикселей, правильно отнесенных к первому классу); TN - истинно отрицательный результат (количество пикселей, правильно отнесенных ко второму классу); FP - ложноположительный результат (количество пикселей второго класса, неправильно отнесенных к первому классу); FN - ложноотрицательный

результат (количество пикселей первого класса, неправильно отнесенный ко второму классу).

Достоинством метода является качественное углубление смыслового содержания промежуточных и итоговых показателей по сравнению с методом определения ошибок первого и второго рода.

Общий недостаток рассмотренных методов определяется тем, что они основаны на данных контрольной выборки, которая формируется, как правило, на основе архивных исходных данных. Архивные исходные данные могут устаревать и не отражать реальной картины. Поэтому актуальной становится задача оперативного обновления содержания контрольной выборки объективными оперативными данными и дальнейшего расчета уточненных значений показателей качества результатов обработки материалов съемки.

Предлагается новый метод расчета уточненных значений показателей качества результатов обработки материалов съемки. Метод предусматривает, в частности, оперативное обновление и уточнение контрольной выборки. Структурная схема предлагаемого метода приведена на рис.1.

3. Возможности алгоритмов кластеризации при выборе дополнительных точек обследований

При реализации предлагаемого метода для формирования уточненной контрольной выборки осуществляется выбор точек дополнительных обследований (рис.1). Данные уточненной контрольной выборки в дальнейшем используются при расчете уточненных значений показателей качества результатов обработки материалов съемки.

Рассмотрим суть предлагаемого метода на примере. Пусть в результате автоматизированной идентификации с помощью одного из алгоритмов автоматизированной обработки предварительно выявлена принадлежность девяти пикселей определенному виду ландшафта, условному кластеру. Возникает задача выбора из этих пикселей точек для проведения дополнительных обследований.

Предлагается выбор дополнительных точек осуществлять на основе анализа степени принадлежности каждого пикселя рассматриваемому кластеру. В качестве степени принадлежности выбрано Евклидово расстояние между условным центром кластера и каждым пикселем внутри кластера в n -мерном пространстве, где n совпадает с количеством спектральных каналов в материалах мультиспектральной съемки.

Введем допущение о том, что пиксели, характеризующиеся невысокой степенью принадлежности кластеру, требуют первостепенного внимания в ходе дальнейших обследований.

Рассмотрим следующие алгоритмы кластеризации: K-средних (k-means) и алгоритм нечеткой кластеризации (FCM) [1, 7]. Выбор алгоритмов обусловлен отсутствием необходимости формирования обучающей выборки и наличием в теле алгоритмов процедуры расчета степени принадлежности каждого пикселя рассматриваемому кластеру. Алгоритм FCM похож на алгоритм k-means, но, в отличие от алгоритма k-means, нечеткая кластеризация обеспечивает на практике возможность обоснованного выбора пикселей, требующих внимания при проведении дальнейших обследований. Это свойство алгоритма FCM достигается благодаря включению в алгоритм параметра «нечеткости»- фаззификатора $m \geq 1$. Роль параметра m можно оценить при анализе кривых, приведенных на рис. 2.

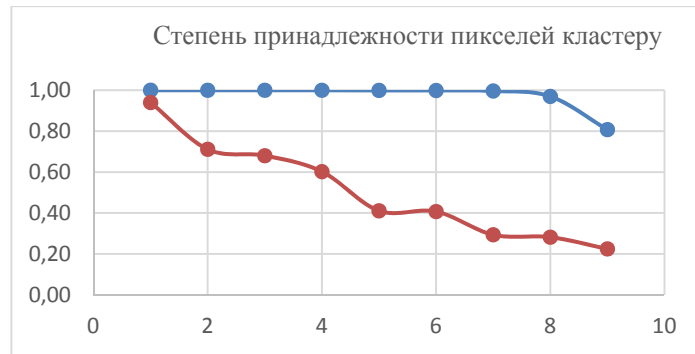


Рис. 2. Данные для обоснования выбора точек дополнительных обследований. Красная кривая соответствует результатам применения алгоритма k-means, синяя – FCM.

В результате применения алгоритмов k-means и FCM получен один и тот же результат - подтверждена принадлежность одних и тех же девяти пикселей рассматриваемому кластеру. На рис. 2 демонстрируются кривые, характеризующие степень принадлежности пикселей кластеру. Благодаря наличию в алгоритме FCM фазсификатора $m=2$ можно отметить, что восьмой и девятый пиксели, с учетом ранее сформулированного допущения, требуют дополнительных обследований. По итогам дополнительных обследований, в случае, если подтвердится принадлежность пикселей рассматриваемому кластеру, они обоснованно могут быть включены в контрольную выборку. Плавная убывающая красная кривая также характеризует степень принадлежности пикселей рассматриваемому кластеру, однако выбрать пиксели, требующие дополнительных обследований, затруднительно.

Таким образом, в докладе представлены основные положения метода оценивания качества результатов обработки материалов аэрокосмической съемки. Практическая реализация метода позволит получить обоснованные оценки качества результатов обработки материалов съёмки. При этом уточненные значения показателей качества обработки могут принимать более высокие или низкие значения в зависимости от содержания уточненной контрольной выборки.

Исследование выполнено в рамках бюджетной темы FFZF-2022-0004.

Список литературы

1. Шовенгердт Р.А. Дистанционное зондирование. Методы и модели обработки изображений. Техносфера. 2010. 560 с.
2. Russell G. Congalton and Kass Green Assessing the Accuracy of Remotely Sensed Data. Principles and Practices. Third Edition. CRC Press, 2019.
3. Рис У.Г. Основы дистанционного зондирования. Москва: Техносфера, 2006. 336 с.
4. Горелик А.Л. Методы распознавания. М.: Высшая школа, 2004. 261 с.
5. Tharwat A. Classification assessment methods // Applied Computing and Informatics. 2018. № 08.
6. Maxwell A.E., Warner T.A. Guillén L.A. Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review // Remote Sens. 2021. Vol. 13. 2450.
7. Ракша С. Python и машинное обучение / Пер. с англ. А.В. Логунова. М.: ДМК Пресс, 2017. 418 с.
8. Книжников Ю.Ф., Кравцова В.И., Тутубалина О.В. Аэрокосмические методы географических исследований. М.: Издательский центр «Академия», 2004. 336 с.
9. Броневиц А.Г., Лепский А.Е. Нечеткие модели анализа данных и принятия решений. М.: Изд. дом Высшей школы экономики, 2022. 264 с.