

ТЕХНОЛОГИЯ ОЦЕНИВАНИЯ ТОЧНОСТИ РЕЗУЛЬТАТОВ ОБРАБОТКИ МАТЕРИАЛОВ МУЛЬТИСПЕКТРАЛЬНОЙ АЭРОКОСМИЧЕСКОЙ СЪЕМКИ

В.Ф. Мочалов

Санкт-Петербургский Федеральный исследовательский центр РАН
Россия, 199178, Санкт-Петербург, 14-я линия В.О., 39
E-mail: vicavia@yandex.ru

Ключевые слова: системное моделирование, автоматизированная обработка, точность результатов обработки, матрица погрешностей, развивающаяся ситуация.

Аннотация: Качество результатов автоматизированной обработки материалов аэрокосмической съемки может оцениваться на основе показателей точности идентификации элементов ландшафта в пределах контролируемой сцены. Представлена обобщенная технология и математический аппарат для оценивания точности идентификации. Показано, что показатели оценивания точности обработки материалов съемки зависят, в том числе, от корректного оперативного учета текущих изменений состояния элементов ландшафта в пределах контрольной выборки. На основе анализа значений частных показателей точности обеспечивается обоснованное управление качеством обработки материалов съемки.

1. Введение

Информационные технологии при управлении развитием территорий все более часто опираются на результаты регулярной автоматизированной обработки материалов мультиспектральной аэрокосмической съёмки. Передовые методы обработки материалов съемки опираются на сложные модели, реализующие комплексное применение алгоритмов классификации, кластеризации и машинного обучения [1-5]. При этом могут приниматься во внимание спектрально-яркостные и пространственно-частотные характеристики элементов ландшафта с учетом временных рядов исходных данных. Спектрально-яркостные характеристики имеют фундаментальное значение для идентификации и оценивания состояний растительных покровов, водных объектов, селитебных территорий. Материалы мультиспектральной съемки обеспечивают возможность учета спектральных отражательных характеристик, как правило, более чем в десяти спектральных каналах видимого и ближнего инфракрасного диапазонов спектра. Пространственно-частотные характеристики обычно лежат в основе мониторинга инфраструктурных объектов. В результате анализа временных рядов исходных данных отслеживается динамика изменений в состоянии природных или антропогенных объектов.

Оценивание точности (качества) результатов обработки материалов съемки осуществляется с целью предоставления сведений о достоверности информации, полученной в результате идентификации элементов ландшафта в пределах рассматриваемой сцены. При этом все более полно задействуется технология,

отражающая реализацию субъект-объектной модели процесса моделирования [6]. В упомянутой модели можно выделить следующие основные элементы: субъект, формулирующий цель и задачи обработки материалов съемки; объект-оригинал, вмещающий элементы ландшафта с изменяющимися характеристиками под воздействием внешних факторов; объект-модель в форме результатов автоматизированной идентификации элементов ландшафта в момент времени, соответствующий обрабатываемым материалам съемки; среда моделирования, включающая спектральный состав исходных данных съемки, промежуточные показатели и их пороговые значения, алгоритмы обработки материалов съемки. Все перечисленные элементы и отношения между ними под действием объективно-субъективных внешних-внутренних причин постоянно изменяются во времени. Мгновенно зафиксированное состояние объекта-модели в виде результатов обработки называется ситуацией, отражающей состояние объекта оригинала, а изменения состояния во времени – развивающейся ситуацией. При обработке материалов съемки идентифицируются только те элементы ландшафта, которые имеют значения при решении конкретной практической задачи. В ходе системного моделирования развивающейся ситуации становится актуальной задача систематического оценивания точности результатов автоматизированной обработки материалов съемки [7].

Целью настоящей работы является краткое описание технологии и математического аппарата для оценивания точности результатов идентификации элементов ландшафта. Описание включает:

- структурную схему технологии, которая предусматривает расчет обобщенных и частных показателей точности обработки материалов съемки, представление данных для управления качеством обработки;
- математический аппарат для расчета обобщенных и частных показателей точности обработки материалов съемки;
- практический пример расчета обобщенного и частных показателей точности обработки материалов съемки в условиях развивающейся ситуации.

2. Методический аппарат и пример его применения

Рассмотрим задачу оценивания точности результатов автоматизированной идентификации элементов ландшафта объекта-оригинала с известными спектрально-яркостными отражательными характеристиками и соответствующим семантическим описанием, обеспечивающим решение практической задачи в пределах рассматриваемой сцены. В результате применения комплекса алгоритмов обработки материалов съемки получены модельные данные о пространственном размещении элементов ландшафта в пределах сцены и контрольной выборки, которые подлежат сравнению с реальными данными контрольной выборки объекта-оригинала. Контрольная выборка объекта-оригинала сформирована заранее на основе архивных данных об объекте и включает достоверные сведения о каждом элементе ландшафта в пределах сцены на момент формирования контрольной выборки.

Показатели точности обработки характеризуют степень совпадения выходных материалов в пределах модели контрольной выборки с объективными данными о контрольной выборке объекта-оригинала.

Вместе с тем данные контрольной выборки объекта-оригинала могут устаревать. Как правило, они отражают ситуацию, предшествующую времени выполнения съемочных работ. Под влиянием внешних факторов данные контрольной выборки могут претерпевать изменения. Более позднее время выполнения съемочных работ обуславливает необходимость учета развивающейся ситуации и внесения

обоснованной корректировки в оригинал контрольной выборки. В связи с этим показатели точности результатов обработки материалов съёмки также будут претерпевать изменения.

На рис. 1 приведена структурная схема технологии оценивания точности результатов обработки, разработанная на основе обобщенной технологии системного моделирования развивающейся ситуации при ее управленческой интерпретации [6].

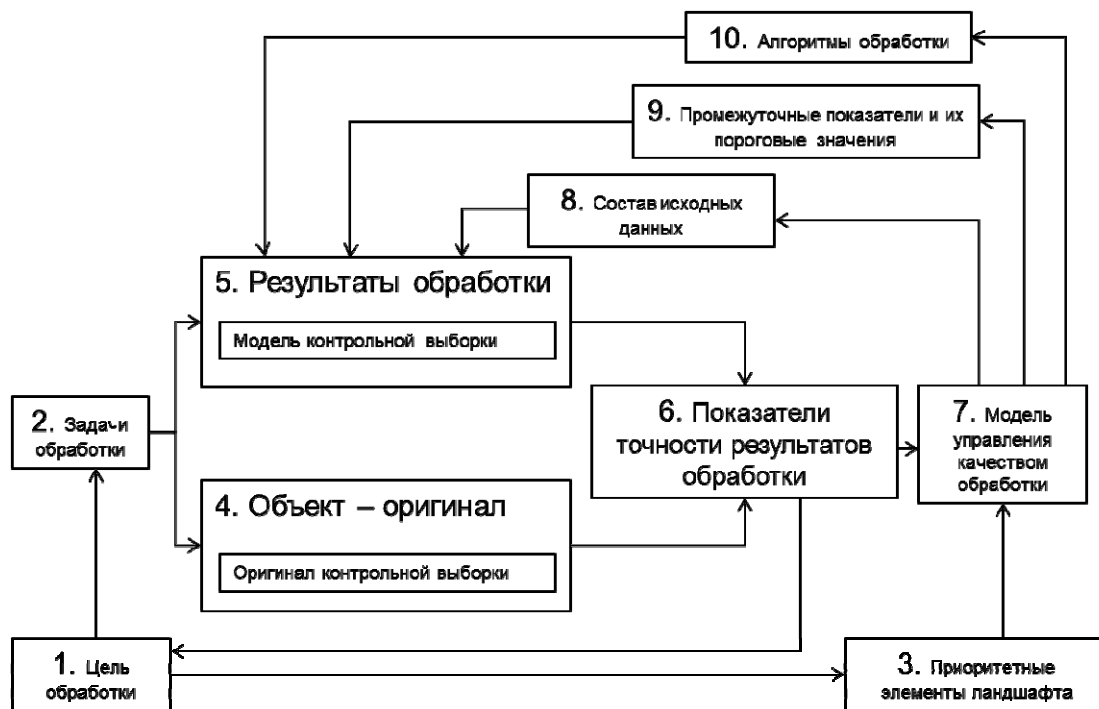


Рис. 1. Схема технологии оценивания точности результатов обработки материалов съёмки.

На схеме приняты следующие условные обозначения: 1 - цель обработки материалов съёмки в пределах сцены, требуемые значения обобщенных и частных показателей точности идентификации элементов ландшафта в пределах сцены; 2 - задачи обработки конкретных материалов съёмки; 3 – перечень приоритетных элементов ландшафта, точность идентификации которых требуется оценить; 4 – объект-оригинал сцены, для которого известны достоверные данные об элементах ландшафта в пределах контрольной выборки; 5 – модель сцены, включающая контрольную выборку, полученную в результате обработки конкретных материалов съёмки; 6 – показатели точности, на основе анализа которых дается заключение о достижении заданных требований к качеству результатов обработки в целом и для конкретных элементов ландшафта в частности; 7 – модель управления качеством обработки, обеспечивающая обоснованный выбор параметров обработки; 8 – состав исходных данных в виде набора наиболее информативных спектральных каналов; 9 – промежуточные показатели обработки и их пороговые значения; 10 – алгоритмы автоматизированной обработки, включающие алгоритмы классификации, кластеризации и машинного обучения.

Математический аппарат для расчета обобщенных и частных показателей точности в большинстве случаев опирается на данные, представленные в матрице погрешностей (Confusion Matrix) [7, 8]. В таблице 1 приведен пример значений матрицы погрешностей, рассчитанных в условиях развивающейся ситуации. Представлены данные для ситуации с условным номером 1 и в скобках для ситуации 2.

Таблица 1. Матрица погрешностей и показатели точности идентификации.

		Данные оригинала контрольной выборки, пиксели				UA, %
		Озеро	Болото	Поле	Лес	
Данные модели контрольной выборки, пиксели	Озеро	187 (150)	40 (42)	7	0	79,9 (75,4)
	Болото	11 (21)	246 (271)	12	9	88,5 (86,6)
	Поле	0	21	239 (236)	39	79,9 (79,7)
	Лес	0	0	49	140 (143)	74,1
РА, %		94,4 (87,7)	80,1 (81,1)	77,9 (77,6)	74,5 (74,9)	
R, %		82,4 (78,1)		86,0 (85,8)		
P, %		94,4 (87,7)		83,0 (82,8)		

Размер контрольной выборки составляет 1000 пикселей. Элементы матрицы, расположенные по главной диагонали, соответствуют корректно идентифицированным пикселям в пределах контрольной выборки, а все элементы матрицы, расположенные вне главной диагонали, соответствуют ошибочно идентифицированным пикселям.

Точность потребителя [7] определяется как:

$$UA = \frac{E_{i,i}}{\sum_{j=1}^N E_{j,i}}$$

Точность производителя определяется по формуле:

$$PA = \frac{E_{i,i}}{\sum_{j=1}^N E_{i,j}}$$

Общая точность идентификации определяется как:

$$W_{\text{общ}} = \frac{\sum_{i=1}^N E_{i,i}}{\sum_{i=1}^N \sum_{j=1}^N E_{i,j}}$$

Также известны показатели «точности» P (Precision) и «чувствительности» R (Recall) идентификации двух поверхностей в пределах сцены. [7, 8]. В таблице 1 приведены значения показателей P и R для двух пар поверхностей «озеро-болото» и «поле-лес», граничащих друг с другом в пределах сцены.

На основе анализа значений таблицы можно отметить, что в ситуации 1 в пределах контрольной выборки 812 пикселей идентифицированы без погрешности. Общая точность идентификации $W_{\text{общ}}$ при этом составляет 81,2 %.

Пусть развивающаяся ситуация сопровождается изменениями во времени объекта - оригинала в части состава контрольной выборки. Изменения произошли под влиянием внешних факторов в период между начальным формированием контрольной выборки и проведением съемочных работ. Изменения были учтены в ходе дополнительных наземных обследований объекта - оригинала. Результаты перехода от ситуации 1 к ситуации 2 и изменений в содержании контрольной выборки представлены в таблице 1 в скобках.

Изменения вызваны объективными факторами и затронули уточнение оригинала контрольной выборки в части количества пикселей озера, болота, поля и леса. На рис. 2 приведены графики, отражающие изменения показателей точности результатов обработки материалов съемки в различных ситуациях.

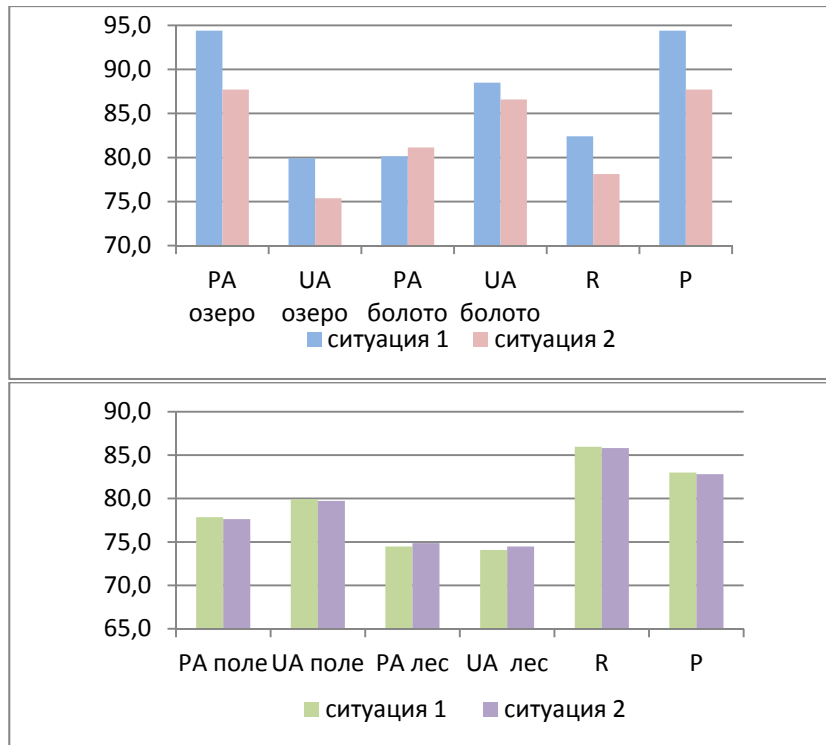


Рис. 2. Изменения показателей точности результатов обработки в результате корректировки оригинала контрольной выборки.

На основе анализа зависимостей, приведенных на рис. 2, можно сделать вывод об изменчивости показателей качества идентификации выбранных пар элементов сцены в условиях развивающейся ситуации. Так точность идентификации пары поверхностей болото-озеро существенно снижена за счет более корректного формирования данных оригинала контрольной выборки. Невысокая точность идентификации пары поле-лес обуславливает необходимость корректировки модели управления качеством обработки материалов съемки.

Предложенная технология и математический аппарат оценивания точности обработки материалов съемки позволяют сделать выводы о необходимости уточнения оригинала контрольной выборки и модели обработки материалов съемки.

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. Ракша С. Python и машинное обучение / Пер. с англ. А.В. Логунова. М.: ДМК Пресс, 2017. 418 с.
4. 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. P. 2450.
5. Горелик А.Л. Методы распознавания. М.: Высш. шк., 2004. 261 с.
6. Микони С.В., Соколов Б.В., Юсупов Р.М. Квалиметрия моделей и полимодельных комплексов: монография. М.: РАН, 2018. 314 с.
7. Рис У.Г. Основы дистанционного зондирования. М.: Техносфера, 2006. 336 с.
8. Tharwat A. Classification assessment methods // Applied Computing and Informatics. 2018. № 8.