

ВЫБОР ПАРАМЕТРОВ АЛГОРИТМА РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ КОЛЛЕКТИВА РЕШАЮЩИХ ПРАВИЛ И ПРИНЦИПА МАКСИМУМА АПОСТЕРИОРНОЙ ВЕРОЯТНОСТИ

Савченко А.В.

Национальный исследовательский университет Высшая школа экономики – Нижний Новгород

Аннотация

Ставится и решается задача выбора параметров алгоритма автоматического распознавания изображений путём построения коллектива решающих правил на основе принципа максимума апостериорной вероятности. Выполнен строгий синтез критерия выбора параметров для информационного рассогласования Кульбака–Лейблера и современного алгоритма распознавания SIFT (Scale-Invariant Feature Transform). Представлены программа и результаты экспериментального исследования в задаче идентификации личности по фотографии лица для известных баз данных (Yale, AT&T). Показано, что применение предложенного критерия позволяет добиться точности распознавания, сравнимой с точностью наилучшего набора параметров, причём не только для рассогласования Кульбака–Лейблера, но и для других популярных расстояний (метрика Евклида, расхождение Кульбака).

Ключевые слова: автоматическое распознавание изображений, коллективы решающих правил, информационное рассогласование Кульбака–Лейблера, принцип максимума апостериорной вероятности.

Введение

Основной проблемой в области автоматического распознавания изображений (АРИ) [1, 2] является недостаточная устойчивость существующих моделей изображений к искажениям (noise) [3] и, как следствие, недостаточная точность традиционных методов классификации [2] в условиях априорной неопределённости (неизвестная заранее освещённость, размер, ракурс, положение распознаваемого объекта на изображении и пр.). Известно [2], что большинство современных математических моделей изображений зависит от ряда параметров: диапазон изменения используемого признака (например, градация яркости), размер сетки, на которую разбивается входной объект с целью учёта вариативности освещения [4], и многие другие. При этом качество АРИ для фиксированной модели может значительно варьироваться в зависимости от различных значений её параметров. Зачастую исследователи ещё на этапе анализа предметной области фиксируют все параметры, полагаясь исключительно на свой опыт и интуицию.

Очевидно, такое решение не всегда можно считать достаточно разумным. Поэтому в данной работе рассматривается логика рассуждений совсем другого рода. Исследователю предлагается зафиксировать лишь некоторое конечное множество потенциально оптимальных параметров. При этом выбор конкретного набора происходит автоматически на основе синтеза коллективов решающих правил (КРП, комитетов классификаторов, алгоритмических композиций) [5, 6]. В работе проводится сравнение КРП, полученных с помощью традиционного усреднения индивидуальных решающих правил [7, 8], с КРП, построенными на основе принципа максимума апостериорной вероятности [9, 10]. Полученные результаты и сделанные по ним выводы рассчитаны на широкий круг специалистов в области распознавания изображений.

Алгоритм SIFT в задаче автоматического распознавания изображений

Пусть задано множество из $L > 1$ полутоновых 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 изображений 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 [11].

Для решения задачи АРИ воспользуемся популярным сейчас методом SIFT (Scale-Invariant Feature Transform) [4]. Вначале выполним предварительную нормировку [12] освещения изображений из базы (например, используя гамма-коррекцию с последующей медианной фильтрацией [13]). В качестве основного признака [14, 15] выбирается направление градиента яркости пикселя, вычисленное по формуле Робертса (применяющегося, например, в наиболее популярном детекторе краёв Кэнни [13])

$$\theta_{u,v}^{(l)} = \frac{\pi}{4} + \arctg \frac{x_{u+1,v+1}^{(l)} - x_{u,v}^{(l)}}{x_{u+1,v}^{(l)} - x_{u,v+1}^{(l)}}. \quad (1)$$

Направление градиента $\theta_{u,v}^{(l)}$ полностью инвариантно к интенсивности освещения (определяющего только модуль градиента).

Далее разобьём каждое изображение квадратной сеткой из $S \times T$ ячеек (по S строк и T столбцов, в оригинале [4] $S = T = 4$). Для каждой ячейки вычис-

лим взвешенную гистограмму направления градиента яркости. В качестве веса используется оценка модуля градиента

$$m_{u,v}^{(l)} = \sqrt{\left(x_{u+1,v+1}^{(l)} - x_{u,v}^{(l)}\right)^2 + \left(x_{u+1,v}^{(l)} - x_{u,v+1}^{(l)}\right)^2}.$$

Для вычисления гистограммы $H_i(s,t)$ направления градиента в ячейке $(s,t), s = \overline{1, S}, t = \overline{1, T}$ разобьём всю область определения $[-3\pi/4; 5\pi/4]$ на N отрезков одинакового размера $2\pi/N$ (в оригинале [4] $N=8$). Тогда элемент гистограммы может быть оценён как

$$h_i^{(l)}(s,t) = \frac{\tilde{h}_i^{(l)}(s,t)}{\sum_{i=1}^N \tilde{h}_i^{(l)}(s,t)}, i = \overline{1, N}. \quad (2)$$

Здесь

$$\tilde{h}_i^{(l)}(s,t) = \sum_{u=U \cdot (s-1)/S+1}^{U \cdot s/S} \sum_{v=V \cdot (t-1)/T+1}^{V \cdot t/T} \frac{m_{uv}^{(l)}}{m^{(l)}(r,c)} \times \left(H\left(\theta_{uv}^{(l)} - \frac{2(i-1) - N}{N} \pi\right) - H\left(\theta_{uv}^{(l)} - \frac{2i - N}{N} \pi\right) \right),$$

а $H(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0 \end{cases}$ – функция Хэвисайда и

$$m^{(l)}(s,t) = \max \left\{ m_{u,v}^{(l)} \mid u \in \left\{ \frac{U \cdot (s-1)}{S} + 1, \dots, \frac{U \cdot s}{S} \right\}, v \in \left\{ \frac{V \cdot (t-1)}{T} + 1, \dots, \frac{V \cdot t}{T} \right\} \right\}.$$

Решение в алгоритме SIFT принимается по критерию минимума некоторой меры близости между гистограммами направления градиента входного изображения и эталонов из БД. Наиболее распространено применение в качестве рассогласования l_2 -метрики (Евклида)

$$\sqrt{\sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^N \left(h_i^{(l)}(s,t) - \bar{h}_i(s,t) \right)^2} \rightarrow \min_l. \quad (3)$$

Заметим, что существуют опциональные этапы алгоритма SIFT, такие как предварительная обработка изображений с помощью DoG (Difference Of Gaussians) фильтра, а также выравнивание направлений градиента относительно среднего направления градиента точек по всему изображению для достижения инвариантности к повороту. Однако мы не будем использовать эти этапы, так как во всех наших экспериментах с распознаванием лиц их применение приводило к значительному (15% и более) ухудшению точности классификации.

Задача выбора параметров алгоритма распознавания

Основная проблема метода SIFT (как, впрочем, и большинства остальных современных алгоритмов АРИ) связана с зависимостью точности классифика-

ции от ряда параметров модели. В частности, для SIFT важными являются следующие параметры – размер сетки для разбиения изображения (S, T), мера близости гистограмм направления градиента яркости, число интервалов в диапазоне изменения направления градиента яркости пикселя (N), размер окна медианного фильтра, необходимость в выполнении опциональных этапов алгоритма. Очевидно, эти параметры могут принимать множество значений, при этом точность распознавания определяется не только свойствами БД $\{X_l\}$, но и характеристиками входного объекта X [16]. Таким образом, даже для одного и того же множества эталонов качество классификации для фиксированного набора параметров может существенно варьироваться в зависимости от распознаваемого изображения. Похоже, невозможно заранее подобрать универсальные значения этих параметров так, чтобы обеспечить оптимальную классификацию для различных задач АРИ.

В условиях полной априорной неопределённости одним из наиболее естественных способов преодоления проблемы выбора наилучших параметров может служить построение КРП [5, 6], в котором каждое решающее правило представляет собой реализацию алгоритма SIFT для конкретного набора параметров. Обзор и анализ публикаций в области обработки данных показывает, что синтез КРП является одним из наиболее эффективных подходов к увеличению точности и устойчивости классификации [17]. В КРП для принятия решения о классификации изображения используется не один, а несколько критериев, каждый из которых самостоятельно присваивает метку класса, после чего на основе некоторого принципа [5] формируется общий результат классификации.

Таким образом, задача состоит в следующем. Вначале исследователем выбирается несколько альтернативных наборов значений параметров, между которыми и требуется осуществить выбор. Можно считать, что эти наборы задают семейства рассогласований

$$P = \{ \rho_k(X_1, X_2) \}, k = \overline{1, K}, \quad (4)$$

где K – количество вариантов (альтернативных наборов параметров алгоритма АРИ). Каждое рассогласование задаёт критерий АРИ, согласно которому решение принимается в пользу класса, соответствующего эталону $X_{l^{(k)}}$, где

$$l^*(k) = \arg \min_{l \in \{1, \dots, L\}} \rho_k(X, X_l). \quad (5)$$

В результате процедуры АРИ объекта X для каждого рассогласования (3) будет получен вектор

$$\bar{\rho}_k = [\rho_k(X, X_1), \rho_k(X, X_2), \dots, \rho_k(X, X_L)].$$

Тогда требуется по набору векторов $\bar{\rho}_k, k = \overline{1, K}$ («мнений экспертов») определить тот $\bar{\rho}_\mu$ из них, который наиболее подходит для АРИ объекта X .

Наиболее простые способы построения КРП [1, 7], такие как голосование членов комитета и усреднение выходов классификаторов ($\rho_k(X, X_i)$) по ансамблю $k = \overline{1, K}$, не подходят для поставленной задачи выбора параметров алгоритма АРИ. Во-первых, они предполагают, что члены комитета в среднем одинаково эффективны (в смысле точности классификации). Однако в реальности для каждого распознаваемого изображения вполне возможна ситуация, когда один критерий подходит лучше всех остальных. И, во-вторых, такое построение комитетов предполагает, что количество членов комитета будет достаточно большим. А на практике уже использование более трёх классификаторов приводит к невозможности реализации системы АРИ в режиме реального времени (например, при распознавании объектов на видеоизображении).

Очевидное решение состоит здесь в использовании более сложных алгоритмов построения КРП, основанных на алгебраическом подходе [5, 18]. Большая часть таких алгоритмов (такие как комитет взвешенного большинства [18], бэггинг и бустинг [17]) требуют достаточной представительной обучающей выборки. К сожалению, во многих задачах АРИ имеющаяся БД содержит недостаточное число эталонов для каждого класса (в худшем случае, один эталон на класс). В настоящей работе предлагается воспользоваться известными статистическими способами синтеза КРП [1], когда тем или иным способом для каждого члена комитета определяются вероятности принадлежности входного объекта к классам из БД. Как следствие, встаёт задача присвоения таких вероятностей на основе только данных $\overline{\rho_k}, k = \overline{1, K}$ о расстоянии от входного изображения до всех эталонов из БД для произвольных мер близости. Далее мы проведём строгий вывод выражения для апостериорной вероятности принадлежности объекта к классу из БД для информационного рассогласования Кульбака–Лейблера [19].

Принцип минимума информационного рассогласования в задаче распознавания изображений

Для применения статистического подхода [1, 15] предполагаем, что $\theta_{u,v}^{(l)}$ является реализацией случайной величины – направления градиента Θ_l изображения-эталона X_l .

Задача сводится в таком случае к проверке L гипотез о распределении [20] $H_l, l = \overline{1, L}$, сигнала изображения на входе H :

$$W_l : H = H_l. \tag{6}$$

Оптимальное решение тогда даёт классический байесовский подход [1, 10] – 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

$$P\{W_l|X\} = \frac{P\{X|W_l\} \cdot P\{W_l\}}{\sum_{i=1}^L P\{X|W_i\} \cdot P\{W_i\}}. \tag{7}$$

Здесь $P\{W_l\}$ – априорная вероятность появления l -го класса, $P\{X|W_l\}$ – правдоподобие, т.е. условная вероятность принадлежности объекта X классу l . В большинстве задач распознавания изображений предполагается, что появление каждого класса равновероятно (полная априорная неопределённость). В этом случае решение (3) принимает наиболее простой вид

$$P\{X|W_l\} \rightarrow \max_l. \tag{8}$$

Учитывая наше предположение о том, что гистограмма H_{Θ_l} является оценкой распределения направления градиента точки изображения Θ для класса l , условная вероятность

$$P\left\{\frac{8(i-1)-3N}{4N} \pi \leq \Theta(s;t) \leq \frac{8i-3N}{4N} \pi \middle| W_l\right\} = h_i^{(l)}(s,t).$$

Далее, делая «наивное» предположение о независимости [21] направлений градиентов яркости в соседних точках изображения, нетрудно показать [20, 21], что условная вероятность $P\{X|W_l\}$ может быть записана как

$$P\{X|W_l\} = \exp\left\{\frac{UV}{ST} \sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^N (h_i(s,t) \ln h_i(s,t))\right\} \times \exp\left\{-\frac{UV}{ST} \rho_{KL}(X|X_l)\right\}. \tag{9}$$

Здесь

$$\rho_{KL}(X|X_l) = \sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^N \left(h_i(s,t) \ln \frac{h_i(s,t)}{h_i^{(l)}(s,t)} \right) \tag{10}$$

– информационное рассогласование Кульбака–Лейблера [19].

В результате получаем, что оптимальное в байесовском смысле решение задачи (то есть эквивалентное (6)) проверки гипотез о распределении дискретной случайной величины (4) даёт принцип минимума информационного рассогласования Кульбака–Лейблера. Поэтому критерий максимума апостериорной вероятности (7) эквивалентен правилу ближайшего соседа [1]

$$l^* = \arg \min_{l=\overline{1,L}} \rho_{KL}(X|X_l). \tag{11}$$

Таким образом, оптимальное в байесовском смысле решение задачи классификации изображений принимается по критерию минимума решающей статистики (10).

Принцип максимума апостериорной вероятности

В настоящей работе для решения задачи выбора параметров алгоритма АРИ (4), (5) предлагается определять наилучший критерий $\bar{\rho}_\mu$ на основе классического принципа максимума апостериорной вероятности [1]

$$\mu = \arg \max_{k \in \{1, \dots, K\}} P\{W_l | X\} \Big|_k \tag{12}$$

Исходя из формулы условной вероятности (9), выражение для апостериорной вероятности $P\{W_l | X\}$ будет иметь предельно простой вид

$$P\{W_l | X\} = \frac{\exp\left\{-\frac{UV}{ST} \cdot \rho_{KL}(X | X_l)\right\}}{\sum_{i=1}^L \exp\left\{-\frac{UV}{ST} \cdot \rho_{KL}(X | X_i)\right\}} \tag{13}$$

В результате имеем следующее решение задачи выбора параметров (4), (5), основанное на принципе максимума апостериорной вероятности

$$\mu = \arg \max_{k \in \{1, \dots, K\}} \frac{\exp\left\{-\frac{UV}{S_k T_k} \cdot \rho_{KL}(X | X_{l^*(k)})\right\}}{\sum_{l=1}^L \exp\left\{-\frac{UV}{S_k T_k} \cdot \rho_{KL}(X | X_l)\right\}} \Big|_k \tag{14}$$

Здесь S_k, T_k – размер сетки для k -го набора параметров.

Заметим, что в выражении (14) используются только компоненты векторов $\bar{\rho}_k, k = 1, \dots, K$. Поэтому окончательно в настоящей работе предлагается следующее решение задачи выбора (4), (5)

$$\mu = \arg \max_{k \in \{1, \dots, K\}} \frac{\exp\left\{-\frac{UV}{S_k T_k} \cdot \rho_k(X, X_{l^*(k)})\right\}}{\sum_{l=1}^L \exp\left\{-\frac{UV}{S_k T_k} \cdot \rho_k(X, X_l)\right\}} \Big|_k \tag{15}$$

Разумеется, это решение будет оптимально (эквивалентно принципу максимума апостериорной вероятности (12)) только в том случае, если семейство (4) состоит только из информационных рассогласований Кульбака–Лейблера. Тем не менее, далее в экспериментальном исследовании мы покажем, что критерий (15) позволяет определить оптимальные наборы параметров для более широкого круга мер близости.

Программа экспериментальных исследований

Для проведения экспериментального исследования эффективности предложенного критерия (15) выбора оптимальных параметров алгоритма АРИ рассмотрим задачу классификации людей по фотографиям лиц [22], являющейся, как известно [23], одной из наиболее сложной в области распознавания образов. В качестве их предварительной обра-

ботки для выделения лиц использовалась библиотека OpenCV. В качестве БД использовались стандартные множества фотографий лиц Yales [24] и AT&T [25].

Как известно [4], качество АРИ для метода SIFT определяется прежде всего размером сетки (S и T) и применяемой мерой близости для сопоставления гистограмм направления градиента (2). Поэтому в настоящем эксперименте производится выбор значений именно этих параметров.

Сетка размером ($S = T = 4$) из оригинального алгоритма SIFT [4] оказалась недостаточной для сложной задачи распознавания лиц и показала низкую точность классификации для всех наших экспериментов. Поэтому далее, кроме оригинальных значений параметров, рассматриваются сетки большего размера: $S = T = 10, S = T = 15$ и $S = T = 20$.

В качестве мер близости в методе ближайшего соседа (11) воспользуемся традиционной метрикой Евклида (3) и теоретически оптимальным рассогласованием Кульбака–Лейблера (10). Кроме них, применяется симметричное информационное расхождение Кульбака [19], которое также рекомендуется использовать для решения задач статистической классификации

$$J(X : X_l) = \sum_{s=1}^S \sum_{t=1}^T \sum_{i=1}^N (h_i(s, t) - h_i^{(l)}(s, t)) \ln \frac{h_i(s, t)}{h_i^{(l)}(s, t)} \tag{16}$$

В эксперименте проводилось сравнение точности АРИ для индивидуальных решающих правил (с фиксированным набором параметров) с точностью КРП, построенных на основе предлагаемого критерия (15), а также с КРП, полученных по алгоритму агрегирования [26]

$$\mu = \arg \min_{n \in \{1, \dots, K\}} \sum_{k=1}^K I(n; k), \tag{17}$$

где $I(n; k) = \sum_{l=1}^L \left(P\{W_l | X\} \Big|_n \cdot \ln \frac{P\{W_l | X\} \Big|_n}{P\{W_l | X\} \Big|_k} \right)$,

а $P\{W_l | X\} \Big|_n$ и $P\{W_l | X\} \Big|_k$ – оценки апостериорной вероятности (13) для наборов параметров с индексами n и k соответственно.

Результаты экспериментальных исследований

Для БД Yale (вариативный параметр – неравномерная освещённость объекта) в БД эталонов помещались 15 изображений (по одной фотографии каждого человека), а тестирование качества распознавания проводилось на остальных снимках этих же людей (181 фотография). Тем самым достигались наиболее жёсткие условия для последующего распознавания (one sample per person [27]).

Для БД AT&T, в которой вариативным параметром является ракурс объекта на изображении, в качестве множества эталонов использовались $L = 110$

изображений 40 различных людей, а тестирование качества АРИ проводилось на других снимках (291 фотография).

Все результаты – оценки вероятности ошибки АРИ – для трёх БД и индивидуальных решающих правил сведены в табл. 1.

Таблица 1. Вероятность ошибки АРИ для мер близости (3), (10) и (16) и БД Yale и AT&T

		Yale	AT&T
Метрика Евклида (3)	$S=T=4$	13,3%	7,5%
	$S=T=10$	7,3%	2,5%
	$S=T=15$	7,3%	4,5%
	$S=T=20$	3%	5,75%
Информационное рассогласование (10)	$S=T=4$	11,5%	5,75%
	$S=T=10$	6,6%	3%
	$S=T=15$	6,1%	4%
	$S=T=20$	2,4%	6%
Информационное расхождение (16)	$S=T=4$	11,5%	5,5%
	$S=T=10$	6,7%	2,7%
	$S=T=15$	6,7%	4%
	$S=T=20$	1,8%	6%

Далее в табл. 2 приведены вероятности ошибки для КРП, построенных при нескольких возможных комбинациях параметров из табл. 1 для БД Yale. Проводится сравнение синтезированного критерия (15) с традиционным агрегированием (17). Кроме того, в столбце «Лучшее РП» (решающее правило) показана наименьшая вероятность ошибки для индивидуальных критериев из комитета.

Проиллюстрируем действие предлагаемого подхода к построению КРП на примере изображений из БД Yale (рис. 1).



Рис. 1. Распознаваемый объект (а), ближайший в смысле рассогласования Кульбака–Лейблера (10) эталон ($S = T = 20$) (б); ближайший в смысле рассогласования Кульбака–Лейблера (10) эталон ($S = T = 10$) (в)

Распознаваемое изображение (рис. 1 а) подавалось на вход критерия (10) для размеров сетки 4, 10, 15 и 20 (вторая строка табл. 2). Несмотря на то, что в целом для этой базы эталонов точность критерия с большой сеткой ($S = T = 20$) в среднем существенно лучше остальных, в этом случае для этого набора параметров было получено неверное решение АРИ (рис. 1 б). И только для сетки ($S = T = 10$) решение в пользу эталона (рис. 1 в) принято верно. Действительно, для $S = T = 10$ в данном случае апостериорная вероятность (13) максимальна. В то же время для решающего правила с $S = T = 20$ рассогласование между входным объектом (рис. 1 а) и эталонами (рис. 1 б и рис. 1 в) практически совпадает: 0,266 и 0,269 соответственно.

Таблица 2. Вероятность ошибки АРИ для КРП (15), (17) и БД Yale

		Лучшее РП	(17)	(15)
Мера близости	Размер сетки	3%	4,8%	2,4%
(3)	$S=T=4$			
(3)	$S=T=10$			
(3)	$S=T=15$			
(3)	$S=T=20$	2,4%	6,1%	2,4%
Мера близости	Размер сетки			
(10)	$S=T=4$			
(10)	$S=T=10$			
(10)	$S=T=15$	1,8%	10,3%	1,8%
(10)	$S=T=20$			
Мера близости	Размер сетки			
(16)	$S=T=4$			
(16)	$S=T=10$	3%	5,5%	3,6%
(16)	$S=T=15$			
(16)	$S=T=20$			
Мера близости	Размер сетки			
(10)	$S=T=10$	2,4%	3%	2,4%
(10)	$S=T=20$			
Мера близости	Размер сетки	1,8%	3,6%	2,4%
(16)	$S=T=10$			
(16)	$S=T=20$			
Мера близости	Размер сетки	6,6%	7,3%	6,6%
(3)	$S=T=10$			
(10)	$S=T=10$			
(16)	$S=T=10$	1,8%	3%	1,8%
Мера близости	Размер сетки			
(3)	$S=T=20$			
(10)	$S=T=20$	1,8%	3%	1,8%
(16)	$S=T=20$			

Поэтому и апостериорная вероятность принадлежности объекта (рис. 1 а) к эталону (рис. 1 б) невелика. В то же время для сетки $S = T = 20$ рассогласования между распознаваемым объектом и эталонами (рис. 1 б, в) заметно различаются: 0,254 и 0,212 соответственно.

В таблице 3 приведены результаты использования КРП для БД AT&T.

По результатам проведённых экспериментов (табл. 1 - 3) можно сделать следующие выводы. Во-первых, точность классификации действительно во многом определяется выбранными параметрами алгоритма АРИ.

Таблица 3. Вероятность ошибки АРИ для КРП (15), (17) и БД АТ&Т

		Лучшее РП	(17)	(15)
Мера близости	Размер сетки	2,5%	4,2%	2,7%
(3)	$S=T=4$			
(3)	$S=T=10$			
(3)	$S=T=20$			
Мера близости	Размер сетки	3%	4,5%	2,5%
(10)	$S=T=4$			
(10)	$S=T=10$			
(10)	$S=T=20$			
Мера близости	Размер сетки	2,7%	6%	2,5%
(16)	$S=T=4$			
(16)	$S=T=10$			
(16)	$S=T=20$			
Мера близости	Размер сетки	2,5%	2,7%	2,7%
(3)	$S=T=10$			
(10)	$S=T=10$			
Мера близости	Размер сетки	5,75%	5,5%	5,7%
(3)	$S=T=20$			
(10)	$S=T=20$			
(16)	$S=T=20$			

При этом набор параметров метода SIFT, предложенный в оригинальной работе [4] (в частности, сопоставление гистограмм в метрике Евклида и небольшая сетка $S=T=4$), оказался недостаточно эффективным в столь сложной задаче распознавания образов, как распознавание лиц. Во-вторых, традиционный способ построения КРП за счёт усреднения результатов индивидуальных решающих правил [1, 7] недостаточно эффективен в задаче выбора оптимальных параметров (по сравнению с наилучшими индивидуальными критериями). И, в-третьих, синтезированный на основе принципа максимума апостериорной вероятности критерий (15) показал высокую эффективность, причём не только для теоретически оптимального рассогласования Кульбака–Лейблера (10), но и для симметричного информационного расхождения (16) и традиционной метрики Евклида (3).

Заключение

Известно [1, 5], что один из наиболее эффективных подходов к увеличению точности и устойчивости классификации основан на синтезе КРП. В них для принятия решения о классификации изображений используется не один, а несколько критериев, каждый из которых самостоятельно присваивает

метку класса, после чего формируется общий результат классификации, например, с помощью простого голосования членов комитета. На большинстве выборок неоднородные КРП, сформированные в известных публикациях, улучшали точность классификации на 3÷10% [26].

К сожалению, как было отмечено выше, большинство традиционных способов построения КРП [5, 17] не могут эффективно применяться в задаче выбора оптимальных параметров алгоритма АРИ. Действительно, они либо предъявляют достаточно жёсткие требования к объёму обучающей выборки, либо накладывают ограничение на равноценность всех параметров-кандидатов. Наиболее подходящим здесь представляется использование статистического подхода [1] и, в частности, классического принципа максимума апостериорной вероятности.

В настоящей работе показано, что критерий (15) обеспечивает максимум апостериорной вероятности принадлежности входного объекта к классу из БД, если во всех индивидуальных правилах (5) используется информационное рассогласование Кульбака–Лейблера. Однако наиболее важным следует признать то, что критерий (15) позволяет получить высокую точность АРИ и для других, более популярных и точных, расстояний – традиционная метрика Евклида (3) и симметричное информационное расхождение (16). Действительно, несмотря на то, что рассогласование Кульбака–Лейблера (10) является статистически оптимальным (в смысле эквивалентности наивному байесовскому правилу), на практике точность АРИ для (10) существенно ниже аналогичного показателя для (16). Это может быть объяснено тем, что само предположение о статистической независимости признаков, лежащее в основе классификатора Байеса, является слишком «наивным» в задаче АРИ. И тем ценнее тот факт, что синтезированный критерий (15) показывает высокие результаты и в реальных практических задачах распознавания лиц.

Литература

1. **Theodoridis, S.** Pattern Recognition (4th Edition) / S. Theodoridis, C. Koutroumbas. – Elsevier Inc., 2009. – 840 p.
2. **Forsyth, D.A.** Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce. – New Jersey: Prentice Hall, 2003. – 693 p.
3. **Zuo, W.** Robust Recognition of Noisy and Partially Occluded Faces Using Iteratively Reweighted Fitting of Eigenfaces / Wangmeng Zuo, Kuanquan Wang and David Zhang // Conference on Advances in Multimedia Information Processing, Lecture Notes in Computer Science. – 2006. – Vol. 4261. – P. 844–851.
4. **Lowe, D.** Distinctive image features from scale-invariant keypoints / D. Lowe // International Journal of Computer Vision. – 2004. – Vol. 60, N 2. – P. 91–110. – ISSN 0920-5691.
5. **Журавлёв, Ю.И.** Об алгебраическом подходе к решению задач распознавания или классификации / Ю.И. Журавлёв // Проблемы кибернетики. – 1978. – Т. 33.– С. 5–68.

6. **Esmaili, M.** Creating of Multiple Classifier Systems by Fuzzy Decision Making in Human-Computer Interface Systems / M. Esmaili, M. Rahmati // Conference IEEE Fuzzy Systems, 2007. – P. 1-7.
7. **Мазуров, В.Д.** Метод комитетов в задачах оптимизации и классификации / В.Д. Мазуров. – М.: Наука, 1990. – 248 с.
8. **Pardo, M.** Learning from data: a tutorial with emphasis on modern pattern recognition methods / M. Pardo, G. Sberveglieri // Sensors Journal, IEEE. – 2002. – Vol. 2(3). – P. 203-217. – ISSN 1530-437X.
9. **Fukunaga, K.** Introduction to Statistical Pattern Recognition; 2nd ed. / K. Fukunaga. – New York: Academic Press, Inc., 1991. – 591 p.
10. **Chow, C.K.** On optimum error and reject trade-off / C.K. Chow // IEEE Transactions on Information Theory. – 1970. – Vol. 16. – P. 41-46. – ISSN 0018-9448.
11. **Цыпкин, Я.З.** Адаптация и обучение в автоматических системах / Я.З. Цыпкин. – М.: Наука, 1968. – 400 с.
12. **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. – ISSN 0134-2452.
13. **Shapiro, L.** Computer vision / L. Shapiro, G. Stockman. – Prentice Hall, 2001. – 752 p.
14. **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.
15. **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. – ISSN 1810-7206.
16. **Горелик, А.Л.** Современное состояние проблемы распознавания: некоторые аспекты / А.Л. Горелик, И.Б. Гуревич, В.А. Скрипкин. – М.: Радио и связь. – 1985. – 160 с.
17. **Tresp, V.** Committee machines / V. Tresp // Handbook for Neural Network Signal Processing. – 2001. – P. 135-151.
18. **Рудаков, К.В.** О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания / К.В. Рудаков, К.В. Воронцов // Доклады Академии наук. – 1999. – Т. 367, № 3. – С. 314-317. – ISSN 0869-5652.
19. **Kullback, S.** Information Theory and Statistics / S. Kullback. – Dover Pub, 1978. – 408 p.
20. **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. – ISSN 1993-8985.
21. **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. – ISSN 0134-2452.
22. **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. – ISSN 0134-2452.
23. **Face Processing: Advanced Modeling and Methods** / edited by W. Zhao, R. Chellappa. – Elsevier: Academic Press, 2005. – 768 p.
24. **The Yale Face database** [Electronical Resource] – <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
25. **The AT&T database** [Electronical Resource] – <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
26. **Савченко, А.В.** Смешение критериев автоматического распознавания изображений на основе принципа минимума информационного рассогласования / А.В. Савченко // Системы управления и информационные технологии. – 2011. – № 2(44). – С. 22-25. – ISSN 1729-5068.
27. **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. – ISSN 0031-3203

References

1. **Theodoridis, S.** Pattern Recognition (4th Edition) / S. Theodoridis, C. Koutroumbas. – Elsevier Inc., 2009. – 840 p.
2. **Forsyth, D.A.** Computer Vision: A Modern Approach / D.A. Forsyth, J. Ponce. – New Jersey: Prentice Hall, 2003. – 693 p.
3. **Zuo, W.** Robust Recognition of Noisy and Partially Occluded Faces Using Iteratively Reweighted Fitting of Eigenfaces / Wangmeng Zuo, Kuanquan Wang and David Zhang // Conference on Advances in Multimedia Information Processing, Lecture Notes in Computer Science. – 2006. – Vol. 4261. – P. 844-851.
4. **Lowe, D.** Distinctive image features from scale-invariant keypoints / D. Lowe // International Journal of Computer Vision. – 2004. – Vol. 60, N 2. – P. 91-110. – ISSN 0920-5691.
5. **Zhuravlev, Yu.I.** On algebraic approach in the problems of pattern recognition or classification / Yu.I. Zhuravlev // Cybernetics problems. – 1978. – Vol. 33. – P. 5-68. – (In Russian).
6. **Esmaili, M.** Creating of Multiple Classifier Systems by Fuzzy Decision Making in Human-Computer Interface Systems / M. Esmaili, M. Rahmati // Conference IEEE Fuzzy Systems, 2007. – P. 1-7.
7. **Mazurov, V.D.** Method of committees in optimisation and classification / V.D. Mazurov. – Moscow: “Nauka” Publisher, 1990. – 248 p. – (In Russian).
8. **Pardo, M.** Learning from data: a tutorial with emphasis on modern pattern recognition methods / M. Pardo, G. Sberveglieri // Sensors Journal, IEEE. – 2002. – Vol. 2(3). – P. 203-217. – ISSN 1530-437X.
9. **Fukunaga, K.** Introduction to Statistical Pattern Recognition; 2nd ed. / K. Fukunaga. – New York: Academic Press, Inc., 1991. – 591 p.
10. **Chow, C.K.** On optimum error and reject trade-off / C.K. Chow // IEEE Transactions on Information Theory. – 1970. – Vol. 16. – P. 41-46. – ISSN 0018-9448.
11. **Tsyppkin, Ya.Z.** Adaptation and training in automated systems / Ya.Z. Tsyppkin. – Moscow: “Nauka” Publisher, 1968. – 400 p. – (In Russian)
12. **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. – ISSN 0134-2452. – (In Russian).
13. **Shapiro, L.** Computer vision / L. Shapiro, G. Stockman. – Prentice Hall, 2001. – 752 p.
14. **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.

15. **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. – ISSN 1810-7206. – (In Russian).
16. **Gorelik, A.L.** Modern state of the art in the recognition problem: several aspects / A.L. Gorelik, I.B. Gurevich, V.A. Scripkin. – Moscow: "Radio i Svyazj" Publisher, 1985. – 160 p. – (In Russian).
17. **Tresp, V.** Committee machines / V. Tresp // Handbook for Neural Network Signal Processing. – 2001. – P. 135-151.
18. **Rudakov, K.V.** On methods of optimization and monotonous correction in the algebraic approach to the pattern recognition problem / K.V. Rudakov, K.V. Vorontsov // Doklady Akademii Nauk. – 1999. – Vol. 367, N 3. – P. 314-317. – ISSN 0869-5652. – (In Russian).
19. **Kullback, S.** Information Theory and Statistics / S. Kullback. – Dover Pub, 1978. – 408 p.
20. **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. – ISSN 1993-8985. – (In Russian).
21. **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. – Vol. 35(3). – P. 385-394. – ISSN 0134-2452. – (In Russian).
22. **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. – ISSN 0134-2452. – (In Russian).
23. **Face Processing: Advanced Modeling and Methods** / edited by W. Zhao, R. Chellappa. – Elsevier: Academic Press, 2005. – 768 p.
24. **The Yale Face database** [Electronical Resource] – <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
25. **The AT&T database** [Electronical Resource] – <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
26. **Savchenko, A.V.** Automatic image recognition criterion combining based on minimum information-discrimination principle / A.V. Savchenko // Control systems and Information Technologies. – 2011. – Vol. 44, N 2. – P. 22-25. – ISSN 1729-5068. – (In Russian).
27. **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. – ISSN 0031-3203.

THE CHOICE OF ALGORITHM PARAMETERS IN IMAGE RECOGNITION ON THE BASIS OF ENSEMBLE CLASSIFIERS AND THE MAXIMUM POSTERIOR PROBABILITY PRINCIPLE

A.V. Savchenko

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

Abstract

The problem of the choice of algorithms parameters in automatic image recognition is put and solved by ensemble classifiers construction using the maximum posterior probability principle. The new criterion of parameters choice is strictly synthesized for Kullback-Leibler information discrimination and modern SIFT (Scale-Invariant Feature Transform) method of object recognition. The program and results of experimental research in a problem of face recognition with widely used databases (Yale, AT&T) are presented. It is shown that the proposed criterion allows to achieve recognition accuracy equal to the algorithm with the best parameters set, and not only for Kullback-Leibler information discrimination, but also for other popular distances (Euclidean metric, Kullback information divergence).

Key words: automatic image recognition, ensemble classifiers, Kullback-Leibler minimum discrimination information principle, maximum posterior probability principle.

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 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.

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

E-mail: 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.

Поступила в редакцию 14 декабря 2011 г.