

## ТЕОРЕТИКО-ВЕРОЯТНОСТНАЯ МОДЕЛЬ ПОЛУТОНОВОГО ИЗОБРАЖЕНИЯ ДЛЯ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ БЕЗ УЧИТЕЛЯ НА ОСНОВЕ МЕТОДА НАПРАВЛЕННОГО ПЕРЕБОРА

Савченко А.В.

*Национальный исследовательский университет «Высшая школа экономики»*

### Аннотация

Ставится и решается задача автоматического распознавания изображений без учителя на основе метода направленного перебора. Исследована новая теоретико-вероятностная модель полутонového изображения. Её применение позволило синтезировать новый критерий распознавания с самообучением на основе оптимального байесовского правила проверки статистических гипотез. Представлены программа и результаты экспериментального исследования для идентификации личности по фотографии лица. Показано, что применение предложенной модели совместно с методом направленного перебора характеризуется существенным сокращением объёма вычислений при сохранении высоких показателей точности распознавания изображений.

**Ключевые слова:** автоматическое распознавание изображений без учителя, самообучение, принцип минимума информационного рассогласования Кульбака-Лейблера, метод направленного перебора.

### Введение

Несмотря на широкую коммерциализацию и большое число программных продуктов автоматического распознавания изображений (АРИ), интенсивность исследований в этой области отнюдь не снижается, т.к., хотя цена существующих систем весьма высока, их надёжность всё ещё не достаточна. И связано это, прежде всего, с острейшей проблемой вариативности [1] – отдельные изображения одного объекта могут существенно варьироваться в зависимости от условий наблюдения: ракурса, расстояния, освещения. Проблема особенно усиливается, если объём базы данных (БД) эталонов составляет тысячи единиц, что приводит к усложнению методов распознавания и, как следствие, невозможности реализации существующих алгоритмов [1, 2] в режиме реального времени. Со всех перечисленных точек зрения несомненный интерес представляет моделирование распознавания изображений на основе теоретико-информационного подхода [3] и общесистемного принципа минимума информационного рассогласования (МИР) Кульбака-Лейблера [4]. Основанная на нём информационная теория восприятия речи показала высокие результаты [5] в задаче автоматического распознавания речи. Между тем, не все преимущества принципа МИР получили необходимое освещение и развитие. В частности, до настоящего времени почти не рассматривалась возможность разработки новых математических моделей изображений, рассчитанных на применение принципа МИР. Исследования в этом актуальном направлении и составляют главное содержание настоящей работы. В ней предложенная модель полутонového изображения сочетается с классическим байесовским подходом для распознавания без учителя, что позволило использовать метод направленного перебора альтернатив (МНП) [6] для сокращения вычислительной трудоёмкости АРИ. Полученные результаты и сделанные по ним выводы расчи-

таны на широкий круг специалистов в области распознавания изображений.

### Задача автоматического распознавания изображений

Пусть задано множество из  $L > 1$  полутонových изображений  $X_l = \|x_{uv}^{(l)}\|$ ,  $l = \overline{1, L}$ ,  $u = \overline{1, U}$ ,  $v = \overline{1, V}$ . Здесь  $U$  – высота изображения,  $V$  – его ширина;  $x_{uv}^{(l)} \in \{1, 2, \dots, x_{\max}\}$  – интенсивность точки изображения с координатами  $(u, v)$ ;  $x_{\max}$  – максимальное значение интенсивности,  $l$  – номер эталона в БД. Задача распознавания состоит в том, чтобы отнести вновь поступающее (на вход) изображение  $X = \|x_{uv}\|$  к одному из классов, заданных эталонами  $X_l$ . Каждый класс характеризуется тем, что принадлежащие ему объекты обладают некоторой общностью или сходством в характеристиках. То общее, что объединяет объекты в класс, и называют образом. Предполагается, что несколько эталонов могут определять один класс (содержать изображение одного и того же объекта), однако априори не задано, к какому классу какой эталон принадлежит. Либо, что более вероятно, известна лишь неполная информация об отношении эталона к классу (например, двум фотографиям одного человека присвоена одинаковая метка, однако сами фотографии отличаются по времени создания на несколько лет). Это типичный пример задачи распознавания образов без учителя.

Процедуры построения решающих правил для поставленной задачи в общем случае делятся на детерминированные и статистические [7]. В настоящее время наиболее часто используется первый, детерминистский, подход. В рамках такого подхода определяется некое расстояние (мера близости) между любыми парами объектов и для классификации используется один из методов типа «ближайших сосе-

дей» [1]. Зачастую для распознавания применяется  $l_1$ -метрика

$$\rho_1(X/X_l) = \frac{1}{U \cdot V} \sum_{u=1}^U \sum_{v=1}^V |x_{uv} - x_{uv}^{(l)}| \rightarrow \min_l. \quad (1)$$

К сожалению, подобный подход не всегда позволяет получить удовлетворительные результаты. Это обстоятельство связано, во-первых, с известной вариативностью зрительных образов, а во-вторых, с наличием во входном изображении  $X$  помех, таких как неопределённая заранее интенсивность источников освещения или просто случайное искажение некоторых точек изображения.

### Теоретико-вероятностная модель полутонового изображения

Во всех перечисленных случаях на помощь приходит второй, статистический, подход. Для каждого изображения-эталона строится гистограмма некоторого признака (цвета, текстуры, формы) [8, 9]. Далее для определённости сосредоточимся на самом простом признаке – интенсивности пикселя изображения.

Выполнив предварительную нормировку [10] освещения изображений из базы (например, используя гамма-коррекцию [8]), поставим каждому эталону  $X_l$  в соответствие гистограмму распределения интенсивности  $H_l = [h_1^{(l)}, h_2^{(l)}, \dots, h_{x_{\max}}^{(l)}]$ :

$$h_x^{(l)} = \frac{1}{U \cdot V} \sum_{u=1}^U \sum_{v=1}^V \delta(x_{uv}^{(l)} - x),$$

где  $\delta(x)$  – дискретная дельта-функция. Такая же процедура построения гистограммы интенсивности  $H = [h_1, h_2, \dots, h_{x_{\max}}]$  применяется и для входного изображения  $X$ .

Традиционным подходом по сравнению гистограмм считается так называемый метод слияния гистограмм (merge histogram) [11, 12]

$$\sum_{x=1}^{x_{\max}} \min\{h_x^{(l)}, h_x\} \rightarrow \max_l. \quad (2)$$

Как известно [13], непосредственно сопоставление гистограмм (2) наталкивается на проблему вариативности освещения – если затемнить/осветлить изображение, то его гистограмма изменится. А традиционная нормировка освещённости [8] зачастую приводит к небольшим сдвигам интенсивности точек изображения, хоть и не заметных человеку, но оказывающих влияние на гистограмму. Именно поэтому после вычисления гистограмм  $H_l$  и  $H$  применяется их динамическое выравнивание [13], что существенно увеличивает объём вычислений.

В настоящей работе предлагается кардинальный способ преодоления указанного недостатка. Можно предположить, что  $H_l$  определяет собой распределение одномерной дискретной случайной величины – интенсивности  $I$  точки изображения  $X_l$ . Задача сводится в таком случае к проверке  $L$  гипотез о

распределении  $H_l, l = \overline{1, L}$ , сигнала изображения на входе  $H$ :

$$W_l : H = H_l.$$

Оптимальное решение тогда даёт классический байесовский подход [1, 7] – критерий максимума апостериорной вероятности  $P\{W_l / X\}$  того, что справедлива гипотеза  $W_l$  при появлении на входе объекта  $X$ , то есть изображение  $X$  принадлежит классу, заданному эталоном  $X_l$ . Эта вероятность вычисляется по формуле Байеса

$$P\{W_l / X\} = \frac{P\{X / W_l\} \cdot P\{W_l\}}{\sum_{i=1}^L P\{X / W_i\} \cdot P\{W_i\}} \rightarrow \max_l. \quad (3)$$

Здесь  $P\{W_l\}$  – априорная вероятность появления  $l$ -го класса,  $P\{X / W_l\}$  – априорная вероятность принадлежности объекта  $X$  классу  $l$ . В большинстве задач АРИ предполагается, что появление каждого класса равновероятно (полная априорная неопределённость). Поэтому, делая традиционное «наивное» предположение о независимости признаков всех пикселей, критерий (3) сводится к упрощённому виду

$$P\{X / W_l\} = \prod_{u=1}^U \prod_{v=1}^V P\{I = x_{uv} / W_l\} \rightarrow \max_l. \quad (4)$$

С учётом нашего предположения о том, что гистограмма  $H_l$  является оценкой распределения интенсивности  $I$  для класса  $l$ , условная вероятность

$$P\{I = x_{uv} / W_l\} = h_{x_{uv}}^{(l)}. \quad (5)$$

Или, используя аппарата дискретных дельта-функций Дирака  $\delta(\cdot)$ , условная вероятность

$$P\{I = x_{uv} / W_l\} = \sum_{x=1}^{x_{\max}} h_x^{(l)} \cdot \delta(x - x_{uv}).$$

После несложных преобразований (4) получаем

$$P\{X / W_l\} = \prod_{x=1}^{x_{\max}} (h_x^{(l)})^{U \cdot V \cdot h_x} \rightarrow \max_l.$$

Удалив из показателя степени константу  $U \cdot V$ , определяемую площадью (в пикселях) входного изображения  $X$ , запишем окончательное решение АРИ по критерию максимального правдоподобия

$$\prod_{x=1}^{x_{\max}} (h_x^{(l)})^{h_x} \rightarrow \max_l. \quad (6)$$

К сожалению, непосредственное сравнение гистограмм (5) наталкивается на очевидную проблему, а именно: близкие значения интенсивности воспринимаются человеком одинаково [9]. Однако если перейти от задачи АРИ к задаче статистической классификации, то, например, компоненты  $h_1$  и  $h_2$  не будут иметь ничего общего, так как события  $I = 1$  и  $I = 2$  в такой «наивной» статистической модели изображения никак между собой не связаны. Безусловно, отмеченное обстоятельство игнорирования схожести близких значений интенсивности ухудшит

качество распознавания, так как ввиду неизбежного применения процедуры нормировки освещения невозможно добиться точного совпадения нормированной интенсивности тестового изображения и эталона из БД, даже если с точки зрения человека эти изображения полностью идентичны.

На практике эта проблема обычно решается дискретизацией значений интенсивности [1, 9]. Очевидно, что это решение может привести к снижению точности распознавания. Эта проблема особенно актуальна, если в качестве признаков используется не интенсивность, а текстура или форма.

Более правильным здесь будет заимствование методов, издавна используемых для сегментации изображений. Например, в хорошо зарекомендовавшем себя графовом разбиении Ши [14] предлагается на множестве значений признаков ввести некий

нормирующий коэффициент  $\exp\left(-\frac{\|F(x_1) - F(x_2)\|}{\sigma_F}\right)$ ,

зависящий от расстояния  $\|\cdot\|$  между ними. Здесь  $F(x_1)$ ,  $F(x_2)$  – значение признака,  $\sigma_F$  – стандартное отклонение признака (обычно задаваемый пользователем параметр). В рассматриваемом случае распознавания изображений по интенсивности  $F(x) = x$  и  $\|F(x_1) - F(x_2)\| = |x_1 - x_2|$ .

В результате значения интенсивности в тестовом изображении определяются не только идентичной интенсивностью в эталоне, но и близкими (в смысле расстояния  $\|\cdot\|$ ) значениями. С точки зрения статистического подхода можно предположить, что условная вероятность  $P\{I = x_{iv} / W_i\}$  определяется не только  $h_{x_{iv}}^{(l)}$ , но и  $h_x^{(l)}$  со значениями интенсивности  $x$ , близкими к  $x_{iv}$ . Более формально,

$$P\{I = x_{iv} / W_i\} = \sum_{x=1}^{x_{\max}} P\{I = x_{iv} / I = x\} \cdot h_x^{(l)}. \quad (7)$$

Здесь набор условных вероятностей  $P\{I = x_{iv} / I = x\}$  определяет вероятность того, что пиксель эталона с интенсивностью  $x$  в ходе предварительной обработки изображения принял значение  $x_{iv}$ . Тогда критерий максимального правдоподобия (4) может быть записан в аналогичной (6) форме

$$\prod_{x_1=1}^{x_{\max}} \left( \sum_{x=1}^{x_{\max}} P\{x_1 / x\} \cdot h_x^{(l)} \right)^{h_{x_1}} \rightarrow \max_l. \quad (8)$$

Вероятности  $P\{x_1 / x\} = P\{I = x_1 / I = x\}$  должны задаваться исследователем для каждой конкретной задачи АРИ с учётом очевидных ограничений – условия регулярности

$$\begin{aligned} \forall x_1, x \in \{1, \dots, x_{\max}\}, \\ P\{I = x_1 / I = x\} \geq 0, \\ P\{I = x / I = x\} \geq P\{I = x_1 / I = x\} \end{aligned} \quad (9)$$

и условия нормировки

$$\forall x \in \{1, \dots, x_{\max}\} \sum_{x_1=1}^{x_{\max}} P\{I = x_1 / I = x\} = 1. \quad (10)$$

Тогда подход по традиционному сравнению гистограмм (6) эквивалентен определению этих вероятностей как

$$P\{I = x_1 / I = x\} = \delta(x_1 - x).$$

Если же применять упомянутый выше подход, используемый для сегментации изображений [14], основываясь на свойствах (9) и (10), получим следующее выражение

$$P\{I = x_1 / I = x\} = \frac{\exp(-k|x_1 - x|)}{\sum_{i=1}^{x_{\max}} \exp(-k|x_1 - i|)}. \quad (11)$$

Здесь  $k = \sigma_F^{-1} > 0$  – параметр, конфигурируемый для конкретной задачи.

Таким образом, процедура распознавания в данном случае реализуется по схеме многоканальной обработки, в которой число каналов определяется количеством изображений-эталонов  $L$ . Решение принимается по критерию минимума статистики из выражения вида (1) – для традиционных методов или из выражения (8) – при использовании метода максимального правдоподобия совместно с предложенной теоретико-вероятностной моделью полутонного изображения (7), (9)-(11).

#### Кластеризация базы данных эталонов

Рассмотрим наиболее актуальный и для теории, и для практики случай  $L \gg 1$ , когда решается задача автоматического распознавания изображений с объёмом БД в сотни и даже тысячи изображений [15]. В указанных условиях практическая реализация оптимального решающего правила (8) по схеме  $L$ -канальной обработки наталкивается на очевидную проблему его вычислительной сложности и даже практической реализуемости. В поиске путей решения указанной проблемы за счёт отказа от сплошного перебора всего множества альтернатив (обучающих выборок) и состоит центральная идея настоящей работы.

Зачастую объём БД  $L$  значительно превышает количество классов изображений  $R$ . Например, в БД может храниться несколько фотографий человека. Конечно, сходные изображения могут быть отображены экспертом вручную. Однако для большого числа эталонов ручное разделение БД является неприемлемым в силу своей трудоёмкости. Кроме того, это решение не является правильным и с точки зрения статистического подхода.

Действительно, математические модели для похожих с точки зрения человека изображений могут существенно различаться. И связано это не столько с несовершенством применяемых моделей, сколько с неопределённостью постановки самой задачи АРИ. Например, одна и та же БД фотографий лиц

может использоваться для решения столь различных задач, как идентификация личности, распознавание гендерной принадлежности, возрастной группы, национальности.

Таким образом, изображение следует рассматривать как неизвестную функцию сразу нескольких факторов. В упомянутом примере фотография лица определяется следующими факторами – биологическими особенностями индивидуума, его гендерной принадлежностью, возрастом, расой, а также условиями наблюдения. Всё это приводит к тому, что задача АРИ на практике становится существенно сложнее задачи статистической классификации, так как наблюдаемый объект и обучающие выборки для одинакового класса представляют собой неизвестную функцию нескольких случайных величин.

В этой интерпретации становится понятна и причина появления ошибок распознавания по методам ближайших соседей вида (1). Такие методы находят эталон, ближайший в некотором смысле к входному изображению по совокупности всех факторов. То есть ошибка происходит, если в БД хранится фотография человека, совпадающая с распознаваемым объектом по большинству факторов (особенно критичны условия съёмки – освещение, ракурс и др.).

Учтёв указанную особенность АРИ – зависимость изображений одновременно от многих факторов – можно только с помощью надлежащего выбора обучающих выборок (БД эталонов), обеспечивая максимально полное покрытие пространства объектов, что приводит к взрывному росту количества эталонов  $L$ . К сожалению, использование здесь математического аппарата факторного анализа невозможно, так как вид функции изображения от отмеченных факторов неизвестен. А непосредственное применение к базе эталонов алгоритмов машинного обучения зачастую приводит к неудовлетворительным результатам. Действительно, такие алгоритмы основаны на сведении классификации к задаче аппроксимации. Но, как известно, аппроксимация неоднородных данных приводит к резкому росту ошибки классификации.

В результате, практически безальтернативным решением здесь будет предварительная редукция БД с применением известных алгоритмов кластеризации [7], таких как k-means или карты Кохонена [16] для выделения центроидов кластеров с последующим перебором в ходе АРИ только выбранных центроидов.

Итак, предполагаем, что множество всевозможных изображений  $\mathbf{X}$ , из которого взяты эталоны  $X_l$ , разбивается на  $R$  непересекающихся классов  $\mathbf{X}_r, r = \overline{1, R}$  со свойствами

- 1)  $\mathbf{X} = \bigcup_{r=1}^R \mathbf{X}_r$ ;
- 2)  $\mathbf{X}_{r_1} \cap \mathbf{X}_{r_2} = \emptyset$  при  $r_1 \neq r_2$ ;
- 3)  $(\forall l)(\exists r) X_l \in \mathbf{X}_r$ .

Предварительная задача состоит в том, чтобы на основе множества изображений-эталонов  $\{X_l\}$  сформировать классы  $\mathbf{X}_r$ . Это типичный пример задачи самообучения (обучение без учителя) [17].

Классические алгоритмы кластеризации [1] основаны на том, что число однородных кластеров  $R$  заранее известно. Для большинства практических приложений задачи распознавания указанное предположение не выполняется, так как в каждом классе имеется большое число эталонов для различных условий съёмки и обычно неизвестно, какие факторы оказывают основное влияние на выбранную математическую модель изображения. Поэтому количество классов требуется определять, используя только информацию, содержащуюся в множестве  $\{X_l\}$ . Решение указанной задачи самообучения проведём на основе подхода информационной теории восприятия речи [5].

### Информационная теория восприятия

Как правило, в математических задачах кластеризации и группировки основное – выбор метрики, расстояния между объектами, меры близости, сходства, различия [1, 18]. Поэтому предварительно преобразуем критерий (8) к виду, приемлемому для применения алгоритмов самообучения. Для этого прологифмируем левую часть (8) и умножим её на  $(-1)$ :

$$\sum_{x_1=1}^{x_{\max}} \left( -h_x \cdot \log \left( \sum_{x=1}^{x_{\max}} P\{x_1 / x\} \cdot h_x^{(l)} \right) \right) \rightarrow \min_l. \quad (12)$$

Добавив к (12) не зависящее от  $l$  слагаемое  $h_x \cdot \log h_x$ , окончательно получим критерий, основанный на минимизации величины

$$\rho_{KL}(X / X_l) = \sum_{x_1=1}^{x_{\max}} h_x \cdot \log \frac{h_x}{\sum_{x=1}^{x_{\max}} P\{x_1 / x\} \cdot h_x^{(l)}}. \quad (13)$$

Статистика  $\rho_{KL}(X / X_l)$  здесь определяет информационное рассогласование по Кульбаку-Лейблеру [4] между наблюдаемым сигналом изображения  $X$  и  $l$ -м эталоном из БД.

Прежде всего, отметим метрические свойства решающей статистики МИР  $\rho_{KL}(X / X_l) \geq 0$  с равенством её нулю лишь в идеальном случае совпадения входного и эталонного сигналов. Основываясь на этом, множество образов  $\mathbf{X}$  разбивается на  $R$  непересекающихся классов  $\mathbf{X}_r = \{X_{r_i}, r_i \leq L\}$ , таких, что либо  $\mathbf{X}_r$  состоит ровно из одного эталона  $X_{r_i}$ , либо

$$(\exists X_{r_j} \in \mathbf{X}_r, r_j \neq r_i) \rho_{KL}(X_{r_j} / X_{r_i}) < \rho_0. \quad (14)$$

Здесь  $\rho_0 = const$  – порог для допустимой величины рассогласований на множестве одноимённых изображений за счёт известной их вариативности. Значение такого порога нетрудно установить опытным путём.

На основе критерия минимума суммы информационных рассогласований [5] в пределах каждого  $r$ -го кластера ( $X_k \in X_r$ )

$$\rho_{KL}^{(r)}(X_k) = \sum_{X_i \in X_r} \rho_{KL}(X_k / X_i),$$

определяется его информационный центр-эталон вида

$$X_r^* = \arg \min_{X_k, X_k \in X_r} \rho_{KL}^{(r)}(X_k). \quad (15)$$

В информационной теории восприятия предполагается [5], что в эталоне  $X_r^*$  содержится существенная информация обо всём классе  $X_r$ . Поэтому последний этап связан с редукцией всего множества эталонов  $\{X_l\}, l = \overline{1, L}$  к множеству информационных центров  $\{X_r^*\}, r = \overline{1, R}, R \leq L$ .

После такого решения задачи самообучения второй этап – распознавание – производится на основе критерия МИР

$$\rho_{KL}(X / X_r^*) \rightarrow \min_r. \quad (16)$$

С учётом определения (14) класса  $X_r$  последнюю формулу можно преобразовать к упрощённому (в её практической реализации) виду [6]

$$\rho_{KL}(X / X_r^*) < \rho_0. \quad (17)$$

По своей сути выражение (17) определяет условие останова при переборе альтернатив в рамках проверочной процедуры по критерию МИР (16).

Заметим, что если рассматривать самообучение как адаптивный процесс [17], то нетрудно заметить эквивалентность выражений (14) и (17). Действительно, при адаптивном подходе классы формируются постепенно следующим образом. Вначале число классов  $R = 0$ . Далее для каждого изображения-эталона  $X_l, l = \overline{1, L}$  ищется тот класс  $r$ , для которого выполняется условие (17), то есть решается обычная задача распознавания изображения  $X = X_l$ . Если такой класс  $v$  найден, то объект  $X_l$  добавляется в класс  $X_v$ , а далее согласно (15) вычисляется его новый центр. И только если  $(\forall v) \rho_{KL}(X_l / X_v^*) \geq \rho_0$ , то создаётся новый,  $(R + 1)$ -й класс  $X_{R+1} = \{X_l\}, X_{R+1}^* = X_l$ .

В такой формулировке задача классификации представляет собой основу адаптивной процедуры самообучения. Поэтому далее сосредоточимся на решении именно задачи классификации (16),(17).

Для решения этой задачи используется предложенный ранее [6] метод направленного перебора, в котором метрические свойства решающей статистики МИР (13) используются в наиболее полной степени.

**Метод направленного перебора**

В качестве предварительного этапа составим  $(R \times R)$ -матрицу  $P = \|\rho_{ij}\|$  попарных рассогласова-

ний между эталонами  $\rho_{ij} = \rho_{KL}(X_i^* / X_j^*)$ . Эту весьма сложную в вычислительном отношении процедуру требуется выполнить лишь раз: на предварительном этапе вычислений и для каждой конкретной БД. Однако отметим, что в контексте решения задача самообучения (14) эта матрица всё равно должна быть вычислена, так как объединение объектов в кластер основывается на попарных расстояниях между ними.

Следуя общей схеме вычислений (16), (17) сведём задачу распознавания изображений  $X$  к проверке  $N$  первых вариантов  $X_1^*, \dots, X_N^*$  из заданного  $R$ -множества альтернатив  $\{X_r^*\}$  при условии  $N \ll R$ . Среди этих  $N$  вариантов определим эталон  $X_{\mu}^*, \mu \leq N$ , ближайший к входному изображению  $X$ . Если, он отвечает требованию условия останова (17), процесс поиска оптимального решения по критерию МИР (16) на нём и завершается. Однако в общем случае можно предположить, что ни одна из первых  $N$  альтернатив проверки (17) на первом шаге не проходит.

Поместим изображения из нашей первой контрольной выборки  $X_1^*, \dots, X_N^*$  в очередь с приоритетом  $Q$ . Выбор значения приоритета для каждого эталона  $X_r^*, r = \overline{1, R}$  можно осуществить различными способами [6, 18]. В наших экспериментах [18] хорошие результаты показал приоритет  $\rho_{KL}(X / X_r^*)$  (ср. с Best-Bin First [19]). На этом завершается первый этап вычислений.

На втором этапе циклически повторяется следующая процедура. Если очередь  $Q$  пуста, то в неё добавляется один наугад выбранный не проверявшийся ранее эталон из БД. Если такой эталон не найден, то алгоритм завершается.

Далее из очереди  $Q$  извлекается эталон  $X_i^*$  с наименьшим приоритетом. Если

$$\rho_{KL}(X / X_i^*) < \rho_{KL}(X / X_{\mu}^*), \quad (18)$$

то номер  $\mu$  заменяется на  $i$  с тем, чтобы всегда ссылаться на ближайший (среди проверенных) к входному объекту эталон как на  $X_{\mu}^*$ .

Затем для выделенного изображения-эталона  $X_i^*$  по матрице  $P$  найдем множество из  $M < R$  изображений  $X^{(M)} = \{X_{i_{N+1}}^*, \dots, X_{i_{N+M}}^*\}$ , такое, что:

$$\begin{aligned} & (\forall_j \in \{1, \dots, L\}) (\forall X_k^* \in X^{(M)}) \\ & (X_j^* \notin X^{(M)}) \Rightarrow (\Delta\rho(X_j^*) \geq \Delta\rho(X_k^*)). \end{aligned} \quad (19)$$

Здесь

$$\Delta\rho(X_j^*) = \left| \rho_{KL}(X_j^* / X_i^*) - \rho_{KL}(X / X_i^*) \right|$$

отклонение рассогласований между входным изображением  $X$  и локальным оптимумом относительно рассогласований между парой изображений  $X_i$  и

$X_i^*$ . После этого добавляем в очередь  $Q$  все непроверенные ранее эталоны из  $X^{(M)}$ .

Все вычисления второго этапа повторяются до тех пор, пока на некотором  $K$ -м этапе для элемента  $X_\mu^*$  не будет выполнено условие

$$\rho_{KL}(X / X_\mu^*) < \rho_0 \quad (20)$$

Решение задачи АРИ принимается в пользу класса с информационным центром  $X_\mu^*$ . В худшем случае, после перебора всех альтернатив из БД  $\{X_r^*\}$ , но в отсутствие решения (20), можно либо сделать вывод о том, что входное изображение  $X$  нельзя отнести ни к одному классу, либо принять решение в пользу найденного «ближайшего соседа»  $X_\mu^*$  с оговоркой на его недостаточную надёжность.

Следуя построению предложенного метода (18) - (20), нетрудно показать, что справедливы следующие теоремы.

*Теорема 1.* Предложенный МНП всегда сходится, причем результат – эталон  $X_\mu^*$  – либо является «ближайшим соседом» к входному объекту  $X$ , либо удовлетворяет условию останова (20).

*Теорема 2.* Если классифицируемым объектом является один из эталонов  $X = X_v, v \in \{1, \dots, R\}$  и

$$\begin{aligned} & (\forall i \in \{1, \dots, N\}, \forall j \in \{1, \dots, R\}) \\ & (i \neq j) \Rightarrow (\rho_{KL}(X_v / X_i) \neq \rho_{KL}(X_v / X_j)). \end{aligned} \quad (21)$$

то количество вычислений рассогласований, выполняемых МНП, не зависит от размера базы данных эталонов.

В общем же случае, суммарное число  $C = N + (M + 1) \cdot K \leq R$  выполняемых согласно (20) проверок может существенно выигрывать по сравнению с объемом  $R$  редуцированной БД. Этот выигрыш обусловлен, в частности, тем обстоятельством, что для рассогласования Кульбака-Лейблера (как, впрочем, и для многих других расстояний, в частности, метрики  $l_1$ ) вероятность  $p$  того, что ближайший к  $X$  эталон  $X^*$  принадлежит множеству  $X^{(M)}$ , как правило, существенно превышает вероятность того, что  $X^*$  будет одним из  $M$  наудачу выбранных эталонов:

$$p = P\{X^* \in X^{(M)}\} \gg p_0 = (M/R). \quad (22)$$

В этом и состоит эффект направленного перебора. А отличия в количестве этапов  $K$  алгоритма для разных экспериментов объясняются тем, что вероятность  $p$  зависит не только от свойств используемой метрики, но и от свойств входного изображения и эталонов в БД.

#### Результаты экспериментальных исследований

Для проведения экспериментального исследования эффективности предложенной теоретико-вероятностной модели изображения рассмотрим задачу

распознавания людей по фотографиям лиц [20, 21]. В качестве их предварительной обработки для выделения лиц использовалась библиотека OpenCV. Для вычисления рассогласований все фотографии предварительно разбивались на 16 (4×4) фрагментов. Общее рассогласование между двумя фотографиями рассчитывалось как сумма рассогласований между фрагментами, вычисленными по формуле (13). Подобная фрагментация позволяет учесть неоднородное освещение изображений – каждый фрагмент нормировался по яркости [9] для ослабления проблемы вариативности освещения входных объектов. Распознавание проводилось на компьютере Pentium-IV (2,9ГГц, 1Гб ОЗУ) средствами системы распознавания людей по фотографиям [22] на базе виртуальной машины Java Runtime Environment 1.6

Исходя из наших исследований [18], параметр МНП  $N$  может выбираться практически произвольно, не оказывая значимое влияние на эффективность метода, если выполнено отмеченное выше условие  $N \ll R$ . В то же время параметр  $M$  значительно важнее. После ряда экспериментов [18] для нескольких БД фотографий лиц наилучшие (с точки зрения наименьшего объема вычислений) значения параметров предложенного метода были выбраны следующим образом:  $N = 1, M = 64$ .

Вначале воспользуемся большой БД фотографий людей Essex [23]. Из 5187 фотографий 400 различных людей с помощью алгоритма самообучения (14), (15) на основе метрики Кульбака-Лейблера (13) были отобраны в качестве эталонов  $R = 954$  наиболее различающиеся изображения. Для алгоритма самообучения без использования МНП количество вычислений расстояний (13) составило чуть более 2,5 млн. Если же, наряду с (14),(15), использовался МНП (18)-(20), то общий объем вычислений снизился на 68% и составил чуть более 0,5 млн. вычисления рассогласований (13). Порог  $\rho_0 = 0,125$  был подобран экспериментально. При этом вероятность того, что в один класс  $X_r$  попали изображения разных людей, составила 1%.

После этого для тестирования точности распознавания по принципу МИР (7), (11). (13) были взяты другие 1200 фотографий тех же людей. Параметр  $k$  (11) был выбран равным 0,125. В результате, в 96,8% случаев было получено точное решение

$X^* = X_r^*$ . Здесь в среднем вычислялись рассогласований до 114 эталонов (11,9% от количества эталонов в редуцированной БД или 2,1% от объема первоначальной БД – до этапа самообучения). Если же для указанной задачи применять только случайный поиск (без использования МНП) на основе критерия останова (17), то в среднем проверяются 55,3% эталонов.

Для критерия (2) точность АРИ оказалась аналогичной предложенной модели и критерию МИР (11), (13) – оценка вероятности ошибки составила 3,1%, в среднем по МНП вычислялись рассогласования (2) до 12% эталонов.

Для сравнения при использовании МНП совместно с метрикой  $l_1$  в среднем вычислялись расстояния (1) до 390 эталонов (23% от объёма редуцированной БД или 7,5% от объёма БД до этапа самообучения). При этом точность классификации составила 94%

Среднее время распознавания одного изображения с помощью метода ближайшего соседа в среднем на распознавание одного изображения составило 1,4 с для полной (нередуцированной БД) и 210 мс для полного перебора редуцированной БД. Среднее же время распознавания одного изображения по МНП на том же компьютере составило 27 мс. Вероятность ошибки полного перебора нередуцированной базы несколько ниже (1,9%), чем для МНП (3,1% - 3,2%). Это обстоятельство вызвано тем, что при оценке по обучающей выборке порога  $\rho_0$  досрочного останова (17) ошибка первого рода (False Reject Rate, FRR) была зафиксирована равной 5%. Её уменьшение приведёт к уменьшению ошибки классификации, но и к увеличению ошибки второго рода (False Accept Rate, FAR) и, в свою очередь, к росту объёма вычислений по МНП.

Все результаты для первого эксперимента сведены в табл. 1.

Таблица 1. Результаты АРИ для критериев (2) и (13) и БД Essex

		(2)	(13)
Сокращение БД за счёт редукции (в %)		17%	18%
Количество вычислений рассогласований при самообучении по МНП (в % от полного перебора)		31%	32%
Оценка вероятности ошибки (в %)	полный перебор БД	1,9%	1,9%
	полный перебор редуцированной БД	2,6%	2,7%
	МНП для редуцированной БД	3,1%	3,2%
Количество вычислений рассогласований при классификации по МНП (в % от полного перебора редуцированной БД)		12%	11%
Среднее время классификации (в мс)	полный перебор БД	1350	1400
	полный перебор редуцированной БД	200	210
	МНП для редуцированной БД	26	27

Кажется, нет никакой разницы между двумя критериями (2) и (13), основанными на сопоставлении гистограмм. Однако это обстоятельство для первого эксперимента обусловлено тем, что все изображения в БД Essex получены в полностью идентичных условиях (освещение, ракурс). Поэтому далее эксперимент был повторён для более сложной БД AT&T

(бывшая ORL) [24]. Из 260 фотографий 40 людей алгоритм самообучения на основе информационной теории восприятия (14), (15) для рассогласований Кульбака-Лейблера (13) выделил 97 кластеров. Для традиционного алгоритма кластеризации потребовалось вычислить около 13000 рассогласований (13), в то время как для МНП – около 6000 (48% от полного перебора). В данном случае это сокращение вычислительной сложности менее существенно, чем в предыдущем эксперименте с базой Essex, так как объём БД AT&T на порядок меньше. Далее для распознавания других 140 фотографий тех же людей вероятность ошибки АРИ по критерию МИР (13) составила 4,2%. Количество вычислений рассогласований по МНП составило 27% по сравнению с полным перебором редуцированной БД.

Для критерия (2) сопоставления гистограмм качество АРИ также несколько ниже, чем для предложенной модели. Точность АРИ упала до 93% при аналогичных показателях быстродействия МНП – 26,5% вычислений (2) по сравнению с полным перебором редуцированной БД.

Результаты для БД AT&T сведены в табл. 2. Отметим, что здесь вероятность ошибки классификации для редуцированной БД даже ниже аналогичного показателя для исходной БД, что в определённой мере показывает важность предварительной кластеризации множества эталонов.

Таблица 2. Результаты АРИ для критериев (2) и (13) и БД AT&T

		(2)	(13)
Сокращение БД за счёт редукции (в %)		36%	37%
Количество вычислений рассогласований при самообучении по МНП (в % от полного перебора)		49%	48%
Оценка вероятности ошибки (в %)	полный перебор БД	6,9%	4,4%
	полный перебор редуцированной БД	5,9%	4,0%
	МНП для редуцированной БД	7%	4,2%
Количество вычислений рассогласований при классификации по МНП (в % от полного перебора редуцированной БД)		26%	27%
Среднее время классификации (в мс)	полный перебор БД	60	63
	полный перебор редуцированной БД	19	21
	МНП для редуцированной БД	9	10

По результатам проведённых экспериментов (табл. 1, 2) можно сделать следующие выводы. Во-первых, гистограммные (и, в частности, предложенная теоретико-вероятностная) модели изображений существенно превосходят традиционное сопоставление матриц интенсивностей пикселей, во всяком

случае, для изображений, полученных при схожих условиях съёмки. Во-вторых, предварительная кластеризация БД зачастую является необходимым инструментом регулирования её размера с целью оптимизации множества эталонов. И, в-третьих, МНП позволяет повысить эффективность как вычислительной процедуры самообучения, так и для последующей классификации, при этом выигрыш тем выше, чем больше объём БД эталонов.

### Заключение

Способ, основанный на вычислении мер близости, является одним из самых эффективных подходов к решению задачи классификации изображений. При этом качество классификации при применении различных мер близости может существенно различаться в зависимости от конкретных особенностей задачи (характера искажений, расположения и др.). В связи с этим в последнее время уделяется повышенное внимание возможности применения мер близости, которые в определённых условиях могут дать лучший результат, чем традиционные метрики (Евклида, Манхэттена). С этой точки зрения исследование эффективности информационного рассогласования Кульбака-Лейблера, проведённое в настоящей работе, представляется весьма перспективным.

Изложенный теоретико-вероятностный подход в задачах АРИ, по-видимому, не имеет серьёзных альтернатив ввиду острейшей проблемы вариативности изображений, связанной с зависимостью одновременно от многих факторов (как признаков изображённого объекта, так и условиями наблюдения). Однако обеспечение достаточно полного покрытия пространства объектов, в свою очередь, наталкивается на другую проблему – недостаточной высокой вычислительной эффективности традиционных методов ближайших соседей. Естественное приемлемое решение – кластеризация БД изображений по критерию МИР, заимствованная из информационной теории восприятия речи, оказалась не только продуктивной в задачах АРИ, но и одновременно послужила своеобразной точкой опоры для разработки нового критерия, основанного на методе направленного перебора альтернатив, с высокими динамическими характеристиками.

### Литература

1. **Theodoridis, S.** Pattern Recognition / S. Theodoridis, C. Koutroumbas. – Elsevier Inc. – 4th Edition, 2009. – 840 p.
2. **Forsyth, D.A.** Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce // Upper Saddle River, NJ: Prentice Hall, 2003. – 693 p.
3. Face Processing: Advanced Modeling and Methods / edited by W. Zhao, R. Chellappa. – Elsevier: Academic Press, 2005. – 768 p.
4. **Kullback, S.** Information Theory and Statistics / S. Kullback. – Dover Pub., 1978. – 408 p.
5. **Савченко, В.В.** Информационная теория восприятия речи / В.В. Савченко // Известия вузов России. Радиоэлектроника. – 2007. – № 6. – С. 3-9.
6. **Савченко, А.В.** Метод направленного перебора альтернатив в задаче распознавания полутоновых изображений / А.В. Савченко // Автотметрия. – 2009. – Т. 45, № 3. – С. 90-98.
7. **Fukunaga, K.** Introduction to Statistical Pattern Recognition, 2nd ed. / K. Fukunaga. – New York: Academic Press, Inc., 1991. – 591 p.
8. **Мясников, В.В.** Анализ методов построения эффективных линейных локальных признаков цифровых сигналов и изображений / В.В. Мясников, А.Ю. Баврина, О.А. Титова // Компьютерная оптика. – 2010. – Т. 34, № 3. – С. 193-201.
9. **Shapiro, L.** Computer vision / L. Shapiro, G. Stockman // Prentice Hall, 2001. – 752 p.
10. **Бибиков, С.А.** Информационная технология коррекции теневых искажений на цветных цифровых изображениях / С.А. Бибиков, А.В. Никоноров, В.А. Фурсов // Компьютерная оптика. – 2010. – Т. 34, № 1. – С. 124-131.
11. **Swain, M.J.** Color indexing / M.J. Swain, D.H. Ballard // International Journal of Computer Vision. – 1991: 11-32.
12. **Wong, K.M.** Dominant Color Image Retrieval using Merged Histogram / K.M. Wong, C.H. Cheung, L.M. Po // Proc. the 2003 Int. Symposium. – 2003. – Vol. 2. – P. 908-911.
13. **Yoo, G.-H.** Content-based image retrieval using shifted histogram / Gi-Hyoung Yoo, Beob Kyun Kim and Kang Soo You // ICCS, LNCS 4489. – 2007. – P. 894-897.
14. **Shi, J.** Normalized cuts and image segmentation / J. Shi, L. Sandford, J. Ponce // IEEE Conference Computer Vision and pattern Recognition. – 1997. – P. 731-737.
15. **Кухарев, Г.А.** Поиск изображений лиц в больших базах данных / Г.А. Кухарев // Мир измерений. – 2009. – № 4(98). – С. 22-30.
16. **Kohonen, T.** Self Organizing Maps / T. Kohonen. – New York: Springer Berlin Heidelberg, 2001. – 500 p.
17. **Цыпкин, Я.З.** Адаптация и обучение в автоматических системах / Я.З. Цыпкин. – М.: Наука, 1968. – 400 с.
18. **Савченко, А.В.** Распознавание образов: метод направленного перебора для задач классификации с большим количеством альтернатив / А.В. Савченко // Lambert Academic Publishing, 2011. – 140 с.
19. **Beis, J.** Shape indexing using approximate nearest-neighbour search in highdimensional spaces / J. Beis, D.G. Lowe // Conference on Computer Vision and Pattern Recognition. – 1997. – P. 1000-1006.
20. **Фурсов, В.А.** Распознавание лиц по показателям сопряженности в пространстве суммирующих инвариантов / В.А. Фурсов, Н.Е. Козин // Компьютерная оптика. – 2008. – Т. 32, № 4. – С. 400-402.
21. **Кухарев, Г.А.** Алгоритмы двумерного анализа главных компонент для задач распознавания изображений лиц / Г.А. Кухарев, Н.Л. Щеголева // Компьютерная оптика. – 2010. – Т. 34, № 4. – С. 119-125.
22. **Савченко, А.В.** Автоматизированная система распознавания людей по фотографиям лиц / А.В. Савченко // Программа для ЭВМ. Свид-во о гос. регистрации № 2009616508 по заявке 2009615314 от 28.09.2009.
23. The Essex Face database: <http://cswww.essex.ac.uk/mv/allfaces/index.html>.
24. The AT&T (ORL) database - [www.uk.research.att.com/public/data/att\\_faces.zip](http://www.uk.research.att.com/public/data/att_faces.zip).



### References

1. **Theodoridis, S.** Pattern Recognition / S. Theodoridis, C. Koutroumbas. – Elsevier Inc. – 4th Edition, 2009. – 840 p.
2. **Forsyth, D.A.** Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce // Upper Saddle River, NJ: Prentice Hall, 2003. – 693 p.
3. Face Processing: Advanced Modeling and Methods / edited by W. Zhao, R. Chellappa. – Elsevier: Academic Press, 2005. – 768 p.
4. **Kullback, S.** Information Theory and Statistics / S. Kullback. – Dover Pub., 1978. – 408 p.
5. **Savchenko, V.V.** Information theory of speech perception / V.V. Savchenko // Izvestia vuzov Rossii. Radioelektronika. – 2007. – Vol. 6. – P. 3-9. – (in Russian).
6. **Savchenko, A.V.**, Method of directed enumeration of alternatives in the problem of automatic recognition of half-tone images / A.V. Savchenko // Optoelectronics, Instrumentation and Data Processing. – 2009. – Vol. 3, N 45. – P. 255-261. – (in Russian).
7. **Fukunaga, K.** Introduction to Statistical Pattern Recognition, 2nd ed. / K. Fukunaga. – New York: Academic Press, Inc., 1991. – 591 p.
8. **Myasnikov, V.V.** Analysis of the methods for construction of linear local features / V.V. Myasnikov, A.U. Bavrina, O.A. Titova // Computer optics. – 2010. – Vol. 34, N 3. – P. 193-201. – (in Russian).
9. **Shapiro, L.** Computer vision / L. Shapiro, G. Stockman. – Prentice Hall, 2001. – 752 p.
10. **Bibikov, S.A.** Correction of shadow artifacts on colorful digital images / S.A. Bibikov, A.V. Nikonorov, V.A. Fursov // Computer optics. – 2010. – V. 34, N 1. – P. 124-131. – (In Russian).
11. **Swain, M.J.** Color indexing / M.J. Swain, D.H. Ballard // International Journal of Computer Vision. – 1991: 11~32.
12. **Wong, K.M.** Dominant Color Image Retrieval using Merged Histogram / K.M. Wong, C.H. Cheung, L.M. Po // Proc. the 2003 Int. Symposium. – 2003. – Vol. 2. – P. 908-911.
13. **Yoo, G.-H.** Content-based image retrieval using shifted histogram / Gi-Hyoung Yoo, Beob Kyun Kim and Kang Soo You // ICCS, LNCS 4489. – 2007. – P. 894–897.
14. **Shi, J.** Normalized cuts and image segmentation / J. Shi, L. Sandford, J. Ponce // IEEE Conference Computer Vision and pattern Recognition. – 1997. – P. 731-737.
15. **Kukharev, G.A.** Find pictures of persons in large databases / G.A. Kukharev // World measurement. – 2009. – Vol. 4(98). – P. 22-30. – (in Russian).
16. **Kohonen, T.** Self Organizing Maps / T. Kohonen. – New York: Springer Berlin Heidelberg, 2001. – 500 p.
17. **Tsyarkin, Y.Z.** Adaptation and training in automated systems // Y.Z. Tsyarkin – Moscow: “Nauka” Publisher, 1968. – 400 p. – (in Russian).
18. **Savchenko, A.V.** Pattern recognition: method of directed enumeration of alternatives in the classification problem with large database / A.V. Savchenko // Lambert Academic Publishing, 2011. – 140 p. – (in Russian).
19. **Beis, J.** Shape indexing using approximate nearest-neighbour search in highdimensional spaces / J. Beis, D.G. Lowe // Conference on Computer Vision and Pattern Recognition. – 1997. – P. 1000-1006.
20. **Fursov, V.A.** Face recognition on the basis of conjugation indexes in the space of summarizing invariants / V.A. Fursov, N.E. Kozin // Computer optics. – 2008. – Vol. 32 (4). – P. 400-402. – (in Russian).
21. **Kukharev, G.A.** Algorithms of two-dimensional principal component analysis for face recognition / G.A. Kukharev, N.L. Schegoleva // Computer optics. – 2010. – Vol. 34(4). – P. 119-125. – (in Russian).
22. **Savchenko, A.V.** Automated system for face recognition / A.V. Savchenko // Computer program. The certificate of state registration N 2009616508 under the demand on 2009615314 from 28.09.2009. – (in Russian).
23. The Essex Face database: <http://cswww.essex.ac.uk/my/allf/aces/index.html>.
24. The AT&T (ORL) database - [www.uk.research.att.com/pub/data/att\\_faces.zip](http://www.uk.research.att.com/pub/data/att_faces.zip).

## PROBABILITY HALFTONE IMAGE MODEL IN A PROBLEM OF UNSUPERVISED PATTERN RECOGNITION BASED ON DIRECTED ENUMERATION METHOD

*A.V. Savchenko*

*National Research University “Higher School Of Economics” – Nizhny Novgorod*

### Abstract

The problem of automatic image recognition with unsupervised learning is put and solved by the direct enumeration method. The new probability-theoretic model of halftone image is investigated. Its application has allowed synthesizing new recognition criterion with self-training on the basis of optimum Bayesian statistical rule. The experimental results in a problem of face recognition are presented. It is shown that the proposed image model combining with directed enumeration method is characterized by frequentative calculations’ reduction with preservation of image recognition accuracy.

**Key words:** automatic image recognition with unsupervised training, self training, minimum discrimination information principle, directed enumeration method.

*Сведения об авторе*

**Савченко Андрей Владимирович**, 1985 года рождения. В 2008 году с отличием окончил Нижегородский государственный технический университет им Р.А. Алексеева (НГТУ) по специальности «Прикладная математика и информатика». Кандидат технических наук (2010 год), работает старшим преподавателем кафедры информационных систем и технологий Национального исследовательского университета Высшая школа экономики – Нижний Новгород (НИУ ВШЭ – Н.Новгород). Область научных интересов: распознавание образов, распознавание изображений, обработка изображений.

Страница в Интернете: <http://www.hse.ru/org/persons/9216523>.

E-mail: [avsavchenko@hse.ru](mailto:avsavchenko@hse.ru).

**Andrey Vladimirovich Savchenko** (b. 1985) graduated with honours (2008) from the Nizhny Novgorod State Technical University, majoring in Applied Mathematics and Informatics. He received his Candidate in Technics (2010) degrees from State University Higher School of Economics. He works as the teacher in the National research university Higher School of Economics, Nizhny Novgorod, department of Information systems and technologies. His research interests are currently focused on pattern recognition, image recognition, image processing.

---

*Поступила в редакцию 28 апреля 2011 г.*