

Методика формирования первичных и уточнённых значений исходных данных

Мочалов В.Ф.

Военно- космическая академия имени А.Ф. Можайского

Цель: формирование первичных и уточненных значений исходных данных в пределах фрагмента сцены для повышения качества результатов обработки материалов съёмки.

В качестве исходных данных рассматривается множество пикселей, представляющих элементы ландшафта $R = \{r_i(x_n, y_m, \lambda_l)\}$.

Рассматривается задача идентификации или семантической сегментации элементов ландшафта в пределах сцены на основе материалов мультиспектральной съёмки.

Вопросы, обсуждаемые в стендовом докладе:

1. Операции и структурная схема методики
2. Состав исходных данных и пример формирования первичных и уточнённых значений исходных данных
3. Порядок расчета базовых и обобщенных показателей качества результатов обработки материалов съёмки
4. Выводы

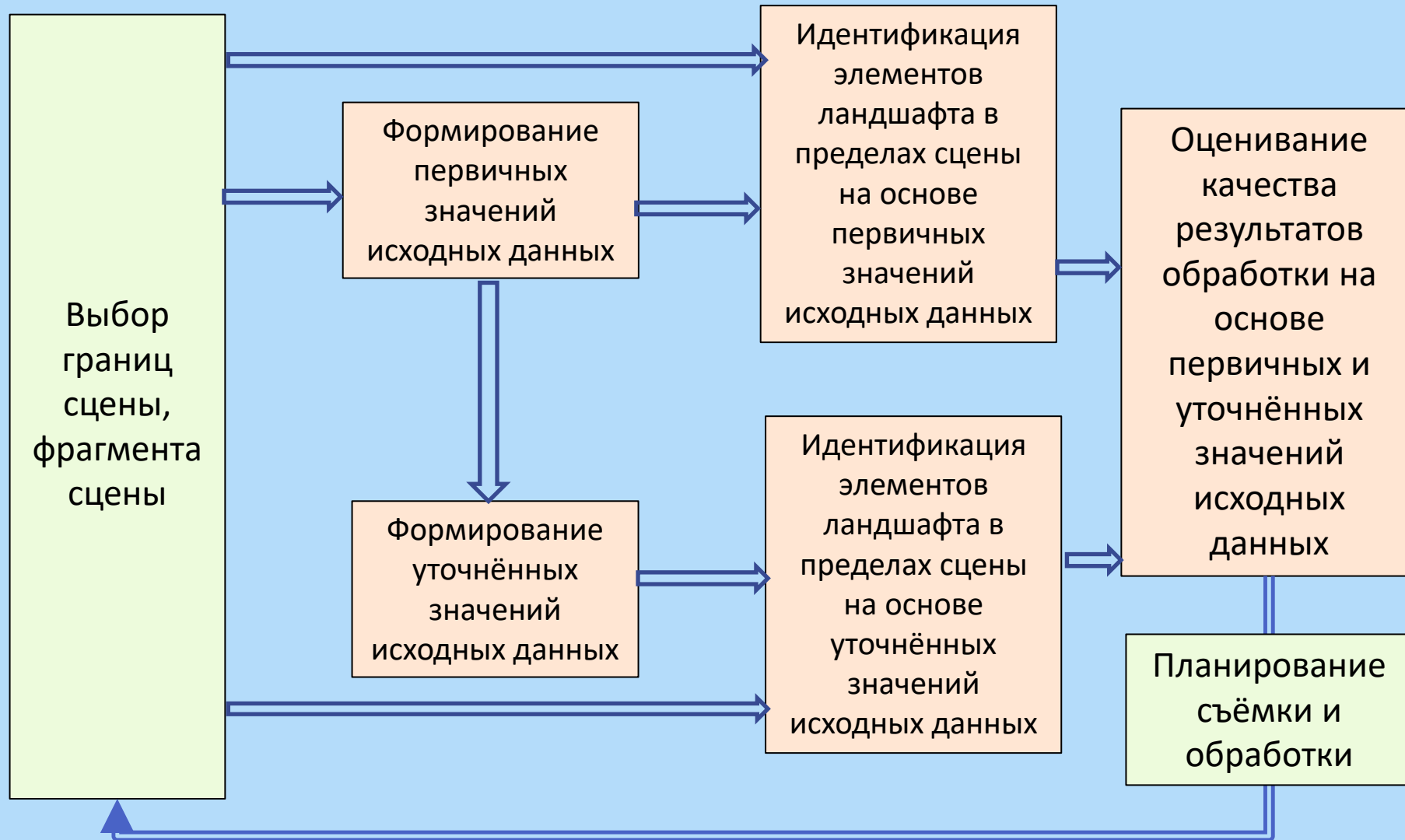
Операции методики

Вопрос 1

- Выбор оператором условных границ сцены, на которой планируется решить задачу автоматизированной идентификации элементов ландшафта, а также фрагмента сцены, содержащего только элементы ландшафта с известными свойствами для формирования исходных данных;
- автоматизированное формирование первичных значений исходных данных для элементов ландшафта в пределах фрагмента сцены в части обучающих и контрольных выборок на основе математического аппарата нечёткой кластеризации (FCM);
- формирование уточнённых значений исходных данных на основе анализа степени принадлежности элементов изображения (пикселей) рассматриваемым кластерам;
- идентификация элементов ландшафта в пределах сцены на основе первичных и уточнённых значений обучающей выборки и одного из алгоритмов машинного обучения;
- оценивание базовых показателей качества результатов обработки материалов съёмки на основе первичных и уточнённых значений исходных данных, анализ полученных результатов.

Структурная схема методики формирования первичных и уточнённых значений исходных данных

Вопрос 1



Состав исходных данных для идентификации элементов ландшафта в пределах сцены

Условный кадр материалов съёмки

Сцена, часть условного кадра – территория, на которой решается задача идентификации элементов ландшафта с известными свойствами, выбирается оператором

Фрагмент сцены, выбранный оператором, и вмещающий только идентифицируемые элементы ландшафта с известными свойствами

$$R = \{r_i(x_n, y_m, \lambda_l)\}$$

Автоматизированная идентификация элементов ландшафта в пределах сцены осуществляется на основе алгоритмов машинного обучения

Элемент ландшафта 1

Обучающая выборка 1

Контрольная выборка 1

Обучающая выборка 2

Контрольная выборка 2

Элемент ландшафта 2

Первичные исходные данные

$$R^{(1)} = \{R_1^{(1)}, R_2^{(1)}\}$$

Вопрос 2

Элемент ландшафта 1

Обучающая выборка 1

Контрольная выборка 1

Обучающая выборка 2

Контрольная выборка 2

Элемент ландшафта 2

Уточнённые исходные данные

$$R^{(2)} = \{R_1^{(2)}, R_2^{(2)}\}$$

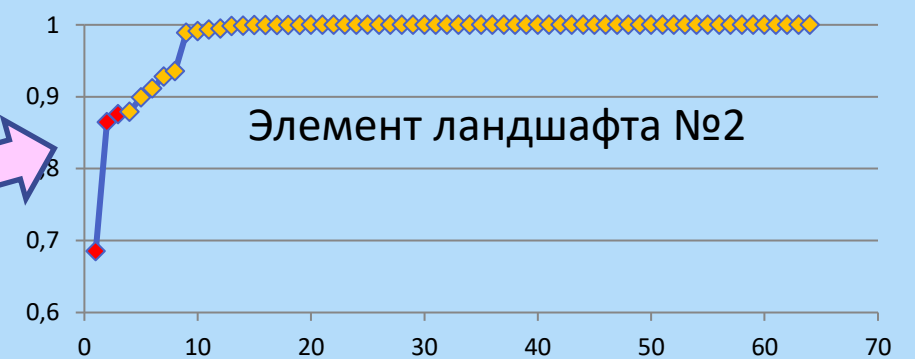
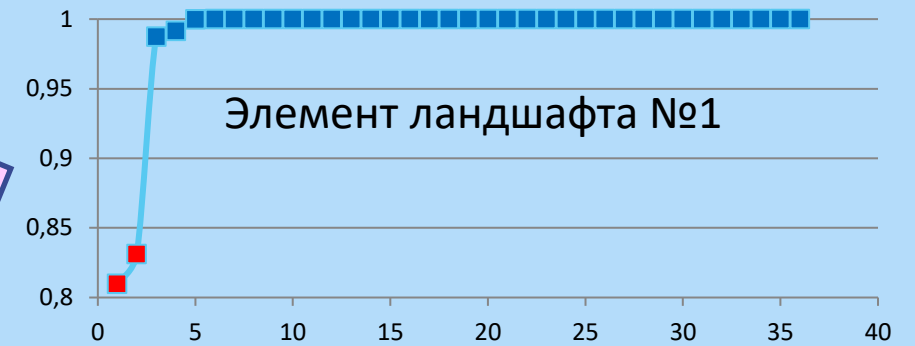
Пример формирования первичных и уточнённых значений исходных данных в пределах фрагмента сцены

Вопрос 2

Результаты кластеризации в рамках фрагмента сцены на основе алгоритма FCM



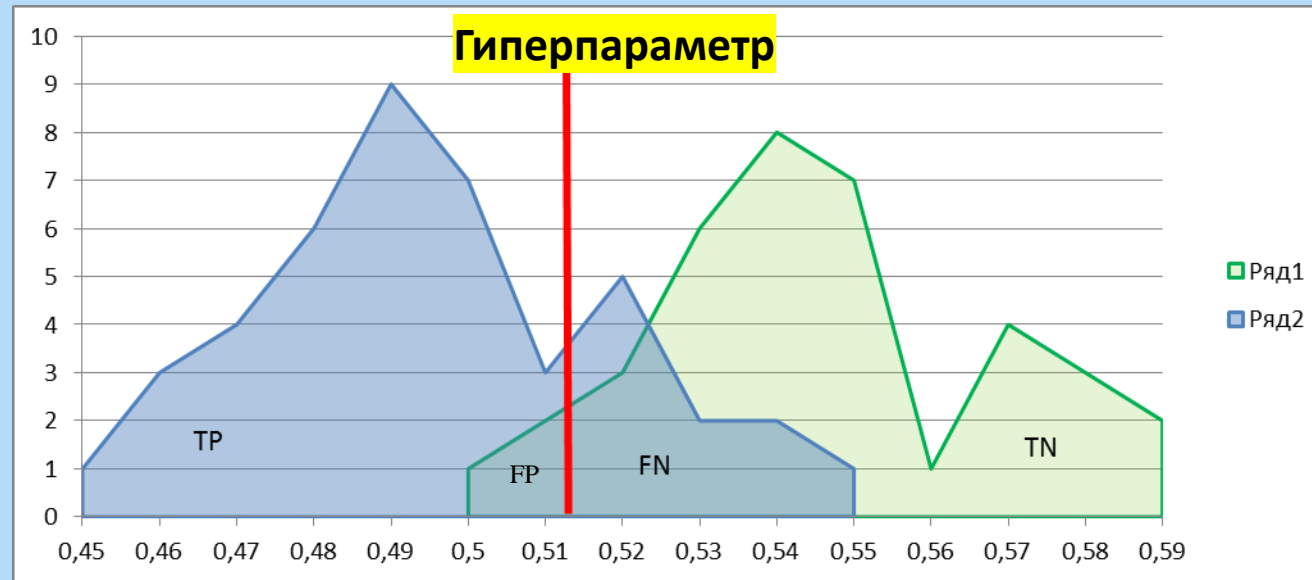
Графики степени принадлежности пикселей соответствующим кластерам



Синие пиксели представляют элемент ландшафта № 1, желтые –элемент ландшафта № 2.

Красными точками отмечены пиксели с невысокой степенью принадлежности рассматриваемым кластерам. Уточнение значений исходных данных осуществляется за счет обоснованного их исключения из обучающих и контрольных выборок.

Пример расчёта базовых показателей качества результатов идентификации элементов ландшафта в пределах сцены



Вопрос 3

TP (True Positive) - количество пикселей истинно принадлежащих элементу ландшафта №1;
TN (True Negative)-количество пикселей истинно принадлежащих элементу ландшафта №2;
FP (False Positive)-количество пикселей ошибочно принадлежащих элементу ландшафта №1;
FN (False Negative)-количество пикселей ошибочно принадлежащих элементу ландшафта №2.

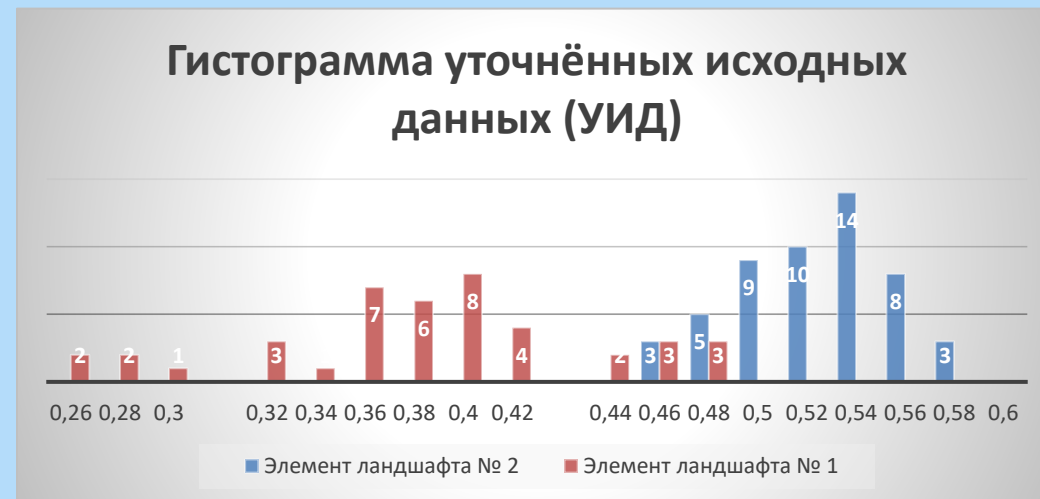
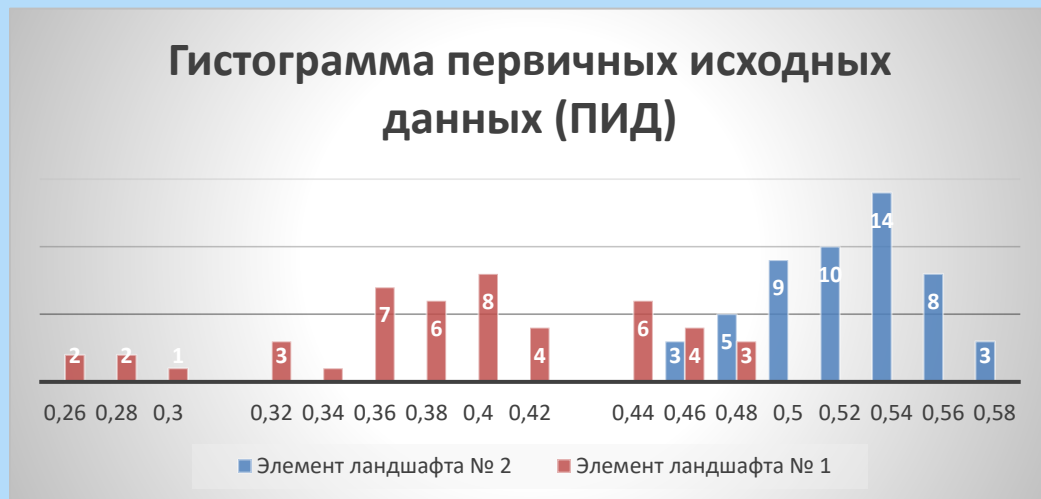
Обобщенные показатели качества результатов обработки материалов съемки

Вопрос 3

Показатель	Описание
$REC = \frac{TP}{TP + FN}$	<p>Чувствительность- доля в контрольной выборке элементов ландшафта, которые характеризуются условно положительными свойствами (элемент 1)</p>
$PRE = \frac{TP}{TP + FP}$	<p>Точность – это отношение между истинно положительными и всеми положительными результатами идентификации.</p>
$OA = \frac{TP + TN}{TP + TN + FP + FN}$	<p>Общая точность привлекаемого метода обработки, полезно знать, когда каждый идентифицируемый элемент ландшафта одинаково важен.</p>
$F1 = 2 \frac{PRE * REC}{PRE + REC}$	<p>Комплексный показатель объединяющий показатели <i>REC</i> и <i>PRE</i> в виде гармонического среднего</p>
$F_{\beta} = \frac{(1 + \beta^2)PRE REC}{\beta^2PRE + REC}$	$\begin{cases} \beta < 1 \rightarrow PRE \\ \beta > 1 \rightarrow REC \end{cases}$

Пример расчета числовых значений показателей качества обработки материалов съёмки

Вопрос 3



Показатели качества при значении гиперпараметра 0,45

	ПИД	УИД	ΔA	$\delta, \%$
TP	40	40		
TN	52	52		
FP	0	0		
FN	7	3		
REC	0,85	0,93	0,08	8,51
PRE	1,00	1,00		
OA	0,93	0,97	0,04	4,04
F1	0,92	0,96	0,04	4,60

Выводы

1. Представлена методика формирования исходных данных, которые могут применяться при идентификации элементов ландшафта в пределах рассматриваемой сцены на основе алгоритмов машинного обучения.
2. Методика основана на применении математического аппарата нечёткой кластеризации в рамках фрагмента сцены и обеспечивает формирование первичных и уточнённых значений исходных данных.
3. Обоснованное формирование уточнённых значений исходных данных сопровождается контролируемым повышением качества результатов автоматизированной обработки материалов съёмки от 4 до 8,5%.
4. Предложенная методика может применяться при обосновании требований к материалам съёмки и прогнозировании показателей качества результатов идентификации элементов ландшафта.