

$$P(\theta_s|W) = \frac{P(W|\theta_s)P(\theta_s)}{P(W)} \quad (1)$$

где  $\theta_s$  – гипотеза о тональности рассматриваемого текста (положительная либо отрицательная).

$W=[w_1, w_2, \dots, w_n]$  – вектор слов, из которых состоит сообщение.

Для каждого слова определяется условная вероятность, того что при его упоминании сообщение будет иметь ту или иную эмоциональную окраску. Затем на основе полученных вероятностей для каждого слова, из которых состоит сообщение, вычисляется вероятность для целого сообщения.

Кроме того, к вычисляемым вероятностям было применено сглаживание Лапласа (аддитивное сглаживание) с целью избежания риска получить нулевые вероятности для слов, которые встречаются впервые, тем самым, не обращая вероятность сообщения в ноль.

В результате осуществлена классификация корпуса сообщений с точностью 76.%.

В рамках задачи отбор и предварительная обработка данных были осуществлены с использованием известной среды RSudio средствами языка R, сам классификатор был написан на языке Python [3].

**Заключение.** В рамках настоящей работы была предпринята попытка построения модели выявления активности, обладающей потенциально негативным воздействием, на основе проведения анализа тональности сообщений на русском языке.

В качестве следующего шага для повышения точности предполагается усложнить механизм отбора исследуемого корпуса сообщений с применением более сложных методов NLP (Natural Language processing). Кроме того, данная задача может быть расширена следующим образом: определение уровня воздействия или, другими словами, уровня покрытия аудитории, что позволит в дальнейшем предпринять попытку построения предсказательной модели о наступлении того или иного события.

#### ЛИТЕРАТУРА

1. Замятин А.В. Введение в интеллектуальный анализ данных: учеб. пособие. – Томск: Издательский Дом государственного университета, 2016. – 120 с.
2. Manning Christopher D., Schütze Hinrich. Foundations of statistical natural language processings. - Cambridge, Massachusetts: The MIT Press, 1999. – 680 с.
3. Grus Joel. Data Science from Scratch. - Sebastopol, CA: O’Reilly Media, 2015. – 311 с.

#### ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ В СЕПАРАЦИИ СЕМЯН ЗЕРНОВЫХ КУЛЬТУР

*А.В. Власов, А.С. Фадеев*  
(г. Томск, Томский политехнический университет)  
e-mail: andark@tpu.ru

#### COMPUTER VISION IN SEPARATION OF CEREAL SEEDS

*A.V. Vlasov, A.S. Fadeev*  
(Tomsk, Tomsk Polytechnic University)

**Abstract.** The most outstanding agricultural problem is a cereal seeds sorting. The sorting reduces the number of impurity seeds, causing the acceleration of growth and amount of the crop. These actions lead to economy prosperity. Today there are a lot of mechanical types of sorting, but they don't provide the required product yield. The paper is concerned with using of the computer vision technology as an additional sorting stage for a better accuracy. It provides a list of sorting criteria and some experimental results. A basic sorting method and some universalized intelligent approaches for unknown cultures are given. The possibility of corrupted or tinted seeds identification is described.

**Key words:** computer technologies, computer vision, color sorting machine, object capture, seeds classification, agriculture

**Введение.** Одной из важных экономических составляющих Российского сельского хозяйства является стабильное производство высококачественного зерна. Основными показателями качества семян является чистота высеваемой культуры. Отсутствие примеси сорных культур влияет на всхожесть и скорость роста, что в следствие сказывается на качестве и объеме урожая. Основной проблемой при подготовке семян к посеву становится получение чистой культуры за счет очистки от примесей.

На сегодняшний день наблюдается широкое применение вычислительных возможностей компьютерной техники для решения и оптимизации многих производственных задач. При этом компьютеризация в сельскохозяйственной отрасли происходит заметно медленнее, чем в других областях человеческой деятельности. Традиционно применяются такие способы очистки, как сепарирующие решетчатые линии, воздушно-решетчатые машины и триеры. Тем не менее, для решения проблемы очистки семян от сорных культур уже возможно применение не только механических методов, но и методов, основанных на использовании компьютерных технологии, таких как компьютерное зрение, используемое в фотосепарации.

**Фотосепарация.** Фотосепарация – процесс разделения обследуемого материала за счет обработки сигналов с камер или фоточувствительных сенсоров. Фотосепаратор представляет собой высокопроизводительное оборудование, предназначенное для извлечения из сыпучих материалов примесей, отличающихся по цвету. Аппарат определяет мельчайшие отклонения в сортируемом продукте и отбраковывает его [1].

Фотосепарация относится к одним из завершающих этапов тонкой очистки. Это связано с тем, что механическая очистка не дает требуемой чистоты культуры сортируемых семян. Механически отсеиваются такие параметры как вес, форма, размер, электропроводимость и аэродинамические свойства. При этом часть сорной культуры обладает точно такими же характеристиками и пропускается всеми предварительными ступенями очистки вместе с основной культурой.

Параметры, наиболее характерные для этапа фотосепарации, это цвет, тип и рельеф поверхности. Цвет наиболее просто и быстро распознается у обследуемого объекта и позволяет сразу отличить, например, темные сорные семена от светлых злаковых культур. Тип поверхности может служить отличительным показателем гляцевых семян и матовых. Рельеф может быть задан как неровным краем, так и впадинами и трещинами в центре объекта, что также может говорить о его принадлежности к искомой культуре.

**Проблемы существующих систем.** На данный момент фотосепараторы имеют ряд недостатков:

- Дороговизна
- Требовательность к условиям сортировки
- Невысокая точность
- Зависимость от наличия шаблонов сортируемой культуры

Практически все разработки являются коммерческими с закрытыми технологическими решениями, а на рынке присутствуют единичные компании представители, что влечет за собой дороговизну приборов. Также многие аппараты требуют теплых помещений и под-

держания влажности на заданном уровне, чтобы проводить процесс сортировки, что напрямую влияет на стоимость технологии.

При использовании приборов для разных сортов семян необходимо вручную точно задавать шаблоны, по которым осуществляется обнаружение сорного составляющего, что не является универсальным способом. При неточном подборе определенного шаблона или его полном отсутствии пропадает возможность сортировать выбранную культуру.

**Применение компьютерного зрения.** В качестве решения задачи о повышении общей точности сортировки культуры семян и устранения недостатков существующих систем предлагается применение компьютерного зрения. На этапе с компьютерной обработкой уже имеются подготовленные семена, так как они уже прошли предварительную механическую очистку, а именно семена одного веса, формы и размера.

На данной стадии требуется получить изображение объекта и обработать его, чтобы определить относится ли он к семенам искомой культуры или нет. Эта задача разделяется на две основные крупные подзадачи:

- Обнаружение и выделение изображения движущегося в видеопотоке объекта
- Классификация статического изображения по ряду признаков

Решение первой задачи возможно с оптимальным подбором скоростной камеры, позволяющей получать изображение движущегося объекта достаточного качества, чтобы переходить ко второму этапу. Изображение выделяется методами сегментации области кадра видеоизображения при наличии в ней движения.

Во втором этапе необходимо классифицировать изображение. Самым первым этапом можно выделить определение цвета. Эксперименты показали, что на изображении (рис. 1) с семенем можно разделить пиксели, относящиеся к фону или к семени, за счет обработки гистограммы зависимости интенсивности пиксела от количества пикселей на изображении [2].

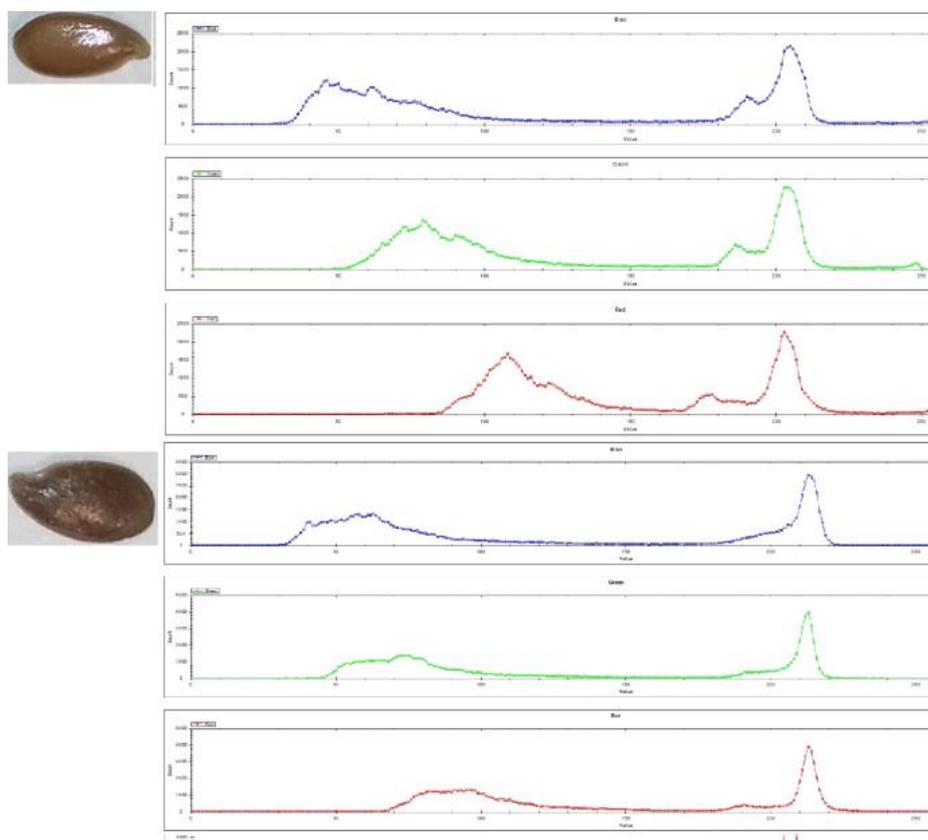


Рисунок 1. Гистограммы интенсивности по каналам RGB пикселей изображений разных тестовых семян

На рисунке видно, что фон выделяется в отдельный пик интенсивности на гистограмме, он является известным и практически неизменным при обработке разных объектов. На втором пике можно наблюдать характеристику интенсивности по каждому из каналов RGB, что позволяет провести соответствие с цветом объекта. У сорных семян форма этой части графика будет сильно отличаться. Помимо цвета можно определить наличие глянцевой поверхности, как у первой семечки, которое выражается явным пиком в области интенсивности близкой к белому фону [3].

Помимо поиска в посевной культуре семян сорных включений существует возможность внести особые критерии сортировки, которые позволят искать некачественные, предположительно невсхожие, больные или пораженные семена искомой культуры. Такой подход позволит не только очистить культуру, но и повысить урожайность [4]. Для этого можно рассмотреть использование нейронной сети, предварительно обученной на распространенных сортах семян [5]. Такой подход позволит в дальнейшем не проводить ручной настройки для каждого типа семян, а также даст возможность расширять диапазон обрабатываемых культур.

**Заключение.** Использование компьютерного зрения позволит выполнять данный этап сортировки быстро и эффективно. В дальнейшем с выделением новых критериев увеличится точность распознавания. Применение интеллектуальной обработки дает возможность изменять только программную часть при этом добиваться оптимизации и улучшения качества сортировки. Предполагается применение обучаемых нейронных сетей, что позволит не вводить шаблонные данные, а только просыпать чистые семена культуры в режиме обучения и не зависеть от известных для программы культур семян и предварительно заданных шаблонов распознавания.

#### ЛИТЕРАТУРА

1. Шаззо А. А., Гюлушанян А. П., Корнена Е. П., Мхитарьянц Л. А. Сравнительная оценка способов сепарирования рушанки подсолнечных семян и способов контроля качества ядра // Новые технологии. – 2011. №3. С.75-79.
2. J. Li, G. Liao, and F. Xiao, "Rapeseed seeds colour recognition by machine vision," *Proc. 27th Chinese Control Conf. CCC*, pp. 146–149, 2008.
3. Шаззо А. Ю., Усатиков С. В. Эффективность распознавания скрытой зараженности зерновок по изображениям в инфракрасном спектре // Известия ВУЗов. Пищевая технология. – 2012. №4. С.105-108.
4. K. Kiratiratanapruk and W. Sinthupinyo, "Color and texture for corn seed classification by machine vision," *2011 Int. Symp. Intell. Signal Process. Commun. Syst. "The Decad. Intell. Green Signal Process. Commun. ISPACS 2011*, pp. 7–11, 2011.
5. R. V Ronge and M. M. Sardeshmukh, "Comparative analysis of Indian wheat seed classification," *Icacci'14*, pp. 937–942, 2014.