

Ваткин М.Е., Ганченко В.В., Дудкин А.А., Петровский А.И.

ИДЕНТИФИКАЦИЯ ЗАБОЛЕВАНИЙ КАРТОФЕЛЯ НА ОСНОВЕ АНАЛИЗА ИЗМЕНЕНИЙ ЦВЕТОВЫХ ХАРАКТЕРИСТИК ЛИСТЬЕВ

Введение

Дистанционные методы мониторинга сельскохозяйственных полей дают возможность оперативно выявить участки, пораженные болезнью. Выявление заболевания на ранних стадиях развития значительно сокращает затраты труда и средств. Находят развитие два основных подхода к решению задачи выявления пораженных участков – спектрометрический и оптический (видеоспектральный) [1,2]. Причем, использование спектральных характеристик позволяет выделить ряд болезней растений на более ранней стадии. К примеру, изменение отражательных характеристик растений в инфракрасной области позволяет идентифицировать фитофтороз картофеля еще до появления визуальных признаков [1, с.409]. Тем не менее, разработка эффективных методов обработки изображений, учитывающих специфику изображений, опирающихся на детальные исследования информативных признаков заболеваний в оптическом диапазоне, позволят подойти к построению систем идентификации, которые как самостоятельно, так и особенно в сочетании со спектральными методами, позволят повысить качество идентификации болезней растений.

Ряд болезней сельскохозяйственных растений проявляется в нарушении фотосинтеза и вызывает поражения отдельных органов или преждевременную гибель растений. Поэтому основной задачей при обработке изображений для идентификации пораженных болезнью участков является сегментация, т.е. выделение однородных по своим цветовым или спектральным признакам объектов изображения (группы пикселей) и отнесение их к тому или иному заранее определенному классу. При этом качество сегментации определяется: а) корректным выделением информативных признаков заболевания и б) алгоритмом сегментации.

В качестве информативных признаков заболеваний используются геометрические и морфологические, а также цветовые характеристики растений или их сочетания с целью сокращения размерности признакового пространства, которое ведет к более простой классификационной схеме и позволяет выполнить более раннюю диагностику определенного заболевания [3-5].

На этапе сегментации используются различные подходы в зависимости от специфики изображений и вида представления результата. В [6] для обработки цифровых аэрофотоснимков предлагается использовать метод наращивания областей в комбинации с применением деформируемых моделей. Цель этого подхода состоит в том, чтобы облегчить решение задачи оцифровки контуров областей и получить их векторное представление.

В [7] предлагается использовать методы математической морфологии для сегментации изображений сельскохозяйственных земель. Алгоритм сегментации полутоновых изображений построен на основе преобразования водораздела и операции полутонового псевдоскелета. Алгоритм характеризуется отказом от использования операции бинаризации, что позволяет значительно повысить качество сегментации, а использование в нем операции полутонового псевдоскелета дает возможность проводить сегментацию полутонового изображения со значительным снижением влияния ошибок квантования и шума на изображении на результат сегментации.

В [8-9] предложена технология автоматического обнару-

жения заданных областей, особенность которой в применении областно-ориентированной классификации на базе географических данных и набора изображений интересующей области за исследуемый отрезок времени с использованием минимального расстояния до среднего. В [10-11] предложена специальная функция расстояния для кластерной сегментации, а также использована операция математической морфологии для улучшения результата кластеризации.

В настоящей статье рассматривается задача построения информативных признаков и анализа последовательности изображений растений, полученных в различные дни, для идентификации заболевания на ранней стадии развития болезни картофеля по цветовым характеристикам листьев (выделение областей заболевания) в лабораторных условиях.

Наблюдению подвергались 3 группы растений [12].

В – Растения, зараженные болезнью из группы *alternaria* (грибные болезни). У пораженных растений на листьях и стеблях появляются темно-бурые пятна, увеличивающиеся в размерах. Листья вянут, желтеют и чернеют, затем засыхают, во влажную погоду загнивают, опадают, а стебли надламываются.

Z – Растения, зараженные бактериальной болезнью *ergwinia* (tzw. czarna pozka). У пораженных растений листья желтеют и свертываются. Нижняя часть стебля и корни загнивают и становятся черными.

K – Здоровые растения (контрольная группа).

В каждой серии наблюдалось по 25 растений. Съемка растений проводилась ежедневно в 8, 10, 12, 14 и 16 часов на протяжении 7 дней.

Методика анализа

Проводилось два вида анализа:

- экспертный;
- автоматический.

Экспертный анализ.

Множество представленных изображений было проанализировано экспертом при помощи программных средств Photoshop.

Цели анализа: 1) создание базы эталонов заболевших и здоровых растений, 2) выявление признаков болезни, 3) формирование критериев условий съемки, 4) формирование метода автоматической обработки изображений на основе выявленных признаков.

Использовались следующие основные операции:

Сегментирование;

Анализ гистограмм оттенка и RGB компонент цвета сегментов.

Автоматический анализ.

Цели анализа: выявление заболевания и мониторинг развития болезни.

Для анализа был разработан многокритериальный пороговый алгоритм сегментации изображений и алгоритм идентификации болезни учитывающий динамику изменения информативных признаков.

Результаты экспертного анализа

Были выявлены следующие закономерности в развитии болезни.

1. Основными признаками проявления болезни является изменение цвета листьев и стебля растения двух видов: 1) пожелтение 2) потемнение (буро-зеленый цвет). Данный факт проиллюстрирован на рис. 1.

Ваткин М.Е., к.т.н., научный сотрудник ОИППИ НАН Беларуси.

Ганченко В.В., аспирант ОИППИ НАН Беларуси.

Петровский А.И., к.т.н., ученый секретарь ОИППИ НАН Беларуси.

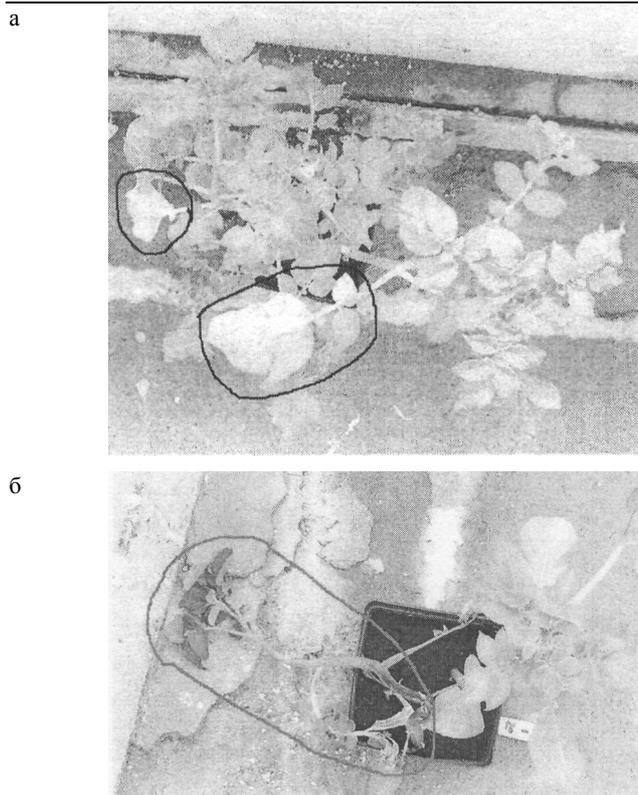


Рис. 1. Растения с изменениями цвета листьев и стебля а) пожелтение листьев б) потемнение листьев и стебля

Эмпирически получены следующие значения диапазонов цвета.

Сегменты зеленого цвета (здоровые сегменты растения):

- диапазон оттенка [1,4; 3,14],
- диапазон насыщенности [40;200].

Сегменты желтого цвета (больные сегменты растения):

- диапазон оттенка [0,9; 1,4],
- диапазон насыщенности [80; 200].

Сегменты буро-зеленого цвета (больные сегменты растения):

- диапазон оттенка [1,5; 1,8],
- диапазон насыщенности [14; 55].

Значение оттенка вычисляется по формуле

$$Hue = \arctan \left(\frac{\sin\left(\frac{2}{3}\pi\right) \cdot g - \sin\left(\frac{2}{3}\pi\right) \cdot b}{r + \cos\left(\frac{2}{3}\pi\right) \cdot g + \cos\left(\frac{2}{3}\pi\right) \cdot b} \right),$$

где r, g, b – значения красной, зеленой и голубой компоненты цвета.

Значение насыщенности вычисляется по формуле

$$Sat = \max(r, g, b) - \min(r, g, b),$$

где r, g, b - значения красной, зеленой и голубой компоненты цвета.

2. Относительно критериев условий съемки получены следующие результаты.

2.1. Появление мелких бурых пятен на зеленом листе (рис. 2). Было замечено, что при съемке с расстояния 1 метр мелкие бурые пятна не обнаруживаются ни при визуальном осмотре, ни при помощи программных средств, т.к. вначале это могут быть области минимальной площади (от одного пикселя, рис. 2,б). Основная задача при этом заключалась в том, чтобы отделить данный признак от фона и случайных помех.

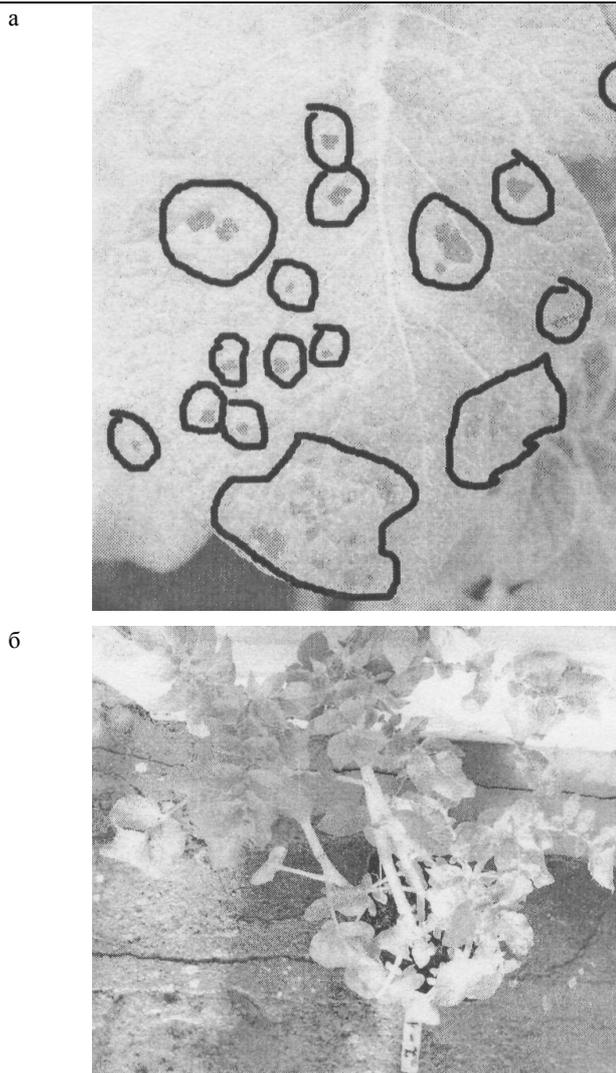


Рис. 2. Пример растения с мелкими бурыми пятнами а) крупный масштаб листа растения; б) съемка с расстояния 1 метр

2.2. Было отмечено, что создаваемые солнцем блики на листьях создавали эффект пожелтения листа (рис. 3). Это вносило дополнительную ошибку при анализе изображения программными средствами. Для устранения влияния бликов на процесс анализа необходимо использовать соответствующие фильтры на этапе съемки изображений.



Рис. 3. Пример растения с бликами

2.3. Сильное влияние на процесс анализа изображения программными средствами оказывали фоновые объекты, совпадающие по цветовым своим характеристикам с цветом выделяемых заболевших сегментов растения. К таким фоновым объектам относятся участки асфальта и грунта имеющие темно бурый цвет, желтый шланг на некоторых снимках. Для устранения их влияния на процесс автоматизированного анализа необходимо в дальнейшем устранить такие объекты с полей.

3. В результате экспертного анализа был предложен метод выявления заболевания растения, состоящий из следующих этапов:

- выделение сегментов желтого и зеленого цвета;
- расчет критериев заболевания, соответствующих соотношениям:
 - пороговый критерий

$$Treash = \frac{100 \cdot N1}{N1 + N2},$$

где $N1$ и $N2$ количество точек для сегментов желтого и зеленого цвета, соответственно;

- процент сдвига среднего оттенка выделенных сегментов в сторону желтого оттенка: первый способ

$$Shift1 = \frac{100 \cdot \left(\frac{1}{N} \sum_{i=1}^N hue_i \right)}{3.14 - 0.9}$$

второй способ

$$Shift2 = \frac{100 \cdot \left(\frac{(3.14 + 1.4)}{2} - \frac{1}{N} \sum_{i=1}^N hue_i \right)}{\frac{(3.14 + 1.4)}{2} - \frac{(1.4 + 0.9)}{2}}$$

- дисперсия оттенка выделенных сегментов (зеленые и желтые сегменты)

$$Disp = \sqrt{\frac{1}{N} \sum_{i=1}^N (hue)^2 - \left(\frac{1}{N} \sum_{i=1}^N hue_i \right)^2},$$

где $N = N1 + N2$.

Эффективность использования предложенных коэффициентов анализировалась на этапе тестирования результатов автоматизированного распознавания заболевания растения.

Алгоритм автоматической обработки

Автоматическая обработка реализует метод выявления заболевания растения, описанный выше, и основана на разработанном нами алгоритме сегментации изображений, а также анализе динамики изменения значений критериев заболевания.

В [10] нами предложен кластерный алгоритм сегментации изображений растений для идентификации областей, пораженных болезнью, на основании цветовых характеристик. Данный алгоритм дает приемлемые результаты сегментации для случая, если на изображении присутствуют сегменты, цвет которых совпадает с цветом выделяемых сегментов. В настоящей статье предложен многокритериальный пороговый алгоритм сегментации, позволяющий работать также с изображениями, на которых нет сегментов искомого цвета. Кроме того, разработанный алгоритм в силу своей простоты обладает большей производительностью.

Алгоритм многокритериальной пороговой сегментации

1. Устанавливаем набор критериев, которым должны удовлетворять пиксели сегментов:

для зеленого цвета – $1.4 < hue < 3.14$ и $40 < sat < 200$;

для желтого цвета – $0.9 < hue < 1.4$ и $80 < sat < 200$.

2. Оставляем на изображении пиксели, удовлетворяющие первому или второму набору критериев.

3. На множестве выделенных пикселей вычисляем значения характеристик $Treash$, $Shift1$, $Shift2$, $Disp$.

Анализ динамики развития болезни

Как показал анализ графиков изменения признаков $Treash$, $Shift1$, $Shift2$, $Disp$ по дням, наиболее эффективным признаком для выявления болезни является $Treash$, с которым имеет корреляцию $Disp$. Поэтому при дальнейшем анализе использовался признак $Treash$.

Для повышения точности идентификации болезни анализировалась динамика изменения значений признака $Treash$.

Автоматизированный алгоритм анализа динамики развития болезни одного растения, состоит из следующих шагов:

1. Сегментация изображений и подсчет размеров сегментов.
2. Расчет признака болезни $Treash$.
3. Усреднение значений признака заболевания для растения, полученных в разное время суток, что позволяет удалить случайные помехи.
4. Построение графика признака заболевания растения по дням наблюдений.
5. Если в течение последних трех дней наблюдения на графике виден рост значения признака болезни, то растение считается заболевшим, а началом болезни считается день, от которого начался рост графика.

Приводятся графики значений $Treash$, на которых отмечены идентифицированные болезни.

В сводной таблице показаны характеристика точности идентификации болезни и день обнаружения болезни. В таблице знаком "+" обозначена уверенная идентификация болезни (100%), знаком "+/-" – 60% идентификация, знаком "-/+" – 40% идентификация, знаком "-" – отмечены растения с не идентифицированной болезнью.

В итоге получаем следующие характеристики алгоритма идентификации болезни:

Среднее время определения болезни по группе В составило 3,9 дня.

Среднее время определения болезни по группе Z составило 4,5 дня.

Среднее время определения болезни по группам В, Z составило 4,2 дня.

Присвоим слабой ошибке идентификации вес 0,5, а группой – 1. В результате на контрольной группе (К) ошибка идентификации болезни составила $100 \cdot (6/25) = 24\%$.

Выводы

1. Предложенный алгоритм автоматического определения заболевания растения показал удовлетворительный результат при использовании признака пожелтения листьев в качестве индикатора заболевания. При использовании признака потемнения листьев алгоритм показал плохие результаты из-за сильного влияния фоновых объектов как асфальт и грунт, темные листья за счет тени, которые имеют те же цветовые характеристики.

2. Экспериментальные результаты показали, что в среднем по всем заболевшим растениям оказалось возможным выявить заболевание на четвертом дне развития.

3. Наибольшее влияние на работу метода оказывает:

- наличие в кадре посторонних предметов (шланг, бирка растения), имеющих желтый цвет;
- наличие солнечных бликов на растении и фоне (создает дополнительные желтые пятна);
- неравномерность освещения растения (может создавать дополнительные бурые пятна);

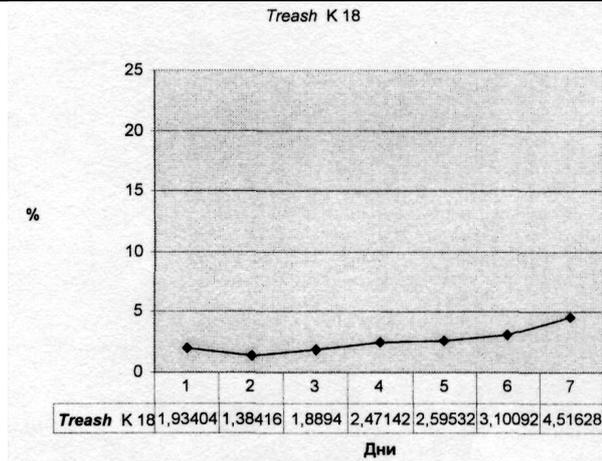
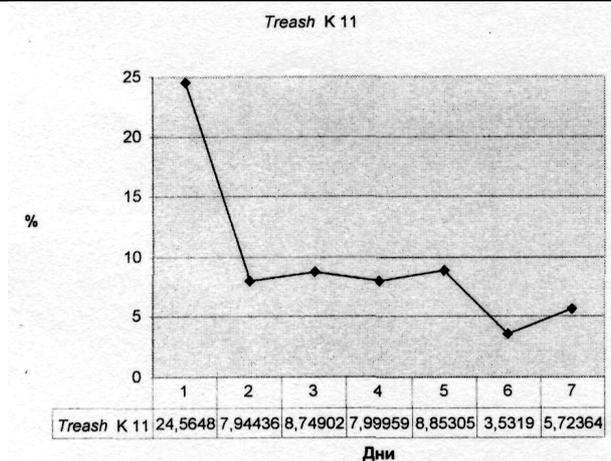


Рис. 4. Пример графиков признака Treash для здоровых растений 11 и 18 из группы К

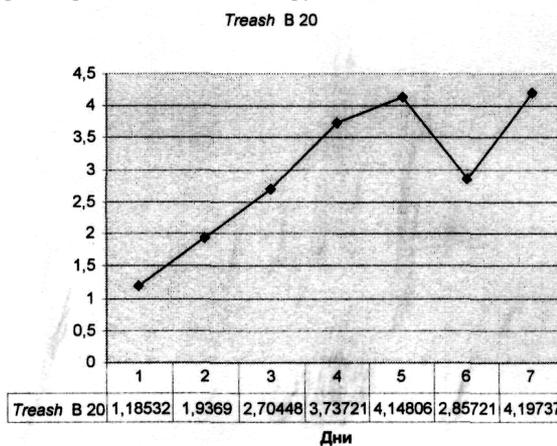
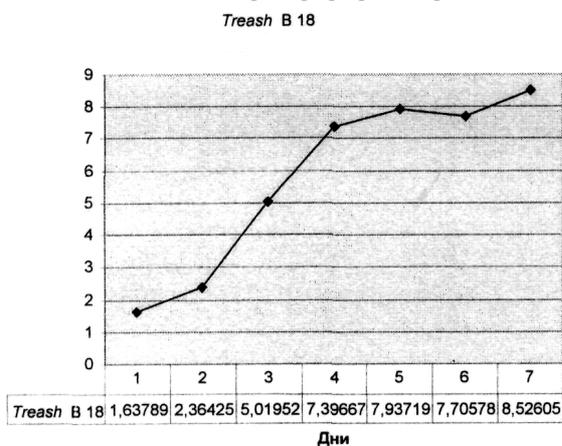


Рис. 5. Пример графиков признака Treash больных растений 18, 20 из группы В

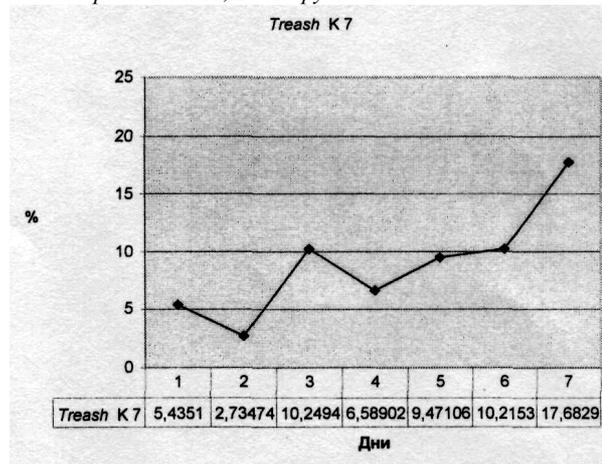
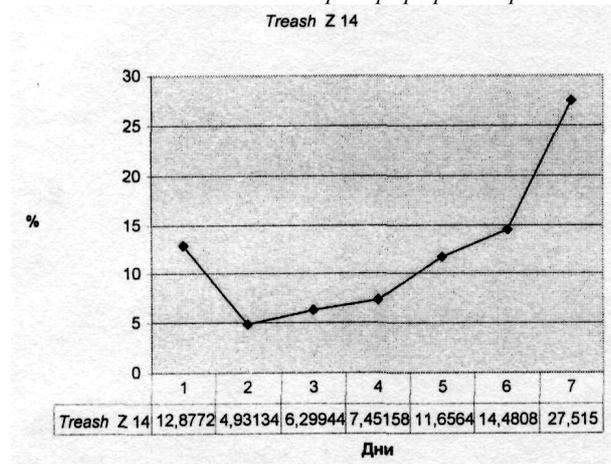


Рис. 6. Пример графиков признака Treash больных растений 14 из группы Z, 7 из – К

- совпадение цветовых характеристик фона с цветовыми характеристиками проявления заболевания (бурый стебель).
- 4. Были сформулированы критерии условий съемки изображений:
 - для устранения бликов съемку необходимо производить в рассеянном освещении, либо используя соответствующие фильтры, либо большем расстоянии от объекта, т.е. в мелком масштабе как при съемке поля в целом;
 - для устранения влияния пожелтения в результате засыхания необходимо производить съемку после полива растений.

- для устранения влияния фоновых объектов необходимо удалить с поля съемки фоновые объекты зеленого и желтого цвета.
- для устранения влияния пожелтения в результате отражения солнечного света (когда спектр солнечного света имеет наибольшую активность в желтом диапазоне), необходимо производить съемку в соответствующее время суток.
- 5. В качестве рекомендаций для дальнейших исследований отметим, что целесообразно провести анализ методов выявления заболеваний за счет изменения спектрального диапазона, в котором производится съемка.

Сводная таблица

№ п/п	Группа-номер в группе	Экспертная оценка наличия пожелтения	Автоматическая идентификация болезни	День идентификации	Примечания
Растения, зараженные болезнью из группы <i>alternaria</i>					
1.	В-2	слабое	+	4	
2.	В-4	слабое	+	5	
3.	В-5	слабое	+	4	
4.	В-6	слабое	+/-	4	
5.	В-7	среднее	+	3	
6.	В-8	не заметно	+	4	
7.	В-10	не заметно	+	5	
8.	В-11	не заметно	+	3	
9.	В-14	сильное	+/-	5	
10.	В-15	среднее	+	3	
11.	В-16	не заметно	+	3	
12.	В-17	не заметно	+/-	5	
13.	В-18	среднее	+	3	
14.	В-20	не заметно	+	3	
15.	В-21	слабое	+	3	
16.	В-22	слабое	+	5	
17.	В-24	не заметно	+	4	
18.	В-25	не заметно	+	4	
Растения, зараженные бактериальной болезнью <i>erwinia</i>					
19.	Z-2	не заметно	+	4	
20.	Z-3	слабое	+/-	5	
21.	Z-4	не заметно	-	-	
22.	Z-5	не заметно	+	4	
23.	Z-7	не заметно	+/-	5	
24.	Z-8	не заметно	+	4	
25.	Z-9	не заметно	+	5	
26.	Z-10	сильное	+	5	
27.	Z-11	сильное	+	5	
28.	Z-12	сильное	+	5	
29.	Z-13	не заметно	+	4	
30.	Z-14	сильное	+	5	
31.	Z-16	не заметно	+	4	
32.	Z-17	не заметно	+	4	
33.	Z-18	не заметно	+	4	
34.	Z-19	не заметно	+	5	
35.	Z-23	сильное	+	4	
36.	Z-25	среднее	+	4	
Здоровые растения (контрольная группа)					
37.	К-1	не заметно	-/+	5	Слабая ошибка идентификации
38.	К-3	сильное	+/-	5	
39.	К-4	сильное	+	4	
40.	К-6	не заметно	-	-	
41.	К-7	среднее	+	7	
42.	К-10	не заметно	-/+	4	Слабая ошибка идентификации
43.	К-11	среднее в начале	-	-	
44.	К-16	не заметно	-/+	5	Слабая ошибка идентификации
45.	К-17	не заметно	-	-	
46.	К-18	не заметно	-	-	
47.	К-19	не заметно	-/+	5	Слабая ошибка идентификации
48.	К-20	не заметно	-	-	
49.	К-21	не заметно	+	4	Грубая ошибка идентификации
50.	К-22	не заметно	+	4	Грубая ошибка идентификации
51.	К-23	не заметно	+	4	Грубая ошибка идентификации
52.	К-25	не заметно	+	5	Грубая ошибка идентификации

Работа частично поддержана Белорусским республиканским фондом фундаментальных исследований, грант Т05МС-060.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Беляев Б.И., Катковский Л.В. Оптическое дистанционное зондирование – Мн.: БГУ, 2006. – 455 с.

2. Chao, K., Y.R. Chen, and M. S. Kim. Machine vision technology for agricultural applications // Elsevier science transactions on computers and electronics in agriculture, 2002. – vol. 36. – P. 173-191.
3. N. Kumar, S. Pandey, A. Bhattacharya, and P. S. Ahuja, "Do leaf surface characteristics affect agrobacterium infection in tea [camellia sinensis (L.) o kuntze]?" J. Biosci., vol. 29, no. 3, pp. 309-317, 2004.
4. P. Soille. Morphological image analysis applied to crop field mapping // Image and Vision Computing.-2000.-vol. 18, no. 13.-P. 1025-1032.
5. Panagiotis Tzionas, Stelios E. Papadakis, Dimitris Manolakis. Plant leaves classification based on morphological features and a fuzzy surface selection technique. // 5th Int. Conf. on Technology and Automation ICTA'05, 15-16 October 2005 Thessaloniki, Greece- 2005. – P. 365-370.
6. Margarita Torre, Petia Radeva. Agricultural-Field Extraction on Aerial Images by Region Competition Algorithm/flnt. Conf. on Pattern Recognition (ICPR'00), September 3-8, 2000, Barcelona, Spain-2000.-vol. 01, no. 1. – P. 1313-1316.
7. A.V. Inyutin. The algorithm of image segmentation by gray-scale pseudo-skeleton // Proc. of the III Int. Conf. on Neural Networks and Artificial Intelligence (ICNNAI2003), November 12-14, Minsk. Belarus. - 2003. – P.263-265.
8. Apan, Armando and Kelly, Rob and Jensen, Troy and Butler, David and Strong, Wayne and Basnet, Badri. Spectral Discrimination And Separability Analysis Of Agricultural Crops And Soil Attributes Using Aster Imagery // In 11th Australasian Remote Sensing and Photogrammetry Conference, 2-6 September, Brisbane, Queensland.- 2002. – P. 396-411.
9. Burks, T.F., S.A. Shearer, and F.A. Payne. Classification of weed species using color texture features and discriminant analysis // Transactions of ASAE.- 2000.- vol. 43(2). – P. 441-448.
10. Analysis of colour images of infected crop field. Part 2. Two-stage algorithm of segmentation and improvement of color images of infected crop field / Alexander A. Doudkin et. // Wybrane zagadnienia ekologiczne we wspolczesnym rolnictwie - PIMR, Poznan, 2005. – P. 118-122.
11. Выделение областей зараженности сельскохозяйственных полей по цветовым характеристикам изображений / М.Е. Ваткин, А.А.Дудкин, А.В.Иньютин и др. // 5-я международная конференция "Обработка информации и управление в чрезвычайных и экстраординарных ситуациях", Минск, Беларусь, 24-26 октября 2006 г. – Мн.: ОИПИ НАН Беларуси, 2006. - Т. 1. – С. 191-195.
12. Boleslaw Sobcowiak, Tadeusz Pawlowsky. Analiza obrazow kolorowych zaifekowanych pqluprawnych. Czesc 1. Przewadzenie testu I ocean wynipow badan // Wybrane zagadnienia ekologiczne we wspolczesnym rolnictwie. - PIMR, Poznan, 2005. – P. 113-117.

Статья поступила в редакцию 28.01.2007

УДК 004.8.032.26

Безобразов С.В., Головки В.А.

НЕЙРОСЕТЕВОЙ ПОДХОД ДЛЯ ФОРМИРОВАНИЯ ДЕТЕКТОРОВ В ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМАХ ДЛЯ ЗАЩИТЫ ИНФОРМАЦИИ

Введение

Традиционный подход в обнаружении компьютерных вирусов, основанный на сигнатурном поиске, имеет существенный недостаток. Сигнатурный поиск не способен обнаруживать неизвестные вирусы, а существующие эвристические алгоритмы далеки от совершенства. Поэтому для успешной борьбы с вредоносными программами необходимо постоянно пополнять антивирусные базы, которые, как правило, располагаются на web-сайте разработчика антивирусного программного обеспечения (ПО). На отслеживание и скачивание новых антивирусных баз тратится какое-то, иногда продолжительное, время. Компьютер с устаревшими антивирусными базами может оказаться бессильным перед угрозой заражения новым вирусом [1].

В силу сложившейся ситуации те, кто разрабатывают вирусы (вирусописатели), постоянно идут на шаг впереди разработчиков антивирусного ПО. Сначала появляется новый вирус. Через некоторое время этот вирус различными путями попадает к разработчикам антивирусного ПО. Затем специалисты анализируют вирус и включают его сигнатуру в антивирусную базу, и только после этого пользователь может скачать обновленную антивирусную базу.

Сегодняшние исследования в области защиты информации направлены на создание такой антивирусной системы, которая позволяла бы обнаруживать неизвестные вирусы. Такая система повысила бы уровень защиты компьютерных систем и избавила бы пользователей от неудобных операций. Практически во всех существующих антивирусных ПО реализован эвристический анализатор. Эвристический алгоритм анализирует набор команд проверяемого файла. Если команды файла предусматривают деструктивные функции, угрожающие со-

хранности и целостности данным, то такой файл считается вирусом. Однако существующие эвристические алгоритмы далеки от совершенства. Зачастую анализатор «пропускает» действительно вирус или принимает за вирус чистый файл.

Позаимствованная у природы и построенная по основным принципам биологической иммунной системы искусственная иммунная система позволяет обнаруживать не только известные ей вирусы, но и неизвестные, как это делает иммунная система человека, ежедневно сталкиваясь с большим количеством чужеродных бактерий и вирусов в организме [2]. Основными элементами искусственной иммунной системы (ИИС), которые несут функцию по обнаружению вирусов, являются детекторы (антитела). На стадиях генерации и отбора детекторы приобретают структуру, схожую с чистыми файлами (позитивная селекция), или различную со структурой чистых файлов (негативная селекция), что в дальнейшем позволяет им различать вирусы от незараженных, чистых файлов [3].

В данной статье представлен разработанный нами метод формирования детекторов на основе LVQ-сетей (нейронные сети для векторного квантования), который позволяет уменьшить временные и вычислительные затраты, связанные с проверкой файлов на наличие вирусов. Также этот метод сочетает в себе преимущества обоих методов (негативная и позитивная селекции) отбора нежелательных детекторов.

В первом разделе статьи представлен механизм обнаружения вирусов при помощи ИИС. Второй раздел содержит описание метода формирования детекторов на основе LVQ-сетей. В третьем разделе содержится описание экспериментальной системы обнаружения компьютерных вирусов. В четвертом разделе рассмотрены результаты исследований.

Безобразова Светлана Владимировна, аспирант кафедры интеллектуальных информационных технологий БрГТУ. Беларусь, Брестский государственный технический университет, 224017, Беларусь, г. Брест, ул. Московская, 267.