Target identification for photon-counting image sensors, inspired by mechanisms of human visual perception

V E Antsiperov

Kotelnikov Institute of Radioengineering and Electronics (IRE) of Russian Academy of Sciences, Mokhovaya 11-7, Moscow, Russia, 125009

e-mail: antciperov@cplire.ru

Abstract. Paper presents the new objects identification/recognition method, designed for modern image sensors. One of the main features of the method proposed is that it is initially focused on presenting the input data as a discrete set of photocounts of image sensors. The developing of the method was based on the machine learning approaches, so the synthesized identification algorithms are principally focused on the modern embedded computations. The results of numerical simulation are given at the end of the paper.

1. Introduction

Besides such elements, as light detectors, modern (digital) image formation systems also contain microcontrollers, microprocessors, and special chips. The latter, apart from managing the detection of radiation falling onto the sensors, storing such recordings, adjusting the tone, maintaining the white balance, etc., also provide giant opportunities for further, more intelligent, processing of the registered images [1].

Essentially, the intelligent processing is analyzing images to extract specific data that may be important for some kinds of application, such as the computerized vision of industrial robots, automatic face or handwriting recognition, navigation of semi-autonomous or unmanned vehicles, etc. In theory, various tasks always require various image analysis methods. In practice, however, most analysis methods always include a number of general elements. Among such elements, object recognition methods should be mentioned first. Generally, recognition is the localization of objects in images and/or classification/identification thereof [2].

Today's recognition methods increasingly either include specific human vision mechanisms or try to imitate its structure [3]. The first group of methods comes with those using various preliminary image processing techniques that the retina is responsible for. Such methods include data compression, noise filtering, contrast, and edge enhancement, etc. Another group incorporates more intelligent methods that imitate cortex-based neurobiological visual information processing mechanisms. Machine learning methods are what should be mentioned first and foremost in this context, namely: subspace methods, slow feature learning, hierarchical feature analysis, etc. It should be noted that considerable results have been achieved so far in this field [5].

Therefore, the existing and known recognition methods implement and imitate human vision mechanisms with some higher or lower accuracy. However, registration data that are supposed to be processed by these methods, usually have different forms in the visual system and artificial image
formation systems. The detected radiation is registered by microprocessors of most devices as
digitalized intensity on discrete elements (pixels) of the imager. Intensity values are approximately
proportional to the overall energy of photons accumulated by a pixel within the frame formation
period and is expressed as an integer number varying across a wide range. Conversely, photoreceptor
cells (rods and cones) of the retina, perceive only single radiation photons and convey registration
signal further, to photosensitive ganglion cells, without accumulation. As a result, the asynchronous
flow of action potentials (nerve impulses), constantly processed by the vision system, is formed [6].

Until quite recently, technological opportunities for creating image formation devices that would be
similar in function to human vision systems were very restrained - this explains such a difference in
data presentation between artificial and natural systems. However, research and development activities
of the last decades led to well-known progress in developing image sensors (operating not only in the
visible range), which opened up the whole new opportunities in this field. The major trend of this
progress, which appeared in late 1960s with invention of charge-coupled devices (CCDs) and then
established with invention of image sensors based on complementary metal–oxide–semiconductors
(CMOS) in early 1990s, was related to the concern to decrease the pixel size (including the sizes lower
than the diffraction limit). Together with enhancing the spatial resolution, this trend ensures the
improvement of other registration features - e.g. increases the frame rate, widens the dynamic range,
decreases the sensor size and lowers power consumption.

It should be noted that progressively as the pixel size decreases, radiation detection becomes more
and more quantum, transforming into the detection of separate photons (photoelectrons) at the limit.
Interestingly, this very limit has been reached by some technologies to date — specifically, production
of photon-counting image sensors [7]. The vivid examples are electron-multiplying charge-coupled
devices (EMCCD) [8], single-photon avalanche diodes (SPAD) [9], Geiger-mode avalanche
photodiodes (GMAPD) [10].

In terms of formed images [11], a crucial feature of photon-counting sensors is that they eliminate
the necessity for long-term exposing — an image is not actually formed; instead, a time-continuous
flow of photoelectrons with random data access is formed [12]. This is what significantly transforms
the basic theoretical and practical concepts of image processing. For example, in traditional
computerized vision tasks, the identification cycle goes step by step: first, photocounts are
accumulated, then images are built with them and then analyzed. The photon-counting sensors allow
combining data registration and processing, by making the identification procedure incremental,
developing synchronously with counts registration. Such a concept eliminates the need for
reconstructing the full image, bringing the new paradigm of "no-image vision" [12].

This paper presents the research results for the aforementioned concept of identifying (classifying)
the registered objects by data regarding photocount flows of their radiation. By using photocount
coordinates as input data, rather than images reconstructed based on them, it is possible to formally
transform the primary task into the problem of the most optimal adjusting the registered data to one of
the statistical descriptions of the previously-registered processes, which are called precedents [13].
This paper shows that the suggested approach can be developed up to the algorithmic implementation
level. The synthesized identification algorithm, which is similar in structure to the algorithms of the
well-known EM-family, is presented below.

2. The statistical description of photon-counting registration data
To make the suggested object identification approach statistically well-grounded, it is required to
formalize how photon-counting sensors register object radiation. This leads to the necessity for a
mathematical model that could adequately, in physical (quantum) terms associate the radiation
intensity \( I(\vec{x}) \), coming on the sensitive surface \( \Omega \) of the sensor with a set of registered photocount
coordinates \( \{\vec{x}_1, ..., \vec{x}_n\} \). Here, vector \( \vec{x} \in \Omega \) denotes point coordinates in some coordinate system in
the plane of registration, \( \vec{x}_i \) are pixel coordinates corresponding to photocounts.

If the sensor pixel sizes could be considered arbitrarily small, it would be an ideal case called
continuous spatial photodetection [14]. In this case photocount coordinates \( \{\vec{x}_i\} \) are the sets of random
vectors (of random but finite number \( n \)), even for the deterministic, non-random intensity \( I(\vec{x}) \). The
randomness of data is often interpreted as quantum/shot noise. As long as quantum noise is almost
always the dominating noise source in the optic range of the spectrum [14], it is not reasonable to represent the intensity \( I(\vec{x}) \) by the object radiation intensity and additive background/noise, including the latter in the object description (if necessary).

The most common mathematical model of photon-counting sensor detection is random point processes [15]. According to the semi-classic photodetection theory, it is widely known that Poisson point processes (PPP) [15] are the appropriate processes that most adequately describe physical mechanisms of photon-counting registration [16]. For PPPs, full statistical description is defined by the only function - photocount intensity \( \lambda(\vec{x}) \), which defines all finite-dimensional distributions of photocounts [15]:

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

(1)

The quantum theory specifies the relation between the PPP photocount intensity \( \lambda(\vec{x}) \) and radiation intensity \( I(\vec{x}) \). This relation is given by simple linear equation \( \lambda(\vec{x}) = \eta T I(\vec{x})/h \nu \), where \( \eta \) is the quantum efficiency of the sensor material, \( T \) is the frame read time, \( \nu \) is the center radiation frequency, \( h \) is the Planck constant [14].

The mapping \( I(\vec{x}) \rightarrow \lambda(\vec{x}) \), together with distributions (1) define the mathematical model of the input-output relation \( I(\vec{x}) \rightarrow \{x_i\} \) of an ideal (continuous) sensor. For real sensors with finite-size pixels, such a relation is obtained through pixel-surface integration of (1) accounting the number of photocounts detected by each of pixels. The real sensor model will be close to the ideal model (1) in cases when probabilities of two and more photocounts of a pixel are too low as compared with the probability of the single photocount. It can be easily shown that his condition is reduced to relation \( \lambda(\vec{x})S = \eta TSI(\vec{x})/h \nu \ll 1 \), where \( S \) is the pixel area. Therefore, during the detection of weak radiations \( I(\vec{x}) \rightarrow 0 \), or at high shooting speed \( T \rightarrow 0 \), or for using the described in the introduction photon-counting sensors \( S \rightarrow 0 \), distributions (1) may be accepted as an adequate photocount detection model.

The primary task can be formalized even more if its statement is extended through clarifying the term of object identification. Supposing the object identification only by the form of their intensity, regardless of their full brightness (refer to [17] for a comprehensive discussion of this problem), the synthesis basis can be not distributions (1) with the arbitrary number of photocounts, but conditional distributions of photocount coordinates \( \{\vec{x}_1, ..., \vec{x}_n\} \), for the given full number thereof \( n \):

\[
\rho(\vec{x}_1, ..., \vec{x}_n | n) = \prod_{i=1}^{n} \rho(\vec{x}_i) , \quad \rho(\vec{x}) = \frac{\lambda(\vec{x})}{\int_{\Omega} \lambda(\vec{x}) d\vec{x}} = \int_{\Omega} \rho(\vec{x}) d\vec{x} .
\]  

(2)

Distributions (2) express the well-known PPP feature: the conditional (for fixed \( n \)) joint distribution of photocount coordinates is defined by the product of the identical distributions of coordinates of each photocount [15]. In other words, for the fixed \( n \), the photocounts \( \{\vec{x}_i\} \) are the sample of independent identically distributed (iid) random vectors. Moreover, the distribution of sample data \( \rho(\vec{x}) \) (2) coincides with the (normed) intensity \( I(\vec{x})/\int\lambda(\vec{x}) d\vec{x} \), i.e. does not depend either on the sensor material \( \eta \), or center radiation frequency \( \nu \), or detection time \( T \). The listed features make the distributions (2) independent of the specific technical sensor implementation details and very suitable for use as the general statistical basis for the approach formulated below.

3. Gaussian mixture model parameters for the precedent description

The suggested object identification approach is focused on machine learning methods [13], i.e. primarily, the precedents are represented as the photocount sets (registered during learning), whose statistics are defined by the distributions (2). However, the practice shows that it is irrational to use the precedent registration data “as is” for identification due to the low efficiency of obtained methods and high computing costs (big data, long target precedent search, etc.). The common way is to use registration data for forming the descriptions of original distributions (normed intensities \( I(\vec{x}) \) (2) of precedents), rather than for identification.

The standard approach to forming such descriptions is the approximation of \( \rho(\vec{x}) \) by model \( p(\vec{x}|\vec{\theta}) \) from some parametrical family of distributions with relatively few parameters \( \vec{\theta} = \{\theta_1, ..., \theta_m\} \). The
family of Gaussian mixtures (GM) is here a flexible general distribution modeling tool. GM are a special subset of the more general family of exponential mixtures, for which the further synthesis can be repeated in the same way. Throughout the rest of this paper, we will use GM, because we do not endeavor neither to provide general formulations, nor to customize the approach to any specific areas. Instead, we will focus on the interpretation clarity, leaving the problem of selecting the most appropriate distributions for further studies.

When selecting the GM as the modeling family, it is supposed that the intensity of each (k-th) precedent can be defined by the mixture of $N_k$ two-dimensional Gaussian components:

$$p_k(\mathbf{x} | \bar{\theta}) = \sum_{j=1}^{N_k} p_{k,j} \frac{1}{2\pi} \sqrt{\det(A_{k,j})} \exp \left\{ -\frac{1}{2} Q_{k,j}(\mathbf{x}) \right\},$$

$$Q_{k,j}(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_{k,j})^T A_{k,j}(\mathbf{x} - \mathbf{m}_{k,j})$$

where parameters $\bar{\theta}$ are: $N_k$ - number of mixture components and the set of $N_k$-triplets \(\{(p_{k,j}, \mathbf{m}_{k,j}, A_{k,j})\}\), which are probabilities $p_{k,j}$ for belonging photocounts to the component $j$ mean and square-form $Q_{k,j}(\mathbf{x})$ matrix of which are, respectively, $\mathbf{m}_{k,j}$ and $A_{k,j}$.

Besides modeling and analyzing convenience, GM (3) are beneficial, because machine learning comprises a set of efficient algorithms that allows finding the parameters of the most likely GM for samples of independent, equally-distributed random points $\{\mathbf{x}_i\}$. This algorithm group includes popular EM-algorithms [19], [20].

For the precedent description forming method that we use, some variant of an EM-algorithm was chosen. In our approach, when initializing learning ($m = 0$) based on the $k$-th precedent $\{\mathbf{x}_i\}$ data, the number of components $N_k$ was selected manually, probabilities $\{p_{k,j}^{(m)}\}$ were deemed equal to $1/N_k$; the centers of components $\{\mathbf{m}_{k,j}^{(m)}\}$ were defined by the random $N_k$ subsample of data $\{\mathbf{x}_i\}$; the square-form matrix $A_{k,j}^{(m)}$ was proportional to the unit $E_2$ matrix $A = N_k E_2 / D^2$, where $D^2$ is the sample variance of registration data $\{\mathbf{x}_i\}$. After that, parameters $\{\{p_{k,j}, \mathbf{m}_{k,j}, A_{k,j}\}\}$ were recurrently calculated by the following scheme (do not confuse the square form $Q_{k,j}^{(m)}(\mathbf{x}_i)$ (3) with the Q-function of EM-algorithm):

$$m = m + 1$$

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

for $j = 1$ to $N_k$

$$Q_{k,j}^{(m)}(\mathbf{x}_i) = (\mathbf{x}_i - \mathbf{m}_{k,j}^{(m)})^T A_{k,j}^{(m)} (\mathbf{x}_i - \mathbf{m}_{k,j}^{(m)})$$

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

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

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

$$p_{k,j}^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} V_{j|i}^{(m+1)},$$

$$\mathbf{m}_{k,j}^{(m+1)} = \frac{1}{np_{k,j}^{(m+1)}} \sum_{i=1}^{n} V_{j|i}^{(m+1)} \mathbf{x}_i;$$

$$\left[A_{k,j}^{(m+1)}\right]^{-1} = \frac{1}{np_{k,j}^{(m+1)}} \sum_{i=1}^{n} V_{j|i}^{(m+1)} (\mathbf{x}_i - \mathbf{m}_{k,j}^{(m+1)}) (\mathbf{x}_i - \mathbf{m}_{k,j}^{(m+1)})^T;$$

The above relations (4) and (5) can be clearly interpreted. Indeed, $V_{j|i}^{(m+1)}$ in (4) is the so-called conditional (posteriori) distribution of the indicator $j_i$ (latent variable) that the photocount $\mathbf{x}_i$ belongs to the component $j$ of the mixture (3). Interpreting $V_{j|i}^{(m+1)}/n$ in (5) as the joint sample distribution of complete date $(\mathbf{x}_i,j_i)$, $\mathbf{x} \in \{\mathbf{x}_i\}$, $s \in \{1,...,N_k\}$, we get that parameters $\{\{p_{k,j}, \mathbf{m}_{k,j}, A_{k,j}\}\}$ coincide, respectively, with the the sample distribution of $j$, the mean and inverse correlation function for the component $j$. 


The last comment on the algorithm (4,5) is as follows. It is known that random initializing of EM-algorithms may lead to critical deficiencies of the resulting solutions. That is why the above initialization procedure should be considered as repeating until the results approvable by the supervisor are obtained. With that, not only parameters \( \{p_{k,j}, \tilde{m}_{k,j}, A_{k,j}\} \) themselves, but their number \( N_k \) are the learning subject. Figure 1 illustrates initialization cases that are less and more approvable by the supervisor.

![Figure 1](image)

**Figure 1.** Results of initializing (left) and applying (right) the EM-algorithm (4,5) for "tree-14" precedent description formation (see Section 5 below). The components of the corresponding mixtures are represented by their centers (points) and lines (ellipses) of fixed-level probability equal to approximately 0.1 of the maximum distribution values. Bottom-line initialization seems to be more approvable than the top-line one.

4. **Identification of registered data by Gaussian mixtures**

As a final step after selecting the form of precedent descriptions (3) and the method for forming such descriptions (4,5) based on discrete photocounts \( \{\tilde{x}_i\} \), we need to determine the type of the identification procedure applicable to these descriptions. For that, we have to select the qualitative applicability measure that would reflect the similarity between the tested PPP photocounts and precedents from the database. Due to (2) and (3), such a measure can be the (logarithmic) likelihood function

\[
L_k(\{\tilde{x}_i\}) = \ln[p_k(\tilde{x}_1, \ldots, \tilde{x}_n|\tilde{\theta}(\cdot))] = \ln[\prod_{i=1}^n p_k(\tilde{x}_i | \hat{\theta}(\cdot))],
\]

where \( \{\tilde{x}_i\} \) are coordinates of photocounts of the tested PPP, and \( \hat{\theta}(\cdot) \) is the description of the \( k \) -precedent obtained with (4,5). In theory, having calculated \( L_k(\{\tilde{x}_i\}) \) (6) for all precedents, the maximum likelihood method could be used as the identification procedure. Considering \( L_k(\{\tilde{x}_i\}) \) (6) as the function of \( k \) for the given data realization \( \{\tilde{x}_i\} \), the tested PPP can be identified with the precedent \( k(\cdot) \) that gives \( L_k(\{\tilde{x}_i\}) \) the maximum value:

\[
k(\cdot) = \arg \max_k L_k(\{\tilde{x}_i\}).
\]

However, such an identification procedure does not seem practically feasible. In the database, the same object must have, formally, different precedents for its different positions, scales, and perspectives. This problem can be solved by identifying data \( \{\tilde{x}_i\} \) of the tested PPP with not a certain precedent description, but a set of such descriptions obtained from each other through some transformations. For instance, if we use the group of affine transformations \( \tilde{x} \rightarrow (\tilde{x} + \tilde{\ell})/s \) only
(displacement by $\tilde{\ell}$ and s-scaling) and set the a priori distribution $\rho_{appr}(\tilde{\ell}, s)$ of parameters, then identification will be possible based on the maximum likelihood (7), but the similarity measure will be the logarithm of the mean over $\rho_{appr}(\tilde{\ell}, s)$ of all descriptions $p_k(s\bar{x} - \tilde{\ell}\hat{\theta}(\cdot))s^2$ obtained from $p_k(\bar{x}|\tilde{\ell}, \hat{\theta}(\cdot))$ by the selected group transformations:

$$L_k(\bar{x}_i) = \ln \left[ \int \int \rho_{appr}(\tilde{\ell}, s) \left\{ \prod_{i=1}^{n} p_k(\bar{x}_i|\tilde{\ell}, s) \right\} \, d\tilde{\ell} \, ds \right] ,$$
$$P_k(\bar{x}|\tilde{\ell}, s) = p_k(s\bar{x} - \tilde{\ell}\hat{\theta}(\cdot))s^2 = \Sigma_{j=1}^{N_k} p_{k,j}^{(s)} \frac{1}{2\pi} \sqrt{\det \{ A_{k,j}^{(s)} \} } \exp \left\{ -\frac{1}{2} Q_{k,j}^{(s)}(s\bar{x} - \tilde{\ell}) \right\} s^2 . \tag{8}$$

Based on the measure (8) and using the variational Bayesian approach, we have synthesized the following EM-like algorithm for similarity measure $L_k(\bar{x}_i)$ calculation (8) that should be used in the identification procedure (7):

$m = m + 1$

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

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

$V_{j}^{(m+1)} = \frac{1}{\Sigma_{j=1}^{N_k} p_{k,j}^{(s)}} \sqrt{\det \{ A_{k,j}^{(s)} \} } \exp \left\{ -\frac{1}{2} Q_{k,j}^{(m)}(\bar{x}_i) \right\} ;$

$\Sigma V = \Sigma_{j=1}^{N_k} p_{k,j}^{(s)} \sqrt{\det \{ A_{k,j}^{(s)} \} } \exp \left\{ -\frac{1}{2} Q_{k,j}^{(m)}(\bar{x}_i) \right\} ; \tag{9}$

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

$\bar{S}^{(m+1)} = \mu_{2}\sigma^{(m+1)} + \mu_{2}\Sigma^{(m+1)} ;$

$\bar{T}^{(m+1)} = \bar{S}^{(m+1)} \bar{x}^{(m+1)} - \bar{M}^{(m+1)} ; \tag{10}$

Essentially, the synthesized algorithm (9,10) recurrently calculates the values of parameters $\tilde{\ell}$ and $s$ - $\bar{T}^{(s)}, \bar{S}^{(s)}$, corresponding to the description of the $k$-precedent $p_k(\bar{x}|\bar{T}^{(s)}, \bar{S}^{(s)})$ (8) of the class determined by the description $p_k(\bar{x}_i|\tilde{\ell}, \hat{\theta}(\cdot))$ (3), which is the maximum likely to the registered photocounts $\{\bar{x}_i\}$. It can be noted that the resulting solutions depend on the initialization $\bar{T}^{(0)}, \bar{S}^{(0)}$. However, generally, supervised learning here, in identification, is unreasonable for many reasons.

The following observation helped us to find the optimal choice of the identification procedure that can lead to adequate solutions. Based on the above presentation of objects by classes of affine descriptions, it is possible to select not the original precedent-based description for each class, but the affine transformation of such a description, i.e. the most convenient (in some sense, standard) representative. To define what is the standard representative, we should note that for the learning algorithm (4,5), regardless of the initialization $\left\{ \left( p_{k,j}^{(m)}, m_{k,j}^{(m)}, A_{k,j}^{(m)} \right) \right\}$ and iteration number $m$, the following invariant relations take place:

$\Sigma_{j=1}^{N_k} p_{k,j}^{(m)} m_{k,j}^{(m)} = \frac{1}{n} \Sigma_{i=1}^{n} \bar{x}_i = \bar{x}_k$

$\Sigma_{j=1}^{N_k} p_{k,j}^{(m)} \left\{ \left( A_{k,j}^{(m)} \right)^{-1} + \left( m_{k,j}^{(m)} - \bar{x}_k \right) \left( m_{k,j}^{(m)} - \bar{x}_k \right)^T \right\} = \frac{1}{n} \Sigma_{i=1}^{n} (\bar{x}_i - \bar{x}_k) (\bar{x}_i - \bar{x}_k)^T \tag{11}$

with right parts depending only on $\{\bar{x}_i\}$. Particularly, the following identity derives from the second relation in (11):

$Tr \Sigma_{j=1}^{N_k} p_{k,j}^{(m)} \left\{ \left( A_{k,j}^{(m)} \right)^{-1} + \left( m_{k,j}^{(m)} - \bar{x}_k \right) \left( m_{k,j}^{(m)} - \bar{x}_k \right)^T \right\} = \frac{1}{n} \Sigma_{i=1}^{n} (\bar{x}_i - \bar{x}_k)^2 \tag{12}$

where $Tr$ is the trace of the matrix. When transforming $\bar{x}' = (\bar{x} + \tilde{\ell})/s$ with $\tilde{\ell} = -\bar{x}_k$ and $s^2 = \frac{1}{n} \Sigma_{i=1}^{n} (\bar{x}_i - \bar{x}_k)^2$, the representative of the same class will be obtained, but with $\bar{x}_k = 0$ and $Tr = 1$ in (12). We define it as the standard representative of the class. The number of its components $N_k$ will be the same as the precedent’s, probabilities $p_{k,j}$ will be the same, but means and square-form matrices will be, respectively, $\bar{m}_{k,j} = (\bar{m}_{k,j} + \tilde{\ell})/s$ and $A_{k,j} = s^2 A_{k,j}$. As $\tilde{\ell}$ and $s$ are defined by only data
\{\hat{x}_i\}, standard representative parameters can be obtained with the algorithm (4.5), but with the further correction to \(\hat{m}_{k,j}\) and \(A'_{k,j}\).

Having related the standard representative to each class, we get to the following decision for initializing \(\vec{T}(0)\), \(\vec{S}(0)\). Having formed values \(\hat{x} = \frac{1}{n} \sum_{i=1}^{n} \hat{x}_i\) and \(D^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{x}_i - \hat{x}_k)^2\) based on photocount coordinates of the tested PPP \(\{\hat{x}_i\}\), and assuming \(\hat{x} = \left(\hat{x}' + \vec{T}(0)\right)/\vec{S}(0)\) in zero approximation, we will get: \(\vec{S}(0) = D^{-1}\) and \(\vec{T}(0) = \vec{S}(0)\hat{x}\). Tests with relatively simple images have proved the effectiveness of the suggested solution.

5. Results of numerical modeling of algorithms for object identification on photocount images

The above similarity measure calculation procedure is the basic tool for identifying the PPP intensity form. It qualitatively characterizes the match between registration data of the identified PPP and the intensity form of a random precedent. The further identification steps include comparing similarity measures, obtained this way, for various precedents from the database, in order to select the most similar to the identified data.

We should note that for large databases, these scenarios can be rather complex, centered around the internal relations between precedents, the type of description similarity, hierarchical relations, etc. Using this information for the optimal search of the maximum-likelihood precedent may significantly reduce the computing costs. However, this paper is not concerned with this important and interesting topic.

Instead, we present the results of the computing experiment for a small artificially-formed database, in order to illustrate the suggested identification method. Descriptions of precedents from this database were obtained based on data of registering small simple-structure objects. The latter were binary (black-and-white) images from the widely-known base “MPEG-7 Core Experiment CE-Shape-1” [21]. This image base is generally used for the purposes of comparing object form recognition algorithms and includes 70 object categories, 20 object images in each.

From the database [21], images of 5 categories — “apple-10”, “butterfly-15”, “cup-19”, “flatfish-10”, “tree-14” - were randomly selected as precedents. The frames for precedents have been unified in size (256x256 pixels), and precedents were centered in them. As the intensity of precedents was very simple (evenly distributed across the object), precedent’s data registration was also simple: from a large number of randomly-generated points evenly distributed across the frame, 1000 of the points that first got within the object boundaries, were picked. The selected precedents and corresponding registration data (1000 photocounts) are given in figure 2.

![Figure 2. Precedents from the database “MPEG-7 Core Experiment CE-Shape-1” [12] as binary images with the size 256 x 256 (top line) and models of data registered -1000 photocounts evenly distributed across the object fields (bottom line).](image-url)

The standard EM-algorithm (4.5) was applied to form the description of the selected precedents as GMs with \(N_k=10\) components (for all precedents). The EM-algorithm was initialized through even
distribution \( p_{k,j} \), selection of \( N_k \) random photocounts from the registration data \( \{\hat{x}_i\} \) as centers of components \( \hat{m}_{k,j} \) and matrices \( A_{k,j} = E_2/2500 \) equal for all components, where \( E_2 \) is the unit 2 \( \times \) 2 matrix, and the denominator 2500 is equal to the squared standard deviation which is \( \sim \) 10 times lower than the frame size.

The tested objects were images “apple-20”, “butterfly-13”, “cup-7”, “flatfish-3”, “tree-9” from the same 5 categories as precedents but different from them. Their frame sizes were unified as 512x512 pixels and images were displaced from the frame center and scaled out. Data registration simulation was constituted of the set of 300 randomly-generated points, evenly distributed within the object boundaries. Test objects and corresponding low-photocount registration data (300 photocounts) are illustrated in Figure 3.

Figure 3. Test objects from the database “MPEG-7 Core Experiment CE-Shape-1” [12] are from the same categories as Figure 2 precedents, but randomly displaced and scaled-out (top line), and modeled registration data as evenly distributed 300 photocounts (bottom line).

For each of the test objects, the precedent-similarity measure array was formed. Essentially, similarity measures calculated by using the suggested procedure, constitute values of the (logarithmic) likelihood function \( L_k(\{\hat{x}_i\}) \) (6), on the hypothesis that \( \{\hat{x}_i\} \) are the result of registering photocounts of the affine transformation of the \( k \)-precedent with maximum-likelihood transformation parameters \( \vec{t} = \vec{T}(\tau) \) and \( s = \vec{S}(\tau) \) (10).

Figure 4. Descriptions of precedents from the formed database (top line) and their descriptions, recalculated by using the obtained maximum-likelihood parameters \( \vec{T}(\tau) \) and \( \vec{S}(\tau) \), for registration data of the test object "tree-9" (bottom line) (see figure 3). The logarithm of the mean likelihood (lnL) is given below each of the cases.
Descriptions $P_k(\vec{x}|\vec{r},s)$ (8) for $k = \text{"apple-10"}, \text{"butterfly-15"}, \text{"cup-19"}, \text{"flatfish-10"}, \text{"tree-14"}$, recalculated by using $\tilde{T}_k^{(*)}$ and $\tilde{S}_k^{(*)}$ (individual for each of the precedents), and values of the logarithm of the mean likelihood $\tilde{L}_k(\{\vec{x}_i\})$ (6), for "tree-9" registration data, are illustrated in figure 4.

It can be clearly seen from figure 4, how values of the mean likelihood $\ln L_h$ are distributed across the precedents. They reach the peak for "tree", and this is what was expected for a test object of this category. What is more interesting is that other categories close to "tree" are "apple", then "cup". The most distant is "butterfly". This observation gets us to suppose that in large precedent databases, optimal scenarios for comparing similarity measures can be built in order to find precedents that would be the most similar to the identified data.

**Table 1.** Similarity measure (mean likelihood $\ln L_h$) values for test objects of various categories.

| Test objects | 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   |

Complete calculation results of the suggested mean-likelihood procedure, which include other tested objects ("apple-20", "butterfly-13", "cup-7", "flatfish-3") (see Figure 3), are given in table 1. We must note that identification based on the target precedent (category) criteria (7) turned out to be correct in all cases. Certainly, the objective and statistically-proper evaluation of the suggested approach requires testing of larger databases, but the actual results allow us to make optimistic forecasts regarding the potential of this approach.

**6. Conclusions**

This paper shows that formalization of photon-counting sensor registration (detection) of radiation by using the Poisson point processes [15] is the most adequate in terms of physics (quantum theory) [14], and exceptionally productive in terms of statistical data analysis [13]. Based on this, using the machine learning principles [13], we have managed to develop an efficient approach to the synthesis of the object identification procedures that belong to the reliable family of EM-algorithms. The obtained algorithmic architecture of the identification system was inspired by widely-known facts about the vision system; as it turned out, the synthesis results make it possible to significantly imitate the neurobiological mechanisms of visual information processing in the cortex [3-5]. Qualitative modeling showed that the developed identification procedure had a high convergence rate - for complex precedent description by Gaussian mixtures with $N_k \sim 10$ components, the recurring computing operations (9,10) converge in $\sim 10$ iterations.

**7. References**

[1] Nakamura J 2005 Image Sensors and Signal Processing for Digital Still Cameras *Optical Science and Engineering* (CRC Press)
[2] Khurana K, Awasthi R 2013 Techniques for Object Recognition in Images and Multi-Object Detection *Int. J. of Advanced Research in Computer Engineering & Technology* 2(4) 1383-1388
[3] Weiren Shi W, Li Z and Zhong Z A 2009 Survey of Biologically Inspired Image Processing for Objects Recognition *Int. J. Image Graphics* 9 495-510
[4] Yin P, Qiao H and Wu W 2018 A Novel Biologically Inspired Visual Cognition Model: Automatic Extraction of Semantics, Formation of Integrated Concepts, and Reselection Features for Ambiguity *IEEE Trans. Cognitive and Developmental Systems* 10(2) 420-431
[5] Cristobal G, Perrinet L, Matthias S and Keil M S 2015 *Biologically Inspired Computer Vision: Fundamentals and Applications* (Weinheim: Wiley-VCH)
Acknowledgments
The author is grateful to the Russian Foundation for Basic Research (RFBR) for the financial support of the work under grant No. 18-07-01295.