

## ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ РАННЕГО РАСПОЗНАВАНИЯ ВИДОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР ПО КОСМИЧЕСКИМ СНИМКАМ

Н.С. Воробьева<sup>1,2</sup>, В.В. Сергеев<sup>1,2</sup>, А.В. Чернов<sup>1,2</sup>

<sup>1</sup> Самарский национальный исследовательский университет имени академика С.П. Королева, Самара, Россия,

<sup>2</sup> Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Самара, Россия

### Аннотация

Статья посвящена разработке информационной технологии раннего распознавания видов сельскохозяйственных культур по набору космических снимков на территориях масштаба региона. Раннее распознавание проводится в первой половине посевного сезона и характеризуется нехваткой наземных и космических данных для настройки алгоритмов распознавания. Предлагаемая в статье технология позволяет генерировать обучающую выборку по данным прошлых лет и использовать её для распознавания культур текущего года. Технология состоит из двух этапов. На первом этапе по данным за прошедшие годы формируются модели временных рядов, сельскохозяйственных культур и агросезонов. На втором этапе производится работа с данными текущего года. Сначала по немногочисленному набору опорных полей с известным видом культуры выбирается подходящая модель агросезона, близкая к развитию культур в текущем году, а затем по характеристикам модели агросезона генерируется обучающая выборка и выполняется распознавание культур текущего сезона. В статье оценивается качество распознавания с использованием предлагаемой информационной технологии и возможность её применения на территории региона. Для экспериментальных исследований используются данные за 2011–2015 годы, полученные на территории Самарской области.

**Ключевые слова:** распознавание сельскохозяйственных культур, космические снимки, вегетационные индексы, временные ряды, NDVI, алгоритм вычисления оценок, модель временного ряда.

**Цитирование:** Воробьева, Н.С. Информационная технология раннего распознавания видов сельскохозяйственных культур по космическим снимкам / Н.С. Воробьева, В.В. Сергеев, А.В. Чернов // Компьютерная оптика. – 2016. – Т. 40, № 6. – С. 929-938. – DOI: 10.18287/2412-6179-2016-40-6-929-938.

### Введение

В настоящее время использование данных дистанционного зондирования (далее – ДДЗ) Земли стало привычным инструментом для решения широкого круга научно-производственных задач [1]. Активное использование космических снимков обусловлено их разнообразием: съемка опико-электронными и радарными системами предоставляет данные с широкими интервалами пространственного и временного разрешений. Сельское хозяйство также является сферой активного развития и внедрения методов, основанных на использовании спутниковых данных – разрабатываются и используются методы оценки и распознавания типов и границ посевов [2], прогнозирования урожайности [3], оценки степени эрозии почвы и множества других важных задач.

Данная статья посвящена распознаванию видов сельскохозяйственных (далее – с/х) культур по космическим снимкам. Актуальность данной задачи не вызывает сомнений. Результаты распознавания применяются для учета и контроля использования сельскохозяйственных земель со стороны региональных органов агропромышленного комплекса (далее – АПК) и сельхозтоваропроизводителей (далее – с/х производителей), контроля севооборота, выявления неиспользуемых земель в течение ряда лет, а также контроля использования субсидий путем выявления расхождений данных мониторинга с декларируемыми с/х производителями данными [4, 5].

При распознавании с/х культур чрезвычайно важно получить результаты как можно раньше с начала посевного сезона – это позволяет быстро принять меры наземного контроля и оперативного управления. Посевной сезон – временной промежуток, соответствующий периоду от начала посевных работ до окончания уборочных работ. Проведение раннего распознавания осложняется малым количеством исходных данных, недостаточных для формирования обучающей выборки, а также невозможностью сформировать по космическим данным признаки, характеризующие вегетационный цикл растений в целом (например, длина цикла, дата максимума).

Исходными данными для распознавания обычно являются сведения, предоставляемые с/х производителями в региональные органы АПК об обрабатываемых полях с привязкой к векторной карте. При этом данные о посеянных на полях с/х культурах появляются только в конце посевного сезона и не могут быть напрямую использованы при раннем распознавании в текущем году. В начале посевного сезона имеется лишь небольшой набор таких данных по опорным полям (несколько десятков полей). Однако в региональных органах АПК имеется набор сведений о засеянных на полях культурах за ряд прошлых лет, так называемые исторические данные, которые разумно использовать при раннем распознавании.

Вопросу решения задачи распознавания видов с/х культур по космическим снимкам посвящено большое количество работ отечественных и зарубежных

авторов [2, 6–10]. Для распознавания культур применяются методы, использующие различное количество космических снимков – один снимок или их разновременный набор. Методы распознавания по одному снимку, в свою очередь, могут использовать космические изображения среднего, высокого или сверхвысокого пространственного разрешения [11, 12]. Такие методы в основном применяются для распознавания видов и оценки состояния с/х культур на территориях отдельных хозяйств (в среднем, от 1 до 50 кв. км). Для решения задачи регионального охвата (от 100 до 1000 тыс. кв. км) в основном используются методы наблюдения за динамикой развития растительности на полях на основе получаемых последовательностей космических снимков среднего и низкого разрешения [9]. В течение посевного сезона с/х культуры проходят различные фазы своего развития, изменяется высота растений, густота покрова, количество хлорофилла в листьях, что, в свою очередь, влияет на значения отраженного от растительного покрова излучения [10, 13]. На рис. 1 приведены типичные графики зависимости значений вегетационного индекса NDVI от относительной даты (рассчитанной от 1 марта), построенные для основных видов с/х культур Самарской области. Каждая кривая получена усреднением по набору полей.

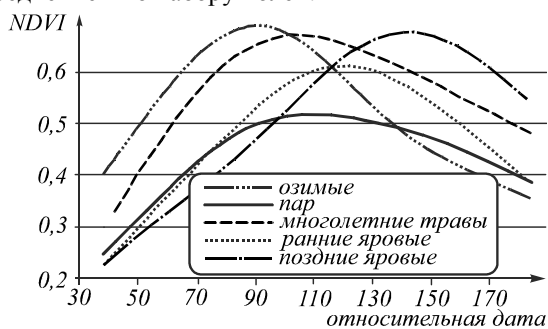


Рис. 1. Усредненные графики NDVI для основных видов с/х культур Самарской области

NDVI (Normalized Difference Vegetation Index) – параметр, наиболее часто используемый для оценки развития растительности на полях, рассчитываемый по значениям в красном и инфракрасном каналах спектра. Как видно из рис. 1, в разные периоды посевного сезона присутствуют и различия, и сходства в значениях NDVI для разных классов, поэтому для распознавания видов с/х культур необходимо наблюдение за развитием растительности в течение некоторого периода времени.

В литературе описывается наблюдение за сезонной динамикой растительности по космическим снимкам среднего [5] и низкого пространственного разрешения [14]. Лучше всего для решения рассматриваемой в работе задачи подходят временные ряды, построенные по данным с космических аппаратов Terra и Aqua (съёмочной аппаратуры MODIS) [8, 9, 15, 16]. Снимки MODIS имеют высокую периодичность (не реже одного раза в день без учета облачности), широкую полосу съёмки и находятся в свобод-

ном открытом доступе [17], что дает возможность проведения регулярных наблюдений за динамикой растительного покрова.

Проведенный авторами анализ существующих методов распознавания показал невозможность прямого использования таких методов для задачи раннего распознавания по причинам нехватки исходных данных и короткого вегетационного цикла. При этом обучающие выборки, сформированные для распознавания с/х культур за прошедшие годы, не могут быть напрямую использованы в процедуре распознавания полей текущего года из-за различий в условиях произрастания растительности.

В статье описывается общий подход, позволяющий выполнять раннее распознавание с использованием последовательности космических снимков, исторических данных и небольшого числа опорных полей текущего сезона. Все эксперименты, проводимые в рамках данной работы, основываются на данных, полученных для территории Самарской области, где с 2009 года происходит активное внедрение геоинформационных технологий в сферу сельского хозяйства [18].

Конкретизируем постановку задачи.

### 1. Постановка задачи

Пусть имеется территория исследования, используемая под растениеводство. Земли, расположенные на территории, используются под посев с/х культур. Для упрощения терминологии будем понимать под с/х культурами в том числе и другие виды использования угодий (пары, залежи, неиспользуемые земли и т.д.). Информационным объектом исследования является поле – однородный участок земли, заданный своими границами и используемый под сев определенной с/х культуры.

Территория исследования характеризуется однородными природно-климатическими условиями по всей своей площади, а значит, внешние условия произрастания с/х культур во всех её частях имеют малые различия.

На территории исследования имеется некоторая система учета полей, осуществляющая сбор и хранение информации и функционирующая в течение ряда лет и в настоящее время. По данным, накопленным за прошедшие годы, сформирована историческая обучающая выборка. В нее включена некоторая часть полей территории в соответствии с определенными критериями отбора.

Требуется провести раннее распознавание классов (видов) с/х культур  $\Omega_l, l=1, L$  в границах полей текущего сезона с использованием алгоритма распознавания, основанного на использовании временных рядов, набору исходных данных для исследования. Состав набора следующий:

1. Исторические данные за каждый год:
  - 1.1. Космические снимки для расчета временных рядов вегетационного индекса.
  - 1.2. Историческая обучающая выборка: векторная карта полей и данные о культурах на полях.

2. Данные за текущий сезон:
  - 2.1. Космические снимки для расчета временных рядов вегетационного индекса.
  - 2.2. Векторная карта полей.
  - 2.3. Данные о культурах на опорных полях.

Раннее распознавание означает, что длина временных рядов, построенных по космическим снимкам для полей, будет ограничена количеством наблюдений, прошедших с начала сезона. Чем раньше производится распознавание, тем короче временные ряды.

Далее рассмотрим информационную технологию для решения поставленной задачи.

### 2. Информационная технология

В информационной технологии используется ряд моделей. Необходимость моделирования, главным образом, связана с зашумленностью временных рядов в связи с влиянием набора случайных факторов, таких как облачность, атмосферные явления, природно-климатические условия. На рис. 2 приведен пример зашумленного временного ряда.

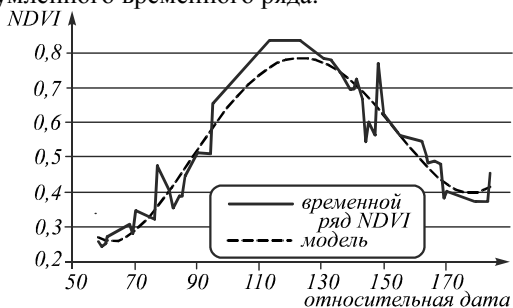


Рис. 2. Временной ряд NDVI и его модель

Временные ряды используются в алгоритме распознавания и в связи с зашумленностью неудобны в работе, поэтому в технологии будут заменены на модели без шумов и пропусков.

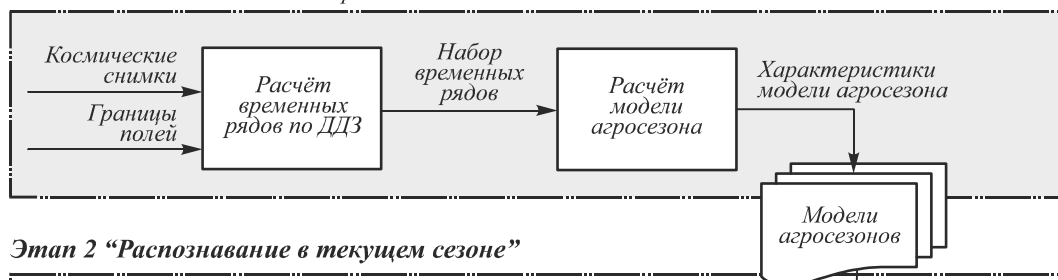
Введем серию моделей:

1. Моделью временного ряда будем называть некоторую функцию  $g_{c_1, \dots, c_m}(x)$ , заданную аналитически, зависящую от вектора детерминированных параметров  $\mathbf{c} = (c_1, \dots, c_m)$  и описывающую временной ряд, рассчитанный для заданного поля в заданном году по набору снимков. На рис. 2 приведена модель зашумленного временного ряда в виде полинома 4-й степени.
2. Совокупность моделей временных рядов для полей, засеянных одним видом с/х культуры  $\Omega_l, l = 1, L$ , будем называть моделью культуры  $\Omega_l$  в некотором году. Статистическое распределение параметров  $P_l(\mathbf{c})$  моделей временных рядов описывает модель культуры  $\Omega_l$ . Для разных лет наблюдения модели одной и той же культуры различны.
3. Объединение моделей всех культур, построенных для некоторого года, будем называть моделью агросезона. Модель агросезона характеризуется объединением распределений  $\prod_{l=1}^L P_l(\mathbf{c})$  моделей культур. Модель агросезона показывает особенности развития всех видов культур в некотором году.

Общая схема информационной технологии раннего распознавания приведена на рис. 3.

#### Этап 1 "Построение моделей"

Выполняется для каждого исторического года



#### Этап 2 "Распознавание в текущем сезоне"

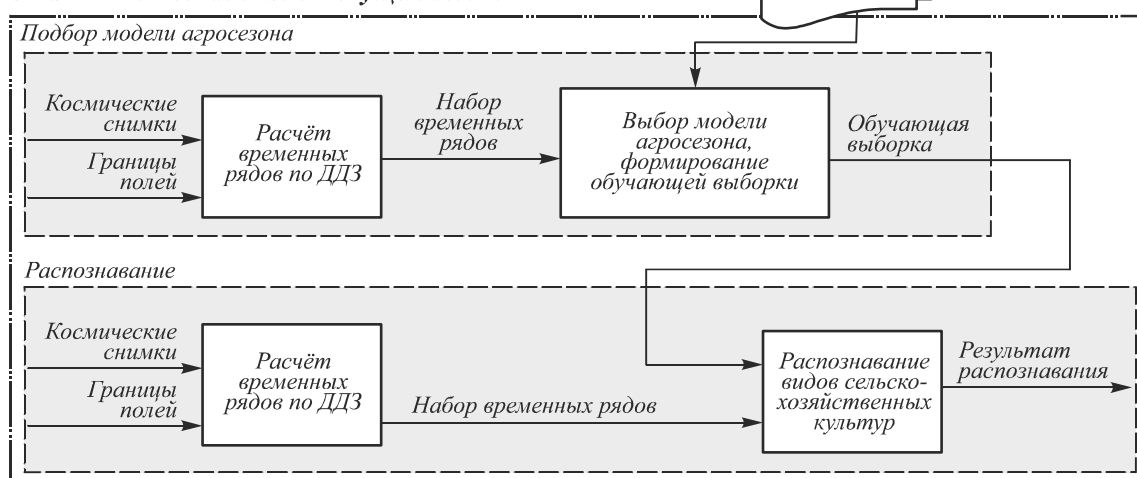


Рис. 3. Общая схема информационной технологии

Технология состоит из двух этапов.

Первый этап моделирования является подготовительным и настраивает технологию по имеющимся в наличии историческим данным: рассчитываются модели временных рядов, модели культур, из последних собираются модели агросезонов.

Второй этап информационной технологии состоит в распознавании видов с/х культур в границах полей текущего года. Из сформированных на первом этапе моделей агросезонов по опорным полям выбирается модель, наиболее близкая к развитию культур в текущем году. Затем по характеристикам модели агросезона генерируется обучающая выборка и выполняется распознавание культуры текущего сезона.

Рассмотрим технологию, её этапы и формирующие их блоки подробнее.

**2.1. Этап 1 «Построение моделей»**

Первый этап информационной технологии выполняется по историческим данным и состоит из двух блоков. Для каждого года повторяется одинаковая последовательность действий, поэтому описание данного этапа будет приведено только для одного года. Рассмотрим блоки, формирующие этап.

**Расчет временных рядов по ДДЗ**

На вход подаются космические снимки и векторные границы полей. Для каждого поля рассчитывается временной ряд вегетационного индекса по набору космических снимков.

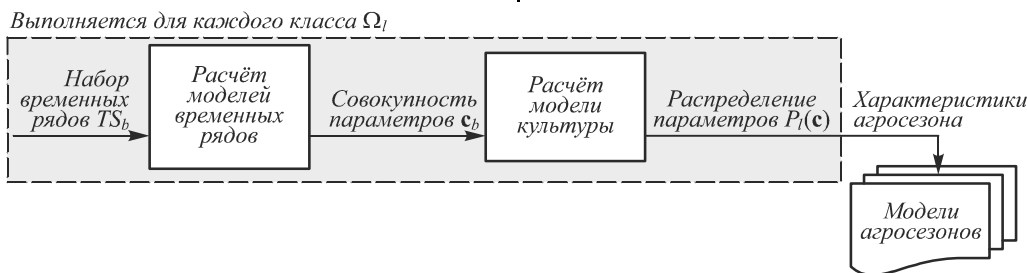


Рис. 4. Расчет модели агросезона

По характеристикам модели агросезона в любой момент можно сгенерировать набор временных рядов любого объема и сформировать по ним обучающую выборку. Генерация выполняется одинаковым способом для каждой модели культуры  $\Omega_l$ , входящей в состав модели агросезона, и состоит из следующих дей-

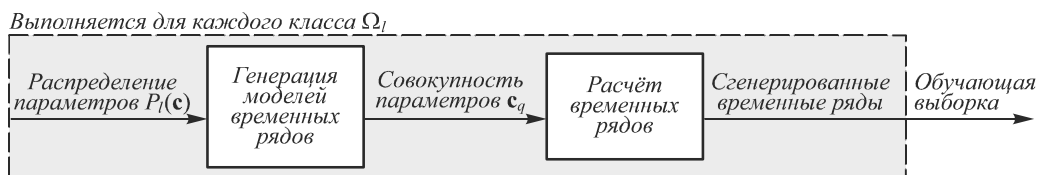


Рис. 5. Генерация обучающей выборки

**2.2. Этап 2 «Расознавание в текущем сезоне»**

Цель этапа – провести распознавание видов культур на полях текущего года с использованием характеристик развития культур, описываемых моделью агросезона.

В связи с влиянием облачности у поля в некоторый день значение вегетационного индекса будет искаженным, если значительная часть площади объекта покрыта облаками. Такие точки наблюдения необходимо исключать из рассчитанного для поля временного ряда (процедура исключения в работе не рассматривается). Выходными данными блока является набор временных рядов.

**Расчет модели агросезона**

В данном блоке производится расчет моделей временных рядов, культур и агросезона. Схема блока приведена на рис. 4. Расчет модели агросезона включает в себя расчет параметров модели каждого класса культур  $\Omega_l, l = 1, L$ . Для каждого класса  $\Omega_l$  выполняется одинаковая последовательность действий:

1. Расчет модели временного ряда  $TS_b, b = \overline{1, C_l}$ ,  $C_l$  – количество полей, засеянных с/х культурой  $\Omega_l$ .
2. Определение характеристик распределения параметров модели культуры  $P_l(c)$  по набору моделей временных рядов.

Набор характеристик моделей всех культур  $\bigcup_{l=1}^L P_l(c)$ , задающий модель агросезона, помещается в некоторое хранилище для дальнейшего использования на втором этапе.

Примеры моделей временного ряда, культуры и агросезона рассматриваются в п. 3.1 «Настройка информационной технологии».

ствий (см. рис. 5). Вначале по распределению параметров  $P_l(c)$  генерируется совокупность параметров  $c_q, q \in \mathbb{N} / \{0\}$  необходимого объема. Затем по совокупности  $c_q$  рассчитываются временные ряды вегетационного индекса.

Этап разбивается на четыре блока. Блоки «Расчет временных рядов по ДДЗ» аналогичны блоку, описанному в п. 2.1, другие два блока описаны ниже.

**Выбор модели агросезона**

В данном блоке из всех моделей агросезонов хранилища, сформированных на первом этапе, выбира-

ется одна, наиболее близкая к развитию культур на опорных полях.

Пример способа выбора модели агросезона будет рассмотрен в 3.1 «Настройка информационной технологии».

Для выбранной модели агросезона генерируется обучающая выборка заданного объема, которая используется в следующем блоке для распознавания видов с/х культур.

#### Распознавание

Рассчитанные временные ряды и полученная из предыдущего блока обучающая выборка поступают на вход блока «Распознавание видов с/х культур», в котором выполняется обнаружение видов с/х культур на полях текущего сезона.

Алгоритм распознавания, используемый в данной работе, будет рассмотрен в 3.1 «Настройка информационной технологии».

### **3. Экспериментальные исследования и примеры применения информационной технологии**

#### 3.1. Настройка информационной технологии

Для запуска информационной технологии необходимо предварительно конкретизировать её настройки.

#### Модели

В данной работе в качестве модели временного ряда использовалось приближение в виде полинома 4-й степени, в котором параметрами ( $c_1, \dots, c_5$ ) являются коэффициенты полинома. Поиск коэффициентов полинома осуществлялся методом наименьших квадратов. Пример аппроксимации временного ряда моделью приведен на рис. 2.

В ходе экспериментов было установлено, что распределение  $P_l(\mathbf{c})$  для каждого класса  $\Omega_l$  представляет собой многомерный нормальный закон, то есть модель культуры  $\Omega_l$  описывается двумя параметрами – математическим ожиданием  $M_l$  и ковариационной матрицей  $B_l$ .

Совокупность пар  $(M_l, B_l), l = \overline{1, L}$  задает модель агросезона для каждого года.

Описание сравнения моделей на основе аппроксимации различными видами функций выходит за рамки статьи и будет опубликовано позднее.

#### Алгоритм распознавания

Способ выбора модели агросезона обусловлен используемым алгоритмом распознавания, поэтому вначале рассмотрим сам алгоритм.

Для распознавания видов с/х культур применялся метод, основанный на алгоритме вычисления оценок (далее – АВО) [19]. Одним из достоинств метода является возможность использования временных рядов с пропусками данных без проведения предварительной интерполяции. АВО представляет собой модель распознающего алгоритма, которую в каждом конкретном случае необходимо конкретизировать. Рассмотрим конкретизацию:

1. *Признаки* – значения временного ряда.

2. Система опорных множеств признаков состоит из единственного множества, включающего в себя все признаки.

3. Функция близости распознаваемого  $\omega^*$  и эталонного  $\omega$  объектов рассчитывается по формуле:

$$\rho_\omega = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (\omega_n - \omega_n^*)^2}, \quad (1)$$

где  $\omega_n^*, n = \overline{0, N-1}$  – признаки объекта  $\omega^*$  (набор значений временного ряда);  $\omega_n, n = \overline{0, N-1}$  – признаки объекта  $\omega$ . Функция близости вычисляется только по тем дням  $n$ , в которые и у объекта  $\omega$ , и у объекта  $\omega^*$  имеются значения временных рядов  $\omega_n$  и  $\omega_n^*$  соответственно.

4. Значение функции близости  $f(\rho_\omega)$  между эталонным объектом  $\omega$  и распознаваемым объектом  $\omega^*$  вычисляется так:

$$f(\rho_\omega) = \begin{cases} 1, & \rho_\omega \leq T \\ 0, & \rho_\omega > T \end{cases}, \quad (2)$$

где  $T$  – пороговое значение близости.

5. Оценка  $\Gamma(\Omega_l)$  близости объекта  $\omega^*$  к некоторому классу  $\Omega_l$  вычисляется так:

$$\Gamma(\Omega_l) = \sum_{\omega \in \Omega_l} f(\rho_\omega). \quad (3)$$

6. Классификация распознаваемого объекта  $\omega^*$  будет произведена в класс  $\Omega_k$  в соответствии с решающим правилом:

$$k = \arg \max_{l=1, L} \Gamma(\Omega_l). \quad (4)$$

#### Выбор модели агросезона

В данной работе предлагается выбирать модель агросезона следующим образом:

1. Сгенерировать обучающую выборку для каждой модели агросезона из хранилища.
2. С использованием полученного набора сгенерированных обучающих выборок найти близкую модель агросезона.

Рассмотрим способ генерации обучающей выборки для одной модели агросезона. Он состоит в генерации  $L$  моделей культур. Каждая модель культуры  $\Omega_l, l = \overline{1, L}$  формируется по своим характеристикам  $(M_l, B_l)$  с помощью генерации набора реализаций многомерного нормального случайного вектора  $\mathbf{c} = (c_1, \dots, c_5)$ . Для каждой реализации вектора  $\mathbf{c}$  рассчитывается модель временного ряда в виде полинома 4-й степени.

Близкая модель агросезона выбирается в результате анализа близости временных рядов опорных полей и временных рядов из обучающей выборки каждой модели агросезона. Близость рассматривается только между полями одного класса. В качестве наиболее близкой модели агросезона будет выбираться та, для которой средняя функция близости (1) от временных рядов всех опорных полей до обучающей выборки является наименьшей.

Обучающая выборка близкой модели агросезона будет применяться в алгоритме распознавания.

**3.2. Данные для исследований**

Для оценки эффективности применения информационной технологии в задаче раннего распознавания видов с/х культур была проведена серия экспериментов по данным, полученным на территории Самарской области за 2011–2015 годы.

Территория области делится на северную, центральную и южную агроклиматические зоны (рис. 6) [20]. Каждая зона характеризуется однородными природно-климатическими условиями. В экспериментальных исследованиях использовались поля центральной зоны.



Рис. 6. Агроклиматическое районирование Самарской области

**Данные дистанционного зондирования**

Расчет временных рядов производился по данным Terra/MODIS, обработанным до уровня продукта MOD09GQ, на временном промежутке, соответствующем весенне-летнему посевному сезону с 1 апреля по 31 августа. Продукт MOD09GQ прошел радиометрическую, атмосферную коррекцию, имеет географическую привязку и представляет собой минимальный уровень комбинирования ежедневных данных разрешения 250 м. При расчетах временных рядов учитывалась маска облачности и облачные пиксели исключались из рассмотрения.

**Геоданные о полях**

В экспериментальных исследованиях использовались векторные карты границ полей и информация о засеянных на них с/х культурах. Список классов с/х культур  $\Omega_l, l = \overline{1, L}$  следующий:

1. Многолетние травы и неиспользуемые земли.
2. Озимые.
3. Пар.
4. Ранние яровые.
5. Поздние яровые.

В первую группу вошли многолетние травы, неиспользуемые земли, залежи. Во вторую – озимые пшеница, рапс, рыжик, рожь, ячмень. В третью – земли, занятые парами. В четвертую – яровой ячмень, овес, рожь. В пятую – подсолнечник, кукуруза, нут, однолетние травы, соя, гречиха. Такое разбиение является довольно стандартным и описывает крупные блоки видов с/х культур, которые интересуют службы

контроля деятельности с/х производителей. Вопрос о разделении культур по группам выходит за рамки данной работы.

Имеющиеся в наличии наземные данные о полях позволили сформировать:

1. Историческую обучающую выборку за 2011 – 2014 годы (см. табл. 1).
2. Набор опорных полей за 2015 год (см. табл. 2).
3. Обучающую выборку за 2015 год (см. табл. 3), которая используется только для оценки качества распознавания в результате применения информационной технологии.

В обучающих выборках за 2011–2015 года имеются экземпляры всех классов видов культур, в наборе опорных полей – только двух классов.

Табл. 1. Историческая обучающая выборка

Год	Кол-во полей	S, тыс. га
2011	1543	248
2012	2316	373
2013	1128	177
2014	2174	356

Табл. 2. Опорные поля за 2015 год

Класс	Кол-во полей	S, тыс. га
Ранние яровые	31	7
Поздние яровые	68	20

Табл. 3. Поля 2015 года для оценки качества распознавания

Год	Кол-во	S, тыс. га
2015	2313	373

**3.3. Экспериментальные исследования**

Эксперименты проводились для оценки качества распознавания с использованием предлагаемой информационной технологии. Качество распознавания характеризует общая вероятность верной классификации.

Учитывая малый объем надежных достоверных данных для исследования, набор опорных полей был разделен на две части. Первая использовалась для выбора модели агросезона, вторая представляла собой контрольную выборку. Поля, входящие во вторую набор, использовались в процедуре распознавания во всех экспериментах.

**Эксперимент 1**

Эксперимент проводился для проверки необходимости замены зашумленных реальных временных рядов модельными данными.

Модельные данные рассчитаны как 4000 реализаций многомерного нормального случайного вектора для каждого класса культур для агросезонов 2011 – 2014 г.

В соответствии с описанной информационной технологией было проведено распознавание полей контрольной выборки. Распознавание проводилось для разной длины временных рядов с шагом примерно в две недели. Назовем эти даты контрольными. В табл. 4 для некоторых контрольных дат приведены результаты выбора наиболее близкой модели агросезона.

Табл. 4. Выбор агросезона в зависимости от контрольной даты

Относительная дата	57	85	113	141	169
Год	2013	2014	2014	2013	2014

Качество классификации контрольной выборки с использованием информационной технологии отражает график черного цвета на рис. 7. Пунктирная кривая отражает качество классификации контрольной выборки по реальным историческим обучающим выборкам. Для каждой контрольной даты год реальной обучающей выборки соответствует табл. 4.



Рис. 7. Качество классификации с использованием моделированных данных и по реальной исторической выборке

Из рис. 7 видно, что в период раннего распознавания в большинстве случаев качество классификации по модельным данным выше.

#### Эксперимент 2

Эксперимент проводился для сравнения качества распознавания по информационной технологии с качеством распознавания по модели текущего агросезона (см. рис. 8).

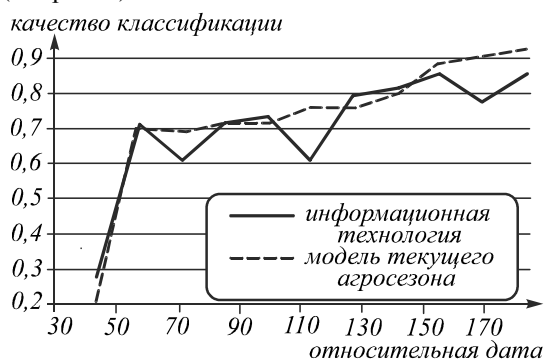


Рис. 8. Качество классификации по информационной технологии и по модели текущего агросезона

Рассмотрим подробнее второй способ распознавания. Текущий агросезон соответствует 2015 году. Распознавание по модели агросезона 2015 года имитирует ситуацию, как если бы у нас в наличии была обучающая выборка текущего года с самого начала посевного сезона, по ней рассчитали модель агросезона и сгенерировали обучающую выборку. В эксперименте реальная обучающая выборка 2015 года соответствует данным из табл. 3.

Вне областей провалов черного графика наблюдается его близость к пунктирному графику, что гово-

рит о возможности использования исторических данных для распознавания полей текущего сезона. Улучшение качества распознавания по информационной технологии можно достичь двумя способами:

1. Накоплением дальнейшей статистики.
2. Использованием данных других двух зон Самарской области для расчета моделей агросезонов и наполнения ими хранилища.

#### 3.4. Достоинства и недостатки технологии

Результаты экспериментальных исследований показывают, что разработанная информационная технология обеспечивает приемлемое качество раннего распознавания с/х культур с возможностью его улучшения.

Перечислим основные достоинства технологии:

1. Неограниченный набор обучающей выборки в виде «гладких» модельных рядов вместо зашумленных исторических данных.
2. Высокая способность к обобщению, адаптивная подстройка к конкретному виду моделей и числу параметров.
3. Алгоритм распознавания использует точки временного ряда, а не вторичные признаки вегетационного цикла (длина, максимум NDVI), что позволяет провести раннее распознавание.
4. Возможность использования временных рядов с пропусками из-за особенностей АВО.
5. Невысокое число необходимых данных об опорных полях.

Основным недостатком технологии является ограниченное число «моделей агросезонов» в исторической выборке при запуске технологии. Этот недостаток может быть частично устранен за счет использования всех агроклиматических зон для наполнения хранилища моделей.

#### Заключение

Разработана информационная технология решения актуальной задачи раннего распознавания видов с/х культур по космическим снимкам на территориях регионального размера. Технология позволяет проводить обнаружение видов с/х культур в ситуации малого количества исходных космических и наземных данных, недостаточных для формирования обучающей выборки. Отличительной особенностью технологии является построение моделей временных рядов и их использование в алгоритме распознавания. Проведена серия экспериментальных исследований на реальных данных территории Самарской области. Результаты работы с 2016 года включены в состав разрабатываемой геоинформационной системы агропромышленного комплекса Самарской области в виде специализированного программного обеспечения.

#### Благодарности

Работа выполнена за счет гранта Российского научного фонда (проект №14-31-00014) «Создание лаборатории прорывных технологий дистанционного зондирования Земли».

*Литература*

1. **Книжников, Ю.Ф.** Аэрокосмические методы географических исследований / Ю.Ф. Книжников, В.И. Кравцова, О.В. Тутубалина. – 2-е изд., перераб. и доп. – М.: Академия, 2011. – 416 с. – ISBN: 978-5-7695-6830-5.
2. **Барталёв, С.А.** Распознавание пахотных земель на основе многолетних спутниковых данных спектрорадиометра MODIS и локально адаптивной классификации / С.А. Барталёв, В.А. Егоров, Е.А. Лупян, Д.Е. Плотников, И.А. Уваров // Компьютерная оптика. – 2011. – Т. 35, №1. – С. 103-116.
3. **Спивак, Л.Ф.** Космический мониторинг сельскохозяйственного производства в Казахстане / Л.Ф. Спивак, Н.Р. Муратова // Земля из космоса. Наиболее эффективные решения. – 2011. – Вып. 9. – С. 57-65.
4. **Султангазин, У.М.** Использование космического мониторинга в планировании и прогнозировании параметров зернового производства / У.М. Султангазин, Н.Р. Муратова, А.Г. Терехов // Современные проблемы дистанционного зондирования Земли из космоса. – 2004. – Т. 1, № 1. – С. 291-297.
5. **Schmedtmann, J.** Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control / J. Schmedtmann, M.L. Campagnolo // Remote Sensing of Environment. – 2015. – Vol. 7(7). – P. 9325-9346. – DOI: 10.3390/rs70709325.
6. **Плотников, Д.Е.** Признаки распознавания пахотных земель на основе многолетних рядов данных спутникового спектрорадиометра MODIS / Д.Е. Плотников, С.А. Барталёв, Е.А. Лупян // Современные проблемы дистанционного зондирования Земли из космоса. – 2010. – Т. 7, № 1. – С. 330-341.
7. **Плотников, Д.Е.** Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости / Д.Е. Плотников, С.А. Барталёв, В.О. Жарко, В.В. Михайлов, О.И. Присянникова // Современные проблемы дистанционного зондирования Земли из космоса. – 2011. – Т. 8, № 1. – С. 199-208.
8. **Wardlow, B.D.** Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains / B.D. Wardlow, S.L. Egbert, J.H. Kastens // Remote Sensing of Environment. – 2007. – Vol. 108. – P. 290-310. – DOI: 10.1016/j.rse.2006.11.021.
9. **Wardlow, B.D.** Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains / B.D. Wardlow, S.L. Egbert // Remote sensing of environment. – 2008. – Vol. 112(3). – P. 1096-1116. – DOI: 10.1016/j.rse.2007.07.019.
10. **Wei, W.** Selecting the optimal NDVI time-series reconstruction technique for crop phenology detection / W. Wei, W. Wu, Zh. Li, P. Yang, Q. Zhou // Intelligent Automation & Soft Computing. Special Section: Intelligent Automation with Applications to Agriculture. – 2016. – Vol. 22(2). – P. 237-247. – DOI: 10.1080/10798587.2015.1095482.
11. **Терехин, Э.А.** Информативность спектральных вегетационных индексов для дешифрирования сельскохозяйственной растительности / Э.А. Терехин // Современные проблемы дистанционного зондирования Земли из космоса. – 2012. – Т. 9, № 4. – С. 243-248.
12. **Castillejo-González, I.L.** Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird / I.L. Castillejo-González, F. López-Granados, A. García-Ferrer // Computers and Electronics in Agriculture. – 2009. – Vol. 68(2). – P. 207-215. – DOI: 10.1016/j.compag.2009.06.004.
13. **Fischer, A.** A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters / A. Fischer // Remote Sensing of Environment. – 1994. – Vol. 48(2). – P. 220-230. – DOI: 10.1016/0034-4257(94)90143-0.
14. **Zhong, L.** A phenology-based approach to map crop types in the San Joaquin Valley, California / L. Zhong, T. Hawkins // International Journal of Remote Sensing. – 2011. – Vol. 32(22). – P. 7777-7804. – DOI: 10.1080/01431161.2010.527397.
15. **Ozdogan, M.** The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis / M. Ozdogan // Remote Sensing of Environment. – 2010. – Vol. 114(6). – P. 1190-1204. – DOI: 10.1016/j.rse.2010.01.006.
16. **Brown, J.C.** Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data / J.C. Brown, J.H. Kastens, A.C. Coutinho, D. de Castro Victoria, Ch.R. Bishop // Remote Sensing of Environment. – 2013. – Vol. 130. – P. 39-50. – DOI: 10.1016/j.rse.2012.11.009.
17. **Justice, C.O.** An overview of MODIS Land data processing and product status / C.O. Justice, J.R.G. Townshend, E.F. Vermote // Remote Sensing of Environment. – 2002. – Vol. 83(1). – P. 3-15. – DOI: 10.1016/S0034-4257(02)00084-6.
18. **Vorobiova, N.S.** How to Use Geoinformation Technologies and Space Monitoring for Controlling the Agricultural Sector in Samara Region / N.S. Vorobiova, A.Yu. Denisova, A.V. Kuznetsov, A.M. Belov, A.V. Chernov, V.V. Myasnikov // Pattern Recognition and Image Analysis. – 2015. – Vol. 25(2). – P. 347-353. – DOI: 10.1134/S1054661815020261.
19. **Журавлев, Ю.И.** Алгоритмы вычисления оценок и их применение / Ю.И. Журавлев, М.М. Камилов, Ш.Е. Туляганов. – Ташкент: Фан, 1974. – 119 с.
20. **Васин, А.В.** Формирование высокопродуктивных поливидовых агрофитоценозов кормовых культур в Среднем Поволжье : дис. ... д-ра с.-х. наук: 06.01.09. / Алексей Васильевич Васин; Самарская гос. с.-х. акад. – Кинель, 2006. – 513 с.

*Сведения об авторах*

**Воробьева Наталья Сергеевна** родилась в 1985 году. В 2008 году окончила Самарский государственный аэрокосмический университет (СГАУ). В настоящее время работает инженером в НИЛ-97 СГАУ. Круг научных интересов включает обработку данных дистанционного зондирования, распознавание образов. E-mail: [natalyavorobiova@gmail.com](mailto:natalyavorobiova@gmail.com).

**Сергеев Владислав Викторович** родился в 1951 году. В 1974 году окончил Куйбышевский авиационный институт. В 1993 году защитил диссертацию на соискание степени доктора технических наук. Основное место работы – заведующий кафедрой геоинформатики и информационной безопасности Самарского национального исследовательского университета имени академика С.П. Королёва (Самарский университет). По совместительству работает



заведующим лабораторией математических методов обработки изображений в Институте систем обработки изображений РАН (ИСОИ РАН) – филиале ФНИЦ «Кристаллография и фотоника» РАН. Круг научных интересов: цифровая обработка сигналов, анализ изображений, распознавание образов, геоинформатика. E-mail: [vsereg@geosamara.ru](mailto:vsereg@geosamara.ru).

**Чернов Андрей Владимирович** родился в 1975 году. В 1998 году окончил Самарский государственный аэрокосмический университет (СГАУ). В 2004 году защитил диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим научным сотрудником в НИЛ-97 СГАУ. Круг научных интересов включает геоинформационные системы, обработку данных дистанционного зондирования, распознавание образов. E-mail: [chernov@geosamara.ru](mailto:chernov@geosamara.ru).

ГРНТИ: 28.23.15, 89.57.35, 89.57.45, 28.17.19.

Поступила в редакцию 17 ноября 2016 г. Окончательный вариант – 28 ноября 2016 г.

## INFORMATION TECHNOLOGY OF EARLY CROP IDENTIFICATION BY USING SATELLITE IMAGES

*N.S. Vorobiova<sup>1,2</sup>, V.V. Sergeyev<sup>1,2</sup>, A.V. Chernov<sup>1,2</sup>*

<sup>1</sup> *Image Processing Systems Institute of RAS – Branch of the FSRC “Crystallography and Photonics” RAS, Samara, Russia,*

<sup>2</sup> *Branch of the FSRC “Crystallography and Photonics” RAS, Samara, Russia*

### Abstract

The paper deals with the development of information technology for early crop identification in the region of interest by using satellite images. With the early identification being performed in the first half of the growing season, it is characterized by the lack of ground-based and space-based data to configure recognition algorithms. The proposed technology allows one to generate a training dataset by using information from the past years and then use it for crop recognition for the current year. The technology consists of two stages. At the first stage, models of time series, crops and agro-seasons are built using data from the past years. At the second stage, data of the current year are processed. First, a suitable model of agro-season, which is close to the development of crops in the current year, is selected by using a tiny set of control fields with known crops. Next, a training dataset is generated based on the characteristics of the agro-season model and the recognition of crops in the current season is performed. The paper assesses the quality of recognition offered by the proposed information technology and the possibility of its application in a particular region. Data collected for the years 2011–2015 in the Samara Region are used for the experimental research.

**Keywords:** crop identification, satellite images, vegetation index, time series, NDVI, time series model.

**Citation:** Vorobiova NS, Sergeyev VV, Chernov AV. Information technology of early crop identification by using satellite images. *Computer Optics* 2016; 40(6): 929-938. DOI: 10.18287/2412-6179-2016-40-6-929-938.

**Acknowledgements:** The work was funded by the Russian Science Foundation, grant No. 14-31-00014.

### References

- [1] Knizhnikov JF, Kravtsova VI, Tutubalina OV. Aerospace methods of geographical research [In Russian]. Moscow: “Academia” Publisher; 2011.
- [2] Bartalev SA, Egorov VA, Loupian EA, Plotnikov DE, Uvarov IA. Recognition of arable lands using multi-annual satellite data from spectroradiometer MODIS and locally adaptive supervised classification [In Russian]. *Computer Optics* 2011; 35(1): 103-116.
- [3] Spivak LF, Muratova NR. Space monitoring of agricultural production in Kazakhstan [In Russian]. *Earth from Space: The Most Effective Solutions* 2011; 9: 57-65.
- [4] Sultangazin UM, Muratova NR, Terehov AG. Use of satellite monitoring in planning and forecasting the parameters of cereal production [In Russian]. *Current Problems in Remote Sensing of the Earth from Space* 2004; 1(1): 291-297.
- [5] Schmedtmann J, Campagnolo ML. Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control. *Remote Sensing of Environment* 2015; 7(7): 9325-9346. DOI: 10.3390/rs70709325.
- [6] Plotnikov DE, Bartalev SA, Loupian EA. The recognition features to map arable lands based on multi-annual MODIS Earth observation data [In Russian]. *Current Problems in Remote Sensing of the Earth from Space* 2010; 7(1): 330-341.
- [7] Plotnikov DE, Bartalev SA, Zharko VO, Mihailov VV, Prosyannikova OI. An experimental assessment of crop types recognisability using time-series of intra-seasonal spectral reflectance measurements by satellite sensor [In Russian]. *Current Problems in Remote Sensing of the Earth from Space*, 2011; 8(1): 199-208.
- [8] Wardlow BD, Egbert SL, Kastens JH. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment* 2007; 108: 290-310. DOI: 10.1016/j.rse.2006.11.021.
- [9] Wardlow BD, Egbert SL. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment* 2008; 112(3): 1096-1116. DOI: 10.1016/j.rse.2007.07.019.
- [10] Wei W, Wu W, Li Z, Yang P, Zhou Q. Selecting the optimal NDVI time-series reconstruction technique for crop

- phenology detection. *Intelligent Automation & Soft Computing. Special Issue: Intelligent Automation with Applications to Agriculture* 2016; 2: 237-247. DOI: 10.1080/10798587.2015.1095482.
- [11] Terekhin EA. The efficiency of spectral Indices for interpretation agricultural vegetation [In Russian]. *Current Problems in Remote Sensing of the Earth from Space* 2012; 9(4): 243-248.
- [12] Castillejo-González IL, López-Granados F, García-Ferrer A. Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture* 2009; 68(2): 207-215. DOI: 10.1016/j.compag.2009.06.004.
- [13] Fischer A. A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Remote Sensing of Environment* 1994; 48(2): 220-230. DOI: 10.1016/0034-4257(94)90143-0.
- [14] Zhong L, Hawkins T. A phenology-based approach to map crop types in the San Joaquin Valley, California. *International Journal of Remote Sensing* 2011; 32(22): 7777-7804. DOI: 10.1080/01431161.2010.527397.
- [15] Ozdogan M. The spatial distribution of crop types from MODIS data: Temporal unmixing using Independent Component Analysis. *Remote Sensing of Environment* 2010; 114(6): 1190-1204. DOI: 10.1016/j.rse.2010.01.006.
- [16] Brown JC, Kastens JH, Coutinho AC, Victoria DC, Bishop CR. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment* 2013; 130: 39-50. DOI: 10.1016/j.rse.2012.11.009.
- [17] Justice CO, Townshend JRG, Vermote EF. An overview of MODIS Land data processing and product status. *Remote Sensing of Environment* 2002; 83(1): 3-15. DOI: 10.1016/S0034-4257(02)00084-6.
- [18] Vorobiova NS, Denisova AYU, Kuznetsov AV, Belov AM, Chernov AV, Myasnikov VV. How to Use Geoinformation Technologies and Space Monitoring for Controlling the Agricultural Sector in Samara Region. *Pattern Recognition and Image Analysis* 2015; 25(2): 347-353. DOI: 10.1134/S1054661815020261.
- [19] Zhuravlev JI, Kamilov MM Tulyaganov ShE. Algorithms for calculating estimates [In Russian]. Tashkent: "Fan" Publisher; 1974.
- [20] Vasin AV. Formation of highly productive polyspecific agrophytocenosis of forage crops in the Middle Volga. Thesis of the doctor of agricultural sciences [In Russian]. Kinel, 2006.

#### Authors' information

**Natalya Sergeyevna Vorobiova** (1985 b.), graduated from Samara State Aerospace University (presently, Samara National Research University or shortly, Samara University) in 2008. Currently, she works at laboratory No. 97 at Samara University and at the Image Processing Systems Institute of RAS – Branch of the FSRC “Crystallography and Photonics” RAS. Areas of research interest include satellite image processing, geoinformatics and pattern recognition. E-mail: [natalyavorobiova@gmail.com](mailto:natalyavorobiova@gmail.com).

**Vladislav Victorovich Sergeyev** (1951 b.), graduated (1974) from S. P. Korolyov Kuibyshev Aviation Institute (presently, Samara University). He received his Candidate's degree in Technical Sciences in 1978 and DrSc degree in Mathematics and Physics in 1993. At present, he is the head of Geoinformatics and Information Security sub-department at Samara University, also holding a part-time position as the head of a laboratory at the Image Processing Systems Institute of the RAS. The areas of research interests include digital signal and image processing, geoinformatics and pattern recognition. E-mail: [vserg@geosamara.ru](mailto:vserg@geosamara.ru).

**Andrey Vladimirovich Chernov** (1975 b.) graduated (1998) from Samara State Aerospace University. He received his Candidate's degree in Technical Sciences in 2004. Currently he is an associate professor at the Geoinformatics and Information Security sub-department at Samara University and the senior researcher in research laboratory of advanced technologies of remote sensing. The areas of research interests include geoinformation systems, processing of remotely sensed data, pattern recognition. E-mail: [chernov@geosamara.ru](mailto:chernov@geosamara.ru).

*Received November 17, 2016. The final version – November 28, 2016.*