

УДК 53.082.53

**ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ  
ВИРУСОВ И БАКТЕРИЙ В ОПТИЧЕСКОМ ДИАПАЗОНЕ ВОЛН****Белый Вячеслав Сергеевич,**

кандидат технических наук,

доцент кафедры физики имени В.А. Фабриканта ФГБОУ ВО «НИУ «МЭИ», г. Москва,

el.belaya2015@yandex.ru

**Аннотация**

В данной научной статье приводятся результаты теоретических исследований возможности распознавания микроскопических объектов в виде вирусов и бактерий, обнаруживаемых методами активной или полуактивной оптической локации на основе использования алгоритма AdaBoost, предложенного в 1996 году израильским учёным в области информатики Й. Фройндом и американским учёным-программистом Р. Шапиром. Данный алгоритм успешно применяется в настоящее время для решения широкого круга задач, связанных с распознаванием объектов по их изображениям. Приведенный в статье пример подтверждает возможность применения данного алгоритма для распознавания в оптическом диапазоне волн объектов наблюдения, имеющих микроскопические размеры.

**Ключевые слова:** оптическая локация, оптико-электронная система, микроскопические объекты наблюдения, вирусы и бактерии, функция-классификатор, распознавание, вектор признаков, метки классификации, алгоритм AdaBoost.

**THEORETICAL ASPECTS OF SOLVING THE PROBLEM OF RECOGNITION  
OF VIRUSES AND BACTERIA IN THE OPTICAL WAVELENGTH RANGE****Vyacheslav S. Belyi,**

Candidate of Technical Sciences,

Associate Professor of the Department of Physics named after V.A. Fabrikant of the National Research University «Moscow Power Engineering Institute», Moscow

**ABSTRACT**

This scientific article presents the results of theoretical studies of the possibility of recognizing microscopic objects in the form of viruses and bacteria detected by active or semi-active optical location methods based on the use of the AdaBoost algorithm proposed in 1996 by the Israeli computer scientist Y. Freund and the American computer scientist R. Shapir. This algorithm is currently being successfully used to solve a wide range of tasks related to the recognition of objects by their images. The example given in the article confirms the possibility of using this algorithm for recognition of observation objects with microscopic dimensions in the optical wavelength range.

**Keywords:** optical location, optoelectronic system, microscopic objects of observation, viruses and bacteria, classifier function, recognition, feature vector, classification labels, AdaBoost algorithm.

**Введение.** Известно [1, с. 20-28], что современные оптико-электронные средства не в полной мере позволяют обеспечить получение высокого качества изображения, необходимого для распознавания малоразмерных и особенно микроскопических объектов, находящихся в определенных фоновых условиях. Это обстоятельство имеет место как при установке оптики на движущемся носителе, так и в стационарных условиях, и имеет соответствующие причины [2, с. 582-591]. Вышеуказанные причины обусловлены наличием шумов в изображениях, отсутствием чётких алгоритмов распознавания, а также необходимостью учитывать ограниченность ресурсов при распознавании в реальном масштабе времени с учётом быстродействия системы.

Целью данной научной статьи является исследование теоретической возможности распознавания микроскопических объектов в виде вирусов и бактерий, обнаруживаемых методами активной или полуактивной оптической локации с использованием фотоприемников, имеющих электрический выход. Исследование теоретической возможности распознавания микроскопических объектов включает в себя научно обоснованный выбор соответствующих методики и алгоритма, реализуемых на основе современных аппаратно-программных средств, а также решение контрольного практического примера, подтверждающего работоспособность и результативность последних.

**Анализ состояния вопроса.** Практическая реализация функций распознавания микроскопических объектов наблюдения методами активной или полуактивной оптической локации требует научно обоснованного подхода по выбору методик и алгоритмов на основе современных аппаратно-программных средств. В настоящее время для решения поставленной задачи широко используются оптико-электронные системы (ОЭС), основу которых составляют матричные фотоприемники с электрическим выходом, посредством которых потребителю становится доступной практически любая информация об объекте наблюдения [3, с. 11-20].

Построение ОЭС для распознавания малоразмерных неподвижных или движущихся объектов, к которым имеют отношение вирусы и бактерии, основано на использовании комбинированных методов обработки отражённых от объектов наблюдения оптических сигналов. Комбинированные методы обработки оптических сигналов одновременно учитывают не только параметры микроскопического объекта, измеренные на основании анализа его изображения, но и параметры, оцениваемые с помощью измерительных датчиков ОЭС. Особый интерес представляет использование ОЭС, работающих в инфракрасной (ИК) области спектра электромагнитного излучения. Это обусловлено наличием так называемых «окон прозрачности» атмосферы в диапазонах длин волн (3,5...4) мкм, (4,5...5) мкм, (7,5...12) мкм.

В состав ОЭС, выполняющей функции обнаружения и распознавания вирусов и бактерий в обязательном порядке должны входить оптические и электрические устройства преобразования сигналов. Принцип функционирования ОЭС основан на приёме и преобразовании электромагнитных волн оптической области спектра в электрические сигналы с последующей их вторичной обработкой. Функцию приёма и преобразования оптических сигналов в виде световых пучков выполняет оптическая система. Задача преобразования электрических сигналов решается посредством применения

электрических схем. Структура ОЭС может подвергаться изменениям в зависимости от реализуемых в ней методов обработки оптико-локационной информации. На рисунке 1 представлена обобщенная структурная схема типовой ОЭС, предназначенной для решения задач обнаружения объектов наблюдения с последующим их распознаванием [4, с. 339-342], [5, с. 12-16].



Рисунок 1 – Обобщенная структурная схема ОЭС обнаружения и распознавания объектов наблюдения

При построении системы распознавания объектов в ИК спектре нужно определиться с выбором наиболее информативных признаков. Ввиду специфичности траекторий движения малоразмерных и микроскопических объектов не целесообразно использовать некоторые геометрические признаки: гистограммы распределения углов, хорд, длин сторон, а также их геометрические размеры. Признаками малоразмерного или микроскопического объекта наблюдения, полученными первоочередным образом в результате анализа его изображения, а также с помощью датчиков ОЭС, могут являться [6, с. 49], [7, с. 84-88]: уровень отражённого от объекта наблюдения сигнала, уровень фонового сигнала, азимут и угол места объекта наблюдения и время создания кадра изображения объекта наблюдения.

Для получения информации о типе объекта наблюдения и его определения используется такие физические величины как яркостная температура объекта наблюдения, характеризующая интенсивность отражённого сигнала. Для расчета яркостной температуры оценивается максимальный и минимальный уровень отражённых от объекта сигналов. Азимут и угол места положения объекта дают представление как о скорости движения объекта, так и о его расстоянии до ОЭС.

Обработка признаков в системах распознавания, как правило, ведется тремя способами – отбором наиболее информативных признаков, образование отношений отдельных признаков и образование линейных комбинаций отдельных признаков. На рисунке 2 представлена обобщенная структурная схема ОЭС с устройством распознавания оптических образов.



Рисунок 2 - Обобщенная структурная схема системы распознавания оптических образов

Оценка и расчет вторичных признаков микроскопического объекта позволяют получить более точный результат при отнесении такого объекта к вполне конкретному классу. Такими признаками являются [8, с. 15-18]: средняя яркость объекта наблюдения, изменение средней яркости объекта наблюдения, сумма уровней сигналов с вычетом и без вычета уровня фона, наименьший и наибольший эффективные размеры и азимутальная и угловая скорости движения объекта наблюдения.

Непосредственно для решения задачи распознавания малоразмерных образов и их классификации могут применяться как статистические, так и детерминированные методы. Однако, в настоящее время не существует комплекса методик распознавания микроскопических объектов наблюдения с высокой степенью достоверности. Поэтому в настоящей статье требуется решить задачу распознавания микроскопического объекта наблюдения для определенных случаев с достаточно высокой для практического использования вероятностью [9, с. 617-621]. В общем случае структура методики вторичной обработки информации в задаче распознавания микроскопических объектов имеет следующий вид, показанный на рисунке 3.

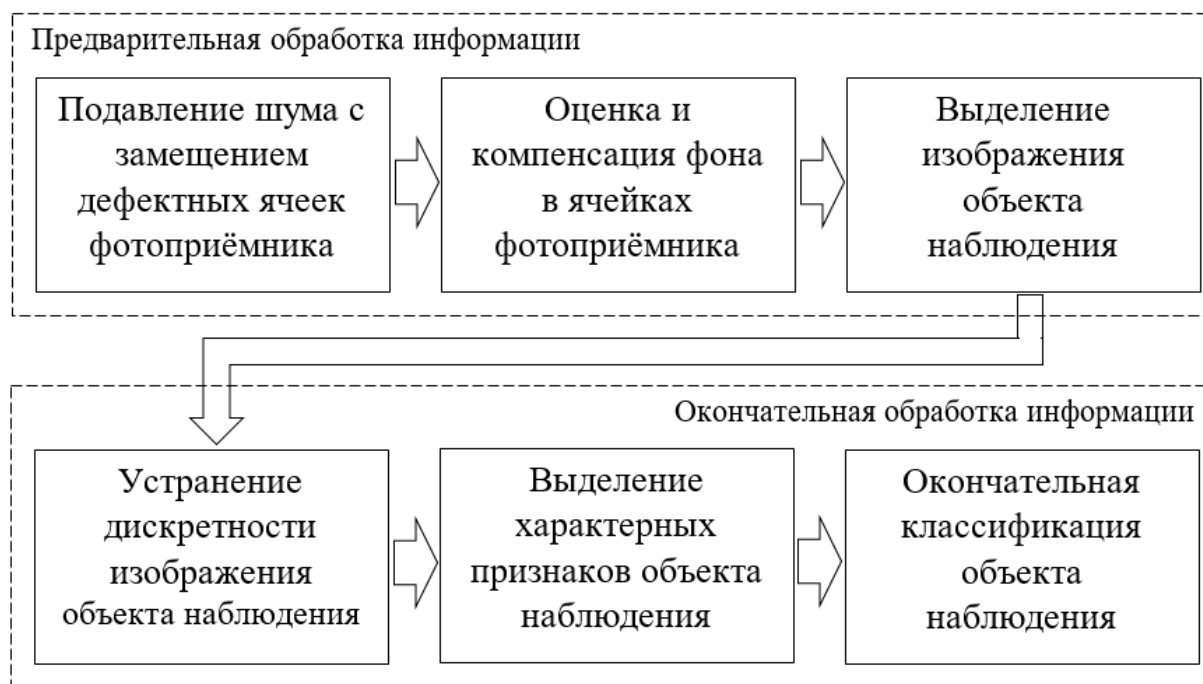


Рисунок 3 - Структура методики вторичной обработки информации для решения задачи распознавания объектов наблюдения

Существует множество алгоритмов распознавания объектов наблюдения по их изображению. Выбор конкретных алгоритмов обусловлен особенностями объектов наблюдения, которые требуется распознать. Иногда задача распознавания формально не ставится, поскольку свойства искомого объекта задаются без строгих математических параметров. Для решения этой задачи необходимо сформулировать такие свойства объектов наблюдения, а также разработать такой метод обнаружения последних, которые хотя бы в общем виде могли соответствовать заданным параметрам. В таблице 1 приведены широко используемые на сегодняшний день алгоритмы распознавания объектов наблюдения с краткой их характеристикой.

Таблица 1

Краткая характеристика алгоритмов распознавания

№	Алгоритм	Сущность	Достоинства, недостатки
1	На основе примитивов Хаара	Сравнение по откликам яркости и контраста двух прямоугольных областей изображения	Независимость откликов от масштабирования яркости изображений
2	Хафа	Обнаружение линий на изображении	Простота реализации при наличии малого числа линий
3	Сравнения с шаблоном	Сравнение полученного изображения с имеющимся шаблоном	Простота реализации, низкая точность
4	AdaBoost	Комбинирование простых классификаторов в более сложные	Высокое качество распознавания, большие вычислительные затраты

Из анализа представленных в таблице 1 данных следует, что при решении задачи распознавания необходимо найти оптимальное соотношение между сложностью вычислений и желаемой точностью результата.

Постановка задачи. Для решения задачи распознавания микроскопического объекта необходимо найти, обобщить и математически сформулировать результаты эмпирических наблюдений. Другими словами, требуется формализовать параметры микроскопического объекта наблюдения. Главная трудность состоит в том, что описать все свойства последних практически невозможно и данные свойства могут соответствовать не всем объектам наблюдения из выборочной совокупности. Поэтому, в процессе математической формализации следует допускать упрощения, которые, в конечном итоге, влияют на снижение качества решения задачи и ухудшение точности получаемого результата.

Формально, задача распознавания микроскопического объекта, представляющего собой вирус или бактерию, и имеющего  $n$  признаков (меток) классификации, сводится к следующему. Пусть некоторому  $n$ -мерному вектору признаков  $X = [x_1, x_2, \dots, x_n]$  объекта наблюдения ставится в соответствие такой же размерности вектор меток классов  $Y = [y_1, y_2, \dots, y_n]$ . Пусть также имеется в наличии некоторый набор простых классифицирующих функций  $h_k$  из множества функций  $H$

$$H: X \rightarrow Y, \quad (1)$$

позволяющий однозначно отобразить переход ряда отдельных признаков объекта наблюдения  $x_k \in X$  в соответствующие им метки классификации  $y_k \in Y, k \in K, k < n$ .

На практике объект наблюдения методами активной или полуактивной оптической локализации может быть охарактеризован количеством признаков  $n$  значительно превышающим количество  $k$  имеющихся в наличии простых классифицирующих функций  $h_k$ , поэтому задача распознавания объекта наблюдения сводится к определению явного вида функций классификации  $f_i$  из множества функций  $F$

$$F: X \rightarrow Y, \quad (2)$$

позволяющих максимально полно отобразить переход каждого признака  $x_i \in X$  в соответствующую ему метку классификации  $y_i \in Y, i \in \overline{1, n}$ .

Существует подход к решению задачи распознавания объектов наблюдения, позволяющий классифицировать даже микроскопический объект наблюдения, основанный на усилении простых классификаторов. Подход основан на комбинировании нескольких простых («слабых») классификаторов в один более сложный («сильный»). Качество решения задачи классификации объекта наблюдения при использовании данного подхода будет определяться показателями эффективности такого классификатора. На основании анализа данных, приведённых в таблице 1, видно, что для решения задачи распознавания микроскопического объекта наиболее пригоден алгоритм AdaBoost, применяемый для задачи построения бинарного классификатора, обеспечивающий комбинирование простых классификаторов в более сложные и гарантирующий высокое качество распознавания. Большие вычислительные затраты следует минимизировать путём рационального выбора количества классифицирующих признаков объекта наблюдения.

Алгоритм распознавания для решения задачи построения бинарного классификатора носит циклический характер. На первой итерации цикла алгоритма определяется первое значение функции классификации  $f_1$ .

1. Сначала, на основании известного значения размерности  $n$  векторов  $X$  и  $Y$ , задаётся «вес» первого шага с помощью выражения вида

$$D_1 = \frac{1}{n}. \quad (3)$$

2. По вычисленному с помощью выражения (3) значению  $D_1$  определяется величина взвешенной ошибки классификации

$$\varepsilon_1 = \frac{1}{2} \cdot D_1 \cdot (h_1 - y_1) \quad (4)$$

с последующим определением значения функции классификатора

$$h_1 = \arg \left\{ \min_{h_1 \in H} \varepsilon_1 \right\}. \quad (5)$$

3. Если выполняется условие вида  $\varepsilon_1 \geq 0,5$ , то никаких дальнейших вычислений не производится. В противном случае, сначала определяется значение коэффициента

$$a_1 = \frac{1}{2} \cdot \ln \left( \frac{1-\varepsilon_1}{\varepsilon_1} \right) \quad (6)$$

и рассчитывается значение функции

$$f_1 = a_1 \cdot h_1. \quad (7)$$

На второй итерации цикла определяется второе значение функции классификации  $f_2$ .

1. Сначала формируется второе значение весового распределения  $D_2$  с помощью выражения

$$D_2 = \frac{D_1}{Z_1} \cdot e^{-y_1 \cdot f_1}. \quad (8)$$

В выражении (8) число  $Z_1$  – нормировочное число, которое выбирается из соображения обеспечения выполнения условия  $\sum_n D_i = 1$ . Значение  $Z_1$  определяется с помощью выражения вида

$$Z_1 = D_1 \cdot e^{-y_1 \cdot f_1}. \quad (9)$$

2. Затем определяется величина взвешенной ошибки

$$\varepsilon_2 = \frac{1}{2} \cdot \sum_{i=1}^2 D_i \cdot (h_i - y_i) \quad (10)$$

и значение классификатора

$$h_2 = \arg \left\{ \min_{h_i \in H} \varepsilon_i \right\}, i \in \overline{1,2}. \quad (11)$$

3. По аналогии с (6) и (7) определяются значения коэффициента  $a_2$  и функции  $f_2$ .

При увеличении числа итераций искомая функция классификации  $F(X)$ , вид которой во многом определяется видами имеющихся в наличии простых функций  $h_k(x_k)$ , усложняется и становится наиболее приближённой к так называемому «сильному» классификатору. «Сильный» классификатор позволяет однозначно с высокой степенью достоверности решить задачу распознавания объектов наблюдения, практически исключая ошибку.

Нормировочное значение  $Z$  на очередной  $n$ -й итерации цикла определяется с помощью выражения

$$Z_n = \sum_{i=1}^n D_n \cdot e^{-y_n \cdot f_n}. \quad (12)$$

Искомую функцию классификации, зависящую от имеющихся в наличии простых функций  $h_k$ , можно представить как кусочно-постоянную функцию следующего вида

$$F(X) = \text{sign}[\sum_{k=1}^K a_k \cdot h_k]. \quad (13)$$

При условии равенства  $K = n$  выражение (13) приобретает вид

$$F(X) = \text{sign}[\sum_{i=1}^n a_i \cdot h_i]. \quad (14)$$

К достоинствам вышеприведенного алгоритма, согласно [], относится относительно высокая скорость его работы и высокая эффективность его применения, где в качестве показателей эффективности могут служить точностные характеристики процесса распознавания. С целью парирования связанного в большими вычислительными затратами недостатка целесообразно использование в составе устройства вторичной обработки информации ОЭС вычислительного устройства на базе микропроцессора.

Приведём практический пример применения алгоритма AdaBoost для решения задачи распознавания микроскопического объекта в виде вируса, имеющего  $n = 7$  характерных признаков. Пусть в устройстве распознавания предусмотрено использование

трёх простых («слабых») классификаторов. Исходные данные для решения задачи представлены в таблице 2.

Таблица 2

Исходные данные

Номера признаков	Результаты работы классификаторов			Метки классов
	№1	№2	№3	
1	1	-1	-1	1
2	-1	1	-1	1
3	-1	-1	1	1
4	-1	1	1	1
5	1	-1	1	1
6	1	1	-1	1
7	1	1	1	1

Ограничившись тремя итерациями, как рекомендовано в [10] и [11], получаются следующие результаты расчётов (см. таблицу 3).

Таблица 3

Результаты расчётов

Номера признаков	Итерации			$\sum_{i=1}^3 a_i \cdot h_i$
	№1	№2	№3	
«Вес»	$D_1$	$D_2$	$D_3$	--
1	0,14	0,13	0,15	-0,23
2	0,14	0,17	0,14	-0,18
3	0,14	0,17	0,2	-0,11
4	0,14	0,17	0,14	0,23
5	0,14	0,13	0,15	0,18
6	0,14	0,13	0,11	0,11
7	0,14	0,13	0,11	0,51
Ошибки классификаторов	$\varepsilon_1 = 0,43$	$\varepsilon_1 = 0,5$	$\varepsilon_1 = 0,49$	--
	$\varepsilon_2 = 0,43$	$\varepsilon_2 = 0,42$	$\varepsilon_2 = 0,2$	
	$\varepsilon_3 = 0,43$	$\varepsilon_3 = 0,42$	$\varepsilon_3 = 0,4$	
Кэф-ты	$a_1 = 0,14$	$a_2 = 0,17$	$a_3 = 0,2$	

**Выводы:**

В результате исследования теоретической возможности распознавания микроскопических объектов в виде вирусов и бактерий, обнаруживаемых методами активной или полуактивной оптической локации с использованием фотоприемников, имеющих электрический выход, в данной научной статье осуществлён обоснованный выбор соответствующих методики и алгоритма, реализуемых на основе современных аппаратно-программных средств.

Работа методики и алгоритма показаны на конкретном практическом примере с получением результата, показывающего возможность распознавания объекта наблюдения по наибольшему, отвечающим свойству  $\sum_{i=1}^3 a_i \cdot h_i > 0$ , количеству классификационных признаков последнего.

**Список литературы:**

1. Алгоритмы распознавания объектов / А. А. Цветков, Д. К. Шорох, М. Г. Зубарева [и др.]. – Текст: непосредственный // Технические науки: проблемы и перспективы: материалы IV Международной научной конференции (г. Санкт-Петербург, июль 2016 г.). – Санкт-Петербург: Свое издательство, 2016. – С. 20-28. – URL: <https://moluch.ru/conf/tech/archive/166/10825/> (Дата обращения: 24.05.2022).
2. Белый, В.С. Исследование процессов прохождения видимого излучения через оптическую систему на базе тонкой собирающей линзы / В. С. Белый, С. С. Воробьев // Вопросы устойчивого развития общества. – 2022. – № 3. – С. 582-591;
3. Березин, В.В. Обнаружение и оценивание координат изображений точечных объектов в задачах астронавигации и адаптивной оптики / В. В. Березин, А. К. Цыцулин // Вестник Тихоокеанского государственного университета. – 2008. – № 1(8). – С. 11-20;
4. Гусев, М.В. Микробиология: Учебник для студентов биологической специальности вузов / М.В. Гусев, Л.А. Минеева. – 4-е изд., стер. – М.: Издательский центр «Академия», 2003. – 464 с.;
5. Новикова, В.А. Рассеяние света и его применение в волоконной оптике / В. А. Новикова, С. В. Варжель; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. – С-Пб: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2019. – 39 с. – EDN AFAXZN;
6. Фираго, В.А. Детектирование слабых световых потоков: Методические указания к лабораторным работам для студентов специальностей 1-31 04 02 «Квантовая радиофизика и лазерные системы» – Мн.: БГУ, 2007. – 56 с.;
7. Шарнин, Л.М. Распознавание малоразмерных объектов инфракрасными системами с матричным фотоприемником / Л. М. Шарнин, А. В. Теплов // Вестник Казанского государственного технического университета им. А.Н. Туполева. – 2014. – № 1. – С. 84-88.
8. Avinash, K. AdaBoost for Learning Binary and Multiclass Discriminations. Purdue University – 2020. – P. 15-18.
9. Parfenov, V. I. Optical signal detection according to the photoelectron flow with unknown density waveform / V. I. Parfenov, V. S. Kirillov // Computer Optics. – 2012. – Vol. 36. – №4. – P. 617-621.
10. Машинное обучение. Взвешенное голосование AdaBoost. Часть 1 [Видеозапись] // YouTube. Режим доступа: [https://www.youtube.com/watch?v=CW6yjsSH\\_AA](https://www.youtube.com/watch?v=CW6yjsSH_AA).
11. Машинное обучение. Взвешенное голосование AdaBoost. Часть 2 [Видеозапись] // YouTube. Режим доступа: <https://www.youtube.com/watch?v=JoF1mfufBok>.

**References:**

1. Algorithms for object recognition / A. A. Tsvetkov, D. K. Shorokh, M. G. Zubareva [et al.]. – Text: direct // Technical sciences: problems and prospects: materials of the IV International Scientific Conference (St. Petersburg, July 2016). – St. Petersburg: Its

- publishing house, 2016. – PP. 20-28. – URL: <https://moluch.ru/conf/tech/archive/166/10825/> (Accessed: 05/24/2022).
2. Bely, V.S. Investigation of the processes of visible radiation passing through an optical system based on a thin collecting lens / V. S. Bely, S. S. Vorobyov // Issues of sustainable development of society. – 2022. – No. 3. – pp. 582-591;
  3. Berezin, V.V. Detection and estimation of coordinates of images of point objects in problems of astronavigation and adaptive optics / V. V. Berezin, A. K. Tsitsulin // Bulletin of the Pacific State University. – 2008. – № 1(8). – Pp. 11-20;
  4. Gusev, M.V. Microbiology: Textbook for students of biological specialties of universities / M.V. Gusev, L.A. Mineeva. – 4th ed., ster. - M.: Publishing Center "Academy", 2003. – 464 p.;
  5. Novikova, V.A. Light scattering and its application in fiber V. A. Novikova, S. V. Varzhel; St. Petersburg National Research University of Information Technologies, Mechanics and Optics. – S-Pb: Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, 2019. – 39 p. – EDN AFAXZN;
  6. Firago, V.A. Detection of weak light streams: Guidelines for laboratory work for students of specialties 1-31 04 02 "Quantum radiophysics and laser systems" – Mn.: BSU, 2007. – 56 p.;
  7. Sharnin, L.M. Recognition of small-sized objects by infrared systems with a matrix photodetector / L. M. Sharnin, A. V. Teplov // Bulletin of Kazan State Technical University named after A.N. Tupolev. – 2014. – No. 1. – pp. 84-88.
  8. Avinash, K. AdaBoost for Learning Binary and Multiclass Discriminations. Purdue University – 2020. – P. 15-18.
  9. Parfenov, V. I. Optical signal detection according to the photoelectron flow with unknown density waveform / V. I. Parfenov, V. S. Kirillov // Computer Optics. – 2012. – Vol. 36. – No. 4. – P. 617-621.
  10. Machine learning. Weighted AdaBoost voting. Part 1 [Video Recording] // YouTube. Access mode: [https://www.youtube.com/watch?v=CW6yjsSH\\_AA](https://www.youtube.com/watch?v=CW6yjsSH_AA) .
  11. Machine learning. Weighted AdaBoost voting. Part 2 [Video Recording] // YouTube. Access mode: <https://www.youtube.com/watch?v=JoF1mfufBok> .