

# Инспирированная механизмами зрительной системы человека идентификация объектов на изображениях, формируемых в режиме счета фотонов

В.Е. Анциперов<sup>1</sup>

<sup>1</sup>Институт радиотехники и электроники (ИРЭ) им. В.А. Котельникова РАН, Моховая 11, корп.7, Москва, Россия, 125009

**Аннотация.** В докладе излагаются результаты синтеза нового метода идентификации / распознавания объектов по реализациям случайных точек – фотоотчетов датчиков формирования изображений в режиме счета фотонов. Синтез существенно использует методы, применяемые в сфере машинного обучения, синтезированные алгоритмы идентификации принципиально ориентированы на современные вычислительные системы. Приведены результаты численного моделирования.

## 1. Введение

Современные системы формирования (цифровых) изображений помимо светочувствительных элементов содержат микроконтроллеры, микропроцессоры, специализированные чипы. Последние, помимо функций управления процессом регистрации падающего на светочувствительные матрицы излучения, хранения результатов регистрации, регулировки тона, баланса белого цвета и т.д., предоставляют большие возможности для дополнительной, более интеллектуальной обработки регистрируемых изображений [1].

Под интеллектуальной обработкой в первую очередь подразумевается анализ изображений с целью извлечения важной для тех или иных приложений информации. Примерами таких приложений являются компьютерное зрение промышленных роботов, автоматическое распознавание лиц или почерка, навигация полуавтономных и беспилотных автомобилей и т.д. Теоретически, большое разнообразие задач предполагает большое количество методов анализа изображений, Практика, однако, показывает, что большинство методов анализа с необходимостью включают ряд универсальных элементов. Среди таких элементов в первую очередь следует отметить методы распознавания объектов. Обычно под распознаванием подразумевается локализация объектов на изображениях и / или их классификация / идентификация [2].

Современные методы распознавания характеризуются тем, что они во все большей степени включают либо отдельные механизмы зрительной системы человека, либо пытаются имитировать ее структуру [3]. К первой группе методов можно отнести те, что используют различные механизмы предварительной обработки изображений, представленные в периферии зрительной системы – в сетчатке. Сюда относятся методы сжатия данных, уменьшения шума, усиления контрастности, подчеркивания краев и т.д. Другая группа более интеллектуальных методов пытается имитировать нейробиологические механизмы обработки визуальной информации в коре головного мозга [4]. Здесь следует отметить в первую очередь методы

машинного обучения: методы сокращения размерности (subspace methods), адаптивное обучение (slow feature learning), анализ иерархических зависимостей (hierarchical feature analysis) и т.д. Отметим, что на этом пути достигнуты заметные результаты [5].

В отношении вышеизложенного следует отметить следующее. Известные методы распознавания реализуют, имитируют механизмы зрительного восприятия с той или иной степенью приближения. Однако, данные регистрации, для обработки которых предназначаются эти методы, имеют, как правило, различное представление в зрительной системе и в искусственных системах формирования изображений. Действительно, в микропроцессорах большинства устройств регистрируемое излучение представлено оцифрованной интенсивностью на дискретных элементах (пикселах) светочувствительной матрицы. Значения интенсивности приближенно пропорциональны суммарной энергии накопленных пикселем за интервал времени формирования кадра (frame) фотонов и выражается изменяющимся в широких пределах целым числом. Наоборот, в сетчатке глаза клетки-фоторецепторы (палочки, колбочки) реагируют скорее на отдельные фотоны излучения и передают сигнал о их регистрации далее в ганглиозные клетки без накопления. В результате формируется несинхронизированный поток нервных импульсов, обрабатываемый зрительной системой в непрерывном режиме [6].

Отмеченное различие в представлении данных искусственными и естественными системами объясняется тем, что до недавнего времени имелись ограниченные технологические возможности реализации устройств формирования изображений подобной зрительной системе. Однако, исследования и разработки в течение последних десятилетий привели к известному прогрессу в области разработки датчиков изображений (не только видимого диапазона), открывающему принципиально новые возможности в этой области. Главная тенденция этого прогресса, проявившаяся уже в конце 1960-х годов при создании приборов с зарядовой связью (CCD) и укоренившаяся с изобретением в начале 1990-х годов светочувствительных матриц на комплементарных МОП структурах (CMOS), связана со стремлением к уменьшению размеров светочувствительных элементов – пикселей (в том числе ниже дифракционного предела). Помимо повышения пространственного разрешения, отмеченная тенденция обеспечивает улучшение и других характеристик регистрации – увеличение частоты кадров, повышение динамического диапазона, миниатюризацию датчиков, уменьшение потребления энергии и т.д. Отметим, что по мере уменьшения размеров пикселей, характер регистрации излучения приобретает все более выраженный квантовый характер и в пределе переходит в регистрацию отдельных фотонов (фотоэлектронов). Замечательным обстоятельством является то, что этот предельный случай на сегодняшний день уже достигнут рядом технологий. Речь идет о технологиях изготовления так называемых фотоотсчётных датчиков (Photon-Counting Image Sensors) [7] – датчиков работающих в режиме счета фотонов. В качестве примера укажем на электронно-умножительные матрицы с зарядовой связью (EMCCD) [8], однофотонные лавинные диоды (SPAD) [9], лавинные фотодиоды в режиме счётчика Гейгера (GMAPD) [10]. Важной особенностью фотоотсчётных датчиков, с точки зрения формируемых ими изображений [11], является то, что здесь исчезает необходимость в длительной экспозиции – изображение не создается в обычном смысле, а вместо него формируется непрерывный во времени поток фотоэлектронов с произвольным доступом к данным [12]. Это обстоятельство существенно видоизменяет основные концепции в теории и практике обработки изображений. Так, например, если в традиционных задачах компьютерного зрения цикл идентификации происходит поэтапно – сначала накапливаются фотоотсчеты, затем по ним реконструируется изображение, а уже затем происходит его анализ, то при использовании фотоотсчётных датчиков процессы регистрации и обработки данных можно совместить – сделать процедуру идентификации инкрементальной, протекающей в синхронном с регистрацией отсчетов режиме. В рамках подобной концепции отпадает необходимость в реконструкции полного изображения и возникает новая парадигма “видения без изображения” [12].

В данной работе приводятся результаты исследования в рамках упомянутой концепции задач идентификации (классификации) регистрируемых объектов по данным потоков фотоотсчетов их излучений. Преднамеренно используя в качестве данных координаты фотоотсчетов, а не

реставрированное по ним изображение, исходная задача формализуется до проблемы наилучшей подгонки регистрируемых данных к одному из статистических описаний зарегистрированных ранее процессов, называемых прецедентами [13]. В работе показано, что предложенный подход может быть доведен до уровня алгоритмической реализации. Синтезированный алгоритм идентификации, по структуре близкий к алгоритмам известного ЕМ-семейства, представлен ниже.

## 2. Статистическое описание данных регистрации в режиме счета фотонов

Для того, чтобы предлагаемый подход к идентификации объектов был бы статистически корректно обоснован, требуется формализовать процесс регистрации фотоотсчётными датчиками падающего на них излучения от объектов. Для этого необходимо подобрать такую математическую модель, которая бы адекватно с физической (квантовой) точки зрения связывала бы интенсивность падающего на чувствительную поверхность  $\Omega$  излучения  $I(\vec{x})$  и последовательность регистрируемых при этом координат фотоотсчетов  $\{\vec{x}_1, \dots, \vec{x}_n\}$ . Здесь вектор  $\vec{x} \in \Omega$  обозначает координаты точки в некоторой системе координат в плоскости регистрации,  $\vec{x}_i$  – координаты пикселей, соответствующих фотоотчётам.

В идеальной ситуации, когда размеры пикселей считаются сколь угодно малыми – в так называемой непрерывной пространственной модели фотодетектирования [14] координаты отсчетов  $\{\vec{x}_i\}$  являются набором случайных векторов (в случайном, но конечном количестве  $n$ ) даже при заданной, неслучайной интенсивности  $I(\vec{x})$ . Эффект случайности в данных часто трактуется как квантовые шумы регистрации (quantum / shot noise). Поскольку в оптической области спектра доминирующим источником шума почти всегда являются квантовые шумы [14], в данной работе мы не будем структурировать интенсивность  $I(\vec{x})$  как интенсивность излучения от объекта и аддитивный фон / шум, включая при необходимости последний в описание объекта.

Наиболее общей математической моделью регистрации фотоотсчётными датчиками излучения являются случайные точечные процессы [15]. Из полуклассической теории фотодетектирования известно, что подходящим классом точечных процессов, наиболее адекватно описывающих физические механизмы генерации фотоотсчетов [16], являются пуассоновские точечные процессы (ППП) [15]. Для ППП полное статистическое описание задается единственной функцией – интенсивностью отсчетов  $\lambda(\vec{x})$  в том смысле, что через нее выражаются все конечномерные распределения вероятностей координат отсчетов [15]:

$$\rho(n, \vec{x}_1, \dots, \vec{x}_n) = \frac{1}{n!} \prod_{i=1}^n \lambda(\vec{x}_i) \times \exp\left\{-\int_{\Omega} \lambda(\vec{x}) d\vec{x}\right\}. \quad (1)$$

Квантовая теория уточняет связь интенсивности отсчетов ППП  $\lambda(\vec{x})$  с интенсивностью падающего излучения  $I(\vec{x})$ . Эта связь задается простым линейным соотношением  $\lambda(\vec{x}) = \eta T I(\vec{x}) / h\bar{\nu}$ , где  $\eta$  – коэффициент квантовой эффективности материала датчика,  $T$  – время считывания кадра,  $\bar{\nu}$  – центральная частота падающего излучения,  $h$  – постоянная Планка [14].

Зависимость  $I(\vec{x}) \rightarrow \lambda(\vec{x})$  вместе с распределениями (1) задают математическую модель связи вход  $I(\vec{x})$  – выход  $\{\vec{x}_i\}$  идеального (непрерывного) датчика. Для реальных датчиков с пикселями конечных размеров соответствующая связь получается интегрированием (1) по поверхности пикселей с учетом чисел регистрируемых каждым из них фотоотсчетов. Модель реального датчика будет близка идеальной модели (1) в тех случаях, когда вероятности двух и более фотоотсчетов пикселя малы по сравнению с вероятностью единственного фотоотсчета. Легко показать, что это условие сводится к оценке  $\lambda(\vec{x})S = \eta T S I(\vec{x}) / h\bar{\nu} \ll 1$ , где  $S$  – площадь отдельного пикселя. Таким образом, при регистрации слабых излучений  $I(\vec{x}) \rightarrow 0$ , при очень быстрой съемке  $T \rightarrow 0$ , или в случае описанных во введении фотоотсчётных датчиков  $S \rightarrow 0$  распределения (1) могут служить адекватной моделью процесса регистрации фотоотсчётов.

В направлении формализации исходной задачи можно сделать еще один шаг, если расширить ее постановку за счет уточнения понятия идентификации объектов. Именно, предполагая идентификацию объектов только по форме их интенсивности вне зависимости от их полной яркости (см. обсуждение этой проблемы в [17]), можно в качестве основы синтеза вместо

распределений (1) выбрать условные распределения координат отсчетов  $\{\vec{x}_1, \dots, \vec{x}_n\}$  при заданном их полном количестве  $n$ :

$$\begin{aligned} \rho(\vec{x}_1, \dots, \vec{x}_n | n) &= \prod_{i=1}^n \rho(\vec{x}_i), \\ \rho(\vec{x}) &= \frac{\lambda(\vec{x})}{\int_{\Omega} \lambda(\vec{x}) d\vec{x}} = \frac{I(\vec{x})}{\int_{\Omega} I(\vec{x}) d\vec{x}}. \end{aligned} \quad (2)$$

Распределения (2) выражают известное свойство ПТП, состоящее в том, что условное, при фиксированном  $n$ , совместное распределение координат отсчетов задается произведением одинаковых распределений координат каждого из них [15]. Другими словами, при заданном  $n$ , отсчеты  $\{\vec{x}_i\}$  представляют собой выборку независимых одинаково распределенных случайных векторов. Более того, при этом распределение выборочных данных  $\rho(\vec{x})$  (2) совпадает с (нормированной) интенсивностью  $I(\vec{x})/\int(\vec{x})d\vec{x}$ , т.е. не зависит ни от материала детектора  $\eta$ , ни от несущей частоты излучения  $\bar{\nu}$ , ни от времени регистрации  $T$ . Ввиду наличия перечисленных свойств распределения (2) обладают универсальным характером и поэтому естественно именно их использовать в качестве статистической основы формулируемого ниже подхода.

### 3. Задание описаний прецедентов параметрами гауссовых смесей

Предлагаемый подход к идентификации объектов ориентирован на методы машинного обучения [13], т.е. предполагается, что исходно прецеденты также представлены зарегистрированными (в процессе обучения) наборами фотоотсчетов, статистика которых задается распределениями (2). Однако, как показывает опыт, использовать непосредственно данные регистрации прецедентов для процедур идентификации нерационально как с точки зрения малой эффективности получающихся методов, так и с точки зрения используемых вычислительных ресурсов (большие объемы данных, большие времена поиска по ним целевых прецедентов и т.д.). Естественный путь здесь – использовать данные регистрации не непосредственно для идентификации, а для формирования описаний породивших их распределений ( $\rho(\vec{x})$  – нормированных интенсивностей  $I(\vec{x})$  (2) прецедентов).

Стандартным подходом к формированию описаний  $\rho(\vec{x})$ , является приближение их моделями  $p(\vec{x}|\vec{\theta})$  некоторого параметрического семейства распределений с относительно небольшим числом параметров  $\vec{\theta} = \{\theta_1, \dots, \theta_m\}$ . Гибким средством моделирования распределений достаточно общего вида является семейство гауссовых смесей (СГС) [18]. Последние являются частным случаем более общего семейства – семейства экспоненциальных смесей, для которых в принципе можно повторить весь последующий синтез. Однако, не преследуя ни цели наиболее общих формулировок, ни, наоборот, цели специализации подхода к конкретным предметным областям, а ориентируясь в первую очередь на ясность интерпретаций, ниже, в данной работе, используем именно модель СГС, оставляя вопрос о лучшем выборе распределений для будущих исследований.

При выборе СГС в качестве моделирующего семейства предполагается, что интенсивность каждого ( $k$ -го) прецедента может быть описана суммой (смесью)  $N_k$  двумерных гауссовых (нормальных) распределений:

$$\begin{aligned} p_k(\vec{x}|\vec{\theta}) &= \sum_{j=1}^{N_k} p_{k,j} \frac{1}{2\pi} \sqrt{\det\{A_{k,j}\}} \exp\left\{-\frac{1}{2} Q_{k,j}(\vec{x})\right\}, \\ Q_{k,j}(\vec{x}) &= (\vec{x} - \vec{m}_{k,j})^T A_{k,j} (\vec{x} - \vec{m}_{k,j}) \end{aligned} \quad (3)$$

где параметрами  $\vec{\theta}$  являются число  $N_k$  – количество компонентов смеси (3) и набор из  $N_k$  троек  $\{(p_{k,j}, \vec{m}_{k,j}, A_{k,j})\}$ , представляющих собой вероятности  $p_{k,j}$  отсчетов принадлежать компоненте с номером  $j$ , среднее и матрица квадратичной формы которой есть, соответственно,  $\vec{m}_{k,j}$  и  $A_{k,j}$ . Выбор СГС (3) помимо удобств моделирования и анализа полезен еще и потому, что в рамках машинного обучения существует ряд эффективных алгоритмов, позволяющих по выборкам независимых, одинаково распределенных случайных точек  $\{\vec{x}_i\}$  находить параметры максимально-правдоподобной (МП) для них гауссовской смеси. В эту группу алгоритмов входят, в частности, популярные EM-алгоритмы [19].

Для использованного нами варианта формирования описаний прецедентов также был выбран некоторый вариант EM-алгоритма. В нашем подходе, при инициализации обучения ( $m = 0$ ) по данным  $i$ -го прецедента  $\{\vec{x}_i\}$  число компонент  $N_k$  подбиралось вручную, вероятности  $\{p_{k,j}^{(m)}\}$  полагались равными  $1/N_k$ , в качестве центров компонент  $\{\vec{m}_{k,j}^{(m)}\}$  выбиралась случайная  $N_k$ -подвыборка из данных  $\{\vec{x}_i\}$  и в качестве матрицы всех квадратичных форм  $\{A_{k,j}^{(m)}\}$  – кратная единичной матрица  $A = N_k E_2 / D^2$ , где  $D^2$  – выборочная дисперсия данных регистрации  $\{\vec{x}_i\}$ . После этого параметры  $\{(p_{k,j}, \vec{m}_{k,j}, A_{k,j})\}$  рекуррентно вычислялись по следующей схеме (не путать квадратичную форму  $Q_{k,j}^{(m)}(\vec{x}_i)$  (3) с традиционной – функцией EM алгоритма !):

$$m = m + 1$$

**E-step.** for  $i = 1$  to  $n$ ,  
for  $j = 1$  to  $N_k$

$$\begin{aligned} Q_{k,j}^{(m)}(\vec{x}_i) &= (\vec{x}_i - \vec{m}_{k,j}^{(m)})^T A_{k,j}^{(m)} (\vec{x}_i - \vec{m}_{k,j}^{(m)}) ; \\ V_{j|i}^{(m+1)} &= \frac{1}{\Sigma_V} p_{k,j}^{(m)} \sqrt{\det\{A_{k,j}^{(m)}\}} \exp\left\{-\frac{1}{2} Q_{k,j}^{(m)}(\vec{x}_i)\right\} ; \\ \Sigma_V &= \sum_{j=1}^{N_k} p_{k,j}^{(m)} \sqrt{\det\{A_{k,j}^{(m)}\}} \exp\left\{-\frac{1}{2} Q_{k,j}^{(m)}(\vec{x}_i)\right\} ; \end{aligned} \quad (4)$$

**M-step.** for  $j = 1$  to  $N_k$

$$\begin{aligned} p_{k,j}^{(m+1)} &= \frac{1}{n} \sum_{i=1}^n V_{j|i}^{(m+1)}, \quad \vec{m}_{k,j}^{(m+1)} = \frac{1}{np_{k,j}^{(m+1)}} \sum_{i=1}^n V_{j|i}^{(m+1)} \vec{x}_i ; \\ [A_{k,j}^{(m+1)}]^{-1} &= \frac{1}{np_{k,j}^{(m+1)}} \sum_{i=1}^n V_{j|i}^{(m+1)} (\vec{x}_i - \vec{m}_{k,j}^{(m+1)}) (\vec{x}_i - \vec{m}_{k,j}^{(m+1)})^T ; \end{aligned} \quad (5)$$

Приведенные выше соотношения (4) и (5) имеют прозрачную интерпретацию. Именно,  $V_{j|i}^{(m+1)}$  в (4) является т.н. апостериорным (условным) распределением индикатора  $s_i$  (скрытой переменной) принадлежности отсчета  $\vec{x}_i$  компоненте  $j$  смеси (3). Интерпретируя  $V_{j|i}^{(m+1)}/n$  в (5) как совместное выборочное распределение полных данных  $(\vec{x}, s)$ ,  $\vec{x} \in \{\vec{x}_i\}$ ,  $s \in \{1, \dots, N_k\}$ , получим, что параметры  $\{(p_{k,j}^{(m+1)}, \vec{m}_{k,j}^{(m+1)}, A_{k,j}^{(m+1)})\}$  являются, соответственно, выборочным распределением  $s$ , выборочным средним и обратной выборочной корреляционной функцией  $\vec{x}$  для компоненты  $s$ .

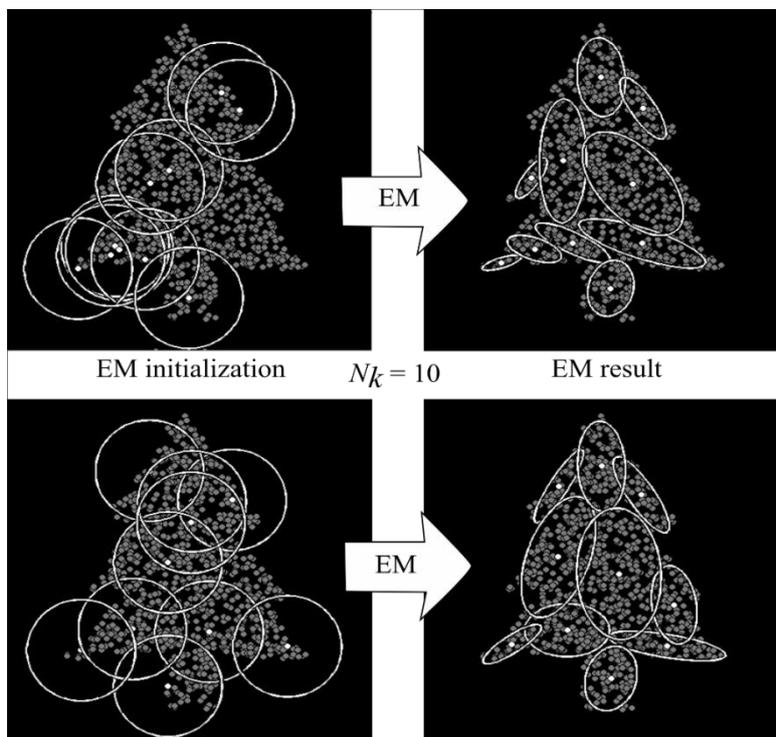
Последнее замечание относительно алгоритма (4,5) следующее. Известно, что инициализация EM-алгоритмов случайным образом чревата серьезными дефектами результирующих решений. Поэтому описанную выше процедуру инициализации следует рассматривать как повторяющуюся до тех пор, пока не будут достигнуты приемлемые с точки зрения учителя (supervisor) результаты. При этом предметом обучения являются не только параметры  $\{(p_{k,j}, \vec{m}_{k,j}, A_{k,j})\}$ , но и их количество  $N_k$ . На Рисунке 1 представлены примеры менее и более удачной (с точки зрения учителя) инициализации.

#### 4. Идентификации зарегистрированных данных по гауссовым смесям

Выбрав форму описаний прецедентов (3) и способ их формирования (4, 5) по реализациям дискретных отсчетов  $\{\vec{x}_i\}$ , в качестве завершающего шага в описании выбранного подхода следует определим вид ориентированной на эти описания процедуры идентификации. Для этого необходимо выбрать количественную меру соответствия, подобия между тестируемыми отсчетами ПТП и прецедентами из сформированной базы данных (БД). Естественным кандидатом на роль такой меры подобия является, ввиду (2) и (3), (логарифмическая) функция правдоподобия

$$L_k(\{\vec{x}_i\}) = \ln[\rho(\vec{x}_1, \dots, \vec{x}_n | n, \vec{\theta}^{(*)})] = \ln\left[\prod_{i=1}^n p_k(\vec{x}_i | \vec{\theta}^{(*)})\right], \quad (6)$$

где  $\{\vec{x}_i\}$  координаты отсчетов тестируемого ПТП, а  $\vec{\theta}^{(*)}$  – полученное с помощью (4,5) описание  $k$  –прецедента.



**Рисунок 1.** Результаты инициализации (слева) и применения (справа) EM–алгоритма (4,5) для формирования описания прецедента “tree-14” (см. раздел 5 ниже). Компоненты соответствующих смесей представлены своими центрами (точками) и линиями (эллипсами) постоянного уровня, приблизительно в 0.1 величины максимального значения всего распределения. Инициализация в нижнем ряду представляется более удачной, чем в верхнем.

Теоретически, вычислив  $L_k(\{\vec{x}_i\})$  (6) для всех прецедентов, можно было бы в качестве процедуры идентификации использовать метод максимального правдоподобия (МП). Именно, рассматривая для заданной реализации  $\{\vec{x}_i\}$  меру  $L_k(\{\vec{x}_i\})$  (6) как функцию  $k$ , можно идентифицировать тестируемый ПТП с тем прецедентом  $k^{(*)}$ , который доставляет  $L_k(\{\vec{x}_i\})$  максимальное значение:

$$k^{(*)} = \arg \max_k L_k(\{\vec{x}_i\}). \tag{7}$$

Однако, с практической точки зрения подобная процедура идентификации не реалистична. Проблема здесь в том, что в БД одному и тому же объекту при разных его положениях, удаленностях, ракурсах формально должны были бы соответствовать разные прецеденты. Решением проблемы является идентификация данных  $\{\vec{x}_i\}$  тестируемого ПТП не с конкретным описанием прецедента, а с целым классом подобных описаний, получаемых друг из друга при помощи некоторой группы преобразований. Например, если ограничиться группой аффинных преобразований  $\vec{x} \rightarrow (\vec{x} + \vec{t})/s$  (смещение на  $\vec{t}$  и изменение масштаба в  $s$  раз) и задать некоторое априорное распределение  $\rho_{apr}(\vec{t}, s)$  параметров, то идентификацию по-прежнему можно осуществлять на основе МП (7), но в качестве меры подобия следует использовать логарифм среднего по  $\rho_{apr}(\vec{t}, s)$  от всех описаний  $p_k(s\vec{x} - \vec{t}|\vec{\theta}^{(*)})s^2$ , полученных из  $p_k(\vec{x}_i|\vec{\theta}^{(*)})$  действиями выбранной группы:

$$\begin{aligned} \bar{L}_k(\{\vec{x}_i\}) &= \ln \left[ \int \int \rho_{apr}(\vec{t}, s) \left\{ \prod_{i=1}^n P_k(\vec{x}_i|\vec{t}, s) \right\} d\vec{t} ds \right], \\ P_k(\vec{x}|\vec{t}, s) &= p_k(s\vec{x} - \vec{t}|\vec{\theta}^{(*)})s^2 = \\ &= \sum_{j=1}^{N_k} p_{k,j}^{(*)} \frac{1}{2\pi} \sqrt{\det \{A_{k,j}^{(*)}\}} \exp \left\{ -\frac{1}{2} Q_{k,j}^{(*)}(s\vec{x} - \vec{t}) \right\} s^2. \end{aligned} \tag{8}$$

Отправляясь от меры (8) и используя вариационный байесов подход нами был синтезирован следующий, EM-подобный алгоритм вычисления меры подобия  $\bar{L}_k(\{\vec{x}_i\})$  (8) для процедуры идентификации (7):

$$m = m + 1$$

**E-step.**

for  $i = 1$  to  $n$

for  $j = 1$  to  $N_k$

$$Q_{k,j}^{(m+1)}(\vec{x}_i) = \left( \bar{S}^{(m)} \vec{x}_i - \bar{T}^{(m)} - \bar{m}_{k,j}^{(*)} \right)^T A_{k,j}^{(*)} \left( \bar{S}^{(m)} \vec{x}_i - \bar{T}^{(m)} - \bar{m}_{k,j}^{(*)} \right);$$

$$V_{j|i}^{(m+1)} = \frac{1}{\Sigma_V} p_{k,j}^{(*)} \sqrt{\det \{A_{k,j}^{(*)}\}} \exp \left\{ -\frac{1}{2} Q_{k,j}^{(m+1)}(\vec{x}_i) \right\}; \quad (9)$$

$$\Sigma_V = \sum_{j=1}^{N_k} p_{k,j}^{(*)} \sqrt{\det \{A_{k,j}^{(*)}\}} \exp \left\{ -\frac{1}{2} Q_{k,j}^{(m+1)}(\vec{x}_i) \right\};$$

**M-step.**

for  $j = 1$  to  $N_k$

(детали вычисления  $\mu_\sigma$ ,  $\mu_\Sigma$ ,  $\sigma^{(m+1)}$ ,  $\Sigma^{(m+1)}$ ,  $\bar{X}^{(m+1)}$ ,  $\bar{M}^{(m+1)}$  по  $V_{j|i}^{(m+1)}$  (9) см. в [17])

$$\bar{S}^{(m+1)} = \mu_\sigma \sigma^{(m+1)} + \mu_\Sigma \Sigma^{(m+1)}; \quad (10)$$

$$\bar{T}^{(m+1)} = \bar{S}^{(m+1)} \bar{X}^{(m+1)} - \bar{M}^{(m+1)};$$

Синтезированный алгоритм (9,10) по существу рекуррентно вычисляет значения параметров  $\vec{t}$  и  $s = \bar{T}^{(*)}, \bar{S}^{(*)}$ , соответствующие максимально правдоподобному для зарегистрированных отсчетов  $\{\vec{x}_i\}$  описанию  $k$ -прецедента  $P_k(\vec{x}|\bar{T}^{(*)}, \bar{S}^{(*)})$  (8) из класса, определяемого описанием  $p_k(\vec{x}_i|\bar{\theta}^{(*)})$  (3). Здесь также уместно замечание о зависимости результирующих решений от инициализации  $\bar{T}^{(0)}, \bar{S}^{(0)}$ . Однако здесь, при идентификации, обучение с учителем по многим причинам, как правило, не оправдано.

Решение для выбора однозначной процедуры идентификации, приводящей, как правило, к адекватным решениям, нами было найдено на основе следующего наблюдения. Отправляясь от приведенной выше мотивации представлений объектов классами подобных (аффинных) описаний, можно для каждого класса выбрать не исходное, полученное по его прецеденту описание, а некоторое аффинное его преобразование, т.е. описание некоторого возможного, в определенном смысле стандартного представителя. Для определения того, что следует считать стандартным представителем заметим, что в алгоритме обучения (4,5) вне зависимости от инициализации  $\left\{ \left( p_{k,j}^{(m)}, \bar{m}_{k,j}^{(m)}, A_{k,j}^{(m)} \right) \right\}$  и от номера итерации  $m$  имеют место следующие инвариантные соотношения:

$$\sum_{j=1}^{N_k} p_{k,j}^{(m)} \bar{m}_{k,j}^{(m)} = \frac{1}{n} \sum_{i=1}^n \vec{x}_i = \bar{X}_k$$

$$\sum_{j=1}^{N_k} p_{k,j}^{(m)} \left\{ \left[ A_{k,j}^{(m)} \right]^{-1} + \left( \bar{m}_{k,j}^{(m)} - \bar{X}_k \right) \left( \bar{m}_{k,j}^{(m)} - \bar{X}_k \right)^T \right\} = \frac{1}{n} \sum_{i=1}^n (\vec{x}_i - \bar{X}_k) (\vec{x}_i - \bar{X}_k)^T \quad (11)$$

с правыми частями зависящими только от  $\{\vec{x}_i\}$ . Из второго соотношения в (11), в частности, следует тождество:

$$Tr \sum_{j=1}^{N_k} p_{k,j}^{(m)} \left\{ \left[ A_{k,j}^{(m)} \right]^{-1} + \left( \bar{m}_{k,j}^{(m)} - \bar{X}_k \right) \left( \bar{m}_{k,j}^{(m)} - \bar{X}_k \right)^T \right\} = \frac{1}{n} \sum_{i=1}^n (\vec{x}_i - \bar{X}_k)^2 \quad (12)$$

где  $Tr$  обозначает след матрицы. При преобразовании  $\vec{x}' = (\vec{x} + \vec{t})/s$  с  $\vec{t} = -\bar{X}_k$  и  $s^2 = \frac{1}{n} \sum_{i=1}^n (\vec{x}_i - \bar{X}_k)^2$  будет получен представитель того же класса, но с  $\bar{X}'_k = \vec{0}$  и  $Tr(\dots)' = 1$  в (12). Определим его как стандартный представитель класса. Число компонент  $N_k$  в нем то же, что у прецедента, вероятности  $p_{k,j}$  те же, но средние и матрицы квадратичных форм будут, соответственно,  $\bar{m}'_{k,j} = (\bar{m}_{k,j} + \vec{t})/s$  и  $A'_{k,j} = s^2 A_{k,j}$ . Поскольку  $\vec{t}$  и  $s$  определяются здесь только данными  $\{\vec{x}_i\}$ , параметры стандартного представителя могут быть по-прежнему найдены с помощью алгоритма (4,5), но с последующей коррекцией до  $\bar{m}'_{k,j}$  и  $A'_{k,j}$ .

Связав описанным образом с каждым классом стандартный представитель, приходим к следующему решению об инициализации  $\bar{T}^{(0)}, \bar{S}^{(0)}$ . Сформировав по координатам отсчетов

тестируемого ПТП  $\{\vec{x}_i\}$  величины  $\vec{X} = \frac{1}{n} \sum_{i=1}^n \vec{x}_i$  и  $D^2 = \frac{1}{n} \sum_{i=1}^n (\vec{x}_i - \vec{X}_k)^2$  и полагая в нулевом приближении  $\vec{x} = (\vec{x}' + \vec{T}^{(0)})/\bar{S}^{(0)}$ , получим:  $\bar{S}^{(0)} = D^{-1}$  и  $\vec{T}^{(0)} = \bar{S}^{(0)}\vec{X}$ . Тесты с относительно простыми изображениями подтвердили работоспособность предложенного решения.

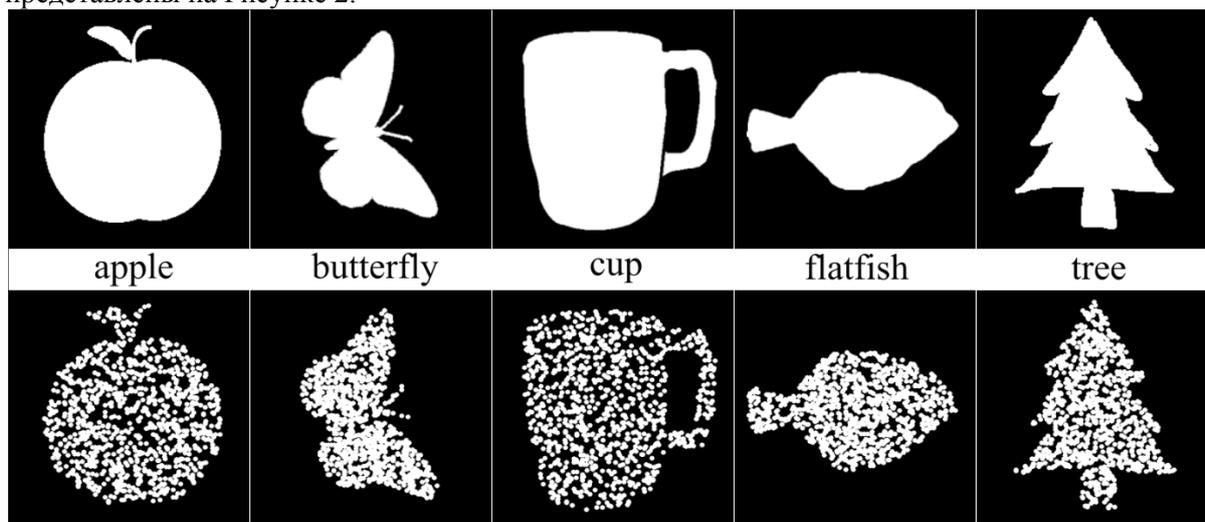
### 5. Результаты численного моделирования алгоритмов идентификации объектов на фотоотсчетных изображениях

Представленная выше процедура вычисления меры подобия является основным, базовым инструментом для идентификации формы интенсивности ПТП. Она количественно характеризует соответствие данных регистрации идентифицируемого ПТП форме интенсивности произвольного прецедента. Дальнейшие шаги при идентификации содержат сценарии сравнения между собой получаемых таким образом мер подобия для различных прецедентов из БД с целью выбора из них максимально подобного идентифицируемым данным.

Заметим, что в случае больших БД эти сценарии могут быть достаточно сложными, основанными на внутренних отношениях между прецедентами, видом близости их описаний, иерархическими связями и т.д. Использование этой информации для оптимального поиска максимально правдоподобного прецедента может существенно экономить вычислительные ресурсы. Однако, эта важная и интересная тема лежит за пределами данной работы.

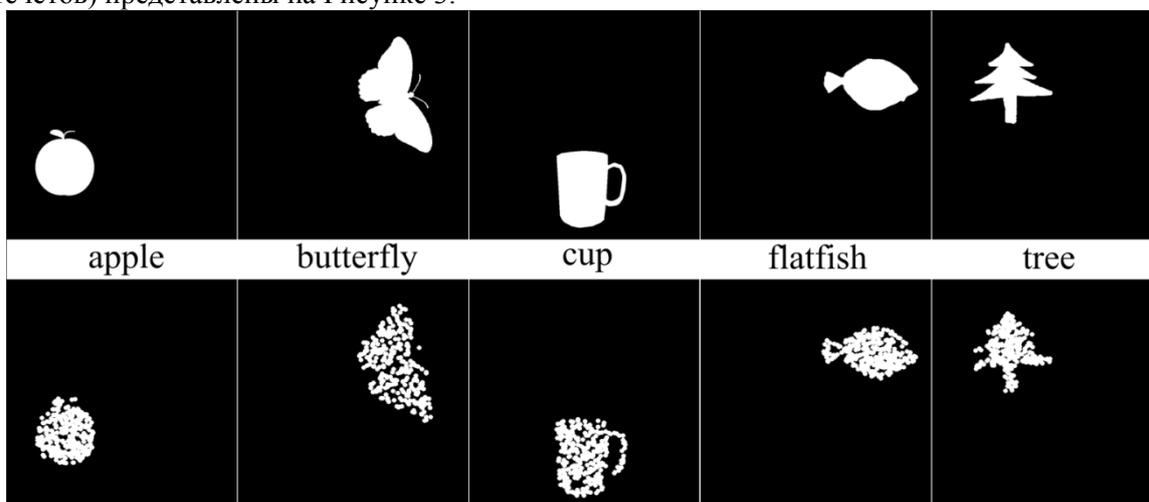
В данной работе для целей иллюстрации предложенного метода идентификации приведены результаты вычислительного эксперимента для небольшой, искусственно сформированной БД. Описания прецедентов в этой БД были получены на основе данных регистрации небольших, простой структуры объектов. В качестве последних выбраны двоичные черно-белые изображения из известной базы “MPEG-7 Core Experiment CE-Shape-1” [20]. Эта база изображений обычно используется для сравнения алгоритмов распознавания формы объектов и включает 70 категорий объектов по 20 изображений объекта в каждой категории.

Из БД [20] в качестве прецедентов были произвольно выбраны изображения 5 категорий – “apple-10”, “butterfly-15”, “cup-19”, “flatfish-10”, “tree-14”. Для унификации они были приведены к размеру кадра  $256 \times 256$  точек и в нем центрированы. Поскольку форма интенсивности на этих изображениях простая – равномерная по объекту, симуляция данных регистрации прецедентов была банальной – из большого числа равномерно распределенных по кадру сгенерированных случайных точек было выбрано 1000 первых, попавших в границы объектов. Выбранные прецеденты и соответствующие им данные регистрации (1000 отсчетов) представлены на Рисунке 2.

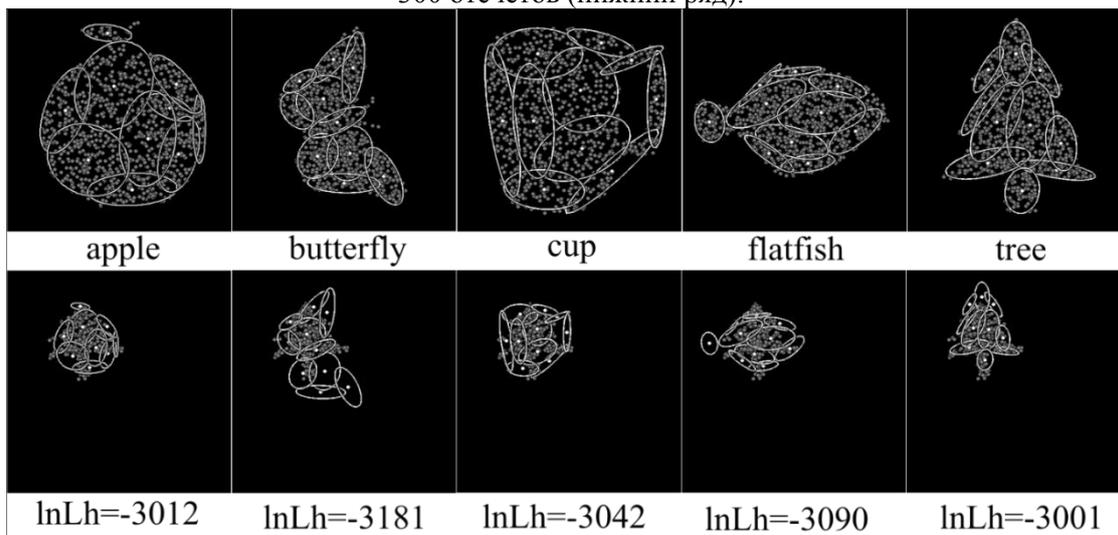


**Рисунок 2.** Прецеденты из БД “MPEG-7 Core Experiment CE-Shape-1” [12] – бинарные изображения размером  $256 \times 256$  (верхний ряд) и модельные данные их регистрации в виде 1000 отсчетов, равномерно распределенных по полю объектов (нижний ряд).

Для формирования описания выбранных прецедентов в виде гауссовых смесей с  $N_k=10$  компонентами (для всех прецедентов) использовался стандартный EM-алгоритм (4,5). Инициализация EM-алгоритма осуществлялась равномерным распределением  $p_{k,j}$ , выбором  $N_k$  случайных отсчетов из данных регистрации  $\{\vec{x}_i\}$  в качестве центров компонент  $\vec{m}_{k,j}$  и одинаковых для всех компонент матриц  $A_{k,j} = E_2/2500$ , где  $E_2$  – единичная  $2 \times 2$  матрица, а знаменатель 2500 равен квадрату стандартного отклонения в  $\sim 10$  раз меньшего размера кадра. В качестве тестируемых объектов были выбраны изображения “apple-20”, “butterfly-13”, “cup-7”, “flatfish-3”, “tree-9” – из тех же 5 категорий, что и прецеденты, но отличные от них. Все они были приведены к одному размеру кадра  $512 \times 512$  точек и случайным образом смещены от центра кадра и уменьшены в масштабе. Симуляция данных регистрации представляла собой выбор 300 сгенерированных случайных точек, равномерно распределенных в границах объектов. Тестовые объекты и соответствующие им малоотсчетные данные регистрации (300 отсчетов) представлены на Рисунке 3.



**Рисунок 3.** Тестовые объекты из БД “MPEG-7 Core Experiment CE-Shape-1” [12] – из тех же категорий, что прецеденты Рисунка 2, но смещенные и уменьшенные случайным образом (верхний ряд) и смоделированные данные их регистрации в виде равномерно распределенных 300 отсчетов (нижний ряд).



**Рисунок 4.** Описания прецедентов из сформированной БД (верхний ряд) и их пересчитанные с помощью найденных максимально-правдоподобных параметров  $\vec{T}^{(*)}$  и  $\vec{S}^{(*)}$  описания (нижний ряд) для данных регистрации тестового объекта “tree-9” (см. Рисунок 3). Внизу для каждого случая приведено значение логарифма среднего правдоподобия lnLh.

Для каждого из тестовых объектов был сформирован массив значений мер подобия с каждым из прецедентов. По существу, вычисленные на основе предложенной процедуры меры подобия представляют собой значения (логарифмической) функций правдоподобия  $L_k(\{\vec{x}_i\})$  (6) в предположении, что  $\{\vec{x}_i\}$  являются результатом регистрации отсчетов аффинного преобразования –прецедента с максимально–правдоподобными параметрами преобразования  $\vec{t} = \vec{T}^{(*)}$  и  $s = \bar{S}^{(*)}$  (10).

Пересчитанные с помощью  $\vec{T}^{(*)}$  и  $\bar{S}^{(*)}$  (своих для каждого прецедента) описания  $P_k(\vec{x}|\vec{t}, s)$  (8) для  $k = \text{“apple-10”}$ ,  $\text{“butterfly-15”}$ ,  $\text{“cup-19”}$ ,  $\text{“flatfish-10”}$ ,  $\text{“tree-14”}$  и значения оценок логарифма среднего правдоподобия  $\bar{L}_k(\{\vec{x}_i\})$  (6), для данных регистрации тестового объекта  $\text{“tree-9”}$  представлены на Рисунке 4.

Из Рисунка.4. видно, как распределены значения среднего правдоподобия  $\ln L_h$  по имеющимся прецедентам. Наибольшего значения они достигают для прецедента  $\text{“tree”}$ , что и ожидалось для тестового объекта из этой категории. Более интересным является тот факт, что другими ближайшими к объекту  $\text{“tree-9”}$  оказываются категории  $\text{“apple”}$ , затем  $\text{“cup”}$ , а наиболее удаленной –  $\text{“butterfly”}$ . Это наблюдение дает повод предполагать, что в больших БД прецедентов действительно могут быть построены оптимальные сценарии по сравнению мер подобия для поиска максимально подобных идентифицируемым данным прецедентов.

**Таблица 1.** Значения меры подобия (среднего правдоподобия) для тестовых объекты различных категорий (объем данных 300 отсчетов).

Тестируемые объекты	Прецеденты				
	apple-10	butterfly-15	cup-19	flatfish-10	tree-14
apple-20	-2907	-3100	-2992	-3026	-2996
butterfly-13	-3202	-3117	-3227	-3320	-3186
cup-7	-3090	-3190	-3029	-3129	-3174
flatfish-3	-3122	-3304	-3175	-2949	-3193
tree-9	-3012	-3181	-3042	-3090	-3001

Полные результаты вычисления на основе предложенной процедуры среднего правдоподобия, включающие другие тестируемые объекты  $\text{“apple-20”}$ ,  $\text{“butterfly-13”}$ ,  $\text{“cup-7”}$ ,  $\text{“flatfish-3”}$  (см. Рисунок 3), представлены в Таблице 1. Заметим, что идентификация на основе критерия (7) целевого прецедента (категории) во всех случаях оказывается правильной. Безусловно, для объективной, статистически достоверной оценки предложенного подхода необходимо его тестирование на гораздо больших базах данных, однако, уже полученные результаты позволяют предположить оптимистические прогнозы в отношении его потенциала.

## 6. Заключение

В работе показано, что формализация процесса регистрации фотоотсчётными датчиками падающего на них излучения моделью пуассоновских точечных процессов [15] является наиболее адекватной с физической (квантовой) точки зрения [14] и исключительно плодотворной с точки зрения статистического анализа [13] данных. На этой основе, используя принципы машинного обучения [13], нам удалось разработать эффективный подход синтеза процедур идентификации (распознавания) объектов, принадлежащих хорошо зарекомендовавшему себя семейству EM-алгоритмов. Полученная алгоритмическая архитектура системы идентификации была изначально мотивирована известными фактами о зрительной системе и, как оказалось, результаты синтеза позволяют в значительной степени имитировать нейробиологические механизмы обработки визуальной информации в коре головного мозга [3-5]. Численное моделирование показало, что синтезированная процедура

идентификации обладает высокой скоростью сходимости – при сложности описания прецедентов гауссовыми смесями с  $N_k \sim 10$  компонентами рекурсивные вычисления (9,10) сходятся за  $\sim 10$  итераций.

## 7. Литература

- [1] Nakamura, J. Image Sensors and Signal Processing for Digital Still Cameras (Optical Science and Engineering). – CRC. Press, 2005.
- [2] Khurana, K. Techniques for Object Recognition in Images and Multi-Object Detection / K. Khurana, R. Awasthi // Int. J. of Advanced Research in Computer Engineering & Technology. – 2013. – Vol. 2(4) – P. 1383-1388.
- [3] Weiren, S.W. A Survey of Biologically Inspired Image Processing for Objects Recognition / S.W. Weiren, Z. Li, Z. Zhong // Int. J. Image Graphics. – 2009. – Vol. 9. – P. 495-510.
- [4] Yin, P. A Novel Biologically Inspired Visual Cognition Model: Automatic Extraction of Semantics, Formation of Integrated Concepts, and Reselection Features for Ambiguity / P. Yin, H. Qiao, W. Wu // IEEE Trans. Cognitive and Developmental Systems. – 2018. – Vol. 10(2). – P. 420-431.
- [5] Cristobal, G. Biologically Inspired Computer Vision: Fundamentals and Applications / G. Cristobal, L. Perrinet, S. Matthias, M.S. Keil. – Wiley-VCH, Weinheim, 2015.
- [6] Matthen, M. Oxford Handbook of the Philosophy of Perception. – Oxford University Press, 2015.
- [7] Fossum, E.R. Photon-Counting Image Sensors / E.R. Fossum, N. Teranishi, A. Theuwissen, D. Stoppa, E. Charbon. – MDPI Books under CC BY-NC-ND license, 2017.
- [8] Robbins, M. Electron-Multiplying Charge Coupled Devices-EMCCDs / M. Robbins, P. Seitz, A.J.P. Theuwissen // Single-Photon Imaging. – Berlin: Springer, 2011 – P. 103-121.
- [9] Dutton, N.A.W. A SPAD-based QVGA image sensor for single-photon counting and quanta imaging / N.A.W. Dutton, I. Gyongy, L. Parmesan // IEEE Trans. Electron Devices. – 2016. – Vol. 63(1). – P. 189-196.
- [10] Aull, B.F. A study of crosstalk in a 256x256 photon counting imager based on silicon Geiger-mode avalanche photodiodes / B.F. Aull, D.R. Schuette, D.J. Young // IEEE Sens. J. – 2015. – Vol. 15(4). – P. 2123-2132.
- [11] Remez, T. A picture is worth a billion bits: Real-time image reconstruction from dense binary threshold pixels / T. Remez, O. Litany, A. Bronstein // Proceedings of the IEEE International Conference on Computational Photography (ICCP). – 2016. – P. 1-9.
- [12] Chen, B. Vision without the Image / B. Chen, P. Perona // J. Sensors. – 2015. – Vol. 16(4). – P. 484.
- [13] Hastie, T. The Elements of Statistical Learning / T. Hastie, R. Tibshirani, J. Friedman // New York: Springer, 2009.
- [14] Goodman, J.W. Statistical Optics // New York: Wiley, 2015.
- [15] Streit, R.L. Poisson Point Processes. Imaging, Tracking and Sensing // New York: Springer, 2010.
- [16] Mandel, L. Optical Coherence and Quantum Optics / L. Mandel, E. Wolf // Cambridge: Cambridge University Press, 1995.
- [17] Antsiperov, V. Precedent-based Low Count Rate Image Intensity Estimation using Maximum Likelihood Distribution Descriptions // Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence (ICPRAI 2018), 2018. – P. 707-711.
- [18] Mengersen, K.L. Mixtures: Estimation and Applications / K.L. Mengersen, C.P. Robert, D.M. Titterton // John Wiley & Sons, 2018.
- [19] Gupta, M.R. Theory and Use of the EM Algorithm // Foundations and Trends in Signal Processing. – 2010. – Vol. 1(3). – P. 223-296.
- [20] Latecki, L.J. Shape Descriptors for Non-rigid Shapes with a Single Closed Contour / L.J. Latecki, R. Lakmper, U. Eckhardt // Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2000.

### **Благодарности**

Автор выражает благодарность Российскому фонду фундаментальных исследований за финансовую поддержку работы в рамках гранта № 18-07-01295.

## **Target Identification for photon counting image sensors, inspired by mechanisms of human visual perception**

**V.E. Antsiperov<sup>1</sup>**

<sup>1</sup>Kotelnikov Institute of Radioengineering and Electronics (IRE) of RAS, Mokhovaya 11-7, Moscow, Russia, 125009

**Abstract.** The report presents the results of the synthesis of a new method of objects identification / pattern recognition, based on the realization of a random point process – the set of photocounts of the photon counting image sensors. The synthesis was carried out entirely on the basis of the methods used in the field of machine learning, the synthesized identification algorithms are principally focused on modern computing systems. The results of numerical simulation are given in the report proposed.