

Разработка пакета функций на языке Python для обработки гиперспектральных изображений и построения модели для предсказания содержания белка в одиночных зерновках пшеницы

Автор: Карнаухова Виктория Олеговна

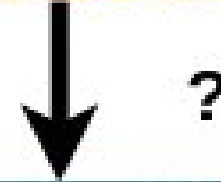
# Зачем нам нужно знать значение белка в зернах пшеницы?

- Для приготовления различных хлебобулочных изделий требуется мука с разным значением белка
- **Традиционный подход** - смешивать партии зерна с высоким и низким содержанием белка
- **Новый подход** - отбирать зерна нужного белкового содержания
- **Задача** - определять показатели белка для зерновки за миллисекунды

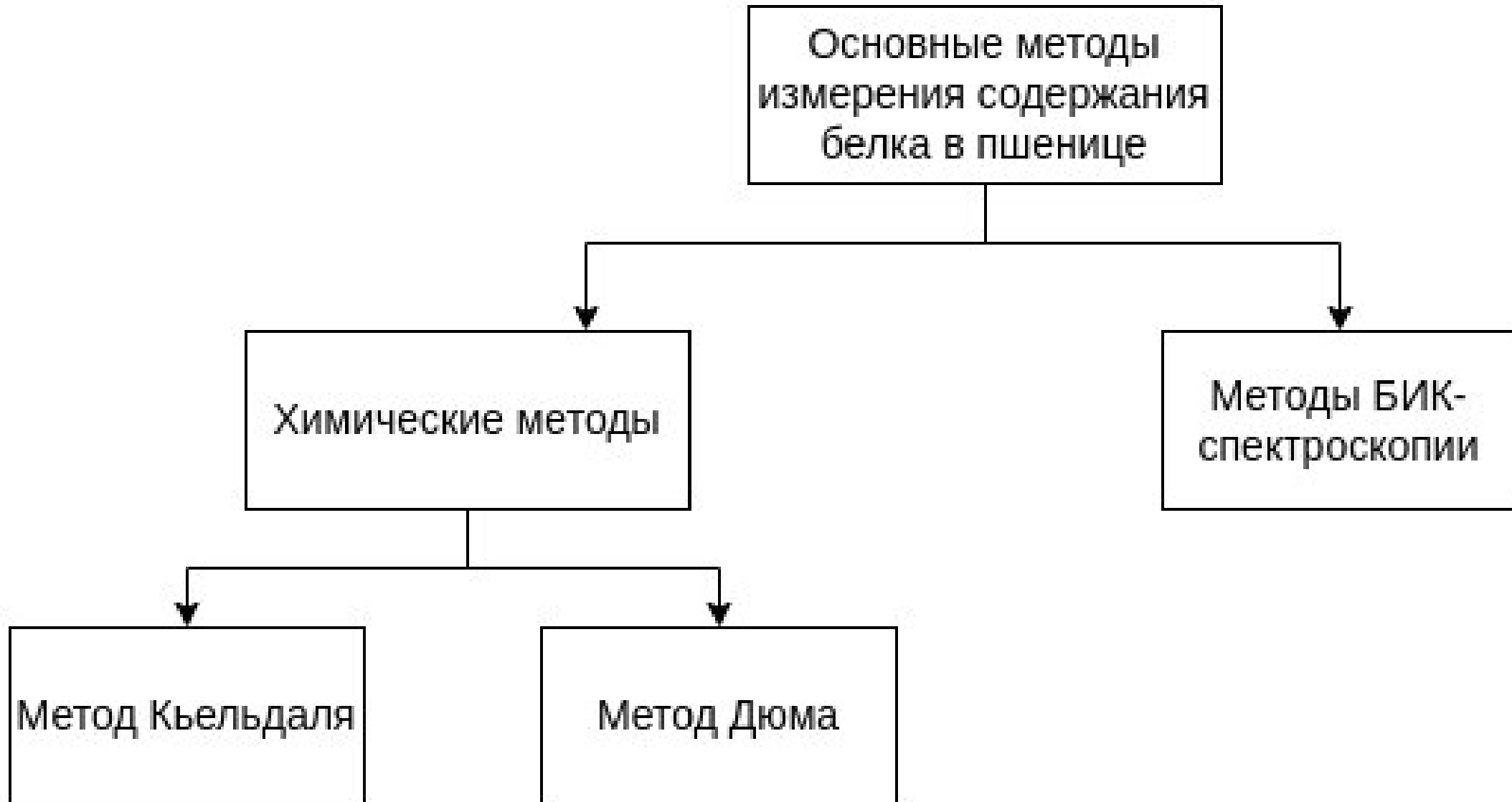


# Зачем нам нужно знать значение белка в зернах пшеницы?

- От количества белка в зерновке зависит ее способность к прорастанию
- Можем заранее убрать зерна с «плохим» содержанием белка

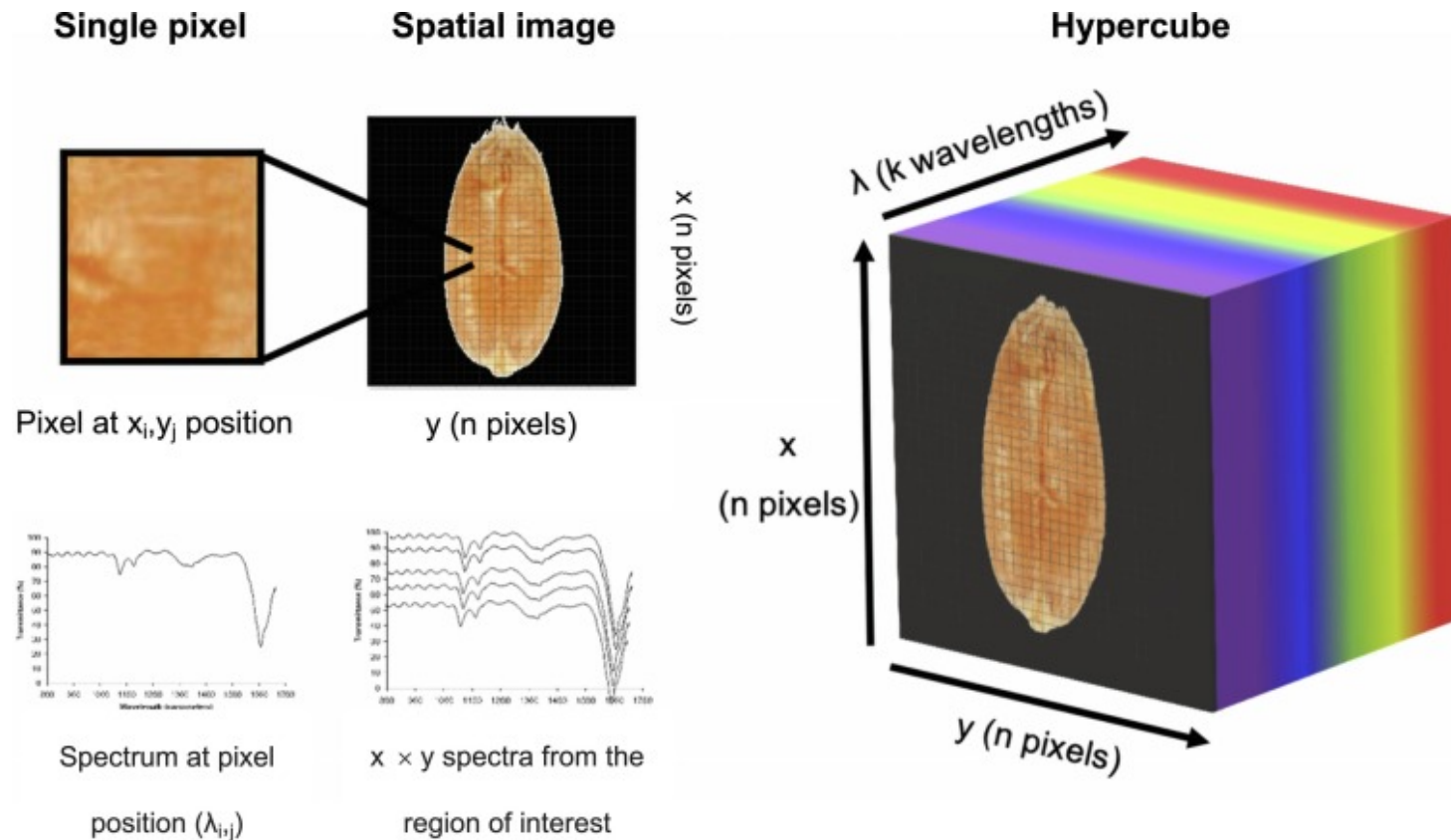


# Основные методы измерения содержания белка в пшенице

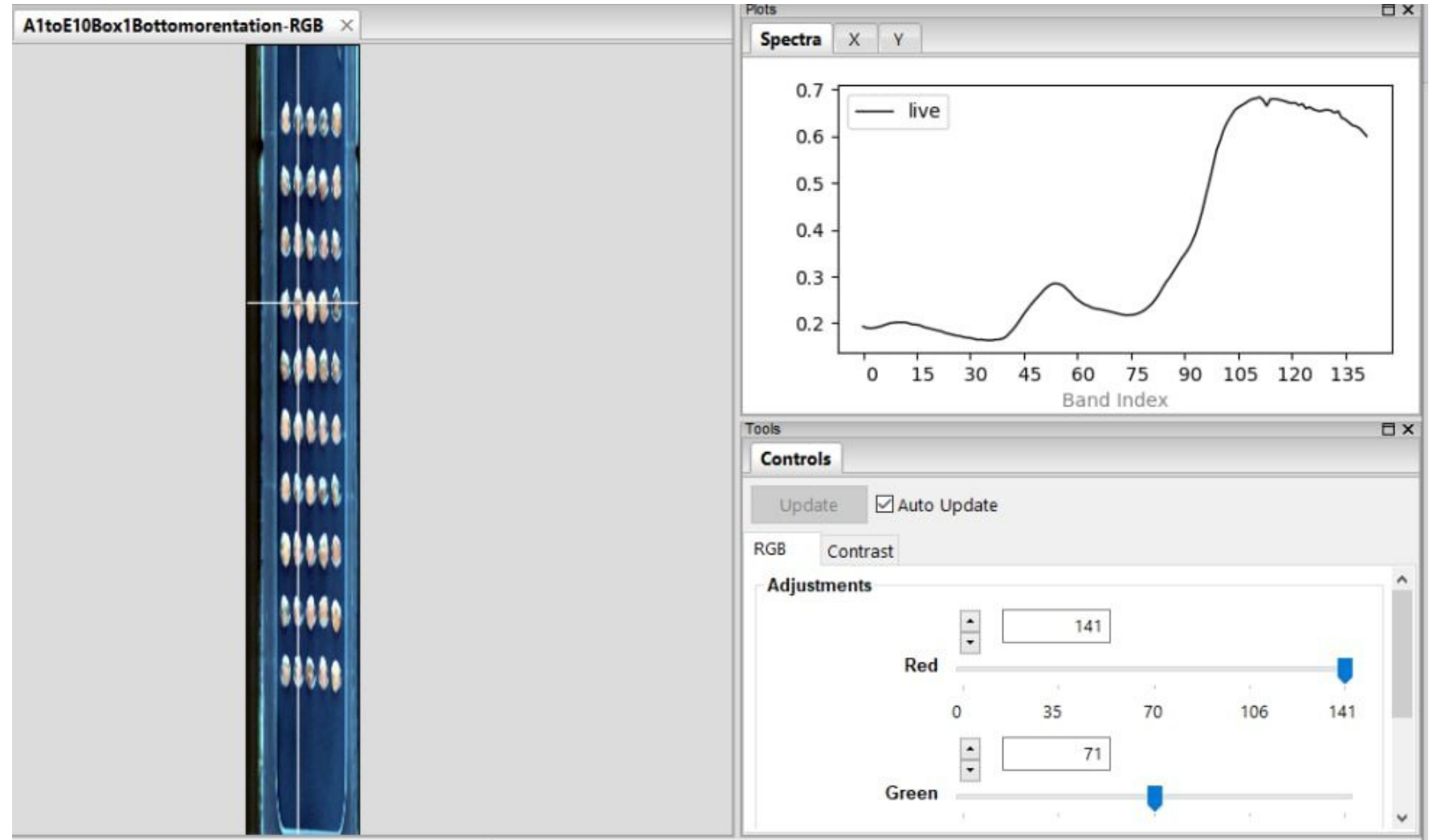
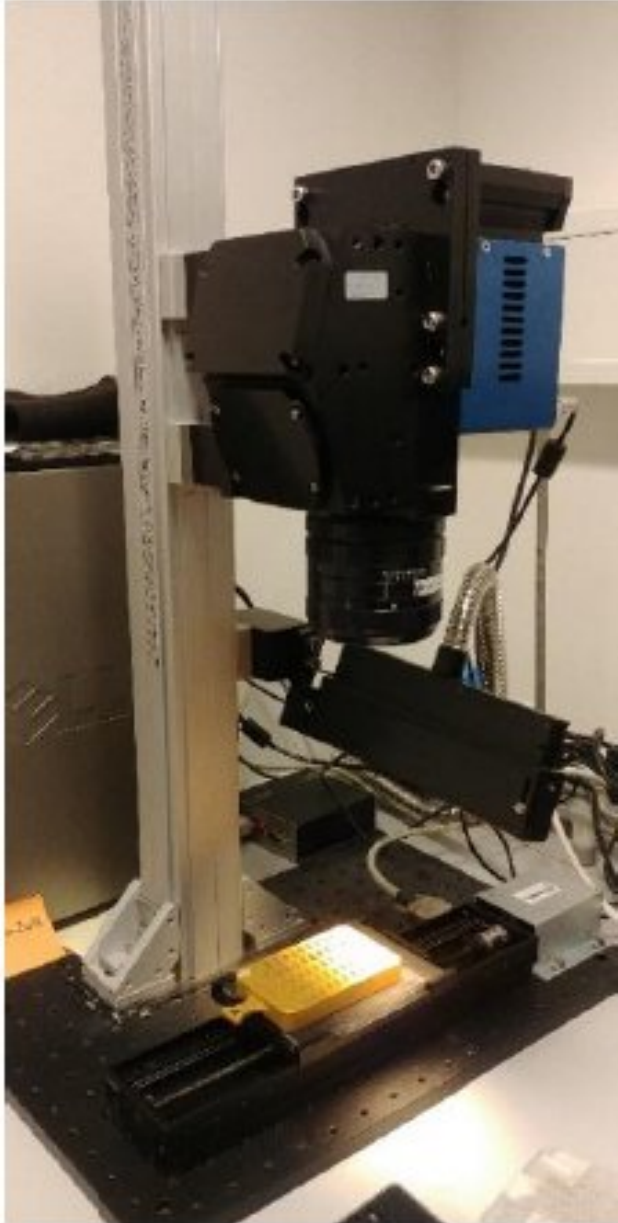


# Гиперспектральные изображения (HSI, hyperspectral image)

Изображения, в которых каждый пиксел представляется набором значений спектральных откликов на определенных длинах волн



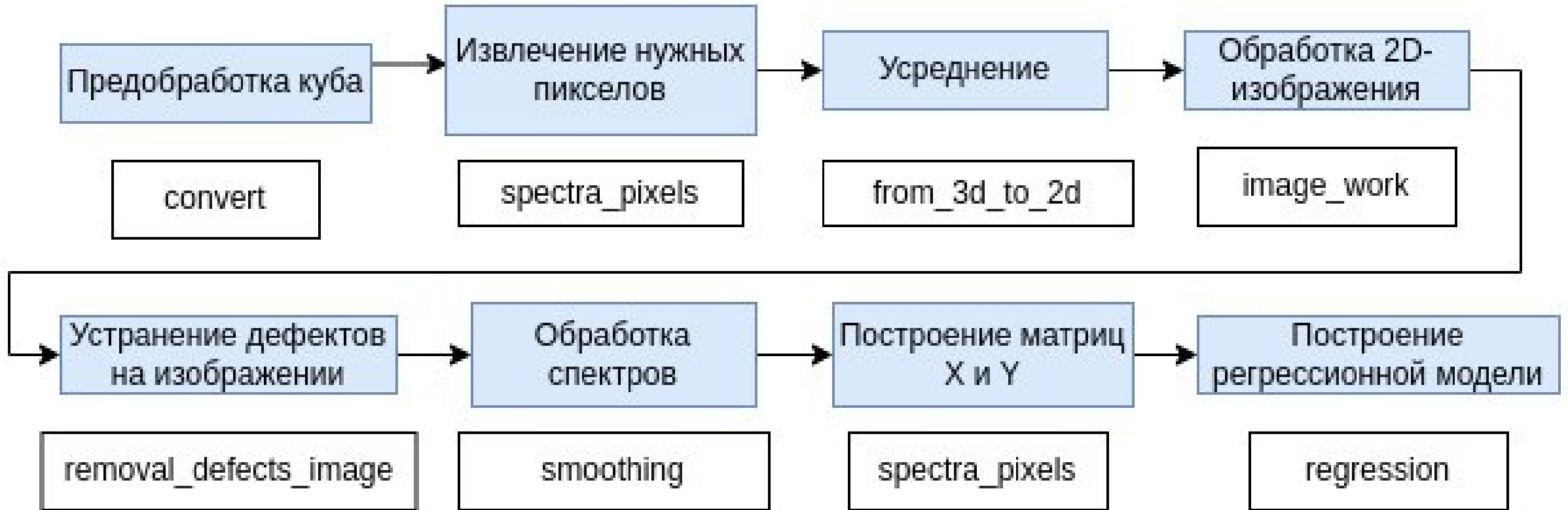
# Входные данные



# Входные данные

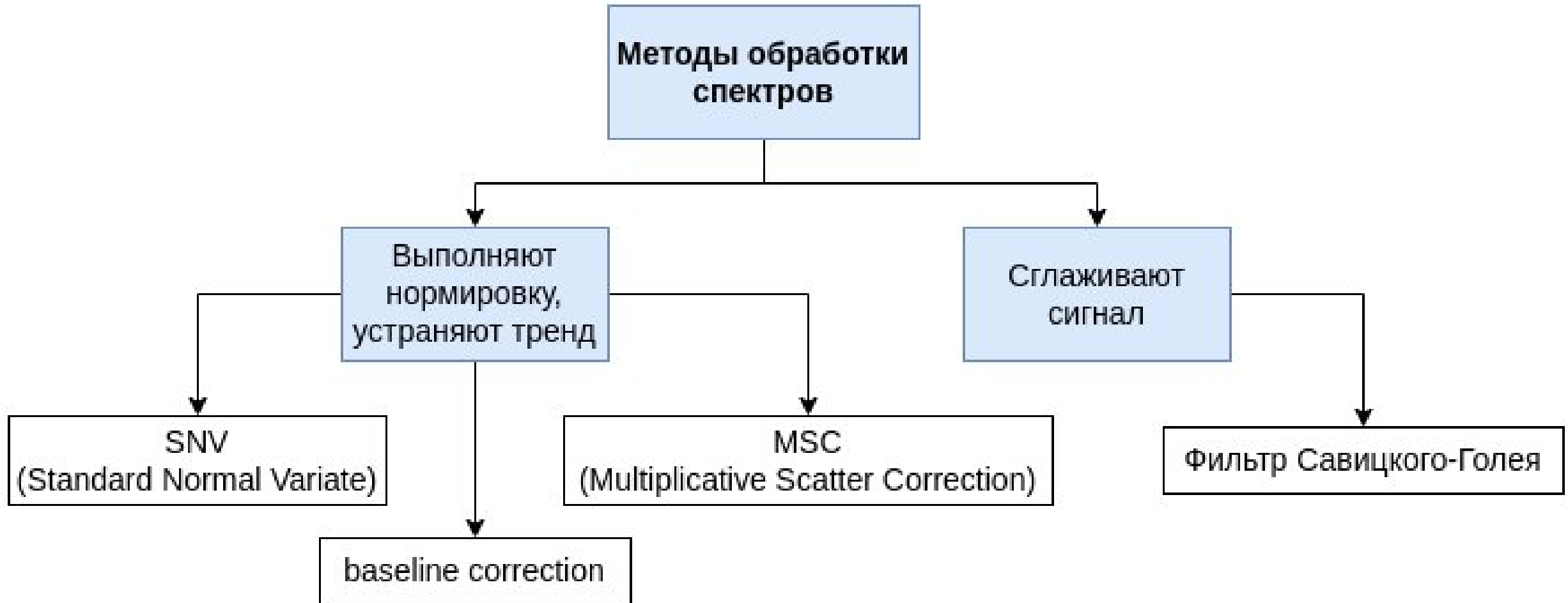
	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	ID #	FILE	[IMAGE]	ORIENT	ROW	COLUMN	[HSI#]	SORT	WEIGHT [MG]	PROTEIN [%]	HARDNESS [N]	H/W [N/MG]	VIR TYPE	VIR I
2	1	A1E10B1T.mat	A1toE10Box1Toporentation	1	1	5	46	OASIS	57,28	8,95	103,2	1,802	C	
3	2	A1E10B1T.mat	A1toE10Box1Toporentation	1	2	5	47	Altzh	48,45	12,05	108,5	2,239	B	1
4	3	A1E10B1T.mat	A1toE10Box1Toporentation	1	3	5	48	Alt105	42,84	10,35	86,6	2,021	B	
5	4	A1E10B1T.mat	A1toE10Box1Toporentation	1	4	5	49	Sibris	32,13	11,04	79,8	2,484	B	1
6	5	A1E10B1T.mat	A1toE10Box1Toporentation	1	5	5	50	Altstep	35,91	9,26	132,4	3,687		
7	6	A1E10B1T.mat	A1toE10Box1Toporentation	1	6	5	41	Pam	38,89	8,87	114,7	2,949	B	6
8	7	A1E10B1T.mat	A1toE10Box1Toporentation	1	7	5	42	Alt530	43,71	9,90	96,5	2,208	C	
9	8	A1E10B1T.mat	A1toE10Box1Toporentation	1	8	5	43	Pam	38,32	9,17	98,1	2,560	C	
10	9	A1E10B1T.mat	A1toE10Box1Toporentation	1	9	5	44	OASIS	56,2	6,28	106,9	1,902	C	
11	10	A1E10B1T.mat	A1toE10Box1Toporentation	1	10	5	45	Pam	55,03	9,40	137,6	2,500	C	
12	11	A1E10B1T.mat	A1toE10Box1Toporentation	1	1	4	34	Alt530	32,19	8,27	115,2	3,579	B	2
13	12	A1E10B1T.mat	A1toE10Box1Toporentation	1	2	4	35	Alt105	43,69	10,47	124,6	2,852	B	7
14	13	A1E10B1T.mat	A1toE10Box1Toporentation	1	3	4	36	Alt325	37,34	8,28	123,5	3,307	C	
15	14	A1E10B1T.mat	A1toE10Box1Toporentation	1	4	4	37	Altzh	54,53	9,61	71	1,302	B	3
16	15	A1E10B1T.mat	A1toE10Box1Toporentation	1	5	4	38	Alt325	50,89	10,37	130,3	2,560	B	
17	16	A1E10B1T.mat	A1toE10Box1Toporentation	1	6	4	39	Alt530	41,97	10,52	117,8	2,807	C	
18	17	A1E10B1T.mat	A1toE10Box1Toporentation	1	7	4	40	Altstep	34,91	9,50	89,7	2,569	A	8
19	18	A1E10B1T.mat	A1toE10Box1Toporentation	1	8	4	31	Alt75	63,54	9,92	135	2,125	B	
20	19	A1E10B1T.mat	A1toE10Box1Toporentation	1	9	4	32	Pam	47,28	8,23	103,8	2,195	C	
21	20	A1E10B1T.mat	A1toE10Box1Toporentation	1	10	4	33	Altzh	44,23	8,36	112,1	2,534	B	
22	21	A1E10B1T.mat	A1toE10Box1Toporentation	1	1	3	30	Alt530	37,91	7,72	147,5	3,891		
23	22	A1E10B1T.mat	A1toE10Box1Toporentation	1	2	3	27	Alt105	48,4	9,80	122,5	2,531	C	1
24	23	A1E10B1T.mat	A1toE10Box1Toporentation	1	3	3	28	Sibris	45,33	11,30	160,5	3,541		

# План решения задачи

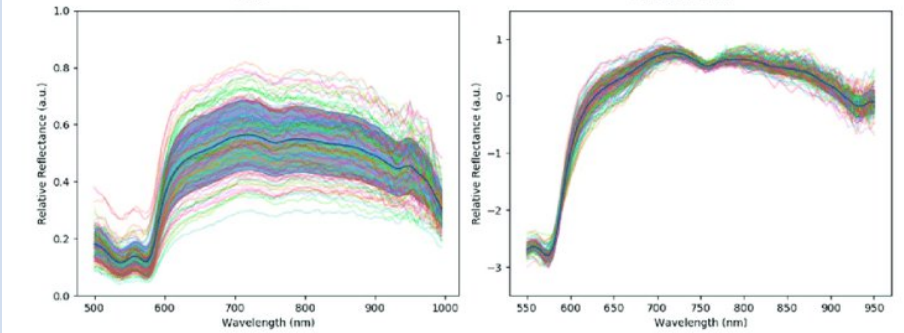
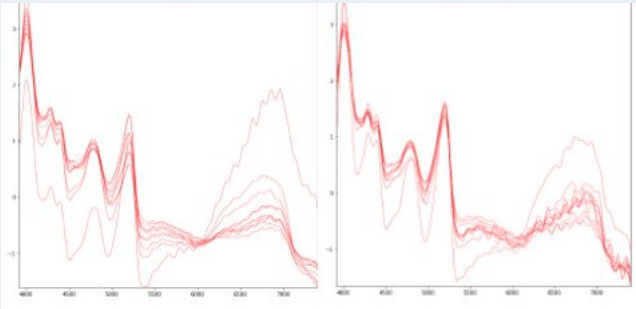
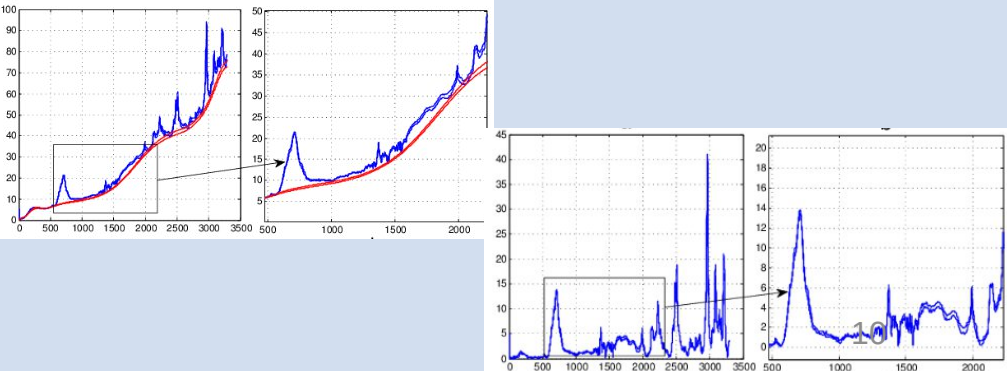




# Методы обработки спектров

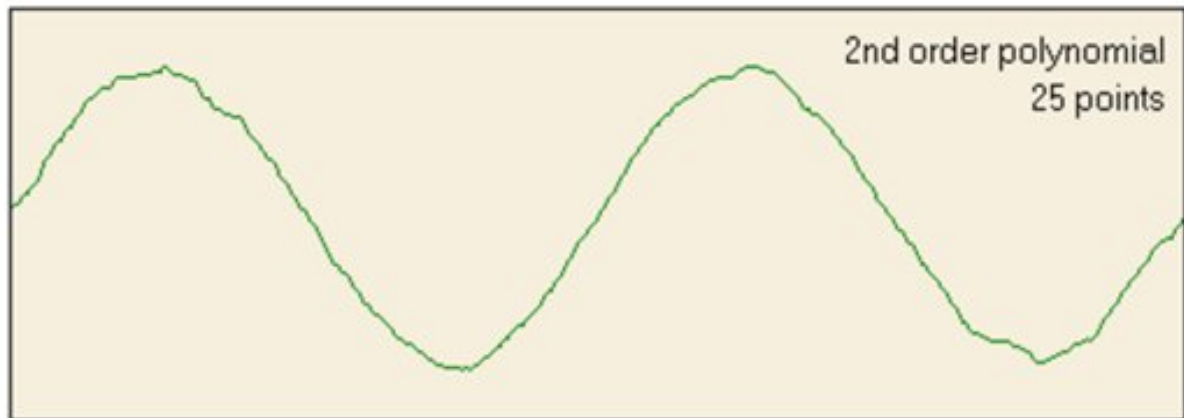
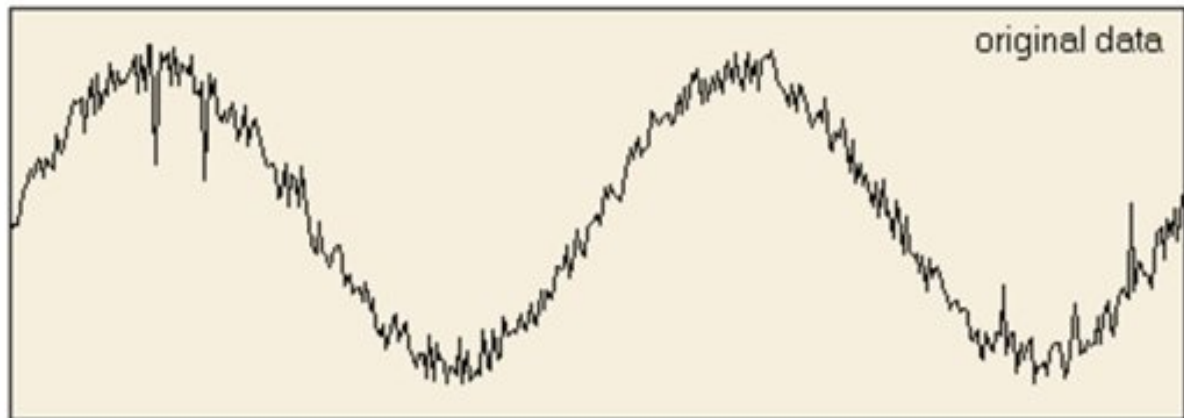


# Методы обработки спектров-1

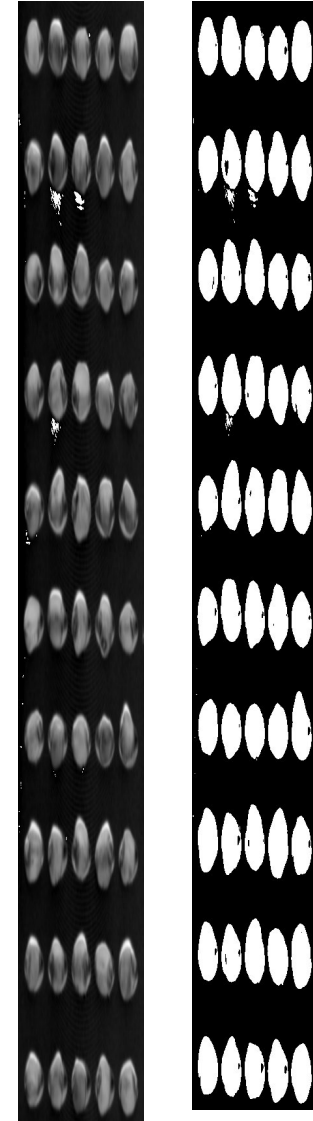
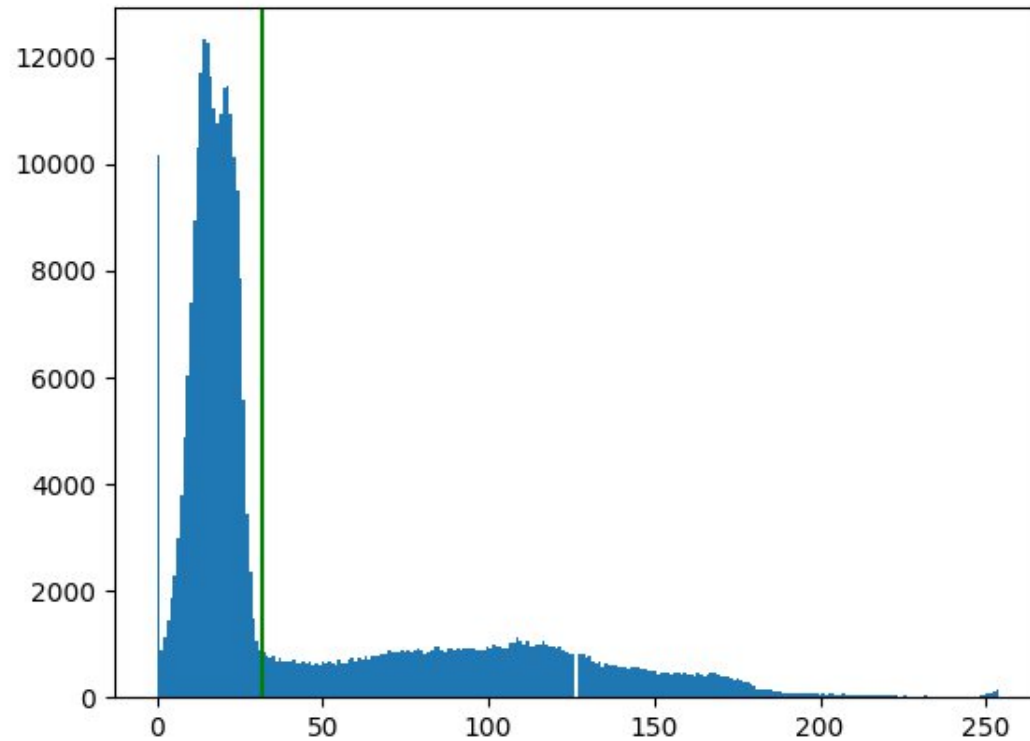
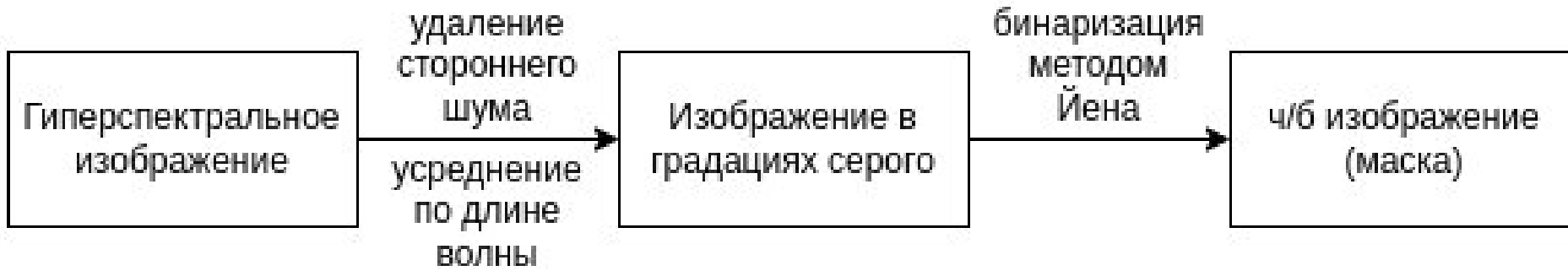
	Зачем используется?	Как работает?	До/После
SNV	Для игнорирования физических особенностей угла отражения спектра	Приводит спектры в единую систему координат (происходит нормировка на разброс)	
MSC	Для устранения переизлучения в спектре – отклика соседних спектров	Убираются тренды	
baseline correction	Для устранения смещения базовой линии, которое могло возникнуть из-за различных факторов	Удаление линейного тренда, смещение базовой линии к нулевому значению	

# Сглаживание спектров с помощью фильтра Савицкого-Голея

- Используется для устранения измерительного шума
- Локальная аппроксимация участка спектра полиномом



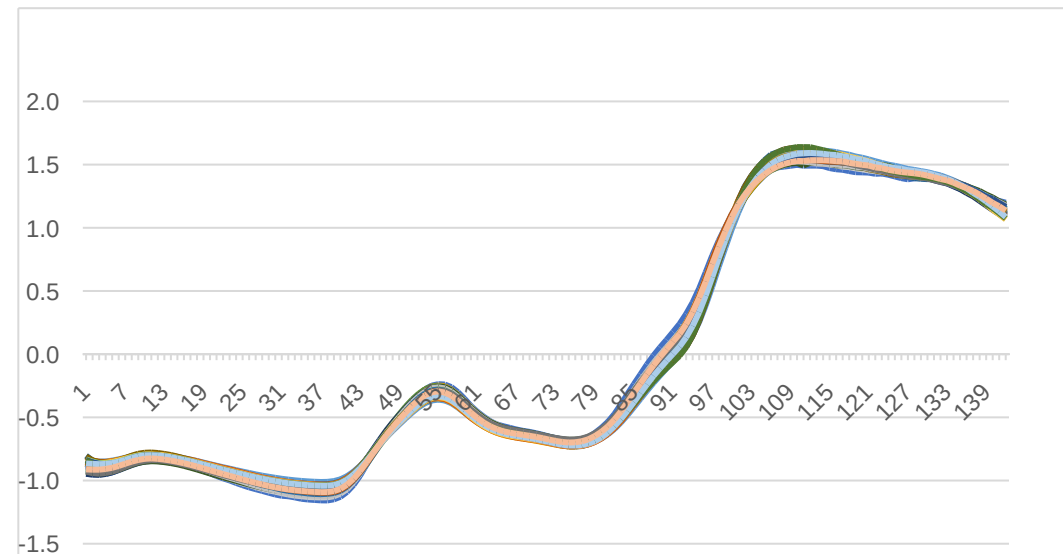
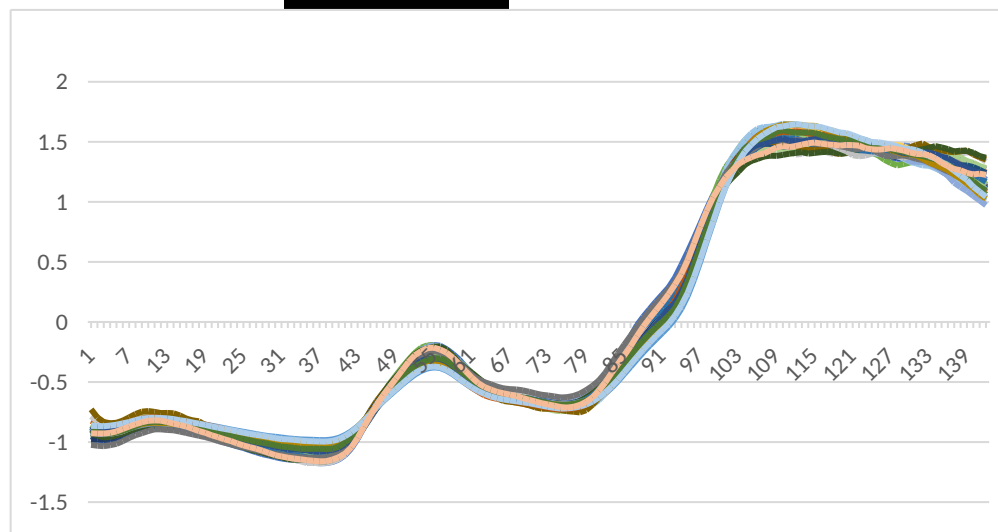
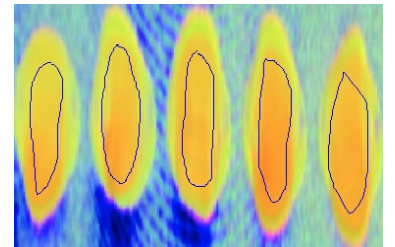
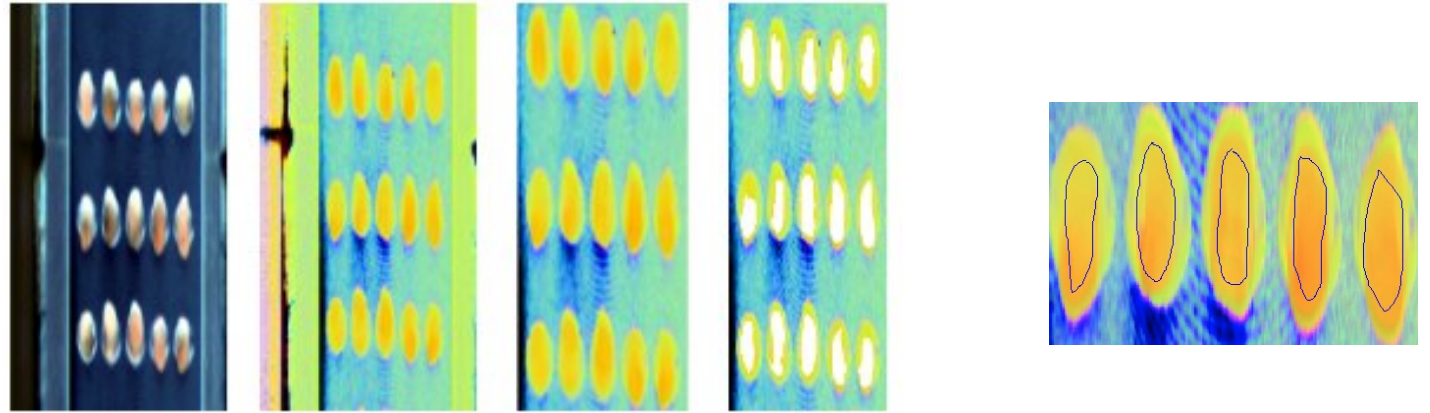
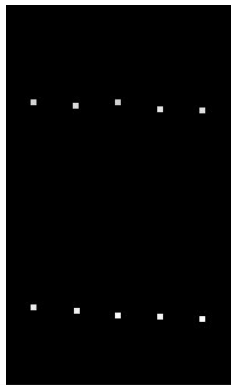
# Получение маски объектов



# Проблема выборки ключевых спектров

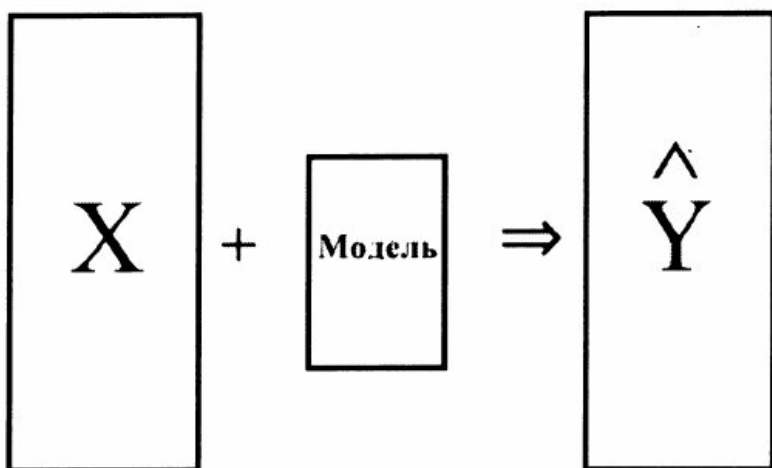
**Способ 1** – извлекать квадрат из центра зерновки со стороной 5

**Способ 2** – использовать специальную маску. Выполним нормализацию Root Mean Square (RMS) и выделим оранжевые пикселы



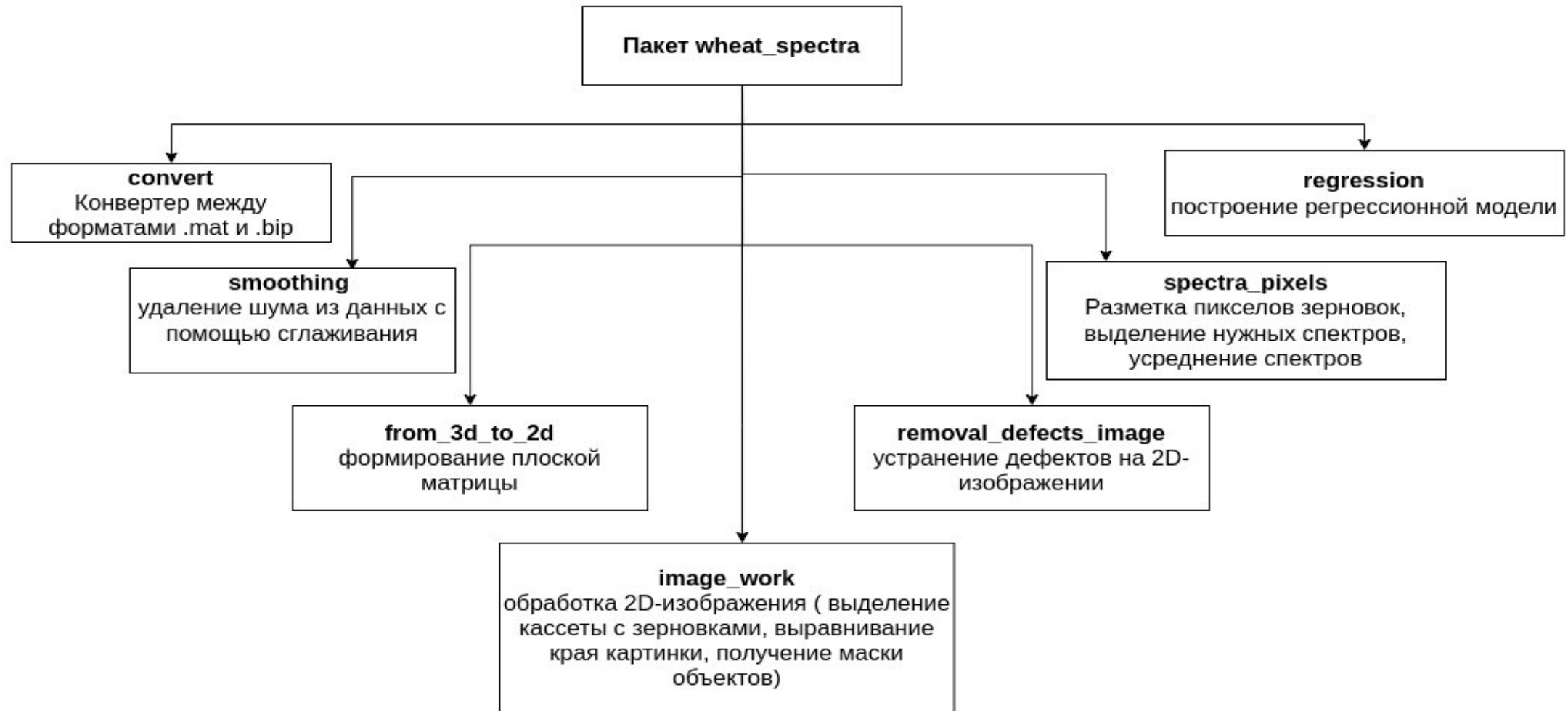
# Построение регрессионной модели

- Для этого применяется PLS-регрессия (Partial Least Squares) - традиционный метод для решения данной задачи
- Использовались такие метрики качества модели, как:
  - $R^2$  – коэффициент детерминации, чем ближе к 1, тем выше качество модели
  - RMSE – среднеквадратичная ошибка (относительная метрика)



```
R^2 для обучающей выборки
0.7732470736225385
R^2 для тестовой выборки
0.532272197201702
```

# Структура пакета wheat\_spectra



Зависимости пакета: scikit-image, OpenCV, PIL (python image library), scipy

# Заключение

Был разработан пакет функций `wheat_spectra` на Python. В настоящее время проводятся эксперименты по построению регрессионной модели с помощью PLS - регрессии

Работоспособность пакета подтверждена справкой о внедрении с предприятия CSort

В рамках данной работы была написана научная статья, принятая к публикации в электронном сборнике материалов II Всероссийской научно-технической конференции «Современные цифровые технологии» 2023

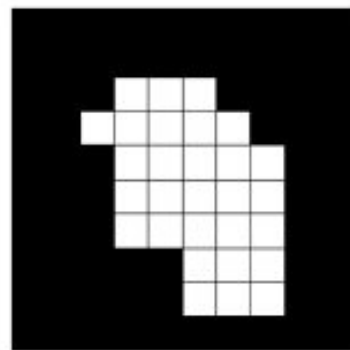


Дополнительные слайды

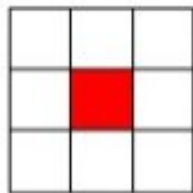
# Устранение структурных дефектов на маске объектов

- Открытие – эрозия + дилатация  
Результат: удаляет мелкие предметы
- Закрытие – дилатация + эрозия  
Результат: удаляет небольшие пустоты

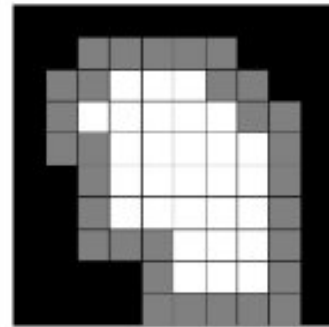
Используется комбинация «открытие + закрытие»



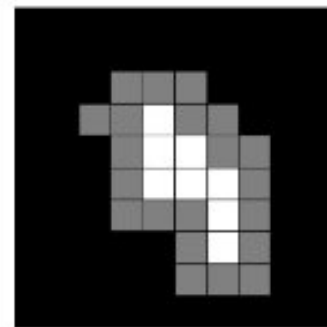
Исходное  
изображение



Шаблон

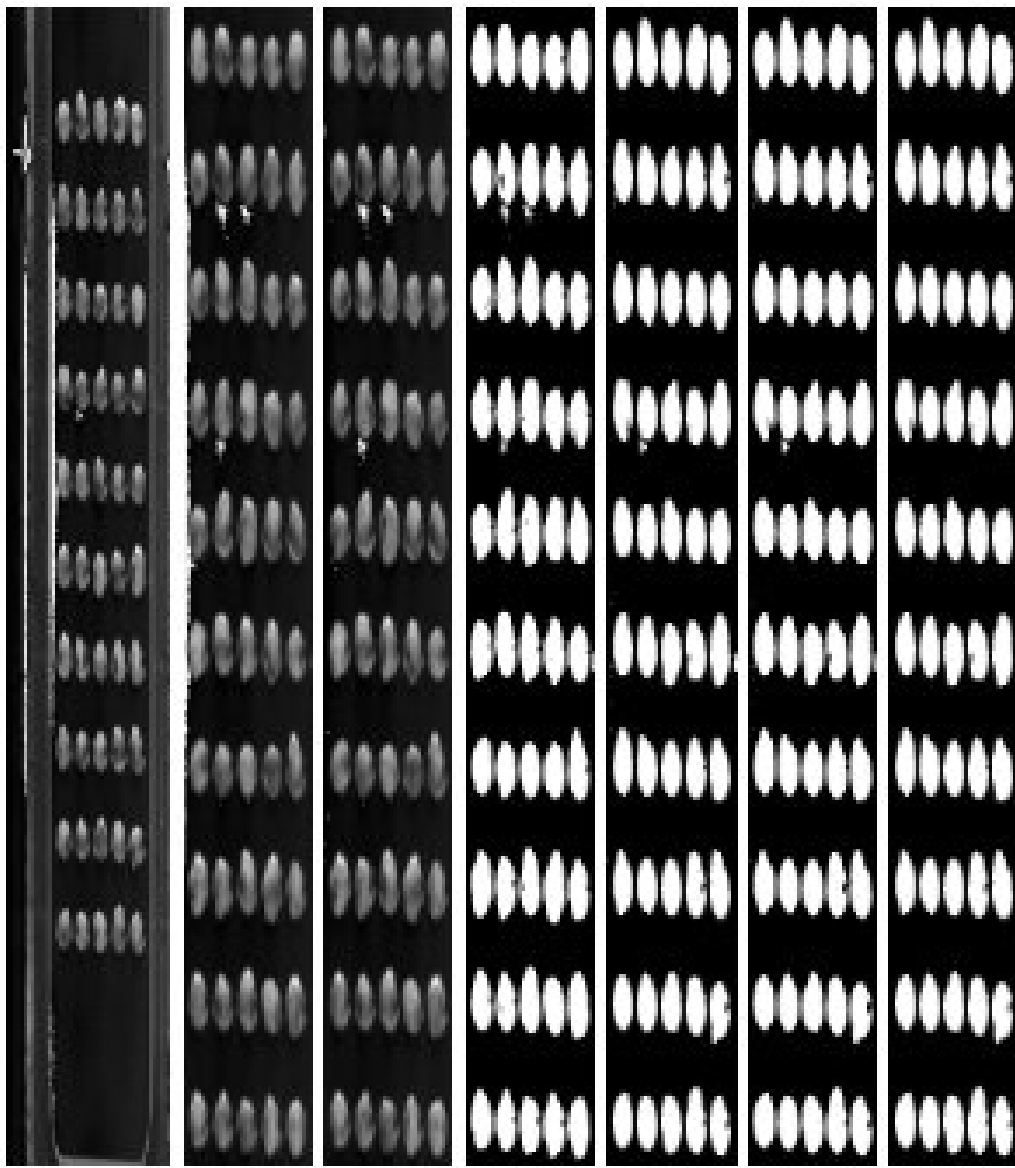


Дилатация



Эрозия

# Преобразование изображения



**Слева направо :**

- изображение в градациях серого, полученное путем преобразования гиперспектрального изображения,
- кадрированное изображение в градациях серого
- изображение с покрашенными краями
- маска объектов по методу Йена
- изображение после выполнения операции «открытие-заккрытие»
- изображение после заливки пустот
- изображение после удаления небольших объектов

# Существующие программные решения

Существующие программные решения не удовлетворяют всем требованиям-либо они требуют достаточно много измерений, либо основаны на более грубых моделях – качественных.

Примеры таких решений:

- Инфракрасный анализатор «ИНФРАЛЮМ ФТ-10» с программой «СпектралЮМ/ПРО»
- Сортировщик зерна VoMill TriQ

# Инфракрасный анализатор «Инфралюм ФТ-10»

Предназначен для экспресс-анализа состава и качества продовольствия, кормов и сырья за минуты без реактивов

## Недостатки данного решения:

- требует для одного измерения много зерновок и выдаёт усредненное значение для всех;
- не позволяет пространственно привязать спектр к отдельным зерновкам
- сам основан на некоторой регрессионной модели, построенной по химическим методам измерения белка



# Сортировщик зерна VoMill TriQ

Промышленная машина, с помощью которой можно исследовать каждое зерно, чтобы определить его относительное качество

Однако данный аппарат основан на классической спектрометрии - для одной зерновки мы можем получить только один спектр, а не целую картинку из спектров. Последний подход потенциально богаче из-за более высокого разрешения.

