УДК 004.056

В. Г. Жуков, М. Н. Жукова, Н. А. Коромыслов

ПРИМЕНЕНИЕ НЕЧЕТКИХ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ В ЗАДАЧЕ ПОСТРОЕНИЯ АДАПТИВНЫХ САМООБУЧАЮЩИХСЯ СРЕДСТВ ЗАЩИТЫ ИНФОРМАЦИИ

Рассмотрено применение принципов функционирования механизмов иммунной системы в задаче построения адаптивного самообучающегося средства защиты информации и использования нечеткой логики для его улучшения.

Ключевые слова: искусственные иммунные системы, нечеткая логика, защита информации.

В современном мире информация играет роль самого востребованного ресурса, а ее потеря или нежелательная модификация может привести к значительному ущербу. Поэтому необходимо применять средства для защиты информации от преднамеренного или непреднамеренного воздействия.

Вопрос построения системы защиты – это ключевой вопрос обеспечения безопасности информации. Одной из основных проблем при проектировании средств защиты информации является проблема постоянного появления новых, ранее неизвестных угроз. Для эффективного противодействия подобным угрозам система защиты информации должна иметь собственные механизмы обучения и генерации новых правил информационной безопасности. В этом случае эффективным решением будет построение самообучающейся системы защиты информации, которая способна реагировать на возникновение новых угроз безопасности информации. Помимо самообучения, немаловажной характеристикой системы защиты информации является адаптивность, так как адаптивная система будет сохранять работоспособность при непредвиденных изменениях свойств управляемого объекта, целей управления или окружающей среды путем смены алгоритма своего функционирования, программы поведения или поиска оптимальных состояний.

Одним из перспективных подходов является разработка алгоритмического обеспечения систем защиты информации. Принцип работы этих систем основан на моделировании механизмов иммунной системы человека, которая обеспечивает защиту организма от разнообразных угроз и представляет собой сложную адаптивную структуру, эффективно использующую различные механизмы защиты. Основная задача иммунной системы заключается в распознавании молекулярных структур организма и классификации их на «своих» и «чужих». Выявленные чужеродные структуры, в том числе и не существующие в природе, например синтезированные в лаборатории, служат сигналом для активации защитного механизма соответствующего типа. Результатом распознавания является обучение и формирование памяти к данной угрозе. В зависимости от частоты и силы конкретной угрозы иммунная система непрерывно изменяется, теряя или усиливая память к разным угрозам, что обеспечивает эффективную защиту в условиях ограниченности ресурсов [1].

Принципы иммунной системы уже использовались в решении задач, связанных с информационной безопасностью, таких как обнаружение компьютерных вирусов [2], организация парольной защиты [3], мониторинг процессов в системе UNIX [4].

Разрабатываемое на базе принципов работы иммунной системы адаптивное самообучающееся средство защиты информации должно иметь следующие компоненты (рис. 1):

– модули-датчики, которые с определенной периодичностью собирают информацию о текущем состоянии системы;

– модуль выявления угроз, который при помощи аппарата искусственных иммунных систем определяет, являются ли события, полученные от модулей-датчиков, инцидентами информационной безопасности [5];

– модуль хранения данных, который содержит информацию о параметрах штатной работы автоматизированной системы, сведения об инцидентах информационной безопасности и хранит протокол журнала с записями о произошедших инцидентах;

– модуль реагирования, который инициирует ответное действие системы при обнаружении инцидента информационной безопасности. Ответное действие определяется в зависимости от политики безопасности и степени опасности инцидента. Это может быть оповещение об инциденте, блокирование части функций системы, прекращение работы системы и другие действия, а также их совокупность [6].

Предлагаемое адаптивное самообучающееся средство защиты информации начинает свою работу с генерации детекторов с тем, чтобы соответствие детекторов и сведений о штатной работе не превышало задаваемого входного значения (рис. 2).

В режиме функционирования средство собирает данные об автоматизированной системе и проверяет их на соответствие сгенерированным ранее детекторам (рис. 3). В случае если какой-либо детектор соответствует полученным данным, эти данные заносятся в журнал записей об инцидентах информационной безопасности и передаются в модуль реагирования.

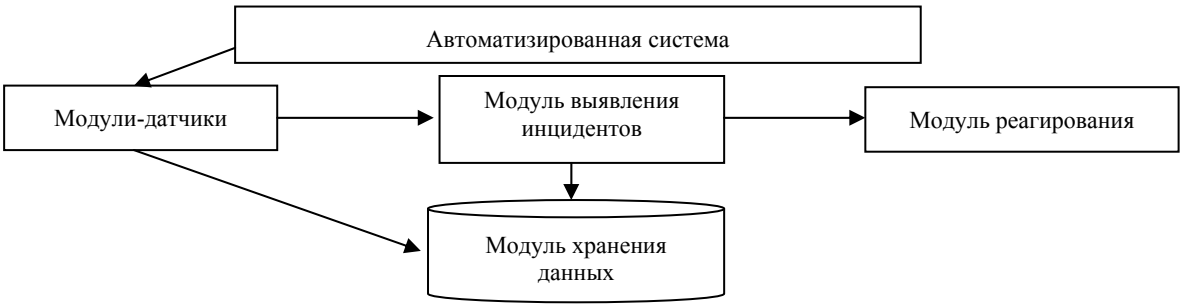

Рис. 1. Обобщенная архитектура адаптивного самообучающегося средства защиты информации

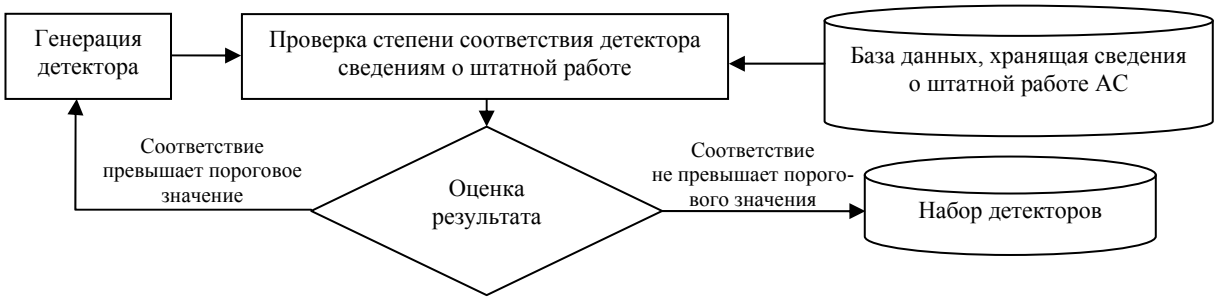

Рис. 2. Обобщенная схема процесса генерации детекторов

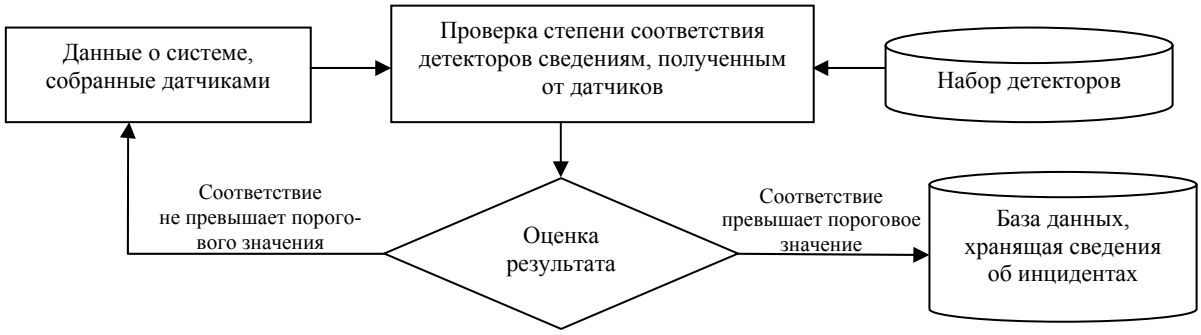

Рис. 3. Обобщенная схема функционирования средства защиты информации

Модуль выявления угроз, как правило, основан на алгоритме отрицательного отбора [2] или его модификациях. Этот алгоритм можно формализовать следующим образом.

Шаг 1. Определим «свое» как совокупность S строк длины l над конечным алфавитом, которую необходимо защищать или контролировать. В качестве S могут выступать программа, файл данных (любое программное обеспечение) или нормальная форма активности, подразделяемые на подстроки.

Шаг 2. Образуем набор детекторов R, каждый из которых не должен соответствовать любой строке из S, при этом вместо точного соответствия используем правило частичного соответствия: две строки соответствуют друг другу, если и только если они совпадают по крайней мере в r следующих друг за другом позициях, где r – целочисленный параметр.

Шаг 3. Проверим S на предмет изменений путем непрерывного сравнения детекторов из R с элементами S. Если хотя бы один из детекторов окажется соответствующим, то это значит, что произошло изменение, поскольку детекторы по определению отобраны так, чтобы не соответствовать любой строке из S (рис. 4).

Описанный алгоритм опирается на три важных принципа:

– каждый вариант алгоритма уникален;

– процесс выявления изменений имеет вероятностный характер;

– надежная система должна обнаруживать не только заранее известные варианты изменений, но и любую чужеродную активность.

Для оценки эффективности работы искусственной иммунной системы на базе алгоритма отрицательного отбора были сгенерированы тестовые данные.

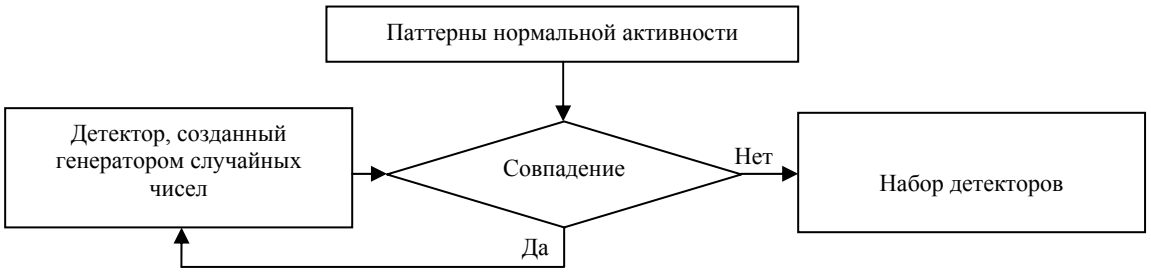

Рис. 4. Схема генерации детекторов в алгоритме отрицательного отбора

Таблица 1

Результаты тестирования алгоритма отрицательного отбора

Число детекторов	5 000	10 000	15 000	20 000	25 000	30 000	35 000	40 000	45 000	50 000
Количество обнаруженных изменений	16	30	41	49	59	61	76	77	81	86

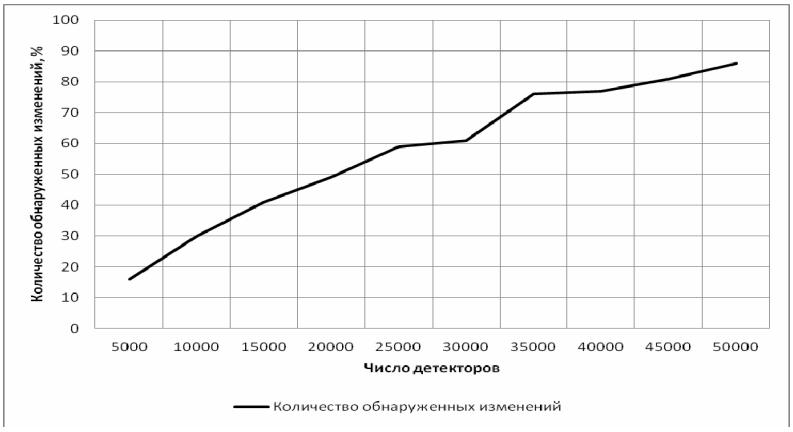

Рис. 5. Результаты тестирования алгоритма отрицательного отбора

В качестве тестовых данных применялись псевдослучайно сгенерированные строки со следующими параметрами:

– алфавит – 0…9;

– размер строки – 20 символов;

– размер частичного соответствия – 40 % от размера строки.

В качестве генератора псевдослучайных чисел использовался генератор на базе алгоритма Блюма–Блюма–Шуба.

Тестирование осуществлялось следующим образом:

– генерировалось множество *S*, мощность множества – 100 строк;

– генерировалось множество *R* различной мощности (табл. 1);

– в каждой строке из множества *S* значение каждого третьего символа менялось на случайно выбранное значение из заданного алфавита;

– подсчитывалось число строк из множества *S*, обнаруженных хотя бы одним детектором из множества *R* в соответствии с заданным размером частичного соответствия.

Результаты исследования эффективности работы искусственной иммунной системы на базе алгоритма отрицательного отбора, полученные путем усреднения результатов работы алгоритма по 100 запускам, приведены в табл. 1.

В классическом алгоритме отрицательного отбора детекторы генерируются случайным образом и зависимость числа обнаруженных изменений от числа сгенерированных детекторов является практически линейной (рис. 5). Это объясняется тем, что из-за большого числа возможных вариантов требуется увеличение количества детекторов для обеспечения заданного пользователем уровня надежности, выраженного в проценте успешно обнаруженных изменений, что в свою очередь приводит к большим затратам ресурсов системы.

Решение данной проблемы может быть найдено с помощью принципов нечеткой логики. Основным понятием теории нечеткой логики является понятие «мера близости к …», например «близко к 1» и т. д. Именно это понятие позволяет в дальнейшем работать с нечеткими, расплывчатыми значениями.

В рассматриваемой нами задаче меру близости предлагается использовать при работе с детекторами, при генерации детекторов и при проверке соответствия в качестве дополнительного средства для определения степени отличия детекторов. Настраиваемый параметр (мера близости) будет указывать на то, в каких пределах детекторы будут считаться схожими. В зависимости от свободных (или доступных) ресурсов система настраивается таким образом, чтобы два детектора (или детектор и данные о системе) считались схожими при различных отличиях. Если система обладает малыми ресурсами, то довольно большое число детекторов будет признаваться близкими друг к другу и отбрасываться при их генерации, что позволит уменьшить число детекторов. Это связано с тем, что чем больше детекторов можно сгенерировать, тем меньшие отличия будут признаваться серьезными. Таким образом, система становится гибкой и настраиваемой в зависимости от доступности ресурсов.

Приведем описание алгоритма отрицательного отбора с применением нечеткой логики.

Шаг 1. Определим «свое» как совокупность S строк длины l над конечным алфавитом, которую необходимо защищать или контролировать. В качестве S могут выступать программа, файл данных (любое программное обеспечение) или нормальная форма активности, подразделяемые на подстроки.

Шаг 2. Образуем набор детекторов R, используя механизмы нечеткой логики.

2.1. Введем функцию принадлежности, которая также является мерой близости детекторов.

2.2. Определим значения мер близости детекторов: малая, средняя, большая.

2.3. Зададим пороговое значение функции принадлежности для каждой меры близости детекторов.

2.4. При генерации детектора при поиске совпадений в r следующих друг за другом позициях вычисляем значение функции принадлежности. Значение функции принадлежности считается равным 1 в том случае, когда найдено r совпадений, где r – целочисленный параметр для каждой меры близости детекторов.

2.5. При значении функции принадлежности, превышающем заданное пороговое значение для данной меры близости, детектор принимается, при меньшем – отбрасывается.

Шаг 3. Проверим S на предмет изменений путем непрерывного сравнения детекторов из R с элементами S. При проверке будем использовать принципы нечеткой логики.

3.1. Введем функцию принадлежности, которая одновременно является мерой близости изменений.

3.2. Определим значения мер близости изменений: превышение размера частичного соответствия на величину менее заданного параметра h; превышение размера частичного соответствия на величину более заданного параметра h, где h – целочисленный параметр.

3.3. Зададим пороговое значение функции принадлежности для каждой меры близости.

3.4. При поиске изменений вычислим значение функции принадлежности.

3.5. При значении функции принадлежности, превышающем заданное пороговое значение для данной меры близости, детектором найдено изменение.

Для оценки эффективности работы искусственной иммунной системы на базе алгоритма отрицательного отбора с применением нечеткой логики была проведена генерация тестовых данных. В качестве тестовых данных использовались строки, псевдослучайно сгенерированные на базе алгоритма Блюма–Блюма–Шуба со следующими параметрами:

– алфавит – 0…9;

– размер строки – 20 символов;

– размер частичного соответствия – 40 % от размера строки.

Параметры нечеткой логики:

– форма представления термов – триангулярные нормы;

– граничные значения для мер близости детекторов: малая – от 15 до 20 % от размера строки; средняя – от 25 до 30 % от размера строки; большая – от 35 до 40 % от размера строки;

– параметр h – 5 % от размера строки;

– пороговые значения для функций принадлежности детекторов и изменений – 0,85 (85 %).

Тестирование осуществлялось следующим образом:

– генерировалось множество S, мощность множества – 100 строк;

– генерировалось множество R различной мощности (табл. 2);

– в каждой строке из множества S значение каждого третьего символа менялось на случайно выбранное значение из заданного алфавита;

– подсчитывалось число строк из множества S, обнаруженных хотя бы одним детектором из множества R.

Результаты исследования эффективности работы искусственной иммунной системы на базе алгоритма отрицательного отбора с применением нечеткой логики, полученные путем усреднения результатов работы алгоритма по 100 запускам, приведены в табл. 2.

Исследования проводились до достижения количества обнаруженных изменений во множестве S, сопоставимых с результатами исследования алгоритма отрицательного отбора (см. табл. 1). Применение нечеткой логики позволило уменьшить не только количество необходимых детекторов (на порядок) (рис. 6), но и количество требуемых ресурсов для работы алгоритма при практически том же проценте обнаруженных изменений.

Таблица 2

Результаты тестирования алгоритма отрицательного отбора с применением нечеткой логики

Число детекторов	500	1 000	1 500	2 000	2 500	3 000	3 500	4 000	4 500	5 000
Количество обнаруженных изменений	25	45	61	70	77	81	85	88	86	90

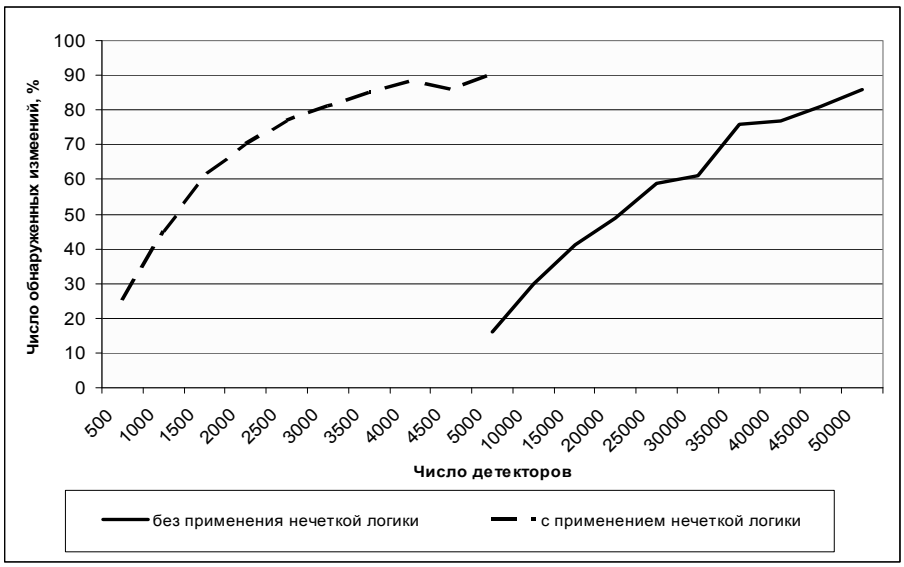

Рис. 6. Сравнительный анализ результатов тестирования

Однако чем бóльшие отличия будут считаться несущественным, тем будет больше ложных срабатываний. Приемлемый уровень ложных срабатываний определяется пользователем системы, так как алгоритмы такого типа обычно используются в системах превентивной защиты, которые применяются в качестве дополнительных методов защиты для систем, построенных на базе сигнатурных алгоритмов обнаружения уже известных угроз, с целью повышения уровня защищенности автоматизированных систем. Поэтому необходимо находить разумный компромисс между экономией ресурсов и точностью обнаружения. В данном случае нечеткая логика применяется для того, чтобы система становилась настраиваемой в зависимости от доступности ресурсов (например, объема используемой оперативной памяти), т. е. более адаптивной.

Для быстрого реагирования на угрозы, которые уже возникали в прошлом, система защиты должна содержать базу данных, хранящую сведения об инцидентах безопасности. В этом случае также целесообразно применение понятий нечеткой логики, особенно для описания угроз, похожих на те, записи о которых уже находятся в базе данных, что позволит ускорить распознавание таких угроз. Помимо скорости реагирования это будет полезно и для экономии ресурсов, так как требуется хранить меньше записей за счет того, что часть из них описывается как близкие к тем, которые уже находятся в базе.

В дальнейшем планируются исследования зависимости эффективности предлагаемого модифицированного алгоритма отрицательного отбора от настроек его параметров (длины проверяемых последовательностей и т. д.) и изучение влияния величины меры близости на процент ошибочных классификаций, а также введение адаптивно настраивающегося параметра «мера близости», который будет изменять свое значение в зависимости от частоты и уровня агрессивности внешних воздействий и режима функционирования системы защиты информации.

Таким образом, разработка адаптивных самообучающихся систем защиты информации позволит повысить эффективность решения задачи обнаружения угроз информационной безопасности и обеспечения заданного уровня безопасности информации, а применение нечеткой логики в дополнение к искусственным иммунным системам существенно улучшит создаваемое адаптивное самообучающееся средство защиты информации.

Библиографические ссылки

1. Искусственные иммунные системы и их применение / под ред. Д. Дасгупты ; пер. с англ. под ред. А. А. Романюхи. М. : Физматлит, 2006.

2. Self-Nonself Discrimination in a Computer / S. Forrest, A. S. Perelson, L. Allen, R. Cherukuri // Proc. of IEEE Symp. on Research in Security and Privacy. Los Alamitos, CA, 1994. P. 202–212.

3. Dasgupta D., Saha S. A Biologically Inspired Password Authentication System // ACM Proc. of 5th Cyber Security and Information Intelligence Research Workshop / Oak Ridge National Lab. Oak Ridge, Tenn, 2009.

4. A Sense of Self for Unix Processes / S. Forrest, S. A. Hofmeyr, A. Somayaji, T. A. Longstaff // Proc. IEEE Symp. on Security and Privacy. Los Alamitos, CA, 1996. P. 120–128.

5. Жуков В. Г., Жукова М. Н., Коромыслов Н. А. О применении искусственных иммунных систем в задаче обнаружения инцидентов информационной безопасности // Решетневские чтения : материалы XIV Междунар. науч. конф. / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2010. Ч. 2. С. 548–549.

6. Коромыслов Н. А. О применении принципов иммунных систем в задаче построения адаптивных самообучающихся систем защиты информации // Актуальные проблемы безопасности информационных технологий : сб. тр. VII науч.-техн. конф. / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2011.

V. G. Zhukov, M. N. Zhukova, N. A. Koromyslov

APPLICATION OF FUZZY ARTIFICIAL IMMUNE SYSTEMS IN THE TASK OF CONSTRUCTION OF ADAPTIVE SELF-LEARNING INFORMATION SECURITY SYSTEM

We consider the application of principles of immune system functioning in the task of construction of adaptive self-organizing security tools and application of fuzzy logic for its improvement.

Keywords: artificial immune systems, fuzzy logic, information security.

УДК 004.932

А. Г. Зотин, А. В. Носов, Д. В. Бузаев

АНАЛИЗ ПРИГОДНОСТИ МЕТОДОВ СЕГМЕНТАЦИИ ДЛЯ ЛОКАЛИЗАЦИИ ОБЪЕКТОВ НА ОСНОВЕ ЦВЕТОВЫХ И СТРУКТУРНЫХ ПРИЗНАКОВ

Рассмотрены существующие методы сегментации по цветовым и структурным признакам. Приведено описание алгоритмов функционирования наиболее широко известных методов сегментации на основе пороговой обработки, графов и каскадов. Представлены результаты анализа пригодности методов сегментации для локализации объектов на примере растений, лиц и кистей рук людей.

Ключевые слова: анализ изображений, сегментация, пороговая обработка, алгоритм Краскала, каскады Хаара.

Развитие компьютерной техники делает возможной сложную обработку видеоданных, приближенную к реальному времени. В обработке изображений при распознавании образов одной из самых важных задач является задача обнаружения и локализации объектов интереса, возникающих в таких сферах, как мониторинг, охрана, управление и т. д. Среди наиболее активно развивающихся направлений можно выделить распознавание лиц людей и их жестов, применяемое в сфере охраны и управления, мониторинг растений и природных ресурсов в сельском хозяйстве.

Существует множество методов сегментации [1]. К наиболее распространенным методам относятся модификации методов пороговой обработки, сегментация на основе теории графов и на основе классификаторов (каскадов). Рассмотрим их более подробно.

В большинстве случаев, когда объекты интереса имеют определенный цвет, для их локализации используется цветовая сегментация, суть которой заключается в выделении областей, имеющих цвет объекта интереса. Наиболее широко известны и легко реализуемы методы, основанные на пороговой обработке, результатом которых является матрица, размеры которой совпадают с размерами изображения [2].

Сегментацию по порогу можно проводить разными способами. Самым простым является сравнение каждого пикселя с некоторым значением (порогом):

$$M(x,y)=\begin{cases}1, \text{если } I(x,y)\geq \text{Threshold},\\ 0, \text{ если } I(x,y)<\text{Threshold},\end{cases}$$

где $I(x, y)$ – яркость изображения в точке (x, y); M – результирующая маска; Threshold – значение порога.

Другой способ заключается в поиске близких по цвету пикселей с эталонным значением. Для этого каждый пиксель изображения рассматривается в выбранных цветовых пространствах, после чего вычисляется расстояние между ним и эталоном и принимается решение о принадлежности пикселя объекту. Однако следует обратить внимание на то, что в зависимости от цветовой модели цвет объекта может иметь различные представления в различных цветовых пространствах. Качество сегментации зависит от того, в какой модели будет проводиться обработка изображений:

$$M(x,y)=\begin{cases}1, \text{если } d(M(x,y),\text{Etalon})\geq \text{Threshold},\\ 0, \text{ если } d(M(x,y),\text{Etalon})<\text{Threshold},\end{cases}$$

где $I(x, y)$ – значение изображения выбранного цветового пространства в точке (x, y); M – результирующая маска; $d(X, Y)$ – функция, вычисляющая расстояние между векторами X и Y; Etalon – значение координат эталонного цвета в рассматриваемом цветовом пространстве; Threshold – значение порога.

Для оценки расстояния между эталонным и проверяемым цветом используют метрики. *Метрика* – это функция, определяющая расстояния в метрическом пространстве, т. е. во множестве, в котором определено расстояние между любой парой элементов. Наиболее популярными метриками являются:

– евклидова метрика – обычное расстояние между двумя точками:

$$d(X,Y)=\sqrt{(X-Y)^T\cdot(X-Y)},$$

где X и Y – сравниваемые векторы;