

## РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ВЕРОЯТНОСТНОЙ НЕЙРОННОЙ СЕТИ С ПРОВЕРКОЙ ОДНОРОДНОСТИ

Савченко А.В.

Национальный исследовательский университет Высшая школа экономики – Нижний Новгород

### Аннотация

Предложено использование вероятностной нейронной сети с проверкой однородности в задаче распознавания изображений. Показано, что это решение является оптимальным в байесовском смысле, если задача рассматривается в терминах статистической проверки однородности выборок признаков входного и эталонных изображений. Рассматривается проблема недостаточной вычислительной эффективности оптимального алгоритма при наличии многих альтернативных классов и большой размерности признакового пространства. Исследуется возможность её преодоления для случая дискретных признаков путём синтеза нового критерия, основанного на сопоставлении гистограмм признаков входного и эталонных изображений. Показано, что частным случаем такого критерия является правило ближайшего соседа с популярными мерами близости хи-квадрат и Йенсена–Шеннона. Приведены результаты экспериментального исследования в задаче идентификации личности по фотографии лица для популярных баз данных AT&T и JAFFE. Продемонстрировано, что предложенный подход существенно превосходит по качеству традиционное решение, основанное на сведении распознавания к задаче статистической классификации.

**Ключевые слова:** автоматическое распознавание изображений, распознавание лиц, вероятностная нейронная сеть, проверка однородности выборок, правило ближайшего соседа.

### Введение

Большинство исследований в области автоматического распознавания изображений (АРИ) [1] сосредоточены на повышении точности, надёжности и вычислительной эффективности существующих решений за счёт применения новых признаков [2, 3], мер близости [4], коллективных решений [5], приближённых методов ближайшего соседа [6] и пр. При этом достаточно часто применяется универсальный статистический подход [7, 8], согласно которому АРИ рассматривается как задача статистической классификации выборки признаков изображения [9]. В настоящее время наиболее часто используются непараметрические методы оценивания распределений классов, такие как нейросетевые методы и, в частности, вероятностные нейронные сети (probabilistic neural network, PNN) [10]. К сожалению, качество получаемого решения всё ещё недостаточно для большинства практически важных приложений (например, распознавание объектов по видео с вариативным освещением, ракурсом, размером и т.п. [11]).

В данной работе для решения задачи АРИ предлагается воспользоваться авторской модификацией PNN [12], основанной на редукции задачи классификации выборки к проверке статистической гипотезы о её однородности [13] с одной из обучающих выборок – множеством признаков эталонного изображения (а не к проверке гипотезы о распределении элементов входной выборки, как это принято в традиционной PNN). Для оценки распределения класса используется совместная выборка – объединение входной и обучающей выборок. В результате применение этой модификации приводит к следующим преимуществам [12] по сравнению с классическим подходом: более быстрая скорость сходимости

к оптимальному решению, существенно меньшая зависимость точности классификации от параметров гауссовой ядерной функции Парзена–Розенблатта, симметрия получающейся меры близости.

Серьёзным недостатком аппарата PNN (и, соответственно, предложенной модификации) [10, 12] является требование к хранению и обработке всех элементов всех обучающих выборок (т.н. memory-based approach), что приводит к большим затратам при реализации этих алгоритмов и делает их физически нереализуемыми для сложных приложений, таких как АРИ. Поэтому актуальной становится задача синтеза критериев, эквивалентных при некоторых, достаточно общих ограничениях оптимальному решению [12], но не требующих хранения и сопоставления всех признаков.

В настоящей работе предлагается исследовать возможность разработки таких алгоритмов при ограничении на дискретность признаков [14]. Указанное ограничение достаточно характерно для обработки изображений, где множества значений признаков чаще всего являются конечными. В этом случае выборка может быть полностью описана своей гистограммой. В результате требования к объёму оперативной памяти и вычислительной сложности уменьшаются пропорционально отношению объёма входной выборки к количеству различных значений признака. В ходе экспериментального исследования в задаче распознавания лиц [15, 16] для популярных баз данных AT&T [17] и JAFFE [18] продемонстрировано повышение точности распознавания по сравнению с традиционными подходами при доступности ограниченного количества эталонов для каждого класса (в худшем случае, один эталон, т.н. one class per person problem [19]).

### Задача распознавания изображений

Пусть задано множество из  $L > 1$  полутоновых изображений  $X_l = [x_{uv}^{(l)}]$ ,  $l = \overline{1, L}$ ,  $u = \overline{1, U_l}$ ,  $v = \overline{1, V_l}$ . Здесь  $U_l$  – высота изображения,  $V_l$  – его ширина;  $x_{uv}^{(l)}$  – интенсивность точки изображения с координатами  $(u, v)$ ;  $l$  – номер эталона. Задача распознавания изображений состоит в том, чтобы отнести вновь поступающее (на вход) изображение  $X = [x_{uv}]$  (с высотой  $U$  и шириной  $V$ ) к одному из классов, заданных эталонами  $X_l$ . Каждый класс характеризуется тем, что принадлежащие ему объекты обладают некоторой общностью, сходством в характеристиках.

В распознавании образов общепринято [1, 11, 20] введение так называемого признакового пространства (feature set), когда исходные матрицы яркостей пикселей вначале преобразуются в набор простейших локальных признаков  $[f_{uv}^{(l)}]$ , где для каждой точки изображения  $(u, v)$  вычисляется признак

$$f_{uv}^{(l)} = f(X_l, u, v), \quad u = \overline{1, U_l}, \quad v = \overline{1, V_l}.$$

Здесь  $f(\cdot)$  – некая функция, выбираемая исследователем заранее. Аналогично строится набор признаков  $[f_{uv}^{(l)}]$  входного изображения  $X$ . В ранних работах в качестве признаков использовались сами отсчёты яркости [21]

$$f(X_l, u, v) = x_{uv}^{(l)}. \quad (1)$$

В настоящее время достаточно часто стали применяться текстурные признаки [2, 3]. Например, хорошие результаты показало направление градиента яркости [3, 22, 23]

$$\theta_{uv}^{(l)} = \arctg \frac{x_{u+1, v+1}^{(l)} - x_{u, v}^{(l)}}{x_{u+1, v}^{(l)} - x_{u, v+1}^{(l)}}. \quad (2)$$

Этот признак характеризуется весьма желательным для многих практических приложений АРИ свойством [1, 2] – инвариантностью к интенсивности освещения. Далее обычно [1–3] на основе локальных признаков вида (1) или (2) путём функциональных преобразований (нормировки, масштабирования и т.п.) и фильтрации (выделения регионов интереса, region of interest (ROI)) [2] формируются более сложные векторы признаков, которые и используются для решения АРИ. Тем не менее в дальнейшем во избежание новых обозначений мы будем предполагать, что имеем дело с простейшими признаками, в которых размерность набора признаков совпадает с размерностью анализируемых изображений.

Для распознавания в условиях небольшого числа эталонов для каждого класса обычно применяют правило ближайшего соседа [20]. Исследователь выбирает некоторую меру близости  $\rho(X, X_l)$  и принимает решение в пользу класса  $X_{l^*}$  такого, что

$$l^* = \arg \min_{l=\overline{1, L}} \rho(X, X_l). \quad (3)$$

Для определения расстояния  $\rho(X, X_l)$  можно сопоставлять непосредственно сами значения призна-

ков. К сожалению, этот подход не всегда позволяет получить удовлетворительные результаты [23], что связано с известной [1] вариативностью зрительных образов, которая не может быть устранена стандартными алгоритмами очистки изображения от шума. Поэтому более перспективным представляется подход [2, 3], в котором каждому изображению ставится в соответствие гистограмма его признаков, при этом дальнейшее решение АРИ (3) основывается на сопоставлении гистограмм [3, 22]. Для построения гистограммы признака его область определения  $[f_{\min} : f_{\max}]$  разобьём на  $N$  отрезков одинакового размера  $\Delta f = \frac{f_{\max} - f_{\min}}{N}$ . Тогда элементы гистограммы  $H^{(l)} = [h_1^{(l)}, h_2^{(l)}, \dots, h_N^{(l)}]$  могут быть вычислены стандартным способом [9, 20]. Аналогично вычисляется и гистограмма  $H = [h_1, \dots, h_N]$  изображения  $X$ . Решение принимается по критерию (3) близости гистограмм, например, в традиционной метрике  $L_2$  Евклида [2]

$$\rho_2(X, X_l) = \sum_{i=1}^N (h_i - h_i^{(l)})^2. \quad (4)$$

Заметим, что зачастую на этапе предварительной обработки после детектирования распознаваемых объектов (например, лиц на фотографиях [15]) их изображения разбиваются на несколько (возможно, пересекающихся) частей [2, 5, 16], а затем решение принимается по принципу минимума суммы рассогласований между соответствующими частями. Однако для упрощения дальнейших выкладок мы будем предполагать, что гистограмма признака вычисляется для всего изображения.

К сожалению, критерий (4) не всегда демонстрирует удовлетворительное качество АРИ [5, 11, 23]. Повышение точности распознавания большинство исследователей связывают с применением статистического подхода, который и будет рассмотрен в следующем разделе.

### Распознавание изображений на основе вероятностных нейронных сетей

Для применения статистического подхода [7, 9] предполагаем, что признаки  $f_{uv}^{(l)}$  в одном фрагменте изображения являются реализацией некой случайной величины  $F_l$ . Как сказано выше, далее для упрощения формул мы предполагаем, что анализируется не отдельный фрагмент, а всё изображение целиком.

В таком случае задача сводится к проверке  $L$  гипотез о распределении  $\mathbf{P}_l$ ,  $l = \overline{1, L}$ , изображения на входе  $\mathbf{P}$ :

$$W_l : \mathbf{P} = \mathbf{P}_l. \quad (5)$$

Оптимальное решение при условии полной априорной неопределённости (равной априорной вероятности появления каждого класса) тогда даёт [7, 20] классический критерий максимального правдоподобия

биа (условной вероятности  $P(X|W_l)$ ) принадлежности объекта  $X$  классу  $l$ )

$$P(X|W_l) \rightarrow \max_l. \tag{6}$$

Далее, делая «наивное» предположение о независимости [5, 7] всех признаков изображений, используем ядерную оценку правдоподобия в (6) на основе классической PNN [10]

$$\hat{P}(X|W_l) = \frac{1}{(U_l V_l)^{UV}} \times \prod_{u=1}^U \prod_{v=1}^V \sum_{u_l=1}^{U_l} \sum_{v_l=1}^{V_l} K(f_{uv}, f_{u_l v_l}^{(l)}). \tag{7}$$

Здесь  $K()$  – ядерная функция (например, гауссовское ядро Розенблатта–Парзена [10]). Использование оценки (7) наталкивается на классические проблемы PNN – невысокая скорость и завышенные требования к доступной памяти, обусловленные тем, что сохраняются все признаки всех эталонов. В результате практическая реализация критерия (6), (7) при  $L, U, V \gg 1$  становится невозможной. Поэтому далее предположим, что случайные величины  $F_l$  являются дискретными с конечным множеством из  $N$  значений  $\{f_1, \dots, f_N\}$ . В этом случае выборки  $[f_{uv}^{(l)}]$  могут быть полностью описаны гистограммой  $H^{(l)}$ . В этом случае критерий (6), (7) может быть записан в виде

$$\prod_{i=1}^N \left( \sum_{j=1}^N K_{ij} h_j^{(l)} \right)^{U V h_i} \rightarrow \max_l. \tag{8}$$

Здесь  $K_{ij} = K(f_i, f_j)$  – значения ядерной функции (предварительно вычисленные) между всевозможными парами значений признаков. Если для каждого эталона  $X_l$  и входного изображения  $X$  заранее вычислить свёртку его гистограммы с ядром

$$h_{K;i}^{(l)} = \sum_{j=1}^N K_{ij} h_j^{(l)}, \tag{9}$$

$$h_{K;i} = \sum_{j=1}^N K_{ij} h_j, \quad i = \overline{1, N},$$

выполнить логарифмирование (9) и вычесть результат из не зависящей от  $l$  величины  $\sum_{i=1}^N h_i \ln h_{K;i}$ , то получим окончательное, основанное на аппарате PNN решение

$$\sum_{i=1}^N h_i \ln \frac{h_{K;i}}{h_{K;i}^{(L)}} \rightarrow \min_l. \tag{10}$$

Заметим, что если здесь в качестве ядра используется дискретная дельта-функция, то  $h_{K;i}^{(l)} = h_i^{(l)}$ ,  $h_{K;i} = h_i$  и выражение (10) будет эквивалентно правилу ближайшего соседа с достаточно часто используемым в АРИ [5, 6] информационным рассогласованием Кульбака–Лейблера [24]

$$\sum_{i=1}^n \ln \frac{h_i}{h_i^{(L)}} \rightarrow \min_l. \tag{11}$$

Таким образом, оптимальное в байесовском смысле решение задачи статистической классификации изображений (5) принимается по критерию минимума решающей статистики (10) или (11). К сожалению, точность таких решений в практических приложениях АРИ особенно, с небольшим количеством эталонов для каждого класса, оказывается иногда [11] даже ниже точности традиционного подхода (4). Поиск причин низкого качества оптимального решения (10) и составляет основную цель настоящей работы.

**Вероятностная нейронная сеть с проверкой однородности в задаче распознавания изображений**

Недостаточную точность решения (10) задачи статистической классификации (5) можно объяснить тем [13], что оно либо требует знания истинных распределений  $P_l$ , либо допускает, что правдоподобие оценивается (7) на основе обучающих выборок  $[f_{uv}^{(l)}]$  большого объёма по сравнению с входной выборкой  $[f_{uv}]$ . Последний факт подчёркивается и асимметрией решающего правила (10), (11).

Поэтому в настоящей работе для решения задачи АРИ воспользуемся известным альтернативным подходом [13], основанным на предложенной нами ранее для задачи классификации выборок независимых одинаково распределённых объектов PNN с проверкой однородности [12]. В настоящем разделе проведём адаптацию этой сети для случая дискретных признаков, выборки которых полностью определяются своими гистограммами.

Согласно известному подходу к распознаванию образов [13] с неизвестными заранее распределениями классов, задачу классификации следует сводить не к проверке гипотез о распределении вида (5), а к проверке статистической однородности обучающей и входной выборок. Тогда задача АРИ сводится к проверке простых гипотез  $W_l$  об однородности выборок признаков входного объекта и  $l$ -го эталона:

$$W_l : \text{выборки признаков } [f_{uv}^{(l)}] \text{ и } [f_{uv}] \text{ однородны (одинаково распределены)}. \tag{12}$$

Её решение даёт классический байесовский подход минимума среднего риска [13], в рамках которого для случая полной априорной неопределённости ищется максимум (по возможному закону распределения входного объекта  $P^*$  и эталонов  $P_j^*$ ,  $j = \overline{1, L}$ ) правдоподобия совместной выборки  $\{X, X_1, \dots, X_L\}$  при справедливости гипотезы  $W_l$ :

$$\sup_{P^*} \sup_{P_j^*, j=\overline{1, L}} P(\{X, X_1, \dots, X_L\} | W_l) \rightarrow \max_l, \tag{13}$$

где  $\sup$  – верхняя граница по всевозможным значениям распределений. Предполагая независимость

всех выборок (признаков входного и эталонных изображений), верхнюю границу в (13) запишем как

$$\begin{aligned} \sup_{\mathbf{P}^*} \sup_{\mathbf{P}_j^*, j=1, \dots, L} P(\{X, X_1, \dots, X_L\} | W_l) = \\ = \sup_{\mathbf{P}^*} P(X | W_l) \cdot \prod_{j=1}^L \sup_{\mathbf{P}_j^*} P(X_j | W_l). \end{aligned}$$

Так как при справедливости  $W_l$  выборки  $X_j, j \neq l$  не зависят от  $W_l$ , то

$$\begin{aligned} \prod_{j=1}^L \sup_{\mathbf{P}_j^*} P(X_j | W_l) = \sup_{\mathbf{P}_l^*} P(X_l | W_l) \cdot \prod_{\substack{j=1 \\ j \neq l}}^L \sup_{\mathbf{P}_j^*} P(X_j) = \\ = \frac{\sup_{\mathbf{P}_l^*} P(X_l | W_l) \cdot \prod_{j=1}^L \sup_{\mathbf{P}_j^*} P(X_j)}{\sup_{\mathbf{P}_l^*} P(X_l)}. \end{aligned}$$

С учётом этих двух равенств, разделив для симметрии (13) на не зависящий от  $l$  множитель

$$\sup_{\mathbf{P}^*} P(X) \cdot \prod_{j=1}^L \sup_{\mathbf{P}_j^*} P(X_j),$$

получаем следующий критерий

$$\frac{\sup_{\mathbf{P}^*} P(X | W_l) \cdot \sup_{\mathbf{P}_l^*} P(X_l | W_l)}{\sup_{\mathbf{P}^*} P(X) \cdot \sup_{\mathbf{P}_l^*} P(X_l)} \rightarrow \max_l. \quad (14)$$

Как известно [12, 13], верхняя граница в (14) достигается при равенстве распределений их максимально правдоподобным оценкам вида (7). Например, для вычисления верхней границы правдоподобия  $P(X | W_l)$  используется объединённая выборка  $\{X, X_j\}$ :

$$\begin{aligned} \sup_{\mathbf{P}^*} P(X | W_l) = \frac{1}{(UV + U_l V_l) UV} \times \\ \times \prod_{u=1}^U \prod_{v=1}^V \left( \sum_{u_1=1}^U \sum_{v_1=1}^V K(f_{uv}, f_{u_1 v_1}) + \right. \\ \left. + \sum_{u_j=1}^{U_l} \sum_{v_j=1}^{V_l} K(f_{uv}, f_{u_j v_j}^{(l)}) \right), \end{aligned}$$

или, переходя к гистограммам выборок,

$$\begin{aligned} \sup_{\mathbf{P}^*} P(X | W_l) = \frac{1}{(UV + U_l V_l) UV} \times \\ \times \prod_{i=1}^N \left( UV \sum_{j=1}^N K_{ij} h_j + U_l V_l \sum_{j=1}^N K_{ij} h_j^{(l)} \right)^{U V h_i}. \end{aligned}$$

Преобразуя последнее равенство, запишем его в упрощённой форме

$$\sup_{\mathbf{P}^*} P(X | W_l) = \prod_{i=1}^N \left( \frac{UV h_{K:i} + U_l V_l h_{K:i}^{(l)}}{UV + U_l V_l} \right)^{U V h_i}. \quad (15)$$

Применяя к оптимальному правилу (14) оценки вида (15), запишем окончательное выражение для PNN с проверкой однородности [12] для дискретных конечных выборок

$$\begin{aligned} UV \sum_{i=1}^N h_i \ln \frac{h_{K:i}}{\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)}} + \\ + U_l V_l \sum_{i=1}^N h_i^{(l)} \ln \frac{h_{K:i}^{(l)}}{\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)}} \rightarrow \min_l, \end{aligned} \quad (16)$$

где

$$\begin{aligned} \tilde{h}_{K:i}^{(l)} &= \frac{U_l V_l}{UV + U_l V_l} h_{K:i}^{(l)}, \\ \tilde{h}_{K:i} &= \frac{UV}{UV + U_l V_l} h_{K:i}. \end{aligned}$$

Если предположить, что размеры всех изображений одинаковы, то последнее выражение для ядра – дискретной дельта-функции – сведётся к минимизации дивергенции Йенсена–Шеннона (Jensen-Shannon), с успехом применяющейся в разнообразных задачах АРИ [22, 25]

$$\sum_{i=1}^N h_i \ln \frac{2h_i}{h_i + h_i^{(l)}} + \sum_{i=1}^N h_i^{(l)} \ln \frac{2h_i^{(l)}}{h_i + h_i^{(l)}} \rightarrow \min_l. \quad (17)$$

Выражения (16), (17) содержат логарифмирование – потенциально «дорогую» (в смысле вычислительной сложности) операцию. Если требования к скорости вычислений являются критическими [6, 9], то зачастую применяют известную [24] аппроксимацию логарифма  $\ln \frac{a}{b} \approx \frac{a^2 - b^2}{2ab}$ . Соответственно, (16) можно приблизить критерием

$$\begin{aligned} UV \sum_{i=1}^N \frac{h_i \left( (h_{K:i})^2 - (\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)})^2 \right)}{\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)}} + \\ + U_l V_l \sum_{i=1}^N \frac{h_i^{(l)} \left( (h_{K:i}^{(l)})^2 - (\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)})^2 \right)}{\tilde{h}_{K:i} + \tilde{h}_{K:i}^{(l)}} \rightarrow \min_l. \end{aligned} \quad (18)$$

А (17), в свою очередь, после ряда упрощений [24] может быть аппроксимировано с помощью расстояния  $\chi^2$  [26]

$$\sum_{i=1}^N \frac{(h_i - h_i^{(l)})^2}{h_i + h_i^{(l)}} \rightarrow \min_l. \quad (19)$$

Таким образом, если сформулировать задачу АРИ не в терминах статистической классификации (5), а свести её к проверке гипотез об однородности выборок признаков, то можно получить не только новые критерии (16), (18), но и при определённых допущениях известные выражения (17), (19), хорошо зарекомендовавшие себя в практических приложениях АРИ [25, 26].

### Результаты экспериментальных исследований

Для проведения экспериментального исследования эффективности предложенного подхода (16), (18) рассмотрим задачу распознавания людей по фотографиям лиц [16, 19], являющуюся, как известно [15], одной из наиболее сложной в области АРИ. В качестве баз дан-

ных использовались стандартные множества фотографий лиц AT&T (бывшая ORL) [17] и JAFFE [18]. Множество AT&T (вариативный параметр – ракурс объекта на изображении) состоит из 400 изображений 40 людей (по 10 фотографий для каждого человека). База данных JAFFE (Japanese Female Facial Expression), в которой вариативным параметром является мимика, содержит 321 фотографию 10 женщин (более 20 фотографий для каждой).

Лица на фотографиях выделялись с помощью библиотеки OpenCV. Далее выполнялась медианная фильтрация с размером окна  $3 \times 3$  пикселя. Для вычисления рассогласований все изображения лиц предварительно разбивались на 16 ( $4 \times 4$ ) и 100 ( $10 \times 10$ ) фрагментов для гистограмм цвета (1) [9, 21, 27] и направлений градиента (2) [3, 22] соответственно. Эти значения показали в наших экспериментах наилучшую точность для рассматриваемых баз данных. Общее рассогласование между двумя фотографиями рассчитывалось как сумма рассогласований между соответствующими фрагментами, причём для учёта неоднородности ракурса и неточности детектирования лица каждая часть распознаваемого изображения сопоставлялась с фрагментами эталонного изображения из окрестности  $3 \times 3$  (более подробно эта процедура и её преимущества описаны в работе [16]).

В ходе эксперимента сопоставлялась зависимость средней точности АРИ от числа  $n_p$  фотографий каждого человека, которые помещаются во множество эталонов. Точность оценивалась следующим образом. Для каждого класса (отдельного человека) наугад выбиралось  $n_p$  фотографий, которые помещались в обучающую выборку. Остальные фотографии формировали тестовое множество. Итоговая точность оценивалась как средняя точность после 20 таких экспериментов.

Исследовались предложенные критерии (16), (18) с их частными случаями (17), (19), а также традиционное расстояние Евклида (4) и оптимальные критерии на основе PNN (10), (11). Для PNN (10) и её модификации (16), (18) использовалось гауссовское ядро Парзена–Розенблатта с параметром сглаживания (среднеквадратичным отклонением)  $\sigma = 0,577$ . Каждый критерий был реализован в нейросетевом (многопоточном) варианте с 8 потоками в среде Microsoft Visual C++ 2010 Express Edition на современном ноутбуке (четырёхядерный процессор Intel Core i7-2630QM, тактовая частота 2 ГГц, объём ОЗУ 6 Гб).

Все результаты – оценки вероятности ошибки АРИ для сопоставления гистограмм яркости (1) и базы данных AT&T сведены в табл. 1. Здесь полужирным шрифтом выделены лучшие результаты (наименьшая вероятность ошибки) по столбцу (для фиксированного  $n_p$ ).

Далее в таблице 2 приведены вероятности ошибки для того же множества фотографий AT&T, но для гистограмм направлений градиента (2).

Наши эксперименты (табл. 1 и 2) показали известный факт: из двух рассматриваемых признаковых

пространств гистограммы направлений градиента существенно превосходят по качеству АРИ гистограммы цвета. Поэтому далее для базы данных JAFFE рассмотрим только более совершенные признаки. Результаты приведены в табл. 3.

Таблица 1. Вероятность ошибки АРИ для гистограмм цвета и AT&T

Критерий \ $n_p$	1	2	3	4
$L_2$ (4)	33,1%	18,5%	10,9%	7,3%
PNN (10)	29,1%	16,2%	9,6%	6,0%
Кульбака–Лейблера (11)	30,9%	17,4%	10,2%	6,6%
Йенсена–Шеннона (17)	26,0%	14,3%	7,6%	<b>4,6%</b>
Хи-квадрат (19)	26,2%	<b>14,0%</b>	<b>7,5%</b>	<b>4,6%</b>
PNN с проверкой однородности (16)	26,0%	14,2%	7,7%	<b>4,6%</b>
Приближённое выражение для PNN с проверкой однородности (18)	<b>25,8%</b>	<b>14,0%</b>	7,7%	4,8%

Таблица 2. Вероятность ошибки АРИ для гистограмм направлений градиента и AT&T

Критерий \ $n_p$	1	2	3	4
$L_2$ (4)	23,5%	11,6%	7,1%	3,4%
PNN (10)	21,9%	10,3%	6,0%	2,8%
Кульбака–Лейблера (11)	24,5%	12,5%	7,5%	3,8%
Йенсена–Шеннона (17)	21,1%	9,7%	5,3%	2,3%
Хи-квадрат (19)	21,0%	9,6%	5,3%	<b>2,2%</b>
PNN с проверкой однородности (16)	20,7%	<b>9,5%</b>	<b>5,2%</b>	<b>2,2%</b>
Приближённое выражение для PNN с проверкой однородности (18)	<b>20,5%</b>	<b>9,5%</b>	5,3%	2,4%

Таблица 3. Вероятность ошибки АРИ для гистограмм направлений градиента и JAFFE

Критерий \ $n_p$	1	2	3	4
$L_2$ (4)	17,4%	8,3%	5,9%	3,6%
PNN (10)	13,6%	5,3%	4,1%	2,0%
Кульбака–Лейблера (11)	16,3%	6,7%	4,6%	2,5%
Йенсена–Шеннона (17)	12,8%	5,3%	4,2%	2,4%
Хи-квадрат (19)	12,9%	5,1%	4,1%	2,3%
PNN с проверкой однородности (16)	11,5%	<b>4,6%</b>	<b>3,5%</b>	<b>1,9%</b>
Приближённое выражение для PNN с проверкой однородности (18)	<b>11,2%</b>	4,7%	3,75%	<b>1,9%</b>

Наконец, в табл. 4 показано среднее время вычисления рассогласования (в микросекундах) между двумя изображениями (для нейросетевой многопоточной реализации).

Таблица 4. Среднее время вычисления меры близости (в мкс)

Критерий	Время (мкс)	
	Гистограмма цвета	Гистограмма направлений градиента
$L_2$ (4)	4	6
PNN (10) / Кульбака–Лейблера (11)	14	40
Йенсена–Шеннона (17)	24	69
Хи-квадрат (19)	6	16
PNN с проверкой однородности (16)	34	101
Приближённое выражение для PNN с проверкой однородности (18)	10	27

Здесь время распознавания для направлений градиентов превосходит аналогичный показатель для гистограмм цвета из-за различий в размере сетки, на которые разбиваются фотографии. Однако пропорция между отдельными критериями остаётся приблизительно одинаковой (за исключением метрики Евклида (4), для которой время классификации является чересчур малым для точного измерения).

По результатам проведённых экспериментов (табл. 1–4) можно сделать следующие выводы. Во-первых, точность классификации критериев (16)–(19), основанных на проверке статистической однородности выборок (12), существенно превосходит как детерминированный подход (4), так и оптимальные решения задачи статистической классификации (5), воплощённые в PNN (10) и принципе минимума информационного рассогласования Кульбака–Лейблера (11). Во-вторых, применение приближённых равенств (18), (19) сопровождается не только резким повышением вычислительной эффективности (что неудивительно, см. таблица 4), но и зачастую понижением вероятности ошибки по сравнению с оригинальными критериями (16) и (17) соответственно. В-третьих, качество АРИ для предложенных критериев (16) и (18) с гауссовским ядром Парзена–Розенблатта оказалась в среднем выше по сравнению с широко используемыми [25, 26] дивергенцией Йенсена–Шеннона и расстоянием хи-квадрат. Таким образом, можно сделать основной вывод: использование подхода, согласно которому задача АРИ сводится к статистической проверке однородности (11), позволяет объяснить недостаточное качество оптимальных решений (10), (11) задачи классификации (5).

### Заключение

В настоящей работе предложена модификация (16) вероятностной нейронной сети с проверкой однородности [12] для случая распознавания выборок дискретных объектов, полностью описываемых своими гистограммами. Эта модификация сделала возможной практическую реализацию в режиме реального времени (см. табл. 4) оптимального решения (13) для задач с большим числом признаков. Действительно, в случае, если множество значений каждо-

го признака является конечным, то появляется возможность сопоставлять не сами признаки (7), а их гистограммы (8), тем самым снижая вычислительную сложность в  $UVU_iV_i / N^2$  раз (ср., например, (7) и (8)). При этом для произвольной ядерной функции объём оперативной памяти, необходимый для реализации критериев (16), (18), лишь в 2 раза превышает аналогичный показатель для традиционных критериев, основанных на сравнении гистограмм, таких как (4), (17) и (19), за счёт сохранения для каждого изображения свёртки его гистограммы с ядром (9). Для большинства приложений это более чем умеренная плата за повышение точности распознавания (см. табл. 2, 3).

Отметим ещё одно потенциальное преимущество [12] использования предложенного подхода (16), (18) модификации (6) в задаче АРИ по сравнению с традиционными реализациями оптимального критерия (10), (11) – синтезированное выражение (16) в отличие от (10) является симметричным (с точностью до множителей  $UV$  и  $U_iV_i$ , характеризующих объём выборки). Действительно, в задаче проверки однородности (12) выборки считаются равнозначными. А при статистической классификации (5) в PNN (10) наблюдается асимметрия между распределением, заданным обучающей выборкой, и распределением входного объекта. Это обстоятельство может служить дополнительным обоснованием для выбора критерия (16), (18), т.к. симметрия является желательным свойством [20] рассогласования между объектами для многих алгоритмов распознавания образов (таких как кластеризация).

Таким образом, в настоящей работе выполнено строгое теоретико-вероятностное обоснование наблюдаемой большинством исследователей АРИ недостаточной точности максимально правдоподобного решения, основанного на принципе минимума информационного рассогласования Кульбака–Лейблера (11), по сравнению с расстоянием хи-квадрат (19) или дивергенцией Йенсена–Шеннона (17). Обычно [5, 11] неоптимальность выражений (10), (11) объясняют тем, что предположение о статистической независимости признаков, лежащее в основе классификатора Байеса, является слишком «наивным» в задаче АРИ. Однако представленная работа наглядно показала, что причина может объясняться и тем, что задачи распознавания образов следует сводить не к статистической классификации (5), а к проверке однородности признаков входного и эталонного изображений (12).

### Благодарности

Исследование осуществлено в рамках Программы «Научный фонд НИУ ВШЭ» в 2013-2014 гг., проект № 12-01-0003.

### Литература

1. **Forsyth, D.A.** Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce. – 2nd ed. – New Jersey: Prentice Hall, 2011. – 792 p.

2. **Lowe, D.** Distinctive image features from scale-invariant keypoints / D. Lowe // *International Journal of Computer Vision*. – 2004. – Vol. 60, N 2. – P. 91-110.
3. **Dalal, N.** Histograms of Oriented Gradients for Human Detection. *Proceedings / N. Dalal, B. Triggs // International Conference on Computer Vision & Pattern Recognition*, 2005. – P. 886-893.
4. **Zuo, W.** Robust Recognition of Noisy and Partially Occluded Faces Using Iteratively Reweighted Fitting of Eigenfaces / W. Zuo, K. Wang, D. Zhang // *Conference on Advances in Multimedia Information Processing, Lecture Notes in Computer Science*. – 2006. – Vol. 4261. – P. 844-851.
5. **Савченко, А.В.** Выбор параметров алгоритма распознавания изображений на основе коллектива решающих правил и принципа максимума апостериорной вероятности / А.В. Савченко // *Компьютерная оптика*. – 2012. – Т. 36, № 1. – С. 117-124.
6. **Savchenko, A.V.** Directed enumeration method in image recognition / A.V. Savchenko // *Pattern Recognition*. – 2012. – Vol. 45, N 8. – P. 2952-2961.
7. **Fukunaga, K.** Introduction to Statistical Pattern Recognition / K. Fukunaga. – 2nd ed. – New York: Academic Press, Inc, 1991. – 591 p.
8. **Webb, A.R.** Statistical Pattern Recognition / A.R. Webb. – 2nd ed. – New York: Wiley, 2002. – 534 p.
9. **Савченко, А.В.** Теоретико-вероятностная модель полутонового изображения для задачи распознавания образов без учителя на основе метода направленного перебора / А.В. Савченко // *Компьютерная оптика*. – 2011. – Т. 35, № 3. – С. 385-394.
10. **Specht, D.F.** Probabilistic neural networks / D.F. Specht // *Neural Networks*. – 1990. – Vol. 3. – P. 109-118.
11. **Savchenko, A.V.** Adaptive Video Image Recognition System Using a Committee Machine / A.V. Savchenko // *Optical Memory and Neural Networks (Information Optics)*. – 2012. – Vol. 21, N 4. – P. 219-226.
12. **Savchenko, A.V.** Statistical Recognition of a Set of Patterns Using Novel Probability Neural Network / A.V. Savchenko // *International Workshop on Artificial Neural Networks and Pattern Recognition, Lecture Notes in Computer Science*. – 2012. – Vol. 7477. – P. 93-103.
13. **Боровков, А.А.** Математическая статистика: дополнительные главы / А.А. Боровков. – М.: Наука, 1984. – 144 с.
14. **Савченко, В.В.** Принцип минимального информационного рассогласования в задаче распознавания дискретных объектов / В.В. Савченко, А.В. Савченко // *Известия вузов России. Радиоэлектроника*. – 2005. – Вып. 3. – С. 10-18.
15. *Face Processing: Advanced Modeling and Methods / edited by W. Zhao, R. Chellappa*. – Elsevier: Academic Press, 2005. – 768 p.
16. **Savchenko, A.V.** Face Recognition in Real-Time Applications: Comparison of Directed Enumeration Method and K-d Trees / A.V. Savchenko // *International Conference on Business Informatics Research, Lecture Notes in Business Information Processing*. – 2012. – Vol. 128. – P. 187-199.
17. AT&T faces dataset, <http://www.cl.cam.ac.uk/research/dtg/atarchive/facedatabase.html> (October 21, 2012).
18. JAFFE dataset, <http://www.kasrl.org/jaffe.html> (October 21, 2012)
19. **Tan, X.** Face recognition from a single image per person: A survey / X. Tan, S. Chen, Z.H. Zhou, F. Zhang // *Pattern Recognition*. – 2006. – Vol. 39, N 9. – P. 1725-1745.
20. **Theodoridis, S.** *Pattern Recognition / S. Theodoridis, C. Koutroubas*. – 4th ed. – Elsevier Inc, 2009. – 840 p.
21. **Yoo, G.-H.** Content-based image retrieval using shifted histogram / G.-H. Yoo, B.K. Kim, K.S. You // *International Conference on Computer Science, Lecture Notes in Computer Science*. – 2007. – Vol. 4489. – P. 894-897.
22. **Лисицын, С.О.** Распознавание дорожных знаков с помощью метода опорных векторов и гистограмм ориентированных градиентов / С.О. Лисицын, О.А. Байда // *Компьютерная оптика*. – 2012. – Т. 36, № 2. – С. 289-295.
23. **Савченко, А.В.** Градиент яркости в задаче распознавания полутоновых изображений на основе статистического подхода / А.В. Савченко // *Вестник компьютерных и информационных технологий*. – 2012. – № 1. – С. 12-16.
24. **Kullback, S.** *Information Theory and Statistics / S. Kullback*. – Dover Pub, 1997. – 399 p.
25. **Martins, A.F.T.** Nonextensive entropic kernels / A.F.T. Martins, M.A.T. Figueiredo, P.M.Q. Aguiar, N.A. Smith, E.P. Xing // *International Conference on Machine Learning*. – 2008. – P. 640-647.
26. **Ahonen, T.** Face recognition with local binary patterns / T. Ahonen, A. Hadid, M. Pietikainen // *European Conference on Computer Vision*. – 2005. – P. 469-481.
27. **Zhang, D.** Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study / D. Zhang, G. Lu // *IEEE International Conference on Multimedia and Expo*. – 2001. – P. 289-293.

### References

1. **Forsyth, D.A.** *Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce*. – 2nd ed. – New Jersey: Prentice Hall, 2011. – 792 p.
2. **Lowe, D.** Distinctive image features from scale-invariant keypoints / D. Lowe // *International Journal of Computer Vision*. – 2004. – Vol. 60, N 2. – P. 91-110.
3. **Dalal, N.** Histograms of Oriented Gradients for Human Detection. *Proceedings / N. Dalal, B. Triggs // International Conference on Computer Vision & Pattern Recognition*, 2005. – P. 886-893.
4. **Zuo, W.** Robust Recognition of Noisy and Partially Occluded Faces Using Iteratively Reweighted Fitting of Eigenfaces / W. Zuo, K. Wang, D. Zhang // *Conference on Advances in Multimedia Information Processing, Lecture Notes in Computer Science*. – 2006. – Vol. 4261. – P. 844-851.
5. **Savchenko, A.V.** The choice of algorithm parameters in image recognition on the basis of ensemble classifiers and the maximum posterior probability principle / A.V. Savchenko // *Computer optics*. – 2012. – Vol. 36(1). – P. 117-124. – (In Russian).
6. **Savchenko, A.V.** Directed enumeration method in image recognition / A.V. Savchenko // *Pattern Recognition*. – 2012. – Vol. 45, N.8. – P. 2952-2961. – ISSN 0031-3203.
7. **Fukunaga, K.** Introduction to Statistical Pattern Recognition / K. Fukunaga. – 2nd ed. – New York: Academic Press, Inc, 1991. – 591 p.

8. **Webb, A.R.** Statistical Pattern Recognition / A.R. Webb. – 2nd ed. – New York: Wiley, 2002. – 534 p.
9. **Savchenko, A.V.** Probability half-tone image model in a problem of unsupervised pattern recognition based on directed enumeration method / A.V. Savchenko // Computer optics. – 2011. – V. 35(3). – P. 385-394. – (In Russian).
10. **Specht, D.F.** Probabilistic neural networks / D.F. Specht // Neural Networks. – 1990. – Vol. 3. – P. 109-118.
11. **Savchenko, A.V.** Adaptive Video Image Recognition System Using a Committee Machine / A.V. Savchenko // Optical Memory and Neural Networks (Information Optics). – 2012. – Vol. 21, N 4. – P. 219-226.
12. **Savchenko, A.V.** Statistical Recognition of a Set of Patterns Using Novel Probability Neural Network / A.V. Savchenko // International Workshop on Artificial Neural Networks and Pattern Recognition, Lecture Notes in Computer Science. – 2012. – Vol. 7477. – P. 93-103.
13. **Borovkov, A.A.** Mathematical statistics: additional chapters / A.A. Borovkov. – Moscow: "Nauka" Publisher, 1984. – 144 p. – (In Russian).
14. **Savchenko, V.V.** Minimum information discrimination principle in the problem of discrete objects / V.V. Savchenko, A.V. Savchenko // Izvestia vuzov Rossii. Radioelektronika. – 2005. – Vol. 3. – P. 10-18. – (In Russian).
15. Face Processing: Advanced Modeling and Methods / edited by W. Zhao, R. Chellappa. – Elsevier: Academic Press, 2005. – 768 p.
16. **Savchenko, A.V.** Face Recognition in Real-Time Applications: Comparison of Directed Enumeration Method and K-d Trees / A.V. Savchenko // International Conference on Business Informatics Research, Lecture Notes in Business Information Processing. – 2012. – Vol. 128. – P. 187-199.
17. AT&T faces dataset, <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (October 21, 2012).
18. JAFFE dataset, <http://www.kasrl.org/jaffe.html> (October 21, 2012).
19. **Tan, X.** Face recognition from a single image per person: A survey / X. Tan, S. Chen, Z.H. Zhou, F. Zhang // Pattern Recognition. – 2006. – Vol. 39, N 9. – P. 1725-1745.
20. **Theodoridis, S.** Pattern Recognition / S. Theodoridis, C. Koutroumbas. – 4th ed. – Elsevier Inc, 2009. – 840 p.
21. **Yoo, G.-H.** Content-based image retrieval using shifted histogram / G.-H. Yoo, B.K. Kim, K.S. You // International Conference on Computer Science, Lecture Notes in Computer Science. – 2007. – Vol. 4489. – P. 894-897.
22. **Lisitsyn, S.O.** Road sign recognition using support vector machines and histogram of oriented gradients / S.O. Lisitsyn, O.A. Bayda // Computer optics. – 2012. – V. 36(2). – P. 289-295. – (In Russian).
23. **Savchenko, A.V.** Gradient Orientation in a Problem of Automatic Halftone Image Recognition Based on Statistical Approach / A.V. Savchenko // Vestnik of computer and information technologies. – 2012. – Vol. 1 – P. 12-16. – (In Russian).
24. **Kullback, S.** Information Theory and Statistics / S. Kullback. – Dover Pub, 1997. – 399 p.
25. **Martins, A.F.T.** Nonextensive entropic kernels / A.F.T. Martins, M.A.T. Figueiredo, P.M.Q. Aguiar, N.A. Smith, E.P. Xing // International Conference on Machine Learning. – 2008. – P. 640-647.
26. **Ahonen, T.** Face recognition with local binary patterns / T. Ahonen, A. Hadid, M. Pietikainen // European Conference on Computer Vision. – 2005. – P. 469-481.
27. **Zhang, D.** Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study / D. Zhang, G. Lu // IEEE International Conference on Multimedia and Expo. – 2001. – P. 289-293.

## IMAGE RECOGNITION ON THE BASIS OF PROBABILISTIC NEURAL NETWORK WITH HOMOGENEITY TESTING

*A.V. Savchenko*

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

### *Abstract*

The usage of the probabilistic neural network with homogeneity testing is proposed in image recognition problem. This decision is shown to be optimal in Bayesian terms if the task is formulated as a statistical testing for homogeneity of query and model images' feature sets. The problem of the lack of computing efficiency with many classes and large dimensions of feature set is discovered. The possibility of its overcoming in the case of discrete features is explored by synthesizing the novel recognition criterion with the comparison of the histograms of query and model images. It is shown that a particular case of this criterion is the nearest neighbor rule with popular measures of similarity, namely, chi-square distance and Jensen-Shannon divergence. The results of experimental research in a problem of face recognition with widely used databases (AT&T, JAFFE) are presented. The proposed approach is demonstrated to achieve better recognition accuracy in comparison with conventional solution with reduction the recognition task to the statistical classification.

Key words: automatic image recognition, face recognition, probabilistic neural network, test for samples nearest neighbour rule.



*Сведения об авторе*

**Савченко Андрей Владимирович**, 1985 года рождения. В 2008 году с отличием окончил Нижегородский государственный технический университет им Р.А. Алексева (НГТУ) по специальности «Прикладная математика и информатика». Кандидат технических наук (2010 год), работает доцентом 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. Автор более 40 научных работ. 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) degree from State University Higher School of Economics - Moscow. 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. He is the author of more than 40 scientific papers. Homepage: <http://www.hse.ru/en/org/persons/9216523>.

*Поступила в редакцию 12 декабря 2012 г.*