

## **НОВЫЙ, ОРИЕНТИРОВАННЫЙ НА МАШИННОЕ ОБУЧЕНИЕ МЕТОД ИДЕНТИФИКАЦИИ ОБЪЕКТОВ НА БИМЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ, ФОРМИРУЕМЫХ ДАТЧИКАМИ СЧЕТА ФОТОНОВ**

**В. Е. Анциперов**

**Институт радиотехники и электроники им. В.А. Котельникова РАН,  
125009, Москва, ул. Моховая 11-7**

Статья поступила в редакцию 15 ноября 2020 г.

**Аннотация.** В работе рассматривается новый подход к решению задач распознавания, или точнее, идентификации объектов в соответствии с формой их интенсивности на изображениях, формируемых датчиками излучения в режиме счета фотонов или близком к нему. Основная проблема, анализируемая в рамках предложенного подхода, связана с вопросами количественной оценки по зарегистрированной совокупности дискретных фотоотсчетов (~фотонов), степени подобия формы интенсивности излучения объекта на изображении, форме интенсивности ранее наблюдавшихся объектов (прецедентов). Показано, что при аппроксимации формы интенсивности смесью компонент экспоненциального семейства распределений для реализации предложенного в работе подхода может быть синтезирован рекуррентный РМ алгоритм оценивания параметров формы интенсивности по связанной с излучением конечной выборке отсчетов. РМ алгоритм оказывается близким аналогом известного алгоритма К-средних, являющегося одним из наиболее популярных в области машинного (статистического) обучения. Основные черты метода и его реализации проиллюстрированы примером применения метода к задаче идентификации изображений из БД COVID–CT–Dataset.

**Ключевые слова:** машинное обучение по прецедентам, распознавание образов, идентификация объектов на изображениях, компьютерная томография (КТ), метод сегментации К-средних, теория фреймов, пуассоновские процессы.

**Abstract.** The article discusses a new approach to the problem of recognition, or rather the identification of objects by the shape of their intensity in images formed by photon counting detectors or those close to them. The main problem analyzed within the framework of the approach proposed is related to the quantitative estimation of similarity between the form of radiation intensity of the object in the image and the forms of intensity of previously observed objects (precedents), providing that all of them are given by the registered sets of discrete photocounts ( $\sim$  photons). It is shown that when the intensity shape is approximated by a mixture of components of an exponential family, for the implementation of the approach proposed in this work a recurrent PM algorithm for estimating the parameters of the intensity shape from a finite sample of counts associated with radiation can be synthesized. The PM algorithm turns out to be a close analogue of the well-known K-means algorithm, which is one of the most popular in the field of machine (statistical) learning. The main features of the method and its implementation are illustrated by the example of applying the method to the problem of image fragments identification for the COVID – CT – Dataset database.

**Key words:** machine learning by precedents, pattern recognition, object identification in images, computer tomography (CT), K-means segmentation method, frame theory, Poisson point processes.

## **Введение**

Важнейшей отличительной чертой современной медицины является увеличение в ее арсенале удельной доли средств биомедицинской визуализации. Биомедицинская визуализация, понимаемая в широком смысле, включает в себя ряд методов и техник обследования при помощи изображений состояния органов человека и их отделов, скрытых, как правило, от непосредственного (визуального) наблюдения. Успехи, достигнутые в медицине с помощью визуализации, привели к лавинообразному расширению сфер применения последней и к росту спроса как в области исследований, так и в лечебной практике, на специализированные методы формирования

биомедицинских изображений. Растущий спрос на современные методы визуализации привел, в свою очередь, к стимулированию разработок в целом ряде (радио) физических и инженерных направлений. Среди наиболее успешных направлений, завоевавших на сегодняшний день признание широкого медицинского сообщества, следует отметить магнитно-резонансную томографию (МРТ), компьютерную (рентгеновскую) томографию (КТ), позитронно-эмиссионную томографию (ПЭТ), однофотонную эмиссионную компьютерную томографию (ОФЭКТ) и т.д. [1].

Параллельное существование целого ряда популярных методов визуализации связано в первую очередь с избирательностью соответствующих техник по отношению к разным видам обследуемых тканей. Так, к примеру, МРТ-изображения чувствительны к мягким тканям, а рентгеновские изображения более чувствительны к твердым и костным структурам. Однако, несмотря на различие физических принципов, лежащих в основе большинства методов визуализации, общей их чертой является достаточно низкий уровень используемого рабочего излучения. Это ведет к тому, что датчики существующих устройств визуализации формируют изображения почти в режиме счета фотонов (photon-counting detectors – PCDs), или в близком к нему режиме накопления фотонов (energy-integrating detectors – EIDs) [2]. Режим PCD в полной мере реализуется при ПЭТ и ОФЭКТ визуализации, в случае же КТ имеет место режим накопления фотонов EIDs. Последний отличается от PCD тем, что в режиме EID для гарантии достаточного отношения сигнал / шум (SNR), требуется примерно  $\sim 10^3$  фотонов на пиксель. Однако, недавние разработки энергочувствительных PCD [3] открыли новые возможности и для КТ. Ожидается, что для получения качественных рентгеновских изображений энергочувствительными PCD будет достаточно  $\sim 10$  фотонов на пиксель (для заданной спектральной зоны), а в перспективе ожидается режим 1:1.

Ввиду отмеченной специфичности PCD изображений, формируемых современными устройствами медицинской визуализации, традиционные методы их анализа и распознавания [4] оказываются недостаточно

эффективными. Соответственно, в этой области остро ощущается потребность в новых методах и подходах. В представленной работе предлагается один из таких, ориентированных в первую очередь на РСД изображения подходов. Он связан с задачами распознавания, или, точнее, с идентификацией объектов на таких изображениях. Идентификация понимается здесь как сопоставление объектов по такой характеристике, как форма интенсивности их излучения на чувствительной поверхности устройства визуализации. Основная проблема, анализируемая в рамках предложенного подхода, связана с вопросами, как по зарегистрированной совокупности дискретных фотоотсчетов (далее отсчетов) могут быть получены обоснованные выводы относительно степени подобия формы интенсивности излучения тестируемого объекта форме ранее наблюдавшихся объектов (прецедентов). Для этих целей далее, в разделе 1, обосновывается статистическая модель РСД изображений заданных выборкой случайных отсчетов. В частности, аргументируется  $n$ -модель совместного их распределения в виде факторизованного произведения распределений отдельных отсчетов, совпадающая с популярной в задачах машинного обучения моделью наивного байесова подхода. В разделе 2 основная проблема формализуется как задача статистической классификации / идентификации выборки отсчетов тестируемого объекта по отношению к уже имеющемуся конечному множеству аналогичных выборок отсчетных данных (прецедентов). На этой основе вводится мера подобия пары выборок отсчетных данных и с ее помощью в общем виде формулируется *метод максимального подобия*. Далее обсуждается уточнение общей формулировки для случая параметрических (байесовых) моделей отсчетных данных и в асимптотическом случае – в случае больших объемов выборок обсуждаются параллели предлагаемого метода с методом максимального правдоподобия Р. Фишера. Для придания методу формы вычислительного алгоритма, в разделе 3 уточняется параметрическая модель вероятностных распределений отсчетов, с помощью которой, в разделе 4 на основе уточненной формулировки метода максимального подобия синтезируются конкретные процедуры идентификации объектов. В разделе 5

основные черты метода и его упрощенной реализации проиллюстрированы примером применения к задаче идентификации изображений из БД COVID–CT–Dataset.

## 1. Представление РСД изображений выборками случайных отсчетов

Для обоснования предлагаемого способа представления изображений выборками случайных отсчетов, напомним основные факты теории взаимодействия излучения с веществом и сформулируем на их основе представление об идеальном изображении. Для простоты изложения воспользуемся полуклассической картиной механизма детектирования излучения материальными (квантовыми) датчиками [5].

Детектирование электромагнитного излучения (света) материальными чувствительными датчиками, например, полупроводниковыми детекторами с  $p - n$  переходами, приводит к появлению электрического тока в виде потока высвободившихся фотоэлектронов, называемых фотоотсчетами или просто отсчетами. В полуклассическом приближении число отсчетов  $K$  случайно и описывается пуассоновским распределением вероятностей [5]:

$$P(K) = \frac{1}{K!} (\alpha W)^K \exp\{-\alpha W\}. \quad (1)$$

где  $W = Nh\bar{\nu} = IAT$  – полная энергия падающего на чувствительную поверхность детектора площади  $A$  за время  $T$  излучения интенсивности  $I$ , что в терминах количества фотонов составляет  $N$  квантов со средней энергией  $h\bar{\nu}$ , падающих на ту же площадку, за то же время, где  $h$  – постоянная Планка,  $\bar{\nu}$  – средняя частота падающего излучения. Размерный параметр  $\alpha$  в (1) с точностью до коэффициента  $(h\bar{\nu})^{-1}$  совпадает с безразмерным параметром квантовой эффективности материала детектора  $\eta$  – отношением среднего числа регистрируемых отсчетов  $\bar{K} = \alpha W$  к числу падающих фотонов  $N$ .

Отметим, что в (1) неявно предполагается постоянство интенсивности  $I$  по поверхности детектора и ее неизменность за время  $T$ . В случае не строго постоянной интенсивности распределение (1) будет выполняться лишь

приближенно, однако приближение будет тем точнее, чем меньше площадь  $A$  и/или короче время экспозиции  $T$ . В частности, это справедливо при  $A \rightarrow 0$  – в модели “точечного детектора”. При этом, однако, к нулю будет стремиться также и среднее число отсчетов  $\bar{K}$ , что приводит к бинарному режиму регистрации, в котором события  $K \geq 2$  имеют исчезающую вероятность по отношению к  $K = 0, 1$ . В бинарном режиме распределение (1) переходит в распределение Бернулли, где событием успеха является регистрация единичного отсчета:

$$P_{\vec{x}}(K) = \begin{cases} \alpha W(\vec{x}), & K = 1; \\ 1 - \alpha W(\vec{x}), & K = 0; \end{cases} \quad (2)$$

В распределении (2) явно указана зависимость энергии падающего излучения  $W(\vec{x}) = I(\vec{x})AT$  от координат  $\vec{x}$  точечного детектора (координат зарегистрированного отсчета). Заметим, что распределение (2) может интерпретироваться как альтернативное определение модели идеального точечного детектора.

С помощью модели идеального точечного детектора (2), может быть определена модель идеального изображения [6]. Именно, идеальное изображение есть результат регистрации за время экспозиции  $T$  излучения интенсивности  $I(\vec{x})$ , падающего на чувствительную поверхность  $\Omega$  идеального устройства визуализации, которое, в свою очередь, представляет собой массив (матрицу) большого числа расположенных вплотную друг к другу в области  $\Omega$  идентичных (с одинаковыми  $A$  и  $\eta$ , в частности) точечных детекторов. Результатом регистрации – идеальным изображением  $X$  является, таким образом, множество  $N$  координат  $\vec{x}_i$  независимых отсчетов, связанных с  $N$  сработавшими точечными детекторами идеального устройства. Далее для краткости совокупность  $X = \{\vec{x}_i\}$  будем называть множеством отсчетов идеального изображения, или просто идеальным отсчетным изображением.



пространственная плотность экспонированных зерен бромистого серебра (зерен металлического серебра  $\sim 40$  нм) кодирует пространственное распределение регистрируемой интенсивности  $I(\vec{x})$ ,  $\vec{x} \in \Omega$  [7].

Принимая во внимание, что при заданной интенсивности  $I(\vec{x})$  отсчеты разных точечных детекторов идеального устройства независимы, стандартными методами, исходя из модели (2), выводится, что плотность совместного распределения вероятностей числа отсчетов  $N$  и их координат  $\{\vec{x}_1, \dots, \vec{x}_N\}$  задается многомерной пуассоновской плотностью [9]:

$$p(N, \vec{x}_1, \dots, \vec{x}_N | I(\vec{x})) = \frac{1}{N!} [\prod_{i=1}^N \alpha T I(\vec{x}_i)] \exp \left\{ -\alpha T \int_{\Omega} I(\vec{x}) d\vec{x} \right\}. \quad (3)$$

Другими словами, отсчеты  $X = \{\vec{x}_i\}$  идеальных изображений представляют собой реализации на  $\Omega$  точечного пуассоновского процесса (ТПП) с функцией интенсивности  $\lambda(\vec{x}) = \alpha T I(\vec{x})$  [10]. Заметим, что маргинальное распределение числа отсчетов  $N$ , получающееся интегрированием (3) по всем координатам  $\{\vec{x}_i\}$ , оказывается в точности пуассоновским (1) (при переобозначении  $N \rightarrow K$ ).

Отметим, что модель (3) не является новой, она с успехом использовалась для обработки изображений в специальных случаях, связанных с низкими интенсивностями регистрируемых излучений  $I(\vec{x})$ . Традиционными сферами применения (3) являются флуоресцентная микроскопия, позитронно-эмиссионная томография (ПЭТ), однофотонная эмиссионная компьютерная томография (ОФЭКТ), оптическая и инфракрасная астрономия и т.д. [11]. Объясняется это тем простым фактом, что в этих случаях изначально справедливо приближение  $\bar{K} = \alpha I A T \rightarrow 0$ , которое, однако, достигается не в пределе  $A \rightarrow 0$ , а за счет  $I \rightarrow 0$ .

Модель идеального изображения (3) имеет достаточно общий характер, однако, в вопросах ее практического использования для случаев низких и средних интенсивностей регистрируемых излучений есть одно важное отличие. Оно заключается в том, что число отсчетов  $N$  для низкоинтенсивного случая может быть хоть и большим, но поддающимся учету, для случая же средних интенсивностей это число огромно (формальным критерием разделения



случаев низкой и средней интенсивности может быть, например,  $N \sim 10^6$  – число отсчетов стандартного ПЭТ изображения). По этой причине практическая работа непосредственно с моделью идеального изображения  $X = \{\vec{x}_i\}$  представляется проблематичной. Предлагаемое нами решение проблемы состоит в следующем – рассматривая модель идеального изображения  $X = \{\vec{x}_i\}$ ,  $i = 1, \dots, N$  только в качестве концепции некоторой “генеральной совокупности” отсчетов, использовать для рабочего представления изображения только относительно небольшую выборку из них  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ , некоторого приемлемого фиксированного размера  $n \ll N$ . Маргинальное распределение выборки заданного размера  $n$  легко находится интегрированием (3) по оставшимся координатам  $\{\vec{x}_{n+1}, \dots, \vec{x}_N\}$  и суммированием полученного результата по  $N \geq n$ . Окончательный результат вычислений имеет следующий вид факторизованного произведения ( $n \ll \bar{N}$ ):

$$\begin{aligned} q_n(\vec{x}_1, \dots, \vec{x}_n | I(\vec{x})) &= \prod_{i=1}^n \rho(\vec{x}_i | I(\vec{x})), \\ \rho(\vec{x} | I(\vec{x})) &= \frac{I(\vec{x})}{\int_{\Omega} I(\vec{x}) d\vec{x}}. \end{aligned} \quad (4)$$

Вероятностная модель (4) представления изображений наборами (выборками) отсчетов  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  обсуждалась ранее в работе [10], где подчеркивался ряд её полезных свойств. Именно, поскольку при заданном  $n$  отсчеты  $\{\vec{x}_1, \dots, \vec{x}_n\}$  составляют набор независимых, одинаково распределенных (iid) случайных векторов на  $\Omega$ , эта выборка является типичным примером исходных данных для многих хорошо разработанных статистических методов, например, для наивного байесова [12]. Более того, поскольку распределения отдельных отсчетов  $\rho(\vec{x} | I(\vec{x}))$  (4) зависит только от формы интенсивности  $I(\vec{x})$  – от ее нормированной версии, оно тем самым независимо ни от квантовой эффективности материала детекторов  $\eta$ , ни от средней частоты  $\bar{\nu}$  излучения, ни от площади  $A$  детекторов, ни от времени экспозиции  $T$ . Ввиду перечисленных свойств, представления изображений в виде выборок отсчетов  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$

имеют универсальный характер и поэтому предлагаются в качестве статистической основы формулируемого ниже подхода.

Ещё одна важная тема, связанная с предложенным представлением изображений в виде выборки отсчетов, касается вопросов формирования этого представления. В общем случае, формально реализация представления  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  состоит в некотором выборе  $n$  значений iid случайных отсчетов, связанных, в соответствии с (4), с нормированной интенсивностью  $I(\vec{x})$  на  $\Omega$ . В случае, когда выборка формируется в процессе регистрации излучения, отсчеты  $X_n$  отбираются из потока фотоэлектронов по мере их эмиссии, пока выборка не достигнет размера  $n$ . Если же изображение уже задано в виде набора пикселей, то известными методами по ним можно интерполировать (восстановить) приближенные значения  $I(\vec{x})$  во всех точках  $\Omega$ . Нормируя это приближение, можно получить аппроксимацию  $\rho(\vec{x}|I(\vec{x}))$  (4), после чего остается произвести процедуру семплирования из него  $n$  отсчетов, чтобы сформировать требуемое представление.

В области машинного обучения существует большая область методов семплирования, объединенных общим названием методов Монте-Карло [13]. Она включает такие известные методы, как семплирование по значимости, семплирование с отклонением, алгоритмы семплирования Метрополиса–Гастингса, Гиббса и т.д. Выбирая подходящий метод можно оптимизировать процедуру семплирования с позиций вычислительной эффективности, репрезентативности, специфики задачи и т.д. Отметим в этой связи работу [14], в которой отмечена интересная ассоциация методов семплирования по значимости с популярным алгоритмом полутонного FM представления изображений – алгоритмом Флойда–Стейнберга [15]. Формально эта аналогия объясняется тем, что полутонное FM представление изображений [16] в определенном смысле аналогично предлагаемому представлению в виде выборок отсчетов  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ , хотя и исходит из совершенно других предпосылок, связанных с особенностями визуального восприятия. Отметим в этой связи, что FM представление, в свою очередь, может служить

дополнительной интерпретацией предложенного представления изображений в виде выборок случайных отсчетов.

## 2. Метод максимального подобия

Формально метод максимального подобия может быть сформулирован безотносительно к происхождению совокупности наблюдаемых данных  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ , представленных  $n$  независимыми, одинаково распределенными (iid) случайными векторами некоторой размерности  $d$ :  $\vec{x}_i \in \mathbb{R}^d$ . Тем не менее, далее в работе будем интерпретировать  $X_n$  как соответствующую некоторому изображению выборку отсчетов, т.е. предполагать  $d = 2$  и  $\vec{x}_i \in \Omega \subset \mathbb{R}^2$ .

Основным статистическим предположением в отношении наблюдений является допущение о существовании некоторого параметрического семейства плотностей распределения вероятностей  $\rho(\vec{x} | \vec{\theta})$ ,  $\vec{\theta} \in \Theta \subset \mathbb{R}^p$ , описывающих случайный характер отдельных данных  $\vec{x}$ . Данное семейство плотностей будем называть моделью наблюдений. Кроме того, следуя байесовской точке зрения, будем предполагать, что параметры  $\vec{\theta}$  также являются случайными величинами, имеющими некоторое априорное распределение с плотностью  $\mathcal{P}(\vec{\theta})$ , точный вид которого, впрочем, не конкретизируется. Оба сделанных предположения позволяют ввести в рассмотрение семейство плотностей совместных распределений отдельных наблюдений  $\vec{x}$  и параметров  $\vec{\theta}$  –  $\rho(\vec{x}; \vec{\theta}) = \rho(\vec{x} | \vec{\theta})\mathcal{P}(\vec{\theta})$ , называемое далее моделью задачи.

Представленная модель задачи  $\rho(\vec{x}; \vec{\theta})$ ,  $\vec{\theta} \in \Theta$  в терминах задач машинного обучения приводит к классу так называемых порождающих (generative) моделей [17], предполагающих соответствие определенных значений параметров  $\vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_k, \dots \in \Theta$  некоторым наблюдаемым объектам ( $\{\vec{\theta}_k\}$  в определенном смысле можно рассматривать как метки классов соответствующих объектов [18]). В случае, когда значения параметров  $\{\vec{\theta}_k\}$  точно не известны, порождающие модели позволяют с помощью данных

наблюдений снизить большую априорную неопределенность в отношении параметров до существенно меньшей их апостериорной неопределенности. В частности, на основе данных наблюдений  $X^k$  от  $k$ -го объекта, могут быть сформированы статистические оценки  $\vec{\theta}_k^{(*)}$  параметров  $\vec{\theta}_k$ , интервалы неопределенности которых существенно меньше, к примеру, ширины распределения  $\mathcal{P}(\vec{\theta})$ . В качестве таких оценок могут использоваться, в частности, оценки апостериорного максимума [17]:

$$\vec{\theta}_k^{(MAP)} = \underset{\vec{\theta} \in \Theta}{\operatorname{arg\,max}} \varrho(\vec{\theta} | X^k) = \underset{\vec{\theta} \in \Theta}{\operatorname{arg\,max}} \frac{\varrho(X^k; \vec{\theta})}{\varrho(X^k)}, \quad (5)$$

или часто близкие к ним оценки максимального правдоподобия  $\vec{\theta}_k^{(ML)}$  [19]:

$$\vec{\theta}_k^{(ML)} = \underset{\vec{\theta} \in \Theta}{\operatorname{arg\,max}} \varrho(X^k | \vec{\theta}) = \underset{\vec{\theta} \in \Theta}{\operatorname{arg\,max}} \frac{\varrho(X^k; \vec{\theta})}{\mathcal{P}(\vec{\theta})}, \quad (6)$$

где  $\varrho(X; \vec{\theta})$  – порождающая модель данных  $X$  и параметров  $\vec{\theta}$ . Если количество наблюдений  $n$  в  $X$  известно и известны их значения:  $X = X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ , то в рассматриваемом случае (iid) случайных наблюдений порождающая модель  $\varrho(X_n; \vec{\theta})$  просто связана с заданной моделью  $\rho(\vec{x} | \vec{\theta})$  отдельных наблюдений  $\vec{x}$ :

$$\varrho(X_n; \vec{\theta}) = \prod_{i=1}^n \rho(\vec{x}_i | \vec{\theta}) \mathcal{P}(\vec{\theta}). \quad (7)$$

Задача идентификации, соотнесения наблюдаемого, тестируемого объекта  $O$  с одним из ранее зарегистрированных, обучающих объектов (называемых в дальнейшем прецедентами) формализуется в контексте представленной модели (7) как задача максимизации некоторой меры подобия  $\mu(X^O, X^k)$  наблюдаемых данных от объекта  $X^O$  с аналогичными, полученными ранее в отношении прецедентов наборами данных  $X^1, X^2, \dots, X^k, \dots$ . Поскольку никаких дополнительных знаний об объекте и прецедентах, кроме данных наблюдений, не предполагается (в том числе о значениях параметров  $\vec{\theta}_O, \vec{\theta}_1, \vec{\theta}_2, \dots, \vec{\theta}_k, \dots \in \Theta$ , характеристиках  $\mathcal{P}(\vec{\theta})$  и т.д.), весьма желательно, чтобы соответствующая мера подобия  $\mu(X^O, X^k)$  была бы выражена в терминах этих и только этих данных.

Достаточно естественной количественной характеристикой согласованности данных  $X^O$  и произвольного набора наблюдений  $X^k$  является плотность их совместного распределения вероятности  $p(X^O, X^k)$ , в предположении, что обе выборки обусловлены одним и тем же неизвестным объектом. С учетом модели (7) (а также с учетом условной независимости наборов  $X^O$  и  $X^k$  при заданном  $\vec{\theta}$ )  $p(X^O, X^k)$  можно записать в следующем виде:

$$p(X^O, X^k) = \int p(X^O, X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} = \int \varrho(X^O | \vec{\theta}) \varrho(X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} = \int \prod_{i=1}^n \rho(\vec{x}_i | \vec{\theta}) \prod_{j=1}^{n_k} \rho(\vec{x}_j | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} = \int \varrho(X^O \cup X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} \quad (8)$$

где  $\{\vec{x}_i\}$  – набор из  $n$  наблюдений  $X^O$ ,  $\{\vec{x}_j\}$  – набор из  $n_k$  наблюдений  $X^k$ ,  $X^O \cup X^k$  – полученный объединением  $X^O$  и  $X^k$  набор из  $(n + n_k)$  наблюдений. Заметим, что соотношение (8) устанавливает равенство  $p(X^O, X^k)$  вероятности объединенного набора наблюдений  $\{\vec{x}_i\} \cup \{\vec{x}_j\}$  в  $(n + n_k)$ -модели (7), см. Рис.2 в качестве иллюстрации эффекта объединения наборов наблюдений от объекта  $O$  и от прецедентов в разной степени ему подобных.

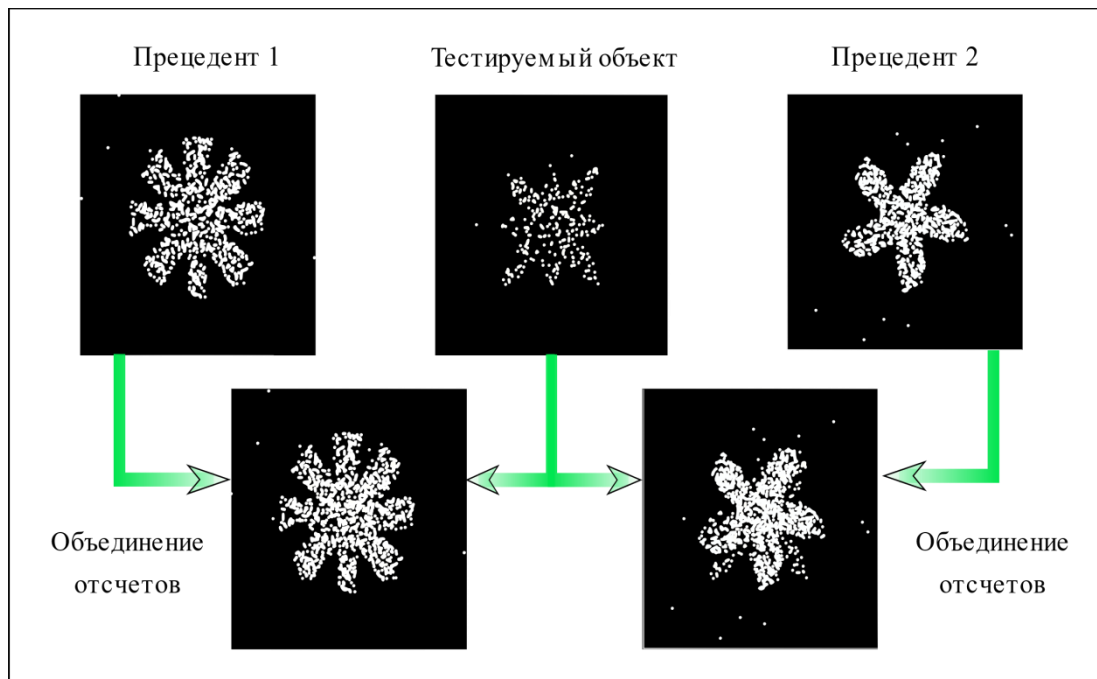


Рис.2. Объединение наборов наблюдений от объекта  $O$  и от прецедентов 1 и 2 в разной степени подобных тестируемому объекту.

Из приведенного замечания следует, что величина  $p(X^0, X^k)$  могла бы служить для сравнения согласованности  $X^0$  с различными  $X^1, X^2, \dots, X^k, \dots$ , если бы последние были одного размера  $n_1 = n_2 = \dots = m$ . Действительно, в этом случае все наборы  $\{(X^0 \cup X^k)\}$  можно было бы считать случайными выборками в рамках одной и той же  $(n + m)$ -модели (7) и по величине их безусловной вероятности  $p(X^0 \cup X^k)$  судить о том, какой из наборов  $X^0 \cup X^k$  более вероятен. Проблема, однако, состоит в том, что в силу произвольности  $n_k$  наборы  $X^0 \cup X^k$  принадлежат к разным моделям и, следовательно, сравнение значений  $p(X^0 \cup X^k)$  для разных  $X^k$  должно быть скорректировано на это обстоятельство.

В предлагаемом *методе максимального подобия* соответствующая коррекция задается посредством нормировки значений  $p(X^0, X^k)$  (8) на вероятности  $p(X^k)$  :

$$\mu(X^0, X^k) = \frac{p(X^0, X^k)}{p(X^k)} = p(X^0 | X^k), \quad (9)$$

т.е. мера подобия  $\mu(X^0, X^k)$  выбирается в виде отношения вероятности выборки  $X^0 \cup X^k$  к вероятности выборки  $X^k$ .

Как следует из (9), выбранное отношение вероятностей формально совпадает с условной вероятностью  $X^0$  при имеющихся наблюдениях  $X^k$ . Другими словами при произвольных  $X^k$  мера  $\mu(X^0, X^k)$  (9) при  $X^0 = \{\vec{x}_1, \dots, \vec{x}_n\}$  определяется в рамках одной и той же  $n$ -модели (7). Последнее ведет к корректному сравнению значений меры на множестве прецедентов  $X^1, X^2, \dots, X^k, \dots$  имеющих произвольные размеры.

Используя выбранную меру подобия наблюдаемых данных объекта и прецедентов  $\mu(X^0, X^k)$  (9), метод максимального подобия можно формализовать как поиск решения следующего уравнения максимального подобия:

$$k_{MS} = \underset{k}{\operatorname{argmax}} \mu(X^0, X^k) = \underset{k}{\operatorname{argmax}} p(X^0 | X^k), \quad (10)$$

Уравнения (10) задают следующую интерпретацию предлагаемого метода: наблюдаемый объект идентифицируется с тем прецедентом, условие наблюдения данных  $X^k$  которого приводят к максимальной вероятности (согласованности) следующих за ними данных  $X^0$  от объекта.

Для соотнесения метода максимального подобия с известными методами, рассмотрим асимптотический случай больших выборок  $n_k \gg 1$  прецедентов. Условная вероятность  $p(X^0 | X^k)$  (9) имеет в параметрической модели (7) вид:

$$p(X^0 | X^k) = \frac{p(X^0, X^k)}{p(X^k)} = \frac{\int \varrho(X^0 | \vec{\theta}) \varrho(X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta}}{p(X^k)}. \quad (11)$$

В Приложении показано, что при  $n_k \gg 1$  для  $\varrho(X^k | \vec{\theta})$  в окрестности оценки максимального правдоподобия  $\vec{\theta}_k^{(ML)}$  (6) имеет место асимптотика (П.6):

$$\varrho(X^k | \vec{\theta}) \approx \varrho(X^k | \vec{\theta}_k^{(ML)}) \exp \left\{ -\frac{n_k}{2} (\vec{\theta} - \vec{\theta}_k^{(ML)})^T I(\vec{\theta}_k^{(ML)}) (\vec{\theta} - \vec{\theta}_k^{(ML)}) \right\}, \quad (12)$$

где  $T$  – операция транспонирования,  $I(\vec{\theta})$  – информационная матрица Фишера для распределения  $\rho(\vec{x} | \vec{\theta})$  (П.4). С учетом остроты пика асимптотики (12) в окрестности  $\vec{\theta}_k^{(ML)}$  числитель (11) можно аппроксимировать посредством:

$$\begin{aligned} \int \varrho(X^0 | \vec{\theta}) \varrho(X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} &\approx \varrho(X^0 | \vec{\theta}_k^{(ML)}) \int \varrho(X^k | \vec{\theta}) \mathcal{P}(\vec{\theta}) d\vec{\theta} = \\ &= \varrho(X^0 | \vec{\theta}_k^{(ML)}) p(X^k) \end{aligned}, \quad (13)$$

что приведет к упрощению  $p(X^0 | X^k)$  (11) и, соответственно, к упрощению меры подобия  $\mu(X^0, X^k)$  (9) до:

$$\mu(X^0, X^k) \cong \varrho(X^0 | \vec{\theta}_k^{(ML)}). \quad (14)$$

Другими словами, в случае больших наборов наблюдений  $X^1, X^2, \dots, X^k, \dots$ ,  $n_k \gg 1$  (число наблюдений  $n$ -набора  $X^0$  идентифицируемого объекта не обязано быть большим) в качестве меры подобия  $\mu(X^0, X^k)$  можно использовать значение плотности условного распределения  $X^0$ , задаваемого в рамках  $n$ -модели задачи (7), при значениях параметра  $\vec{\theta} = \vec{\theta}_k^{(ML)}$ . Отметим, что это обстоятельство явным образом демонстрирует сделанное выше замечание о  $\mu(X^0, X^k)$  как о мере, определяемой на основе общей  $n$ -модели задачи.

С прикладной точки зрения полученное выражение (14) для меры подобия  $\mu(X^0, X^k)$ , очевидно, более конструктивно, чем общее выражение (9). Если связать метод максимального подобия с этой мерой, то он примет следующую форму:

$$k_{MS} = \underset{k}{\operatorname{argmax}} \varrho \left( X \mid \vec{\theta}_k^{(ML)} \right) = \underset{\vec{\theta} \in \{\vec{\theta}_k^{(ML)}\}}{\operatorname{argmax}} \varrho(X \mid \vec{\theta}), \quad (15)$$

переходящую в исходную общую формулировку (10) в асимптотическом пределе  $n_k \gg 1$ . Заметим, что в этой альтернативной формулировке метод имеет сходную с методом максимального правдоподобия Р. Фишера (6) форму с тем лишь отличием, что максимизация производится не по всем  $\vec{\theta} \in \Theta$ , а лишь по конечному набору оценок  $\vec{\theta}_1^{(ML)}, \vec{\theta}_2^{(ML)}, \dots, \vec{\theta}_k^{(ML)}, \dots \in \Theta$ .

Преимуществом формулировки (15) метода максимального подобия является также то обстоятельство, что он не требует хранения наборов наблюдений прецедентов  $X_1, X_2, \dots, X_k, \dots$  в полном объеме – достаточно хранить лишь построенные по ним оценки параметров (статистики)  $\{\vec{\theta}_k^{(ML)}\}$ .

### 3. Параметрическая модель смесей экспоненциального типа для интенсивностей регистрируемых излучений от объектов

С тем, чтобы на основе предлагаемого метода (15) синтезировать алгоритмические (компьютерные) процедуры идентификации объектов необходимо явно определить параметрическую модель наблюдений  $\rho(\vec{x} \mid \vec{\theta})$ . Поскольку, согласно модели идеального изображения (4) из раздела 1, плотность распределения вероятностей  $\rho(\vec{x} \mid \vec{\theta})$  совпадает с нормированной версией интенсивности регистрируемого излучения, эта задача сводится к выбору параметрического семейства интенсивностей. В качестве такого семейства – модели интенсивностей – предлагается использовать достаточно широкое семейство положительных смесей компонент (аддитивных наборов кадров) экспоненциального семейства, которое строится следующим образом.



Пусть дано изображение, на котором тестируемый объект  $O$  представлен результатами регистрации падающего от него на фоточувствительную поверхность  $\Omega \subset \mathbb{R}^2$  излучения с интенсивностью  $I_O(\vec{x})$ . Аппроксимируем  $I_O(\vec{x})$  параметрической моделью  $I(\vec{x}|\vec{\theta}_O)$ ,  $\vec{\theta}_O \in \Theta \subset \mathbb{R}^P$ , представляющей собой сумму, смесь из  $K$  перекрывающихся кадров, компонент вида  $\{F_j(\vec{x}|\vec{\mu}_j, \vec{\theta}_O)\}$ ,  $j = 1, \dots, K$ , расположенных в узлах  $\{\vec{\mu}_j\}$  некоторой воображаемой регулярной решетки, накрывающей  $\Omega$ :

$$I(\vec{x}|\vec{\theta}_O) = \sum_{j=1}^K F_j(\vec{x}|\vec{\mu}_j, \vec{\theta}_O). \quad (16)$$

В качестве компонент  $F_j(\vec{x}|\vec{\mu}_j, \vec{\theta}_O)$  в (16) возьмем смещенные в узлы решетки  $\{\vec{\mu}_j\}$  копии некоторого базового кадра  $F(\vec{x}|\vec{\xi}) \geq 0$ ,  $\vec{\xi} \in \mathbb{R}^p$ ,  $p = P/K$ , заданные с точностью до значений параметров  $\vec{\xi}$ . Область  $\Delta \subset \mathbb{R}^2$ , в которой  $F(\vec{x}|\vec{\xi}) \neq 0$  будем называть носителем базового кадра или базовым носителем. Носитель  $\Delta$  предполагается симметричным в том смысле, что он содержит начало координат  $\vec{x} = \vec{0}$  и вместе с каждым  $\vec{x} \in \Delta$  содержит  $-\vec{x}$ . Простейшим примером подобного носителя является квадратная область с центром в начале координат, как это показано на Рис.3.

Обозначим посредством  $\vec{\xi}_j$  значения параметров кадра–компоненты смеси, расположенной в узле  $j$ . С учетом сделанных предположений модель (16) теперь уточняется в виде:

$$I(\vec{x}|\vec{\theta}_O) = \sum_{j=1}^K F(\vec{x} - \vec{\mu}_j|\vec{\xi}_j), \quad (17)$$

где полный набор параметров  $\vec{\theta}_O$  представлен совокупностью  $\{\vec{\xi}_1, \dots, \vec{\xi}_K\}$  и учтено, что  $j$ -я компонента смеси зависит только от части параметров – от  $\vec{\xi}_j$ .

В модели (17) носители компонент смеси  $\{\Delta_j\}$  – смещенные на  $\{\vec{\mu}_j\}$  копии базового носителя  $\Delta$ , предполагаются частично перекрывающимися. Это обеспечивается, в частности, требованием  $D > d$ , где  $D$  – характерный размер  $\Delta$ , а  $d$  – шаг решетки. Считая последнее требование выполненным, получим,

что объединение носителей  $\{\Delta_j\}$  полностью покрывают область  $\Omega$  изображения. Простейшим примером подобного покрытия  $\Omega$  являются квадратные носители  $\{\Delta_j\}$ , центры которых расположены в узлах  $\{\vec{\mu}_j\}$  прямоугольной решетки, как это представлено на Рис.3.

При полном покрытии  $\Omega$  совокупностью  $\{\Delta_j\}$  каждая точка  $\vec{x} \in \Omega$  принадлежит по крайней мере одному из носителей. Поэтому множество узлов, носители которых содержат  $\vec{x}$ , не пусто. Обозначим множество индексов этих узлов посредством  $\delta_{\vec{x}} = \{j\}$  и будем называть его решеточным окружением  $\vec{x}$ . В силу симметричности базового носителя  $\Delta$ , узлами в  $\delta_{\vec{x}}$  являются те, которые содержатся в области, полученной смещением в точку  $\vec{x}$  базового носителя  $\Delta$  (см. Рис.3). Решеточные окружения  $\delta_{\vec{x}}$  будут использованы при построении процедур идентификации в следующем разделе.

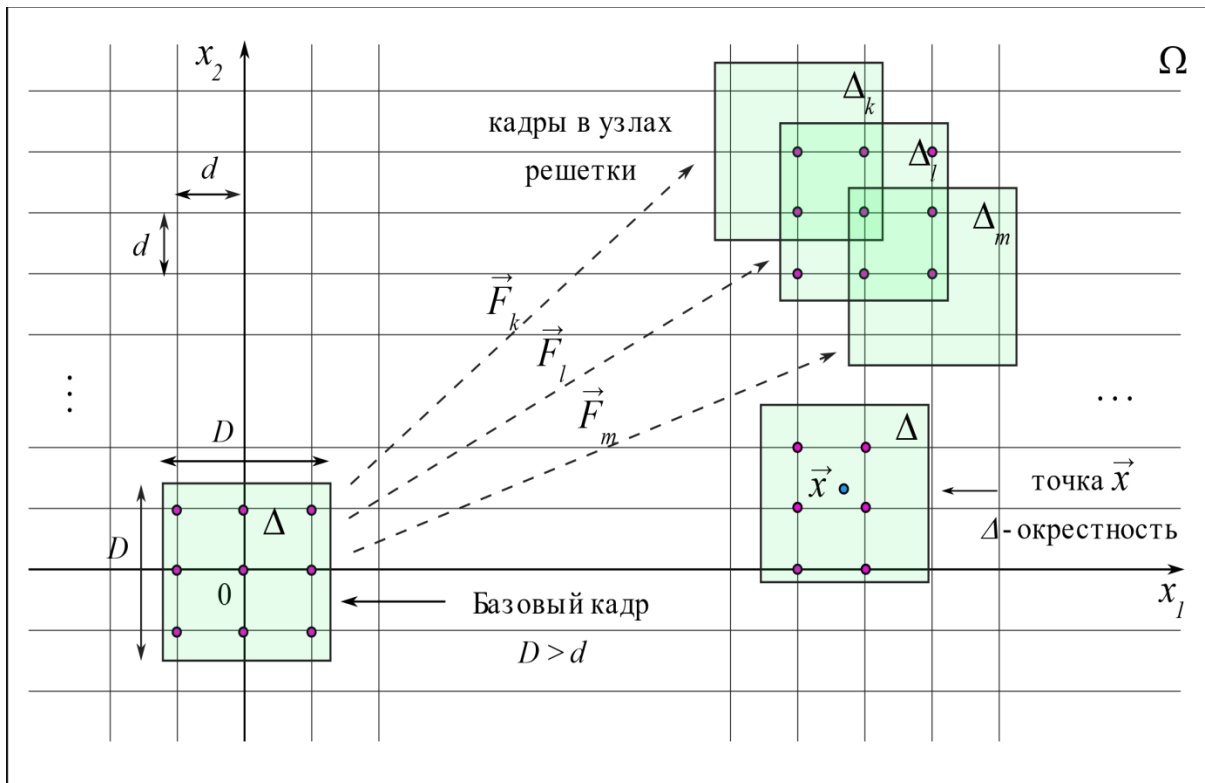


Рис.3. Простейшая геометрия расположения носителей компонент  $\{\Delta_j\}$  в узлах  $\{\vec{\mu}_j\}$  покрывающей  $\Omega$  решетки. Показана  $\Delta$ -окрестность произвольной точки  $\vec{x} \in \Omega$  и принадлежащие ей узлы решетки.

Дальнейшее уточнение модели (17) связано с выбором вида зависимости базового кадра  $F(\vec{x} | \vec{\xi})$  от параметров  $\vec{\xi} = (\xi_0, \xi_1, \dots, \xi_{p-1})^T$ . Простейшим видом такой зависимости могла бы быть линейная комбинация с коэффициентами  $\vec{\xi}$  элементов некоторого функционального базиса  $\{\varphi_0(\vec{x}), \varphi_1(\vec{x}), \dots, \varphi_{p-1}(\vec{x})\}$  по образу, например, принятому в теории фреймов [20]. Однако, в этом случае для обеспечения положительности базового кадра  $F(\vec{x} | \vec{\xi}) > 0$  потребовались бы существенные ограничения как на базис  $\{\varphi_q(\vec{x})\}$ , так и на параметры  $\vec{\xi}$ . Поэтому более подходящим представляется разложение по функциональному базису не самого кадра, а его логарифма:

$$F(\vec{x} | \vec{\xi}) = I_0 \exp\left\{\sum_{q=0}^{p-1} \xi_q \varphi_q(\vec{x})\right\}, \vec{x} \in \Delta, \quad (18)$$

где множитель  $I_0$  введен для обеспечения правильной размерности кадра. Имея в виду последующую нормировку интенсивности, в качестве  $I_0$  можно взять, например, среднюю интенсивность излучения объекта на поверхности  $\Omega$ :

$$I_0 = \frac{1}{\Sigma_\Omega} \iint_\Omega I_0(\vec{x}) d\vec{x}, \quad (19)$$

где  $\Sigma_\Omega$  – площадь области  $\Omega$ .

Естественно потребовать от параметрического семейства  $F(\vec{x} | \vec{\xi})$  (18), чтобы оно содержало как минимум все константы  $I(\vec{x}) \equiv I > 0$ . Для этого следует допустить в качестве одной из базисных функций, например  $\varphi_0(\vec{x})$ , постоянную:  $\varphi_0(\vec{x}) \equiv 1$ . При равенстве всех прочих параметров  $\vec{\xi}$  кроме  $\xi_0$  нулю из (19) следует, что  $I(\vec{x}) \equiv I = I_0 \exp\{\xi_0\}$  действительно задает постоянную интенсивность, а  $\xi_0$  задает ее нормировку по отношению к  $I_0$  (19). Ввиду последующего перехода к нормированной версии интенсивности  $I(\vec{x} | \vec{\theta}_0)$  (17), удобно вместо  $\xi_0$  ввести другой параметр  $w = w(\vec{\xi})$  (являющийся функцией всех  $\vec{\xi}$ ), имеющий смысл доли от общей энергии  $W_0 = I_0 \Sigma_\Omega$ , падающей за единицу времени на  $\Omega$ , той энергии, которая приходится на кадр (18) при заданных значениях параметров  $\vec{\xi}$ :

$$\begin{aligned}
 w(\vec{\xi}) &= \frac{1}{W_0} \iint_{\Delta} F(\vec{x} | \vec{\xi}) d\vec{x} = \\
 &= \exp\{\xi_0\} \frac{1}{\Sigma_{\Omega}} \iint_{\Delta} \exp\{\sum_{q=1}^{p-1} \xi_q \varphi_q(\vec{x})\} d\vec{x} =, \\
 &= \exp\{\xi_0\} \exp\{A(\xi_1, \dots, \xi_{p-1})\}
 \end{aligned}
 \tag{20}$$

где введена связанная с базисом  $\{\varphi_q(\vec{x})\}$  вспомогательная (кумулянтная) функция

$$A(\xi_1, \dots, \xi_{p-1}) = \ln \left\{ \frac{1}{\Sigma_{\Omega}} \iint_{\Delta} \exp\{\sum_{q=1}^{p-1} \xi_q \varphi_q(\vec{x})\} d\vec{x} \right\}.
 \tag{21}$$

Осуществляя в соответствии с (20) замену параметров  $\vec{\xi} \rightarrow (w, \eta_1, \dots, \eta_{p-1})$ ,  $w = w(\vec{\xi})$ ,  $\eta_q = \xi_q$ ,  $q = 1, \dots, p-1$ , приходим к следующему представлению базового кадра (18):

$$F(\vec{x} | w, \vec{\eta}) = I_0 w \exp\{\vec{\eta} \vec{\varphi}(\vec{x}) - A(\vec{\eta})\} P_{\Delta}(\vec{x}),
 \tag{22}$$

где введены обозначения  $\vec{\eta} = (\eta_1, \dots, \eta_{p-1})^T$ ,  $\vec{\varphi}(\vec{x}) = (\varphi_1(\vec{x}), \dots, \varphi_{p-1}(\vec{x}))^T$  и для автоматического учета ограничения  $\vec{x} \in \Delta$  введена индикаторная функция базового носителя:  $P_{\Delta}(\vec{x}) = 1$ ,  $\vec{x} \in \Delta$  и  $P_{\Delta}(\vec{x}) = 0$ ,  $\vec{x} \notin \Delta$ .

С учетом выбранной структуры базового кадра  $F(\vec{x} | w, \vec{\eta})$  (22), параметрическая модель интенсивности  $I(\vec{x} | \vec{\theta}_0)$  (17) принимает следующий окончательный вид:

$$I(\vec{x} | \vec{\theta}_0) = I_0 \sum_{j=1}^K w_j \exp\{\vec{\eta}_j \vec{\varphi}(\vec{x} - \vec{\mu}_j) - A(\vec{\eta}_j)\} P_{\Delta}(\vec{x} - \vec{\mu}_j),
 \tag{23}$$

где  $w_j$  и  $\vec{\eta}_j$  – вес и нормальные параметры  $j$ -й компоненты,  $\vec{\theta}_0 = \{(w_j, \vec{\eta}_j)\}$  – полный набор параметров модели регистрируемой интенсивности,  $I_0$  и  $A(\vec{\eta})$  определены, соответственно, выражениями (19) и (21).

Отметим, что формально интеграл по  $\Omega$  от правой части (23) дает с учетом (21) величину  $I_0 \Sigma_{\Omega} \sum_{j=1}^K w_j$ , откуда следует свойство нормированности на единицу весов  $\{w_j\}$ , совпадающее с их трактовкой как распределения суммарной энергии излучения  $W_0 = I_0 \Sigma_{\Omega}$  по отдельным компонентам.

#### 4. Метод максимального подобия в случае моделирования интенсивностей регистрируемых излучений моделями смесей экспоненциального типа

Используя принятую параметрическую модель интенсивности регистрируемого излучения (23), запишем плотность распределения вероятности (4) отдельного отсчета–наблюдения из представляющей изображение объекта  $O$  выборки  $X^O = X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  в развернутом виде:

$$\begin{aligned} \rho(\vec{x}|\vec{\theta}_O) &= \frac{I_O(\vec{x})}{\int_{\Omega} I_O(\vec{x})d\vec{x}} = \frac{I(\vec{x}|\vec{\theta}_O)}{I_O\Sigma_{\Omega}} = \\ &= \frac{1}{\Sigma_{\Omega}} \sum_{j \in \delta_{\vec{x}}} w_j \exp\{\vec{\eta}_j \vec{\varphi}(\vec{x} - \vec{\mu}_j) - A(\vec{\eta}_j)\} \end{aligned} \quad (24)$$

где учтено, что при заданном  $\vec{x}$  формальное суммирование по всем компонентам  $j$  от 1 до  $K$  в (23) фактически сводится к суммированию по ненулевым в точке  $\vec{x}$  компонентам – по решеточному окружению  $\delta_{\vec{x}} = \{j\}$ , для которого  $P_{\Delta}(\vec{x} - \vec{\mu}_j) = 1$ .

Отметим, что плотность  $\rho(\vec{x}|\vec{\theta}_O)$  (24) с параметрами смеси  $\vec{\theta}_O = \{(w_j, \vec{\eta}_j)\}$  совпадают по структуре со смесями плотностей распределений экспоненциального семейства, хорошо изученными в теории статистических выводов [17]. В этой связи отметим тот факт, что для плотностей экспоненциального семейства наряду с нормальными параметрами  $\vec{\eta}_j$  иногда используются и двойственные им параметры средних  $\vec{\psi}_j$ , связанные с преобразованием Лежандра  $A(\vec{\eta}_j) \rightarrow \hat{A}(\vec{\psi}_j)$  [17] кумулянтной функции:

$$\begin{cases} \hat{A}(\vec{\psi}_j) = \vec{\eta}_j \vec{\psi}_j - A(\vec{\eta}_j), \\ \vec{\psi}_j = \nabla_{\vec{\eta}} A(\vec{\eta}_j) = \frac{1}{\Sigma_{\Omega}} \iint_{\Delta} \vec{\varphi}(\vec{x}) \exp\{\vec{\eta}_j \vec{\varphi}(\vec{x}) - A(\vec{\eta}_j)\} d\vec{x} \end{cases} \quad (25)$$

В [21] показано, что для минимальных (с регулярными базисами  $\{\varphi_q(\vec{x})\}$ ) экспоненциальных семейств связь параметров  $\vec{\eta}_j$  и  $\vec{\psi}_j$  является взаимно-однозначной (обратимой). Более того, сами параметры  $\vec{\eta}_j$  являются двойственными к  $\vec{\psi}_j$  по отношению к преобразованию Лежандра  $\hat{A}(\vec{\psi}_j) \rightarrow A(\vec{\eta}_j)$ , т.е. выражаются через них в виде  $\vec{\eta}_j = \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j)$ .

Соответствующая (24)  $n$ -модель для выборок отсчетов изображений  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  получается непосредственной подстановкой  $\rho(\vec{x}_i | \vec{\theta}_0)$  (24) на место сомножителей  $\rho(\vec{x}_i | I(\vec{x}))$  в (4). После перемножения всех  $K$  слагаемых во всех  $n$  сомножителях, плотность совместного распределения вероятностей отсчетов  $\{\vec{x}_i\}$  принимает вид:

$$\begin{aligned} \varrho(X_n | \vec{\theta}_0) &= \prod_{i=1}^n \rho(\vec{x}_i | \vec{\theta}_0) = \\ &= \frac{1}{\Sigma_{\Omega}^n} \sum_{j_1 \in \delta_{\vec{x}_1}} \dots \sum_{j_n \in \delta_{\vec{x}_n}} \prod_{i=1}^n [w_{j_i} \exp\{\vec{\eta}_{j_i} \vec{\varphi}(\vec{x}_i - \vec{\mu}_{j_i}) - A(\vec{\eta}_{j_i})\}] \end{aligned} \quad (26)$$

где индексы  $j_i \in \delta_{\vec{x}_i}$  ассоциируют отсчеты  $\vec{x}_i$  с узлами  $j$  решеточного окружения  $\delta_{\vec{x}_i}$ , суммирование производится по всем возможным  $n$ -наборам  $\{j_1, \dots, j_n\} \in \delta_{\vec{x}_1} \times \dots \times \delta_{\vec{x}_n} = \delta_X$ .

Заметим, что каждый из  $n$ -наборов  $\{j_1, \dots, j_n\} \in \delta_X$  задает разбиение выборки  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  на  $K$  не пересекающихся поднаборов  $\{\pi_j\}$ , также ассоциированных с узлами  $j$ :  $\pi_j = \{\vec{x}_i | j_i = j\}$ ,  $X_n = \bigcup_{j=1}^K \pi_j$ . Если в слагаемых-произведениях (26) объединить показатели экспонент и привести подобные относительно параметров  $w_j$  и  $\vec{\eta}_j$  члены, то плотность  $\varrho(X_n | \vec{\theta}_0)$  (26) можно переписать в виде суммы по всем разбиениям:

$$\varrho(X_n | \vec{\theta}_0) = \frac{1}{\Sigma_{\Omega}^n} \sum_{\{j_1, \dots, j_n\} \in \delta_X} \left[ \exp \left\{ -H_{\{\pi_j\}}(w_1, \dots, w_K) + \sum_{j=1}^K \bar{p}_j \vec{\eta}_j \vec{\Phi}_{\pi_j} - \Gamma_{\{\pi_j\}}(\vec{\eta}_1, \dots, \vec{\eta}_K) \right\} \right]^n, \quad (27)$$

где  $\bar{p}_j = n_j/n$  – выборочные веса узлов,  $n_j = |\pi_j|$  – число отсчетов  $\vec{x}_i$  в части  $\pi_j$  разбиения,  $\sum_{j=1}^K n_j = n$ ,  $\sum_{j=1}^K \bar{p}_j = 1$  и  $H_{\{\pi_j\}}$ ,  $\{\vec{\Phi}_{\pi_j}\}$ ,  $\Gamma_{\{\pi_j\}}$  определяются разбиением  $\{\pi_j\}$  следующим образом:

$$\begin{aligned} H_{\{\pi_j\}}(w_1, \dots, w_K) &= - \sum_{j=1}^K \bar{p}_j \ln w_j \\ \vec{\Phi}_{\pi_j} &= \frac{1}{n_j} \sum_{\vec{x}_i \in \pi_j} \vec{\varphi}(\vec{x}_i - \vec{\mu}_j) \quad , \\ \Gamma_{\{\pi_j\}}(\vec{\eta}_1, \dots, \vec{\eta}_K) &= \sum_{j=1}^K \bar{p}_j A(\vec{\eta}_j) \end{aligned} \quad (28)$$

Теоретически, при заданной, представляющей изображение тестируемого объекта выборке  $X_n$  и известных максимально-правдоподобных оценках параметров прецедентов  $\{\vec{\theta}_k^{(ML)}\}$ ,  $n$ -модель (26) / (27) вполне однозначно задает

схему вычисления меры подобия  $\mu(X^0, X^k) \cong \varrho(X_n | \vec{\theta}_0 = \vec{\theta}_k^{(ML)})$  (14). Однако, практическая реализация этой схемы сталкивается с проблемой вычисления большого  $\sim |\delta_X| \approx k^n$  числа слагаемых, где  $k > 1$  – средний размер решеточного окружения  $\delta_{\vec{x}}, \vec{x} \in \Omega$ . В случае, когда размер базового носителя  $D$  заметно превосходит шаг решетки  $d$ ,  $k \approx (D/d)^2 \gg 1$ , число слагаемых становится чрезвычайно большим, а схема – *EXP*-сложной по  $n$  проблемой [22].

В связи с отмеченной проблемой, актуальной представляется задача аппроксимации  $\varrho(X_n | \vec{\theta}_0)$  (27) более простым для вычислений выражением. Одна из возможностей, очевидно, состоит в аппроксимации суммы (27) отдельным слагаемым. Основной вопрос заключается в том, какое из слагаемых для этого выбрать. Поскольку каждое слагаемое находится во взаимно-однозначном соответствии с  $n$ -набором индексов отсчетов  $\{j_1, \dots, j_n\} \in \delta_X$ , а каждый такой набор – во взаимно-однозначном соответствии с разбиением  $\{\pi_j\}$  выборки  $X_n$ , вопрос о аппроксимирующем слагаемом эквивалентен вопросу о выборе подходящего разбиения  $\{\pi_j^*\}$ , задаваемого набором  $\{j_1^*, \dots, j_n^*\}$ .

Однако, прежде чем решать вопрос о выборе подходящего разбиения  $\{\pi_j^*\}$ , приведем необходимые сведения о оптимальных значениях параметров  $\vec{\theta}_0$ , максимизирующих слагаемые в (27). Максимум любого слагаемого в (27) по параметрам  $\{w_j\}$  достигается в точке минимума перекрестной (cross) энтропии [21]  $H_{\{\pi_j\}}(w_1, \dots, w_K) \geq 0$ , которая, в свою очередь имеет условный минимум в точке  $\{\bar{p}_j\}$  выборочных весов (при нормировке  $\{w_j\}$  на единицу). Таким образом, соответствующие значения оптимальных весов  $\{w_j\} \subset \vec{\theta}_0$  равны выборочным весам разбиения  $\{\pi_j\}$ :  $w_j = \bar{p}_j = |\pi_j|/n, j = 1, \dots, K$ . Оптимальные значения нормальных параметров  $\{\vec{\eta}_j\}$  также можно получить из условия минимальности выражений в показателе экспонент слагаемых (27). Именно, если параметры  $\{\vec{\eta}_j\}$  не связаны ограничениями, уравнения для их оптимальных значений можно получить дифференцированием соответствующих выражений и приравнованием нулю соответствующих градиентов. С учетом (25) и (28)

легко вывести, что эти уравнения имеют вид уравнений для средних параметров  $\vec{\psi}_j = \vec{\Phi}_{\pi_j}$ , откуда, с учетом двойственности  $\vec{\eta}_j = \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j)$  могут быть найдены и значения нормальных оптимальных параметров  $\{\vec{\eta}_j\} \subset \vec{\theta}_0$ . Таким образом, с каждым разбиением  $\{\pi_j\}$  связаны оптимальные значения параметров  $\vec{\theta}_{\{\pi_j\}}$ , доставляющих максимум по  $\vec{\theta}_0$  соответствующему слагаемому в (27). Отметим, что вычисление этих параметров по разбиению  $\{\pi_j\}$  осуществляется очень простыми средствами.

Опираясь на приведенные выше сведения, будем осуществлять выбор подходящего разбиения  $\{\pi_j^*\}$  таким образом, чтобы связанные с ним оптимальные параметры  $\vec{\theta}_{\{\pi_j^*\}}$  доставляли бы соответствующему слагаемому в (27) самый большой максимум среди максимумов всех других слагаемых:

$$\{\pi_j^*\} = \arg \max_{\{j_1, \dots, j_n\} \in \delta_X} S_{\{\pi_j\}}(\vec{\theta}_{\{\pi_j\}}), \quad (29)$$

$$S_{\{\pi_j\}}(\vec{\theta}) = \exp \left\{ \begin{array}{l} -H_{\{\pi_j\}}(w_1, \dots, w_K) + \\ + \sum_{j=1}^K \bar{p}_j \vec{\eta}_j \vec{\Phi}_{\pi_j} - \Gamma_{\{\pi_j\}}(\vec{\eta}_1, \dots, \vec{\eta}_K) \end{array} \right\}.$$

Заметим, что в (29) можно обойтись без явного вычисления параметров  $\vec{\eta}_j = \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j)$ , если выразить показатели экспонент  $S_{\{\pi_j\}}(\vec{\theta})$  через средние параметры  $\vec{\psi}_j$  и заменить кумулянтные функции  $A(\vec{\eta}_j)$  преобразованиями Лежандра  $\hat{A}(\vec{\psi}_j)$  (25):

$$\begin{aligned} \sum_{j=1}^K \bar{p}_j \vec{\eta}_j \vec{\Phi}_{\pi_j} - \Gamma_{\{\pi_j\}}(\vec{\eta}_1, \dots, \vec{\eta}_K) &= \sum_{j=1}^K \bar{p}_j \left[ \vec{\eta}_j \vec{\Phi}_{\pi_j} - A(\vec{\eta}_j) \right] = \\ &= \sum_{j=1}^K \bar{p}_j \left[ \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j) \left( \vec{\Phi}_{\pi_j} - \vec{\psi}_j \right) + \hat{A}(\vec{\psi}_j) \right] = \\ &= - \sum_{j=1}^K \bar{p}_j D_{B\hat{A}} \left( \vec{\Phi}_{\pi_j}, \vec{\psi}_j \right) + \sum_{j=1}^K \bar{p}_j \hat{A} \left( \vec{\Phi}_{\pi_j} \right) \end{aligned} \quad (30)$$

где введена дивергенция Брегмана [23]:

$$D_{B\hat{A}}(\vec{\chi}_j, \vec{\psi}_j) = \hat{A}(\vec{\chi}_j) - \hat{A}(\vec{\psi}_j) - \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j)(\vec{\chi}_j - \vec{\psi}_j). \quad (31)$$



Обоснованием предлагаемого выбора подходящего разбиения  $\{\pi_j^*\}$  (29) является тот факт, что при  $n \gg 1$  соответствующее этому разбиению слагаемое в некоторой окрестности своего оптимального  $\vec{\theta}_{\{\pi_j^*\}}$  определяет асимптотику всей плотности  $\varrho(X_n | \vec{\theta}_0)$  (27). Действительно, если выделить из суммы (27) слагаемое, соответствующее  $\{\pi_j^*\}$ , то в обозначениях (29) получим:

$$\frac{\varrho(X_n | \vec{\theta}_0)}{[S_{\{\pi_j^*\}}(\vec{\theta}_0)/\Sigma_\Omega]^n} = 1 + \sum_{\{\pi_j\} \neq \{\pi_j^*\}} [S_{\{\pi_j\}}(\vec{\theta}_0)/S_{\{\pi_j^*\}}(\vec{\theta}_0)]^n. \quad (32)$$

Поскольку в некоторой окрестности  $\vec{\theta}_{\{\pi_j^*\}}$  в соответствии с (29):

$$S_{\{\pi_j\}}(\vec{\theta}_0) < S_{\{\pi_j\}}(\vec{\theta}_{\{\pi_j\}}) < S_{\{\pi_j^*\}}(\vec{\theta}_{\{\pi_j^*\}}) \sim S_{\{\pi_j^*\}}(\vec{\theta}_0), \quad (33)$$

то в сумме (32) все слагаемые есть  $n$ -ые степени не превосходящих  $\varepsilon_{\{\pi_j\}} = S_{\{\pi_j\}}(\vec{\theta}_{\{\pi_j\}})/S_{\{\pi_j^*\}}(\vec{\theta}_{\{\pi_j^*\}}) < 1$  отношений. Отсюда следует, что если обозначить среднее степени  $n$  от всех  $k^n$  величин  $\{\varepsilon_{\{\pi_j\}}\}$  посредством  $\varepsilon_n < 1$ , то сумма в (32) может быть оценена сверху величиной  $[k\varepsilon_n]^n$ . При  $k\varepsilon_n < 1$  с ростом числа отсчетов в выборке  $X_n$  эта сумма будет стремиться к нулю и, следовательно, будет иметь место асимптотика  $\varrho(X_n | \vec{\theta}_0)/[S_{\{\pi_j^*\}}(\vec{\theta}_0)/\Sigma_\Omega]^n \sim 1$ .

Что касается самой процедуры выбора подходящего разбиения  $\{\pi_j^*\}$ , как решения задачи (29), то понятно, что непосредственное сравнение максимумов всех  $k^n$  слагаемых не приемлемо, поскольку также представляет собой *EXP*-сложную задачу [22]. К счастью, для решения задач оптимизации, подобных (29), существуют вполне эффективные процедуры [24]. Ниже предлагается одна из таких процедур – рекуррентная сегментация выборки  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ , подобная известному методу сегментации *K*–средних (*K*–means) [24]. Эта процедура состоит в последовательных (рекуррентных) итерациях двух основных шагов – шага *P* уточнения подходящего разбиения  $\{\pi_j^{(v)}\}$  и шага *M* – вычисления соответствующих оптимальных параметров  $\vec{\theta}_{\{\pi_j^{(v)}\}}$ .

Уточнение разбиения  $\{\pi_j^{(v)}\}$  на шаге Р при найденных на предыдущей итерации параметрах  $\vec{\theta}_{\{\pi_j^{(v-1)}\}}$  осуществляется как решение для каждого отсчета  $\vec{x}_i \in X_n$  следующей задачи на максимум:

$$j_i = \arg \max_{j \in \delta_{\vec{x}_i}} \left[ w_j^{(v-1)} \exp \left\{ \vec{\eta}_j^{(v-1)} \vec{\varphi}(\vec{x}_i - \vec{\mu}_j) - A(\vec{\eta}_j^{(v-1)}) \right\} \right]. \quad (34)$$

Решение  $j_i$  каждой из задач (34) выделяет в сумме (26) подмножество тех слагаемых, у которых  $i$ -й сомножитель больше, чем у остальных. После того, как найдены индексы для всех отсчетов  $\{\vec{x}_1, \dots, \vec{x}_n\}$ ,  $n$ -набор  $\{j_1, \dots, j_n\} \in \delta_X$  задаст некоторое уточненное разбиение  $\{\pi_j^{(v)}\}$ , соответствующее слагаемое которого для данных значений параметров  $\vec{\theta}_{\{\pi_j^{(v-1)}\}}$  будет максимальным.

Отметим, что если оптимальные параметры  $\vec{\theta}_{\{\pi_j^{(v-1)}\}}$  заданы через средние параметры  $\vec{\psi}_j^{(v-1)}$ , то заменяя  $A(\vec{\eta}_j^{(v-1)})$  на её преобразование Лежандра  $\hat{A}(\vec{\psi}_j^{(v-1)})$  (25), можно получить двойственную формулировку задач (34):

$$j_i = \arg \max_{j \in \delta_{\vec{x}_i}} \left[ w_j^{(v-1)} \exp \left\{ \vec{\eta}_j^{(v-1)} \left[ \vec{\varphi}(\vec{x}_i - \vec{\mu}_j) - \vec{\psi}_j^{(v-1)} \right] + \hat{A}(\vec{\psi}_j^{(v-1)}) \right\} \right]. \quad (35)$$

где  $\vec{\eta}_j^{(v-1)} = \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j^{(v-1)})$ .

При найденном на шаге Р разбиении  $\{\pi_j^{(v)}\}$  нахождение на шаге М соответствующих оптимальных параметров  $\vec{\theta}_{\{\pi_j^{(v)}\}}$  уже обсуждалось выше и в компактном виде представляет собой вычисление для каждого  $j = 1, \dots, K$  величин:

$$\begin{aligned} w_j^{(v)} &= \bar{p}_j^{(v)} = |\pi_j^{(v)}|/n, \\ \vec{\psi}_j^{(v)} &= \vec{\Phi}_{\pi_j^{(v)}} = \frac{1}{|\pi_j^{(v)}|} \sum_{\vec{x}_i \in \pi_j^{(v)}} \vec{\varphi}(\vec{x}_i - \vec{\mu}_j), \\ \vec{\eta}_j^{(v)} &= \nabla_{\vec{\psi}} \hat{A}(\vec{\psi}_j^{(v)}). \end{aligned} \quad (36)$$

Пересчитанные значения параметров  $\vec{\theta}_{\{\pi_j^{(v)}\}}$  (36) задают для соответствующего слагаемого значение максимума  $[S_{\{\pi_j^{(v)}\}}(\vec{\theta}_{\{\pi_j^{(v)}\}})/\Sigma_\Omega]^n$ , которое в силу пошагового увеличения значений будет больше или равно предыдущего  $[S_{\{\pi_j^{(v-1)}\}}(\vec{\theta}_{\{\pi_j^{(v-1)}\}})/\Sigma_\Omega]^n$ .

Отметим, что вычисление нормальных параметров  $\vec{\eta}_j^{(k)}$  на шаге М (36) можно не проводить, а перенести их на шаг Р следующей итерации, при этом можно воспользоваться двойственной формулировкой (35) задачи на максимум.

При вычислительной организации представленной рекуррентной процедуры отдельное внимание следует уделить процессу ее инициализации – заданию некоторого начального допустимого разбиения  $\{\pi_j^{(0)}\}$ . Очевидно, в зависимости от задачи, тестируемого объекта, имеющихся прецедентов и т.д. здесь возможен очень большой выбор вариантов. Простейшим, возможно не лучшим, будет случайный выбор индексов  $j_i$  из соответствующих решеточных окружений  $\delta_{x_i}$  отсчетов для инициализации  $\{j_1, \dots, j_n\} \in \delta_X$ .

Вопрос о завершении процедуры решается так же, как в случае сегментации K–средних, т.е. определяется критерием стабилизации разбиений:  $\{\pi_j^{(v-1)}\} = \{\pi_j^{(v)}\} = \{\pi_j^*\}$ .

Аргументация вопросов сходимости разбиений  $\{\pi_j^{(v)}\} \rightarrow \{\pi_j^*\}$  также во многом повторяет обоснование сходимости метода K–средних [21] и основывается на монотонном возрастании соответствующих максимальных значений  $[S_{\{\pi_j^{(v)}\}}(\vec{\theta}_{\{\pi_j^{(v)}\}})/\Sigma_\Omega]^n$ . Вместе со сходимостью разбиений происходит и сходимость параметров  $\vec{\theta}_{\{\pi_j^{(v)}\}} \rightarrow \vec{\theta}_{\{\pi_j^*\}} = \vec{\theta}^*$ , однако из сходимости не следует, что предел совпадает с оценкой максимального правдоподобия  $\vec{\theta}_O^{(ML)}$  (6). С уверенностью можно утверждать лишь то, что  $[S_{\{\pi_j^*\}}(\vec{\theta}^*)/\Sigma_\Omega]^n$  будет в

некотором смысле лучшей оценкой снизу величины максимального правдоподобия  $\varrho(X_n | \vec{\theta}_0^{(ML)})$ .

После того, как на основе представленной рекуррентной процедуры (34) / (35), (36), найдено подходящее разбиение  $\{\pi_j^*\}$  и соответствующие ему параметры  $\vec{\theta}^*$ , однозначно находится и аппроксимирующая плотность  $\varrho(X_n | \vec{\theta}_0)$  (26) / (27) слагаемое, которое с учетом всех сделанных выше замечаний, может быть записано как произведение следующих сомножителей:

$$\begin{aligned} \varrho(X_n | \vec{\theta}_0) \sim & \left[ \frac{1}{\Sigma_\Omega} \exp \left\{ \hat{\Gamma}_{\{\pi_j^*\}} - \hat{H}_{\{\pi_j^*\}} \right\} \right]^n \times \\ & \times \left[ \exp \left\{ -D_{KL}(\{w_j^*\}, \{w_j\}) - D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j\}) \right\} \right]^n, \end{aligned} \quad (37)$$

где  $\hat{\Gamma}_{\{\pi_j^*\}}$  представляет собой, подобно  $\Gamma_{\{\pi_j\}}(\vec{\eta}_1, \dots, \vec{\eta}_K) = \sum_{j=1}^K \bar{p}_j A(\vec{\eta}_j)$  (28), среднее с весами  $\{w_j^*\}$  значений преобразований Лежандра  $\hat{A}(\vec{\psi}_j)$  в точках  $\{\vec{\psi}_j^*\}$ :

$$\hat{\Gamma}_{\{\pi_j^*\}} = \sum_{j=1}^K \bar{w}_j^* \hat{A}(\vec{\psi}_j^*), \quad (38)$$

$\hat{H}_{\{\pi_j^*\}}$  – энтропия Шеннона [13] для распределения выборочных весов  $\{w_j^*\}$ :

$$\hat{H}_{\{\pi_j^*\}} = - \sum_{j=1}^K w_j^* \ln w_j^*, \quad (39)$$

$D_{KL}(\{w_j^*\}, \{w_j\})$  – дивергенция Кульбака-Лейблера [21], равная перекрестной энтропии  $H_{\{\pi_j\}}(\{w_j\})$  (28) за вычетом энтропии Шеннона  $\hat{H}_{\{\pi_j^*\}}$  (39):

$$D_{KL}(\{w_j^*\}, \{w_j\}) = - \sum_{j=1}^K w_j^* \ln \frac{w_j}{w_j^*}, \quad (40)$$

$D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j\})$  – среднее с весами  $\{w_j^*\}$  значение дивергенций Брегмана  $D_{B\hat{A}}(\vec{\chi}_j, \vec{\psi}_j)$  (31) на множестве пар  $(\vec{\psi}_j^*, \vec{\psi}_j)$ :

$$D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j\}) = \sum_{j=1}^K w_j^* D_{B\hat{A}}(\vec{\psi}_j^*, \vec{\psi}_j). \quad (41)$$

В отношении найденной аппроксимации (37) сделаем важное замечание. Для ее вычисления, как это следует из (38–41), сверх значений найденных оптимальных параметров  $\vec{\theta}^* = \{(w_1^*, \vec{\psi}_1^*), \dots, (w_K^*, \vec{\psi}_K^*)\}$  более ничего не требуется – эти параметры полностью задают статистическое описание задачи.

В отношении же оптимальных параметров  $\vec{\theta}^*$  заметим, что, будучи связаны с подходящим разбиением  $\{\pi_j^*\}$  в силу уравнений (36) шага М, они сами однозначно определяют это разбиение, для чего достаточно повторить Р-шаг (34) / (35) и учесть, что  $\{\pi_j^{(v+1)}\} = \{\pi_j^*\}$ . Тем самым устанавливается эквивалентность способов задания аппроксимации (37) как с помощью подходящего разбиения  $\{\pi_j^*\}$ , так и с помощью оптимальных параметров  $\vec{\theta}^*$ .

Аппроксимация (37) для плотности распределения  $\varrho(X_n | \vec{\theta}_0)$  автоматически задает, согласно (14), подходящую аппроксимацию основного инструмента метода максимального подобия – меру подобия  $\mu(X^0, X^k)$ . Этого, в принципе, достаточно для реализации вычислительной схемы метода (15). Однако, поскольку речь идет об эффективных вычислениях, обратим внимание на следующее. В конечном счете, для предлагаемого метода (15) не столь важны значения меры подобия  $\mu(X^0, X^k)$  сами по себе, сколько важен характер их изменения – возрастания / убывания по отношению к параметрам  $\vec{\theta}_k^{(ML)}$  прецедентов. С этой точки зрения поиск максимально-подобного прецедента  $X^k$  эквивалентным образом можно осуществлять на основе всякой другой меры, являющейся монотонной функцией от  $\mu(X^0, X^k) \cong \varrho(X_n | \vec{\theta}_k^{(ML)})$ . Вопросы же предпочтения какой-то одной из них привносятся исходя из дополнительных требований, которыми могут быть и требования эффективных вычислений.

Для выбора более эффективной с вычислительной точки зрения меры подобия, нежели (37), построим ее монотонную функцию  $\tilde{\mu}(X^0, X^k)$  следующим образом. Отбросим первый, вообще не зависящий от параметров  $\vec{\theta}_0 = \{(w_1, \vec{\psi}_1), \dots, (w_K, \vec{\psi}_K)\}$  сомножитель в (37), а от второго возьмем натуральный логарифм, после чего сократим его на общий множитель  $n$ . В результате получается следующая, эквивалентная  $\mu(X^0, X^k)$  мера подобия:

$$\tilde{\mu}(X^0, X^k) = -[D_{KL}(\{w_j^*\}, \{w_j^k\}) + D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j^k\})], \quad (42)$$

где  $\vec{\theta}_k^{(ML)} = \{(w_1^k, \vec{\psi}_1^k), \dots, (w_K^k, \vec{\psi}_K^k)\}$

Заданная посредством (42) мера подобия  $\tilde{\mu}(X, X_k)$  представляет собой сумму с обратным знаком двух дивергенций – Кульбака-Лейблера  $D_{KL}(\{w_j^*\}, \{w_j\})$  (40) и среднюю Брегмана  $D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j\})$  (41). Поскольку обе дивергенции всегда положительны, за исключением равного нулю минимума при равенстве своих аргументов, мера  $\tilde{\mu}(X, X_k)$ , напротив, является не положительной и имеет максимум при  $\vec{\theta}_k^{(ML)} = \vec{\theta}^*$ .

Таким образом, при моделировании интенсивностей регистрируемых излучений моделями смесей экспоненциального типа, метод максимального подобия сводится к выбору минимальной суммарной дивергенции между оптимальными параметрами тестируемого объекта  $\vec{\theta}^*$  и максимально-правдоподобными параметрами прецедентов  $\vec{\theta}_k^{(ML)}$ . Обозначая сумму дивергенций (40), (41) посредством  $\bar{\mu}(X, X_k) = -\tilde{\mu}(X, X_k)$ , можно переформулировать метод максимального подобия (15) для данного случая следующим образом:

$$\begin{aligned} k_{MS} &= \underset{k}{\operatorname{argmin}} \bar{\mu}(X, X_k) = \\ &= \underset{k}{\operatorname{argmin}} [D_{KL}(\{w_j^*\}, \{w_j^k\}) + D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j^k\})], \end{aligned} \quad (43)$$

который буквально представляет собой критерий минимального различия.

Отметим, что в минимизируемом выражении  $\bar{\mu}(X, X_k)$  (43) группы параметров  $\{w_j^k\}$  и  $\{\vec{\psi}_j^k\}$ , распределенны по разным слагаемым. Это позволяет организовать оптимизацию (43) в два этапа: сначала на основе приемлемых значений дивергенции Кульбака–Лейблера  $D_{KL}(\{w_j^*\}, \{w_j^k\})$  отобрать узкий класс допустимых прецедентов, а затем уже среди них на основе средней дивергенции Брегмана выбрать максимально подобный тестируемому объекту. Такая организация процедуры идентификации, очевидно, также позволяет заметно сократить вычисления.

Другое направление экономии вычислений связано с тем, что как дивергенция Кульбака–Лейблера (40) так и средняя дивергенция Брегмана (41) являются взвешенными суммами по всем узлам решетки. В тех узлах, где веса

$w_j^* \approx 0$ , соответствующие слагаемые обращаются в ноль и для фактического вычисления обеих дивергенций требуется меньше слагаемых, чем число  $K$  узлов решетки. Поэтому дополнительной экономии вычислений можно добиться, если с самого начала провести цензурирование узлов в соответствии со значениями параметров  $\{w_j^*\}$ .

Подводя итогу рассмотрений текущего раздела целесообразно подчеркнуть, что использование совместно с методом максимального подобия модели интенсивностей в виде смесей экспоненциального типа позволяет не только получить явные уравнения для схемы алгоритмической реализации метода, но и определить пути их эффективных вычислений. Важность последнего обстоятельства трудно преувеличить – вычислительная неэффективность ряда методов машинного обучения, например, популярного EM–алгоритма, является одной из главных проблем его широкого использования [25].

Для иллюстрации особенностей метода максимального подобия и предложенной его реализации, в следующем разделе приведен простой пример идентификации GGO объектов в базе данных COVID-CT-Dataset реальных КТ изображений [26].

## **5. Результаты численного моделирования идентификации объектов методом максимального подобия на PCD изображениях в модели смесей экспоненциального типа**

Для иллюстрации метода максимального подобия и его алгоритмической реализации рассмотрим простую модель интенсивности  $I(\vec{x}|\vec{\theta}_0)$  (23) в виде смеси из  $K = L^2$  компонент с квадратными носителями размеров  $D \times D$ , расположенными в узлах прямоугольной  $L \times L$  решетки с расстояниями между соседними узлами  $d$ . То же самое можно выразить математически более точно, определив базовый носитель как прямое произведение полуинтервалов  $\Delta = (-D/2, D/2] \times (-D/2, D/2]$  и носители компонент как его копии, центры которых расположены в точках  $\vec{\mu}_j = (\{l_1 - L \setminus 2\}d, \{l_2 - L \setminus 2\}d)$ , где  $l_1, l_2 =$

$1, \dots, L$  и  $\setminus$  является операцией целочисленного деления. В целом геометрия расположения носителей компонент в выбранной модели подобна изображенной на Рис.3.

Выберем стороны базового носителя кратными шагу решетки:  $D = ld$  с некоторым целым  $l > 1$ . При этом условии для всех точек  $\vec{x} \in \Omega$  легко определяются узлы решеточного окружения  $\delta_{\vec{x}}$  и их количество  $k$ . Именно, если некоторая точка  $\vec{x}$  имеет в  $\Omega$  координаты  $(x_1, x_2)$ , то областью, полученной смещением в  $\vec{x}$  базового носителя  $\Delta$  будет  $\Delta + \vec{x}$ , в которой содержится  $k = l^2$  узлов решетки. Все эти узлы  $\vec{\mu}_j$  имеют координаты  $(\{X_1 + l_1\}d, \{X_2 + l_2\}d)$ , где  $l_1, l_2 = 1, \dots, l$ , а величины  $X_1 = (x_1 - D/2) \setminus d$  и  $X_2 = (x_2 - D/2) \setminus d$  являются результатом целочисленного деления координат левого нижнего угла области на  $d$ . Целесообразно уже на этапе инициализации рекуррентной РМ–процедуры вычисления оптимальных параметров найти по приведенным формулам решеточные окружения для всех отсчетов представляющей изображение объекта  $O$  выборки  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$ .

Форму базового кадра (22) зададим с помощью функционального базиса  $\{\varphi_q(\vec{x})\}$  только из двух линейных по  $\vec{x}$  базисных функций –  $\varphi_1(\vec{x}) = x_1$  и  $\varphi_2(\vec{x}) = x_2$ . Другими словами, параметрическая модель базового кадра задается в следующем виде:

$$F(\vec{x} | w, \vec{\eta}) = I_0 w \exp\{\vec{\eta}\vec{x} - A(\vec{\eta})\} \Pi_{\Delta}(\vec{x}), \quad (44)$$

где полное число параметров  $p = 3$  включая  $w$  и  $\vec{\eta} = (\eta_1, \eta_2)$ . Для упрощения последующих вычислений удобно индикаторную функцию базового носителя  $\Pi_{\Delta}(\vec{x})$  аппроксимировать гауссовой функцией  $\exp\{-\vec{x}^2/2\sigma^2\}$ , где  $\sigma = D/\sqrt{2\pi}$  выбирается из условия равенства интеграла от гауссоиды величине  $D^2$ .

Согласно (21) кумулянтная функция  $A(\vec{\eta})$  для модели кадра (44) есть:

$$A(\vec{\eta}) = \ln \left[ \frac{1}{\Sigma_{\Omega}} \int \int \exp\{\vec{\eta}\vec{x} - \vec{x}^2/2\sigma^2\} dx_1 dx_2 \right] = \frac{1}{2} (\sigma\vec{\eta})^2 - \ln N, \quad (45)$$

где  $N = \Sigma_{\Omega}/D^2 = (L/l)^2$  – минимально необходимое число носителей компонент для полного покрытия области  $\Omega$  изображения объекта.



С помощью  $A(\vec{\eta})$  (45) легко находятся в явном выражения связывающие  $\vec{\eta}$  с параметрами средних  $\vec{\psi}$ , и преобразование Лежандра  $\hat{A}(\vec{\psi})$  с  $\vec{\psi}$  (25):

$$\vec{\psi} = \nabla_{\vec{\eta}} A(\vec{\eta}) = \sigma^2 \vec{\eta}, \quad \hat{A}(\vec{\psi}) = \vec{\eta} \vec{\psi} - A(\vec{\eta}) = \frac{1}{2} (\vec{\psi} / \sigma)^2 + \ln N, \quad (46)$$

Используя явный вид  $\hat{A}(\vec{\psi})$  (46), можно легко найти и выражение для дивергенции Брегмана (31):

$$D_{B\hat{A}}(\vec{\chi}, \vec{\psi}) = \frac{1}{2} \left( \frac{\vec{\chi} - \vec{\psi}}{\sigma} \right)^2. \quad (47)$$

С учетом принятой аппроксимации и выражений (45), (46) параметрическая модель базового кадра может быть записана в виде:

$$F(\vec{x} | w, \vec{\psi}) = W_0 \frac{w}{2\pi\sigma^2} \exp \left\{ - (\vec{x} - \vec{\psi})^2 / 2\sigma^2 \right\}, \quad (48)$$

где использовано обозначение  $W_0 = I_0 \Sigma_\Omega$  для общей энергии регистрируемого излучения, падающей в единицу времени на  $\Omega$ .

С учетом явного вида (45), (46) всех входящих в вычисления РМ алгоритма выражений, его рекуррентная схема (34) / (35), (36) принимает следующую форму.

Инициализация: сначала для всех отсчетов выборки  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  вычисляются решеточные окружения  $\{\delta_{\vec{x}_i}\}$ . Затем, устанавливается равномерное распределение  $w_j^{(0)} = K^{-1}$  и  $\vec{\psi}_j^{(0)} = \vec{0}$  для всех  $j = 1, \dots, K$ . Устанавливается счетчик итераций  $\nu$  в начальное значение 1 и осуществляется переход к рекуррентным итерациям.

Итерация  $\nu$ :

Шаг Р – уточнение разбиения – в соответствии с (34) для каждого отсчета  $\vec{x}_i$  находится индекс  $j_i$  в  $\delta_{\vec{x}_i}$ , ассоциирующий отсчет с узлом решетки  $\vec{\mu}_{j_i}$ :

$$j_i = \arg \max_{j \in \delta_{\vec{x}_i}} \left[ w_j^{(\nu-1)} \exp \left\{ - \frac{1}{2\sigma^2} (\vec{x}_i - \vec{\mu}_j - \vec{\psi}_j^{(\nu-1)})^2 \right\} \right] \quad (49)$$

(не существенный общий множитель  $N$  опущен). После того, как сформирован полный  $n$ -набор индексов  $\{j_1, \dots, j_n\}$ , задающий уточненное разбиение  $\{\pi_j^{(\nu)}\}$ ,

он сравнивается с набором индексов предыдущей итерации. Если наборы совпадают, процедура завершается, если нет – переход к шагу М.

Шаг М – вычисление оптимальных параметров  $\vec{\theta}_{\{\pi_j^{(v)}\}}$  в соответствии с общими уравнениями (36), которые в рассматриваемом случае есть:

$$w_j^{(v)} = n_j/n, \quad \vec{\psi}_j^{(v)} = \frac{1}{n_j} \sum_{\vec{x}_i \in \pi_j^{(v)}} (\vec{x}_i - \vec{\mu}_j), \quad (50)$$

где  $n_j = |\pi_j^{(v)}|$  – число отсчетов в  $\pi_j^{(v)}$ . После того, как параметры найдены, инкрементируется счетчик  $v \rightarrow v + 1$  и осуществляется переход к шагу Р следующей итерации.

В целом процедура идентификации в упрощенной модели имеет следующую вычислительную схему. После того, как некоторым образом сформирована представляющая изображение тестируемого объекта  $O$  выборка отсчетов  $X_n = \{\vec{x}_1, \dots, \vec{x}_n\}$  (см. раздел 1), на ее основе с помощью рекуррентного РМ–алгоритма (48), (49), (50) находится подходящее разбиение  $\{\pi_j^*\}$  и соответствующие ему оптимальные параметры  $\vec{\theta}^*$ . После этого осуществляется собственно решение задачи оптимизации (43), направленной на поиск наименьшего отклонения  $\bar{\mu}(X, X_k)$  – суммарной дивергенции между  $\vec{\theta}^*$  и максимально–правдоподобными параметрами  $\vec{\theta}_k^{(ML)}$  прецедентов. При этом, с учетом сделанных в предыдущем разделе замечаний по поводу вычислительной эффективности процедуры (43), целесообразно при реализации схемы воспользоваться следующими рекомендациями.

Как отмечалось выше, целесообразно сразу после определения оптимальных параметров  $\vec{\theta}^*$ , до организации поиска максимально подобного прецедента (43), провести цензурирование узлов решетки согласно найденным значениям оптимальных весов  $\{w_j^*\}$ . При этом должны быть исключены из рассмотрения те узлы  $j$ , веса которых  $w_j^*$  в определенном смысле малы. Конкретная процедура цензурирования может заключаться в следующем: вначале из найденных весов  $\{w_j^*\}$  формируется вариационный ряд  $w_{j_1}^* \leq \dots \leq$

$w_{j_K}^*$ , а затем из него исключаются первые  $K'$  узлов  $\{j_1, \dots, j_{K'}\}$ . Количество  $K'$  исключаемых узлов может быть оценено величиной их суммарного веса, не превосходящего, например, некоторого малого  $\varepsilon < 1$ , принимаемого в качестве параметра метода:  $\sum_{i=1}^{K'} w_{j_i}^* \leq \varepsilon$ . Эквивалентно, данная процедура может быть переформулирована как исключение узлов с наименьшими элементами разбиения – наименьшими  $n_j = |\pi_j^*|$ , сумма которых не превосходят  $\varepsilon n$ .

Следующее замечание об организации поиска целевого прецедента (43) относится к отмеченной выше поэтапной минимизации суммы дивергенций. Именно, оценивая на первом этапе с помощью дивергенции Кульбака–Лейблера  $D_{KL}$  степень близости выборочных весов тестируемого объекта  $\{w_j^*\}$  и соответствующих максимально–правдоподобных весов  $\{w_j^k\}$  прецедентов, можно существенно сузить круг потенциальных целей. Тогда на втором этапе можно уже только для этих прецедентов вычислять среднюю дивергенцию Брегмана  $D_{B\hat{A}}$ . Критерий отбора прецедентов на первом этапе можно задать следующим образом. Если потребовать, чтобы ко второму этапу допускались только прецеденты  $X_k$  у которых относительная невязка весов не превосходит некоторую величину  $\gamma < 1$ , именно  $|w_j^* - w_j^k|/w_j^* < \gamma$ , то для соответствующей им дивергенции Кульбака–Лейблера (40) получается следующее неравенство:

$$D_{KL}(\{w_j^*\}, \{w_j^k\}) = -\sum_{i=K'}^K w_{j_i}^* \ln \frac{w_{j_i}^k}{w_{j_i}^*} < -\sum_{i=K'}^K w_{j_i}^* \ln(1 - \gamma) < \ln \frac{1}{1 - \gamma}, \quad (51)$$

где использовано свойство положительности весов  $\{w_j^*\}$  и что сумма оставшихся после цензурирования весов меньше единицы. Таким образом, задавшись некоторым значением параметра  $\gamma$ , также рассматриваемого в качестве параметра метода, отбраковку прецедентов на первом этапе можно осуществлять на основе превышения величиной их дивергенцией Кульбака–Лейблера (51) порогового значения  $\ln(1 - \gamma)^{-1}$ .

На втором этапе поиска целевого прецедента (43) следует провести оптимизацию по уже отобраннным прецедентам, причем для вычисления

$\bar{\mu}(X, X_k)$  достаточно в дополнение к уже найденной дивергенции Кульбака–Лейблера (51) вычислить только среднюю дивергенцию Брегмана параметров  $\{\vec{\psi}_j^*\}$  и  $\{\vec{\psi}_j^k\}$  (41), которая в рассматриваемом случае, с учетом (47), имеет вид:

$$D_{B\hat{A}}(\{\vec{\psi}_j^*\}, \{\vec{\psi}_j^k\}) = 6 \sum_{i=K'}^K w_{j_i}^* \left( \frac{\bar{\psi}_{j_i}^* - \bar{\psi}_{j_i}^k}{D} \right)^2. \quad (52)$$

Отметим, что в рассматриваемой модели средняя дивергенция Брегмана (52) имеет форму средне–квадратичной невязки параметров  $\{\vec{\psi}_j^*\}$  и  $\{\vec{\psi}_j^k\}$  по  $K - K'$  используемым после цензурирования узлам решетки.

Изложенная упрощенная алгоритмическая схема метода была применена к идентификации объектов со следами “матового стекла” (GGO – ground-glass opacities) на КТ изображениях органов грудной клетки из базы данных COVID–CT–Dataset [26]. БД является открытым источником, имеющим целью способствовать исследованиям и разработкам в области компьютерного тестирования COVID–19 и содержащим 349 КТ изображений пациентов с подтвержденными диагнозами на COVID–19 и 397 КТ изображений с не подтвержденными, все файлы изображений представлены в формате PNG .

Потенциальные характеристики предложенного метода максимального подобия ниже иллюстрируются на примере идентификации фрагментов пары изображений А и D из БД [26]. Характеристики выбранных изображений представлены в Таблице 1, сами изображения и выбранные фрагменты представлены на Рис.4.

Из выбранных КТ изображений были выделены фрагменты размером 50×50 пиксел – из изображения А фрагмент  $O$  с помутнением типа матового стекла средней степени в качестве тестируемого объекта и из изображения D пять фрагментов с помутнением разной степени (2,4,5) и без него (1,3) в качестве набора прецедентов.

Таблица 1

Изображения / характеристики	2020.02.22.20024927-p19-68%0.png	2020.02.22.20024927-p19-68%3.png
Размер в пикселах	300×188	310×212
Градации серого	32	24
Идентификатор	A	D
Пациент	76-летняя женщина из группы выживших после COVID-19	72-летняя женщина из группы не выживших после COVID-19
КТ-признаки COVID-19	помутнения в виде матового стекла с преобладающим периферическим распределением в средней и нижней зонах легких	воздушная бронхограмма с обширными уплотнениями и помутнениями матового стекла

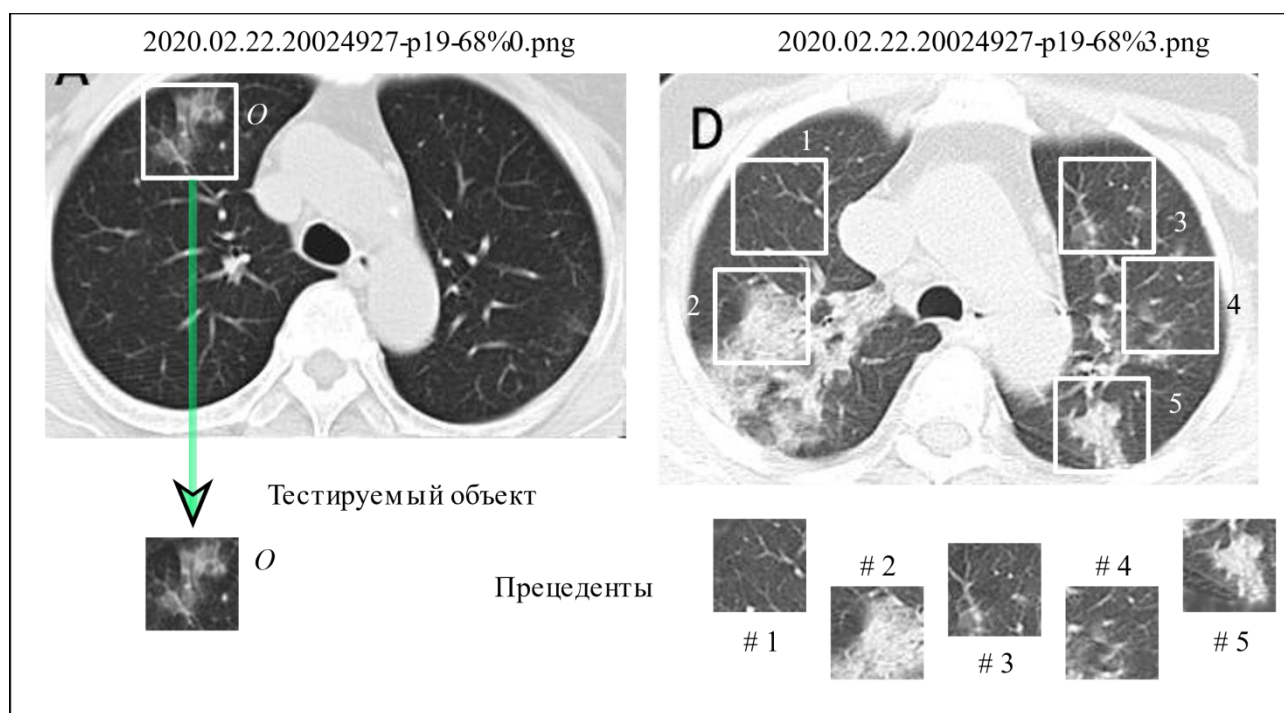


Рис.4. КТ изображения A и D двух пациентов (Таблица 1) из COVID-CT-Dataset [26] и заданные на них тестируемый объект O и пять различных прецедентов #1–5.

В соответствии с принятым методом идентификации для представления всех фрагментов были сформированы выборки  $X_n$ , содержащие по  $n = 10000$  случайных отсчетов. Выборки формировались методом сэмпирования с отклонением [13]. В качестве основного распределения вероятностей использовались пиксели фрагмента, нормированные в сумме на единицу, а в качестве вспомогательного – равномерные на фрагменте распределение с константой, равной максимуму основного распределения. Результаты формирования выборки отсчетов фрагмента  $O$  тестового объекта представлены на Рис.5.

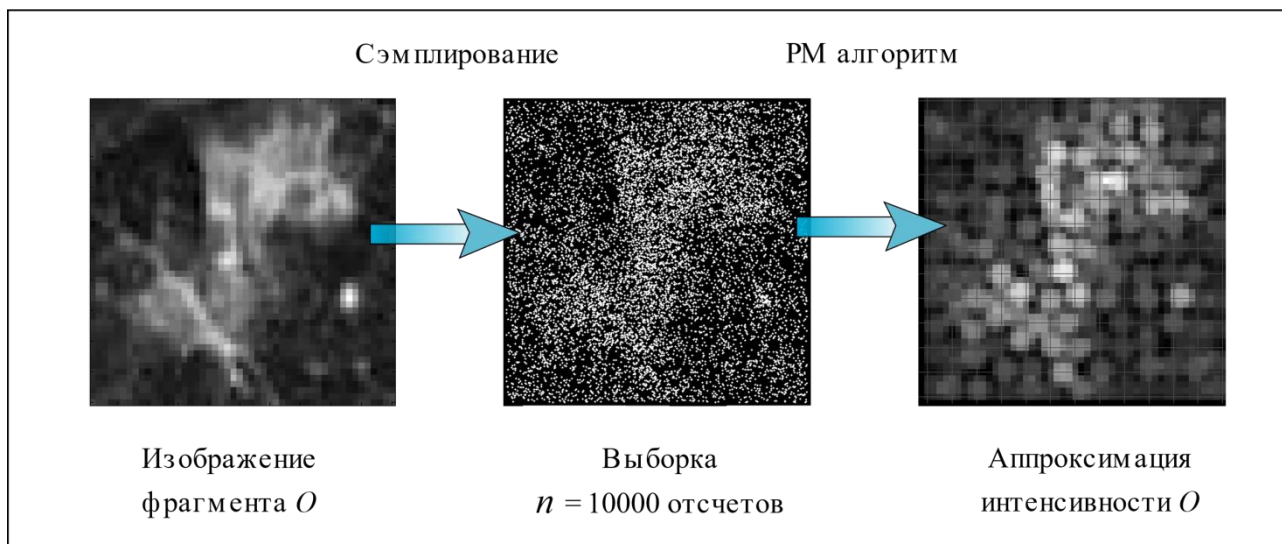


Рис.5. Конвейер подготовки фрагмента тестируемого объект  $O$  для идентификации – сэмпирования – вычисления оптимальных параметров  $\vec{\theta}^*$ .

Для реализации РМ алгоритма параметры модели интенсивности были выбраны следующими. Число компонент  $K = 225$ , т.е. компоненты располагались в узлах решетки  $15 \times 15$  с расстоянием между соседними узлами  $d = 4$  пиксела. Сторона базового носителя была выбрана в  $D = 8$  пикселей, таким образом, параметр  $l = 2$ . В этой геометрии у каждого отсчета в его решеточном окружении имелось по  $k = 4$  узла, все они легко определяются по координатам  $(x_1, x_2)$  отсчета.

Вычисления параметров  $\vec{\theta}^*$  осуществлялись в среде MatLab, релиз R2013b. Среднее время сэмпирования выборки из  $n = 10000$  отсчетов для фрагмента

изображения  $50 \times 50$  пиксел составляло меньше секунды, время выполнения на этой выборке РМ алгоритма составляло  $\sim 10$  сек и  $\sim 30$  итераций до стабилизации разбиений.

Собственно идентификация тестируемого фрагмента  $O$  изображения  $A$  с пятью фрагментами–прецедентами изображения  $D$ , представленными на Рис.4, осуществлялась вычислением дивергенций (51), (52) между оптимальными параметрами  $\vec{\theta}^*$  фрагмента  $O$  и аналогичными параметрами прецедентов, выбранных в качестве максимально–правдоподобных  $\vec{\theta}_k^{(ML)}$ . Поскольку сами параметры были подготовлены заранее, время процедуры идентификации оказалось пренебрежимо малым. Результаты вычисления дивергенции Кульбака–Лейблера  $D_{KL}(\{w_j^*\}, \{w_j^k\})$  (51) представлены в Табл. 2.

Таблица 2

Прецеденты	#1	#2	#3	#4	#5
дивергенция Кульбака– Лейблера	0.173	0.159	0.177	0.185	0.190

Из Таблицы 2 следует, что средняя относительная невязка весов  $\overline{|w_j^* - w_j^k|/w_j^*} \sim D_{KL}^{1/2}(\{w_j^*\}, \{w_j^k\})$  составляет около 0.44, что составляет довольно большую величину и не позволяет говорить об уверенной идентификации. Тем не менее, минимум дивергенции (0.159) приходится на прецедент #2, имеющий большую область помутнения. Этот факт говорит об адекватности полученных результатов даже в очень упрощенном варианте предложенного метода максимального подобия.

## Заключение

Рассмотренный пример аппроксимации формы интенсивности объектов смесью гауссовых компонент и результаты соответствующего численного моделирования показали адекватность метода максимального подобия задачам

анализа PCD изображений. Даже на изображениях невысокого качества соответствующий предложенному методу алгоритм дает приблизительно правильную идентификацию объектов. Отметим, что реализация алгоритма, аналогичная аналогичную известному методу сегментации K-средних, является весьма эффективной в вычислительном отношении даже для смесей с  $\sim 200$  компонент в среде MatLab, релиз R2013b (вычислительная неэффективность ряда методов машинного обучения, например, популярного EM-алгоритма, является одной из главных проблем его использования при большом количестве компонент модели [25]).

Что же касается собственно метода максимального подобия, то простота его интерпретации, которой посвящена первая половина работы, и естественность его алгоритмической реализации, которой посвящена вторая половина, делает его очень привлекательным как в теоретическом, так и в практическом плане, особенно в контексте современных, ориентированных на машинное обучение подходов. В определенном смысле для задач машинного обучения предложенный метод является адаптацией широко используемого в традиционной статистике метода максимального правдоподобия Р. Фишера. Плодотворное использование последнего, как известно, привело к огромному числу важных статистических результатов. В этой связи можно надеяться, что предложенный метод максимального подобия также окажется полезным при решении широкого круга современных задач машинного обучения.

Работа выполнена при финансовой поддержке РФФИ, грант № 18-07-01295 А.

**Приложение. Асимптотика  $n \gg 1$  параметрических плотностей распределения вероятности совокупности  $n$  независимых, одинаково распределенных (iid) случайных величин**

Пусть задана совокупность из  $n$  независимых, одинаково распределенных (iid) случайных величин  $X = \{\vec{x}_1, \dots, \vec{x}_n\}$ , общее распределение каждой из



которых задается параметрической плотностью  $\rho(\vec{x} | \vec{\theta})$ ,  $\vec{\theta} \in \Theta \subset \mathbb{R}^p$ . Пусть действительное значение  $\vec{\theta}$  не известно, но может быть оценено по реализации данных  $X$  статистикой  $\vec{\theta}^* = \vec{\theta}^*(X) \in \Theta$ . В частности, оценкой максимального правдоподобия  $\vec{\theta}^{(ML)}$  является статистика, максимизирующая функцию правдоподобия (как функцию  $\vec{\theta}$ )  $L(\vec{\theta}; X) = \varrho(X | \vec{\theta}) = \prod_{i=1}^n \rho(\vec{x}_i | \vec{\theta})$ , либо ее логарифм:

$$l(\vec{\theta}; X) = \ln L(\vec{\theta}; X) = \ln \prod_{i=1}^n \rho(\vec{x}_i | \vec{\theta}). \quad (\text{П.1})$$

В обозначениях (П.1), уравнением для  $\vec{\theta}^{(ML)}$  является  $\nabla_{\vec{\theta}} l(\vec{\theta}^{(ML)}; X) = 0$ .

Разложим логарифм функции правдоподобия (П.1) в окрестности  $\vec{\theta}^{(ML)}$  по  $\Delta\vec{\theta} = (\vec{\theta} - \vec{\theta}^{(ML)})$ :

$$\begin{aligned} l(\vec{\theta}; X) &= l(\vec{\theta}^{(ML)}; X) + \nabla_{\vec{\theta}} l(\vec{\theta}^{(ML)}; X) \Delta\vec{\theta} + \frac{1}{2} Q(\Delta\vec{\theta}) + \dots = \\ &= l(\vec{\theta}^{(ML)}; X) + \frac{1}{2} \Delta\vec{\theta}^T \hat{Q} \Delta\vec{\theta} + \dots \end{aligned}, \quad (\text{П.2})$$

где матрица квадратичной формы  $Q(\Delta\vec{\theta})$  при  $X = \{\vec{x}_1, \dots, \vec{x}_n\}$  имеет вид:

$$\hat{Q} = \left\| \frac{\partial^2 l(\vec{\theta}^{(ML)}; X)}{\partial \theta_k \partial \theta_l} \right\| = n \left\| \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln \rho(\vec{x}_i | \vec{\theta}^{(ML)})}{\partial \theta_k \partial \theta_l} \right\|. \quad (\text{П.3})$$

Считая вторые производные  $\partial^2 \ln \rho(\vec{x} | \vec{\theta}^{(ML)}) / \partial \theta_k \partial \theta_l$  в (П.3) случайными величинами – функциями случайных  $\vec{x}$  и применяя к их выборочному среднему закон больших чисел, в асимптотике  $n \gg 1$  получим:

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln \rho(\vec{x}_i | \vec{\theta}^{(ML)})}{\partial \theta_k \partial \theta_l} \approx \int \rho(\vec{x} | \vec{\theta}) \frac{\partial^2 \ln \rho(\vec{x} | \vec{\theta}^{(ML)})}{\partial \theta_k \partial \theta_l} d\vec{x}. \quad (\text{П.4})$$

Далее, учитывая, что асимптотически  $\vec{\theta}^{(ML)} \rightarrow \vec{\theta}$ , можно (П.4) записать в виде:

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln \rho(\vec{x}_i | \vec{\theta}^{(ML)})}{\partial \theta_k \partial \theta_l} \approx \int \rho(\vec{x} | \vec{\theta}) \frac{\partial^2 \ln \rho(\vec{x} | \vec{\theta})}{\partial \theta_k \partial \theta_l} d\vec{x} = -I(\vec{\theta}), \quad (\text{П.4})$$

где  $I(\vec{\theta})$  – информационная матрица Фишера [18].

Собирая все вместе, асимптотику  $n \gg 1$  для (П.2) можно записать в виде:

$$l(\vec{\theta}; X) \approx l(\vec{\theta}^{(ML)}; X) - \frac{n}{2} \Delta\vec{\theta}^T I(\vec{\theta}) \Delta\vec{\theta}. \quad (\text{П.5})$$

или, повторно используя асимптотическое свойство  $\vec{\theta}^{(ML)} \sim \vec{\theta}$ :

$$q(X|\vec{\theta}) \approx q(X|\vec{\theta}^{(ML)}) \exp \left\{ -\frac{n}{2} (\vec{\theta} - \vec{\theta}^{(ML)})^T I(\vec{\theta}^{(ML)}) (\vec{\theta} - \vec{\theta}^{(ML)}) \right\}. \quad (\text{П.6})$$

Из (П.6) следует, что при  $n \gg 1$  плотность  $q(X|\vec{\theta})$ , как функция  $\vec{\theta}$ , имеет узкий, ширины  $\sim \sigma_k/\sqrt{n}$  пик в определяемой данными  $X$  точке  $\vec{\theta}^{(ML)}$  ( $\sigma_k$  – обратные квадратные корни из собственных значений  $I(\vec{\theta})$  по  $k$ -координатам).

### Литература

1. Darby M.J., Barron D., Hyland R.E., editors. *Oxford Handbook of Medical Imaging*. Oxford, Oxford University Press. 2011.
2. Leng S., et. al. Photon-counting Detector CT: System Design and Clinical Applications of an Emerging Technology. *RadioGraphics*. 2019. Vol.39. No.3. P.729-743.
3. Willeminck M.J., et. al. Photon-counting CT: technical principles and clinical prospects. *Radiology*. 2018. Vol.289. No.2. P.293-312.
4. Gonzalez R.C., Woods R.E. *Digital Image Processing*, 3rd ed. Prentice Hall. 2010.
5. Goodman J.W. *Statistical Optics*. 2nd ed. New York. Wiley. 2015.
6. Pal N.R., Pal S.K. Image model, poisson distribution and object extraction. *J. Patt. Recogn. Artif. Intell.* 1991. Vol. 5. No.3. P.459–483.
7. Yang F., Lu Y. M., Sbaiz L., Vetterli M. Bits from Photons: Oversampled Image Acquisition Using Binary Poisson Statistics. *IEEE Transactions on Image Processing*. 2012. Vol.21. No.4. P.1421-1436.
8. Fossum E.R., Teranishi N., et. al., editors. *Photon-Counting Image Sensors*. MDPI Books under CC BY–NC–ND license. 2017.
9. Streit R. L. *Poisson Point Processes. Imaging, Tracking and Sensing*. New Yor. Springer. 2010.
10. Antsiperov V. Machine Learning Approach to the Synthesis of Identification Procedures for Modern Photon-Counting Sensors. *8th International Conference*

- on Pattern Recognition Applications and Methods ICPRAM Proceedings*. 2019. Vol.1. P.19–21.
11. Bertero M., Boccacci P., Desidera G., Vicidomini G. TOPICAL REVIEW: Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*. 2009. Vol.25. No.12. P.123006. <https://doi.org/10.1088/0266-5611/25/12/123006>
  12. Barber D. *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge Univ. Press. 2012.
  13. MacKay D.J.C. *Information Theory, Inference, and Learning Algorithms*. Cambridge. Cambridge Univ. Press. 2003.
  14. Szirmay-Kalos L., Szecsi L., Penzov A.A. Importance Sampling with Floyd-Steinberg Halftoning. *Eurographics Short Papers*. 2009. P. 69–72.
  15. Floyd R., Steinberg L. An adaptive algorithm for spatial gray scale. *Proceedings of the Society of Information Display*. 1976. Vol.17. No.2. P.75–77.
  16. Gooran S., Yang L. Basics of Tone Reproduction. In: Kriss M., editor. *Handbook of Digital Imaging*. 2015.
  17. Jebara T. *Machine Learning: Discriminative and Generative*. Dordrecht. Kluwer. 2004.
  18. Wasserman L. *All of statistics: a concise course in statistical inference*. New York. Springer. 2004.
  19. Efron B. Maximum likelihood and decision theory. *Ann. Statist.* 1982. Vol.10. P.340–356.
  20. Gröchenig K. *Foundations of Time-Frequency Analysis*. Boston. Birkhäuser. 2001.
  21. Murphy K.P. *Machine learning: a probabilistic perspective*. Cambridge, MA. MIT Press. 2012.
  22. Du D, Ko K. *Theory of Computational Complexity*, 2nd ed. John Wiley & Sons. 2014.
  23. Bregman L.M. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR*

*Computational Mathematics and Mathematical Physics*. 1967. Vol.7. No.3. P.200–217.

- 24.Sun S., Cao Z., Zhu H., Zhao J. A Survey of Optimization Methods from a Machine Learning Perspective. *IEEE Transactions on Cybernetics*. 2020. Vol.50. No.8. P.3668-3681.
- 25.Nguyen H.D., Forbes F., McLachlan G.J. Mini-batch learning of exponential family finite mixture models. *Statistics and Computing*. 2020. Vol.30. P.731–748.
- 26.Yang X., He X., Zhao J., Zhang Y., Zhang S. COVID-CT-Dataset: A CT san dataset about COVID-19. [arXiv:2003.13865](https://arxiv.org/abs/2003.13865), 2020.

**Для цитирования:**

Анциперов В.Е. Новый, ориентированный на машинное обучение метод идентификации объектов на биомедицинских изображениях, формируемых датчиками счета фотонов. Журнал радиоэлектроники [электронный журнал]. 2020. №11. <https://doi.org/10.30898/1684-1719.2020.11.14>