

DOI 10.25741/2413-287X-2019-02-2-044

УДК 548.734:633.1

МАТЕМАТИЧЕСКИЙ ПОДХОД К АВТОМАТИЗАЦИИ РЕНТГЕН-АНАЛИЗА ЗЕРНА И СЕМЯН

А. ЖЕЛУДКОВ, канд. физ.-мат. наук, ГНУ Агрофизический НИИ**Н. ПОТРАХОВ**, д-р тех. наук, Санкт-Петербургский государственный электротехнический университет**С. БЕЛЕЦКИЙ**, канд. тех. наук, ФГУП НИИПХ Росрезерва

E-mail: grain-miller@yandex.ru

Рентгенографический метод часто является незаменимым при экспресс-анализе целого ряда показателей семян и продовольственного зерна, выражающихся в изменении внутренней структуры зерновки.

Ключевые слова: продовольственное зерно, семена, интроскопические показатели, рентгенография, аппаратура, съемка, снимок, искажения, компенсация, анализ, выделение объектов, трещиноватость, крупные дефекты, зараженность.

Рентгенографический метод анализа зерна и семян уже в достаточной мере показал свою эффективность. Он основывается на использовании оборудования для рентгеновской съемки, которое, будучи надежным, компактным и безопасным, позволяет получать четкие рентгеновские снимки зерновок, по которым могут распознаваться и интерпретироваться их внутренние дефекты (рис. 1).

Применение рентгенографического метода для анализа качества семян по показателям скрытой зараженности насекомыми отражено в отечественных и международных стандартах. Помимо скрытой зараженности данный метод эффективно выявляет и другие дефекты, а также позволяет определять ряд показателей качества зерновки. Широкое его внедрение затрудняется прежде всего отсутствием адекватных стандартизованных методов автоматического

X-ray method is often indispensable for the rapid analysis of a broad number of indicators of seeds and food grains, expressed in changes in the internal structure of the bruchid weevil.

Keywords: food grains, seeds, introspective indicators, X-ray, equipment, shooting, snapshot, distortion, adjustments, analysis, object selection, fracturing, major defects, contamination.

анализа рентгенообразов зерна и семян. Однако в процессе автоматизации рентгенографического метода и численной оценки рентгенографических показателей становится очевидным, что для точного и хорошо обоснованного его применения нужно решить целый ряд теоретических и практических задач обработки и анализа рентгенообразов зерновок, а также интерпретации результатов.

ПОДГОТОВКА ИЗОБРАЖЕНИЯ

Необходимыми этапами предварительной цифровой обработки изображения для получения адекватного цифрового образа зерновки являются: масштабирование динамического диапазона яркостей изображения; подавление шума видеосигнала и общее улучшение качества изображения; компенсация искажений, обусловленных погрешностями и конструктивными особенностями аппаратуры съемки.

Основная цель масштабирования динамического диапазона — его унификация, необходимая для надежного анализа снимка и адекватного последующего вычисления его количественных характеристик. Такое масштабирование целесообразно применять путем приведения характеристик гистограммы изображения (динамического диапазона, яркости фона, распределения яркостей объектов) к заранее заданным эталонным значениям.

Подавление шума и общее улучшение качества изображения необходимо также для его унификации, в частности для локальных градиентов поля яркости. Требуемое улуч-

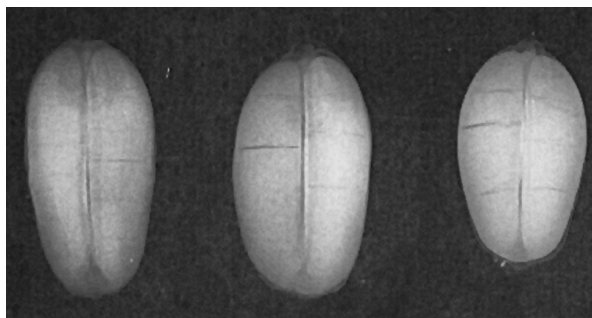


Рис. 1. Внутренняя структура зерновок на рентгенограмме

шенное качество изображения может быть достигнуто, например, сглаживанием изображения низкочастотным фильтром Гаусса с последующим увеличением резкости путем добавления результата перемножения лапласиана и сглаженного градиента яркостей изображения.

Для типичного аппаратного комплекса рентгенсъемки искажения, обусловленные погрешностями и конструктивными особенностями аппаратуры съемки, можно разделить на искажения значений яркости и геометрические искажения образа.

Искажения значений яркости могут быть обусловлены как неравномерностью светового потока для различных сеансов съемки и для различных участков одной рентгенограммы, так и неравномерностью подложки, на которой расположены зерновки.

По закону Бугера-Ламберта-Бэра интенсивность излучения, прошедшего через объект, выражается формулой:

$$J = J_0 \chi \cdot e^{-\mu \cdot h},$$

где J_0 — интенсивность излучения на поверхности объекта;
 χ — фактор накопления;
 μ — оптическая плотность объекта;
 h — толщина объекта.

Если подложка, на которой располагаются анализируемые объекты, достаточно однородна, то для компенсации искажений несложно оценить изменения интенсивности J_0 для различных участков фона, считая постоянными значения μ , h , χ , и, интерполируя значения интенсивности для областей зерновок, скорректировать соответствующим образом значения яркостей для всех точек снимка. Если же неравномерной является толщина и ее оптическая плотность, то, интерполируя аналогичным образом значение $\mu \cdot h$, яркость снимка компенсируется аналогичным образом.

Еще один вид оптических искажений — наличие на снимке яркостных артефактов износа и прочих дефектов чувствительных пластин (рис. 2). В процессе решения дан-

ной задачи получен метод программного удаления артефактов, включающий их обнаружение, основанное на высокочастотной цифровой фильтрации образа и последующей билинейной интерполяции внутренних областей обнаруженных яркостных артефактов для их удаления со снимка.

Вызванные геометрией съемки искажения могут повлечь за собой искажения размеров объектов и характеристик их внут-

ренних деталей. В случае зерновок это могут быть относительные размеры зародыша, интенсивность механических повреждений и т.д. В качестве решения этой проблемы целесообразно реализовать программно-геометрические преобразования изображения, компенсирующие искажения размеров объектов в зависимости от их расстояния от источника излучения.

ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ

Задавая пороговое значение яркости объекта и области объектов, где яркость превышает пороговое значение, можно представить по строкам изображения кодами длин серий с заданием между ними отношения смежности по вертикали. Таким образом, данные коды длин серий можно рассматривать как вершины неориентированного графа, а задачу выделения отдельных объектов на снимке целесообразно решать как задачу выделения его компоненты связности. Для численного решения этой задачи был разработан и программно реализован достаточно эффективный и быстрый численный метод. Из-за упомянутой пространственной неравномерности потока излучения при съемке данную процедуру необходимо проводить итеративно с уточнением пороговых значений яркости объектов и параметров светового потока на каждом шаге.

После того как объекты выделены, необходимо определить пространственную ориентацию объекта. Для этого есть несколько способов, в частности, для зерновок: в силу их формы, идеально подходит способ, основанный на построении для каждой из них эллипсоида инерции. Для дальнейшего анализа зерновку можно развернуть по главным осям эллипсоида инерции. Для зерновок пшеницы на основе их контурных характеристик показана также возможность выделения зоны зародыша.

ДЕФЕКТЫ ЗЕРНОВКИ

Скрытая травмированность (трещиноватость)

Скрытая травмированность (трещиноватость) зерновки, обнаруживаемая рентгенографическим методом, является важным показателем, характеризующим как качество зерновки, так и влияние техногенных факторов, к которым относятся воздействия механического и термического характера при уборке, вымолачивании, сушке и хранении, а также устойчивость зерна к действию данных факторов.

Для качественной и количественной оценки трещиноватости может быть предложен следующий алгоритм анализа изображения зерновки. Здесь и далее за единицу измерения яркости изображения принимается шаг дискретизации яркостной компоненты сигнала изображения в диапазоне от 0 до 255, задаваемый аппаратурой съемки и оцифровки. Дефектные области, соответствующие области трещины на снимке, идентифицируются как узкие горизонтальные темные полосы, расположенные на изображении зерновки. Для более надежного выделения

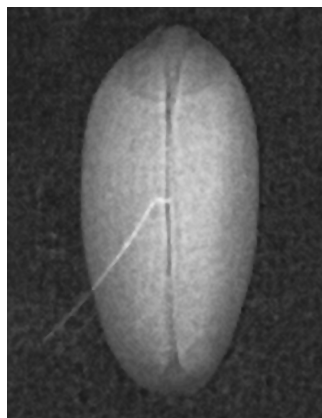


Рис. 2. Артефакты износа чувствительных пластин («царапина»)

границ области трещины на снимке изображение может быть подвергнуто предварительной фильтрации. Для идентификации областей трещин каждой точке изображения зерновки соответствует значение падения яркости, равное разности яркости этой точки и точки, расположенной непосредственно под ней, и значение увеличения яркости, равное разности яркости этой точки и точки, расположенной непосредственно над ней.

Точка считается потенциально принадлежащей верхней границе трещины, если значение падения яркости превышает некоторое пороговое значение. Точка считается потенциально принадлежащей нижней границе трещины, если значение увеличения яркости превышает некоторое пороговое значение.

Точка изображения считается потенциально принадлежащей области трещины при условиях: если выше нее находится точка, потенциально принадлежащая верхней границе трещины, а ниже нее — точка, потенциально принадлежащая нижней границе трещины, и расстояние между этими двумя точками не превосходит некоторого порогового значения. Наконец, множество точек, фактически принадлежащих области трещины, выделяется из построенного выше множества точек, потенциально принадлежащих области трещины, как множество точек, формирующее узкие горизонтальные кластеры на изображении зерновки.

Суть метода количественной оценки трещиноватости зерновки при выделенных согласно приведенному выше алгоритму точках, принадлежащих области трещины, состоит в следующем. На первом этапе для всех точек дефектной области реконструируют значения яркости, которые имели бы эти точки, если бы дефект отсутствовал. Эта реконструкция значений яркости точек дефектной области (области трещины) производится путем интерполяции по значениям яркости граничных точек области. На втором этапе интенсивность дефекта определяется как функция суммы отклонений реальных значений яркости точек от реконструированных на первом этапе.

При этом суммарное отклонение яркости будет:

$$D(A) = \sum_{(x_0, y_0) \in A} |P(x_0, y_0) - I(x_0, y_0)|,$$

где A — множество точек, принадлежащих области трещины;

$I(x_0, y_0)$ — реконструированное значение яркости точки с координатами (x_0, y_0) (в пикселях);

$P(x_0, y_0)$ — яркость точки с координатами (x_0, y_0) .

Каждой точке (x_0, y_0) из множества A соответствует две точки (x_0, y_1) и (x_0, y_2) , где координаты y_1, y_2 , вычисляются по формулам:

$$y_1 = \arg \min_{(x_0, y) \in A, y < y_0} (|y - y_0|);$$

$$y_2 = \arg \min_{(x_0, y) \in A, y > y_0} (|y - y_0|).$$

Реконструированное значение яркости точки, принадлежащей области трещины:

$$I(x_0, y_0) = \frac{(y_2 - y_0)P(x_0, y_1) + (y_0 - y_1)P(x_0, y_2)}{(y_2 - y_1)}.$$

Окончательная количественная характеристика трещиноватости:

$$T(A) = \frac{100 \cdot D(A)}{S(A)},$$

где $S(A)$ — площадь области A (в пикселях).

На рисунке 3 изображены зерновки с различными значениями показателя T . Вычисленные автоматически значения уровня скрытой травмированности хорошо согласуются и с визуальной оценкой трещиноватости зерновок.

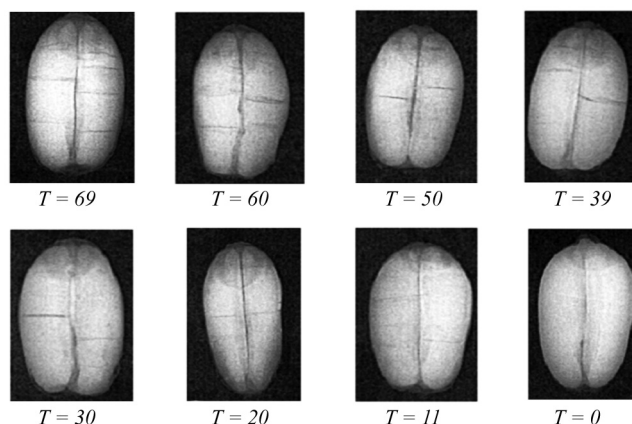


Рис. 3. Зерновки с различными степенями скрытой травмированности

Выявление и оценка крупных дефектов

Рентгенографический метод позволяет эффективно выявлять, в частности, следующие внутренние дефекты зерновки: энзимомикозное истощение эндосперма; скрытая зараженность насекомыми (наличие личинок внутри зерновки); поражение сосущими насекомыми (в частности, клопом вредная черепашка); дефекты зародыша; щуплость зерновки.

Очевидно, что щуплые зерновки могут быть идентифицированы несложным анализом геометрических характеристик контура и гистограммы распределения яркостей образа зерновки. Для идентификации и анализа других крупных дефектов требуются более сложные методы, включающие в себя этапы локализации и идентификации дефекта.

В общем виде локализацию крупных дефектов зерновки можно проводить следующим образом. По контуру зерновки и значениям яркости в опорных точках ее внутренности проводится реконструкция образа зерновки без дефектов. Затем потенциально дефектные области отделяются как участки наибольшего отклонения от реконструированного

нормального образа. Такой подход позволяет не только локализовать дефектные области, но и измерить их количественную характеристику как суммарное отклонение яркости от построенного эталона, масштабированное по суммарной яркости зерновки.

После того, как установлено наличие в зерновке дефектной области и эта область локализована описанным выше методом, можно проводить идентификацию дефекта. Такая идентификация сводится к классификации в пространстве признаков, вычисляемых для дефектной области. Такими признаками являются: показатели яркостной резкости контура дефекта; величина падения яркости и ее дисперсия в области дефекта; коэффициент формы дефекта, а также ряд нетривиальных характеристик, относящихся к геометрии аномальной области и ее локализации внутри объекта. Согласно численным экспериментам данная задача может быть достаточно эффективно решена применением современных методов параметрической классификации.

На рисунке 4 приведена программная визуализация результатов описанных методов локализации дефектных областей и формальные признаки основных обнаруживаемых дефектов. Аномальное затемнение в зоне зародыша зерновки указывает на дефекты зародыша (рис. 4а). Характерные затемнения, имеющие овальную форму (в случае камеры насекомого) или гладкую внутреннюю структуру и резкие границы (в случае ходов насекомых), резкие края затемнения, обширные участки гладкости функции яркости в зоне дефекта — на наличие насекомых и их ходов (рис. 4б). Немного размытые, интенсивные затемнения, имеющие характерную зернистую внутреннюю структуру, — на поражение клопом вредная черепашка (рис. 4в). Локализация дефекта преимущественно на краях зерновки, характерная форма и распределение функции яркости в зоне дефекта — на энзимомикозное истощение эндосперма (рис. 4г).

Как и для оценки трещиноватости, в случае крупных локализуемых дефектов за количественную оценку дефекта можно принять отношение суммарного отклонения яркости в области дефекта к площади зерновки.

ОЦЕНКА ПАРТИИ ПО ИССЛЕДОВАННОЙ ПРОБЕ

Количественные показатели

Получив значения интересующих показателей для каждой зерновки в отдельности, нетрудно вычислить соответствующие общие показатели для всей проанализированной выборки: средний размер зерновки, среднюю количественную характеристику скрытой травмированности, содержание зерновок с тем или иным дефектом. Однако одной из конечных целей автоматического анализа пробы должна являться оценка характеристик не только для данной пробы, но и для всей партии в целом. Естественно, эту задачу целесообразно решать в предположении, что методика отбора проб обеспечивает достаточную репрезентативность выборки. Также необходимо определять достоверность такой оценки для

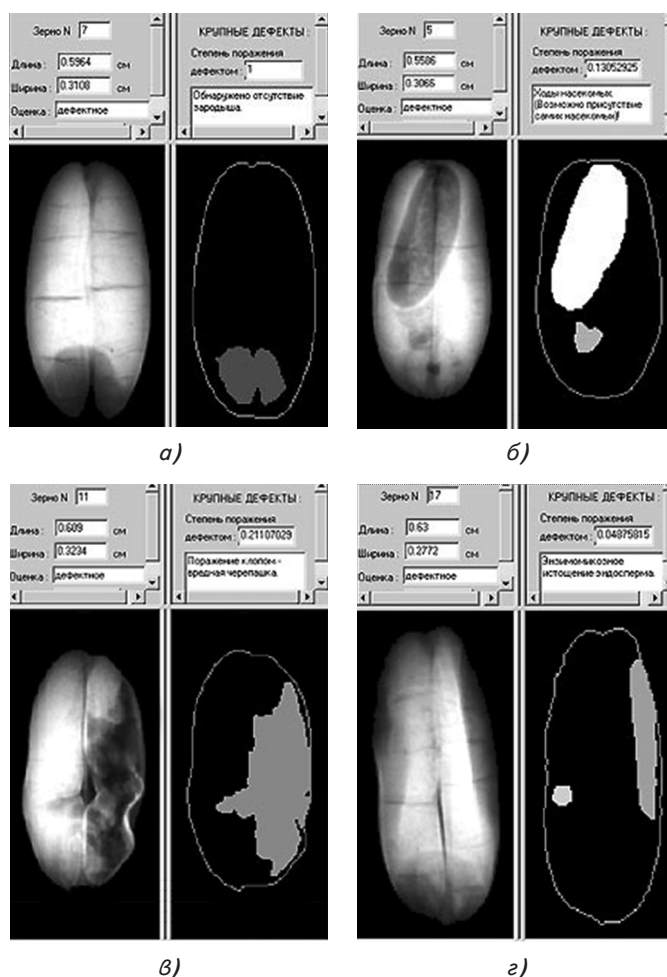


Рис. 4. Программная визуализация:

а — отсутствие зародыша; б — поврежденность насекомыми; в — поражение клопом вредная черепашка; г — энзимомикозное истощение эндосперма

анализируемой партии и достаточность данных, накопленных в процессе анализа пробы.

Для показателей, оцениваемых количественно для каждой зерновки (размеры, степень скрытой травмированности и т.д.) эту задачу можно решить, вычисляя доверительные интервалы для параметров и считая оценку достоверной, если размеры доверительных интервалов, вычисляемых с заданной надежностью оценки, не превосходят заданных значений.

На рисунке 5 приведен пример анализа доверительных интервалов для показателя трещиноватости при последовательном анализе 10 рентгенограмм, по сто зерновок в каждой. После каждой проанализированной рентгенограммы вычислялся с надежностью оценки 0,95 размер доверительного интервала среднего значения показателя трещиноватости для рентгенограммы (верхний график) и для всего проанализированного массива данных, состоящего из текущей рентгенограммы и всех предыдущих (нижний график). Эмпирически установлено, что для среднего значения показателя трещиноватости критиче-

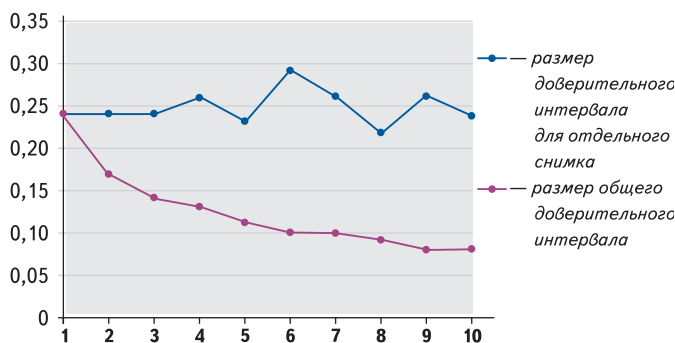


Рис. 5. Размеры суммарного и локального доверительных интервалов для показателя трещиноватости при последовательном анализе 10 рентгенограмм

ским значением размера 0,95 доверительного интервала можно принять величину 0,1. Как видно из графиков, для обеспечения требуемой достоверности оценки скрытой травмированности партии в этом случае достаточно проанализировать 7–8 рентгенограмм (700–800 зерновок).

Для оценки достоверности выборки также целесообразно контролировать однородность проанализированных рентгенограмм и считать результаты анализа достаточно достоверными, если данные проанализированных рентгенограмм удовлетворяют на заданном уровне значимости известным статистическим критериям однородности выборок (достаточно адекватные результаты обеспечивает применение критерия однородности χ^2 на уровне значимости 0,05).

Скрытая зараженность вредителями

По действующим в Российской Федерации нормативам характеристикой зараженности зерна вредителями является суммарная плотность зараженности (СПЗ), измеряемая в экзemplарах (вредителей) на 1 кг зерна. При этом для каждого вида вредителей установлены максимально допустимые уровни (МДУ) зараженности. Так, для насекомых, дающих скрытую зараженность, особенно эффективно выявляемую рентгенографическим методом, показатели МДУ составляют: для рисового долгоносика — 7,5 экз/кг, для амбарного долгоносика — 15 экз/кг. В силу очевидной корреляции между содержанием насекомых в зерне и количеством зерновок с выявленной рентгенографическим методом поврежденностью насекомыми, данный показатель можно рассматривать в той же терминологии (понимая под СПЗ количество поврежденных вредителями зерновок) и с теми же значениями МДУ.

Для оценки зараженности партии по проанализированной пробе выявление зараженных вредителями зерновок будем моделировать схемой независимых испытаний Бернулли. Считая долю зараженных зерновок в партии, равной d_0 (очевидно, она равна отношению СПЗ к количеству зерновок в килограмме), вероятность обнаружения m зараженных зерновок в пробе из N зерновок в этом случае будет равна:

$$P(m, d_0) = C_N^m d_0^m (1 - d_0)^{N-m},$$

а вероятность встретить хотя бы одну зараженную зерновку:

$$P = (1 - (1 - d_0)^N).$$

Используя данные выражения, несложно оценить точность рентгенографического метода выявления зараженности вредителями по количеству зерновок в пробе для различных фактических значений плотности заражения.

Подобный подход также возможен при оценке суммарной плотности заражения партии. Предположим, что из N проанализированных зерновок t оказались зараженными. Не превышает ли СПЗ партии заданный уровень S ? Эту задачу можно рассматривать как задачу проверки статистической гипотезы $СПЗ = S$ против альтернативной $СПЗ > S$. Используя известные критерии оценки успехов в испытаниях Бернулли, несложно получить, что суммарную плотность заражения можно считать не превышающей S на уровне значимости α , если выполняется неравенство:

$$\frac{\frac{t}{N} - \frac{S \cdot M_{1000}}{1000^2}}{\sqrt{\frac{S \cdot M_{1000}}{1000^2} \cdot \left(1 - \frac{S \cdot M_{1000}}{1000^2}\right)}} \cdot \sqrt{N} < \Phi^{-1}\left(\frac{1 - 2 \cdot \alpha}{2}\right),$$

где M_{1000} — масса 1000 семян, г;

$\Phi(x)$ — интегральная функция Лапласа.

Адекватная и эффективная автоматизация анализа зерна (семян) рентгенографическим методом — комплекс задач, для решения которых требуется системный подход и комплексное применение биологических, технических, компьютерных и математических методов. Предложенные методы позволяют эффективно решать сопутствующие задачи автоматического анализа изображений, интерпретировать полученные результаты и применять рентгенографический метод в самых различных задачах оценки качества зерна и семян.

Литература

1. Архипов, М. В. Микрофокусная рентгенография растений / М. В. Архипов, Н. Н. Потрахов. — СПб. : ГНУ АФИ РАСХН, 2008. — 192 с.
2. Белецкий, С. Л. Современные рентгендиагностические комплексы для экспресс-оценки качества сырья / С. Л. Белецкий, Е. В. Иванова, Т. В. Фешина // Практические аспекты исследования и мониторинга качества сырья и продуктов питания для обеспечения продовольственной безопасности России : сб. докл. науч.-практ. семинара. — М. : ФГУ НИИПХ Росрезерва, 2010. — С. 78–83.
3. Анализ качества семян по рентгенограмме с помощью ЭВМ / В. П. Якушев [и др.] // Доклады РАСХН. — 2001. — № 1. — С. 16–17.
4. Методика рентгенографии в земледелии и растениеводстве / М. В. Архипов [и др.]. — М. : РАСХН, 2001. — 102 с. ■