

# Многоклассовое распознавание образов в пространстве представлений с многоуровневым разрешением\*

М. М. Ланге, С. Н. Ганебных, А. М. Ланге

lange\_mm@ccas.ru, sng@ccas.ru

ФИЦ «Информатика и управление» РАН, г. Москва, ул. Вавилова, 44/2

Рассматривается метрическая схема распознавания образов, порождаемых многоклассовым источником изображений, в пространстве древовидных представлений с многоуровневым разрешением. Для различных уровней разрешения на множестве представлений введено семейство мер различия образов и разделяющих функций в форме правдоподобий по классам. Решение базируется на критерии голосования значений разделяющих функций и допускает отказ. Разработана процедура обучения, включающая отбор эталонов и оптимизацию параметров. Построен параметрический решающий алгоритм, включающий стратегии иерархического и переборного поиска решения в многоуровневой сети эталонов. Получена аналитическая оценка вычислительной сложности решающего алгоритма. Эффективность классификаторов с различными параметрами демонстрируется ROC-кривыми и эмпирическими зависимостями доли ошибок от вычислительной сложности для объединенного источника изображений подписей, жестов руки и лиц.

**Ключевые слова:** *древовидное представление; разделяющая функция; многоклассовое распознавание образов; многоуровневая сеть эталонов; ROC-кривая; иерархический поиск; вычислительная сложность*

DOI: 10.21469/22233792.2.1.06

## 1 Введение

Во многих приложениях эффективность методов распознавания образов целесообразно оценивать в терминах соотношения характеристик качества и трудоемкости, которые определяются вероятностью ошибок и вычислительной сложности алгоритмов принятия решения. Увеличение информации о распознаваемых объектах, как правило, позволяет уменьшить вероятность ошибки, но сопряжено с ростом вычислительной сложности решающих алгоритмов, что приводит к снижению быстродействия распознающих устройств (классификаторов). Острота проблемы существенно возрастает с увеличением числа классов и, в частности, в случае распознавания биометрических объектов, когда число классов определяется количеством персон, данные о которых хранятся в памяти классификатора, и может составлять десятки и сотни тысяч. Поэтому в рамках многоклассового распознавания важной задачей является уменьшение вычислительной сложности решающего алгоритма при сохранении малой вероятности ошибок. Для решения этой задачи естественно использовать структурные методы представления данных с многоуровневым разрешением, которые позволяют строить решающие алгоритмы на основе иерархических процедур направленного поиска решения.

Структурные представления изображений относятся к описаниям, которые ориентированы на ускорение процедур анализа изображений. Среди таких описаний особое место

---

\*Работа выполнена при частичной финансовой поддержке РФФИ, проекты №№ 15-01-04671, 15-07-09324 и 15-07-07516.

занимают древовидные и пирамидальные представления с многоуровневым разрешением [1–3], которые позволяют ускорить процедуры сегментации и выделения информативных объектов на изображениях. Древовидные и пирамидальные представления растровых изображений не инвариантны к аффинным преобразованиям объектов изображения и избыточны, поскольку содержат описания как информативных объектов (образов), так и фона.

Отмеченные недостатки, как правило, отсутствуют у представлений образов, выделенных на изображениях. Известны представления геометрических форм в форме графов [4] и скелетов [5]. Такие представления инвариантны к преобразованиям поворота и смещения объекта в поле изображения. Однако существует проблема построения меры на множестве таких представлений, связанная с неопределенностью установления соответствия фрагментов представлений. Развитие структурных представлений образов для эффективного распознавания в терминах соотношения «качество–сложность» заключается в построении инвариантных древовидно-структурированных описаний образов, допускающих быстрое вычисление мер различия и сходства таких представлений. В работе [6] предложен метод представления образов, заданных полутоновыми изображениями, в форме деревьев эллиптических примитивов и введена мера различия образов на множестве таких представлений. Показано, что рассмотренные древовидные представления позволяют построить эффективный метрический классификатор на основе голосования эталонов с использованием решающего алгоритма, вычислительная сложность которого существенно меньше сложности переборного алгоритма. Сокращение вычислительной сложности решающего алгоритма достигнуто за счет многоуровневого разрешения используемых представлений и сужения зоны поиска решения (вычисления меры) на последовательных уровнях базы эталонов. Быстрый алгоритм вычисления меры с использованием деревьев рассмотрен также в работе [7].

В наиболее полной постановке, задача многоклассового распознавания с возможностью отказа исследована в работе [8]. В указанной работе для серии источников образов приведены сравнительные доли ошибок при различных методах распознавания в векторных пространствах признаков. Однако в этой работе отсутствуют оценки вычислительной сложности решающих алгоритмов, которые использованы в рассмотренных методах. Метод многоклассового распознавания образов в пространстве древовидных представлений с метрическим решающим правилом, не предусматривающим отказа, исследован авторами в работе [9] в терминах соотношения «качество–сложность». В настоящей работе дано обобщение этого метода для решающего правила с отказом и продемонстрирована его эффективность для распознавания лиц, жестов и подписей. Характеристики эффективности приведены в терминах ROC-кривых (Receiver Operating Characteristics) [10] и зависимостей доли ошибок от вычислительной сложности решающего алгоритма.

## 2 Источник образов и модель распознавания

Пусть источник порождает множество изображений, а выделенные на них образы образуют множество  $\mathbf{X} = \{\mathbf{x}\}$  объектов, которые могут быть представлены в форме векторов, деревьев, графов и т. п. Множество  $\mathbf{X}$  содержит объекты, относящиеся к множеству классов  $\Omega = \{\omega_0, \omega_1, \dots, \omega_c\}$ ,  $c \geq 1$ , в котором каждый положительный класс из подмножества  $\Omega_c = \{\omega_i, i = 1, \dots, c\}$  включает объекты с одинаковой семантикой, определяемой номером  $i$ , а нулевой класс  $\omega_0$  включает все прочие объекты. На множестве классов  $\Omega$  задано априорное распределение  $\{p(\omega_i) \geq 0, i = 0, \dots, c : \sum_{i=0}^c p(\omega_i) = 1\}$ , которое дает вероятности  $p(\Omega_c) = \sum_{i=1}^c p(\omega_i) = \theta \geq 0$  и  $p(\omega_0) = 1 - \theta$ ; на множестве объектов  $\mathbf{X}$  определена

мера различия  $d(\mathbf{x}, \hat{\mathbf{x}}) \geq 0$  любой пары объектов  $\mathbf{x} \in \mathbf{X}$ ,  $\hat{\mathbf{x}} \in \mathbf{X}$ . Предполагается, что распределение  $\{\theta, 1 - \theta\}$  неизвестно, но известно условное распределение на подмножестве  $\Omega_c : \{p(\omega_i|\Omega_c) = p(\omega_i) / \sum_{i=1}^c p(\omega_i), i = 1, \dots, c\}$ .

Будем считать, что задано обучающее множество

$$\mathbf{X}^{\text{train}} = \{\mathbf{X}_i^{\text{train}} = \{\mathbf{x}_{ij}\}_{j=1}^{m_i}\}_{i=0}^c \subset \mathbf{X}$$

мощности  $\|\mathbf{X}^{\text{train}}\| = \sum_{i=0}^c m_i = m$ , в котором кластеры с положительными номерами образуют подмножество  $\mathbf{X}_{\Omega_c}^{\text{train}} = \{\mathbf{X}_i^{\text{train}}\}_{i=1}^c$ , а кластер с нулевым номером дает подмножество  $\mathbf{X}_{\omega_0}^{\text{train}} = \mathbf{X}_0^{\text{train}}$ . На подмножестве  $\mathbf{X}_{\Omega_c}^{\text{train}}$  отобраны наборы эталонных объектов, образующие множество

$$\hat{\mathbf{X}} = \{\hat{\mathbf{X}}_i = \{\hat{\mathbf{x}}_{ij}\}_{j=1}^{\hat{m}_i} \subseteq \mathbf{X}_i^{\text{train}}\}_{i=1}^c \quad (1)$$

мощности  $\|\hat{\mathbf{X}}\| = \sum_{i=1}^c \hat{m}_i = \hat{m}$ , где  $\hat{m}_i \leq m_i$  и  $\hat{m} \leq m$ . Для пары объектов  $\mathbf{x} \in \mathbf{X}$  и  $\hat{\mathbf{x}} \in \hat{\mathbf{X}}$  вводится ядро

$$K_d(\mathbf{x}, \hat{\mathbf{x}}) = e^{-d(\mathbf{x}, \hat{\mathbf{x}})} \quad (2)$$

по мере  $d(\mathbf{x}, \hat{\mathbf{x}})$ . Ядро (2) и нормированные по классам подмножества  $\Omega_c$  весовые коэффициенты эталонов

$$\mathbf{W} = \{\mathbf{W}_i = \{w(\hat{\mathbf{x}}_{ij}) \geq 0\}_{j=1}^{\hat{m}_i} : \sum_{j=1}^{\hat{m}_i} w(\hat{\mathbf{x}}_{ij}) = 1\}_{i=1}^c \quad (3)$$

порождают разделяющие функции [11]

$$g_i(\mathbf{x}) = \sum_{j=1}^{\hat{m}_i} w(\hat{\mathbf{x}}_{ij}) K_d(\mathbf{x}, \hat{\mathbf{x}}_{ij}), i = 1, \dots, c, \quad (4)$$

которые эквивалентны потенциалам в точке  $\mathbf{x}$ , создаваемым наборами эталонов  $\hat{\mathbf{X}}_i, i = 1, \dots, c$  с весами (3). Разделяющие функции (4) и набор порогов  $\Delta = \{\Delta_i, i = 1, \dots, c\}$  дают следующее решающее правило определения номера класса для объекта  $\mathbf{x} \in \mathbf{X}$ :

$$i^*(\mathbf{x}) = \max_{i=1}^c [g_i(\mathbf{x}) \geq \Delta_i] \arg \max_{i=1}^c (g_i(\mathbf{x}) [g_i(\mathbf{x}) \geq \Delta_i]), \quad (5)$$

где  $[f]$  – индикатор, принимающий значение 1, если условие  $f$  выполняется, и значение 0 в противном случае.

В общем случае решение (5) соответствует критерию голосования разделяющих функций [11], дополненному отказом от распознавания. В настоящей работе исследуются классификаторы по критерию (5), в котором используются наборы эталонов (1) и весовые коэффициенты

$$w(\hat{\mathbf{x}}_{ij}) = [\hat{\mathbf{x}}_{ij} = \arg \min_{k=1}^{\hat{m}_i} d(\mathbf{x}, \hat{\mathbf{x}}_{ik})], i = 1, \dots, c,$$

которые приводят к модификациям критерия ближайшего соседа. В любом случае при выбранном множестве весов эталонов  $\mathbf{W}$  вида (3), качество решения (5) определяется условными вероятностями ошибок  $\varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W}|\Omega_c)$  и  $\varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W}|\omega_0)$  соответственно по «своим» (из подмножества классов  $\Omega_c$ ) и «чужим» (из класса  $\omega_0$ ) объектам. Полная вероятность ошибки определяется математическим ожиданием

$$\varepsilon_\theta(\hat{\mathbf{X}}, \Delta, \mathbf{W}) = \theta \varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W}|\Omega_c) + (1 - \theta) \varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W}|\omega_0). \quad (6)$$

Полагая, что  $\theta$  является случайной величиной с плотностью  $p(\theta) = 1, 0 \leq \theta \leq 1$ , и интегрируя функцию (6) по  $\theta$  с указанной плотностью, получим функционал ошибок

$$\varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W}) = \int_0^1 \varepsilon_\theta(\hat{\mathbf{X}}, \Delta, \mathbf{W}) p(\theta) d\theta = \frac{1}{2} \left( \varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \Omega_c) + \varepsilon(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \omega_0) \right). \quad (7)$$

В рамках рассматриваемой модели, процедура обучения сводится к отбору множества эталонов  $\hat{\mathbf{X}}$  вида (1) с параметрами  $\hat{m}_i < m_i, i = 1, \dots, c$ , и выбору пороговых значений  $\Delta$ , которые минимизируют функционал ошибок  $\varepsilon^{\text{train}}(\hat{\mathbf{X}}, \Delta, \mathbf{W})$  вида (7) на объектах обучающего множества  $\mathbf{X}^{\text{train}}$ . Оптимизация классификатора сводится к нахождению пары

$$(\hat{\mathbf{X}}^*, \Delta^*) = \arg \min_{\hat{\mathbf{X}}, \Delta: \hat{m} < m} \varepsilon^{\text{train}}(\hat{\mathbf{X}}, \Delta, \mathbf{W}). \quad (8)$$

Множество эталонов и набор порогов, удовлетворяющие соотношению (8), определяют оптимальный классификатор с решающим правилом (5), которое допускает точную или приближенную реализацию на алгоритме  $a_\alpha$  с параметром  $\alpha$ . Эффективность классификатора с оптимальными параметрами (8) и алгоритмом  $a_\alpha$  определяется двумя характеристиками: вероятностью ошибок  $\varepsilon_\alpha^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})$  вида (7), вычисляемой на тестовом множестве  $\mathbf{X}^{\text{test}} = \mathbf{X} \setminus \mathbf{X}^{\text{train}}$ , и вычислительной сложностью  $C_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})$ , измеряемой количеством операций, затрачиваемых алгоритмом  $a_\alpha$  на один объект. Полагая, что с увеличением сложности  $C_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})$  (путем варьирования параметром  $\alpha$ ) вероятность ошибок  $\varepsilon_\alpha^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})$  не возрастает и исключая  $\alpha$  при заданной сложности  $C^* > 0$ , алгоритм  $a_\alpha$  порождает функцию «качество–сложность»

$$\varepsilon(C^* | \hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}) = \varepsilon_\alpha^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}) : \alpha = \arg(C_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}) = C^*) \quad (9)$$

для классификатора с параметрами  $\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}$ .

Ниже исследуется классификатор с решающим правилом (5), реализующим критерий ближайшего соседа в пространстве древовидных представлений объектов с многоуровневым разрешением. Дано описание способа построения древовидных представлений объектов, и на множестве представлений построена мера различия объектов, используемая в ядре (2). В заданном пространстве представлений определены разделяющие функции вида (4) и построен параметрический решающий алгоритм, который допускает существенное уменьшение вычислительной сложности по сравнению с алгоритмом полного перебора при достаточно малом увеличении доли ошибок. Для источников изображений лиц, жестов руки и подписей при различных способах отбора эталонов построены оценки функции «качество–сложность» (9) путем варьирования параметром решающего алгоритма.

### 3 Представление образов и решающий алгоритм

В этом разделе рассматривается способ древовидного представления образов с многоуровневым разрешением, вводится семейство мер различия образов и соответствующие этим мерам разделяющие функции на различных уровнях разрешения. Используя разделяющие функции всех уровней, строится иерархический алгоритм, реализующий решающее правило (5).

### 3.1 Многоуровневое описание образов примитивами

Множество объектов источника  $\mathbf{X}$  включает образы, выделенные на полутоновых изображениях и удовлетворяющие некоторым ограничениям. Налагаемые ограничения формируют множество допустимых образов, удовлетворяющих следующему определению.

**Определение 1.** Образ, заданный на изображении односвязным или многосвязным набором пикселей, яркости которых соответствуют их «массам», считается допустимым, если набор пикселей объекта образует компактное или распределенное двумерное твердое тело с однозначно идентифицируемой декартовой системой собственных координат.

Для множества образов  $\mathbf{X}$ , удовлетворяющих определению 1, в работе [6] предложен способ их представления бинарными деревьями эллиптических примитивов. Способ базируется на дихотомическом разбиении любого образа  $\mathbf{x} \in \mathbf{X}$  на сегменты (наборы пикселей) и аппроксимации каждого сегмента с номером  $n$  эллиптическим примитивом  $Q_n$ , который соответствует вершине бинарного дерева. Каждый делимый сегмент с номером  $n$  порождает пару новых сегментов следующего уровня с номерами  $2n+1$  и  $2n+2$  и соответствующую пару аппроксимирующих примитивов. Нумерация сегментов и примитивов производится с учетом ориентации оси разбиения делимого сегмента.

Корневой примитив  $Q_0$  аппроксимирует исходный образ  $\mathbf{x}$  и имеет номер  $n = 0$ . Концевые вершины дерева примитивов соответствуют неделимым сегментам с числом пикселей не более заданного значения (например, сегменты, состоящие из одного пикселя). Аппроксимирующие примитивы строятся в декартовой системе собственных координат аппроксимируемых сегментов. В случае отсутствия у сегмента с номером  $n$  однозначно идентифицируемой системы собственных координат примитив  $Q_n$  строится в системе координат сегмента-родителя с номером  $\lfloor (n-1)/2 \rfloor$ . Таким образом, для построения дерева примитивов достаточно существования идентифицируемой системы собственных координат исходного образа, который является корневым сегментом с номером  $n = 0$ .

Введем следующие характеристики дихотомического разбиения образа  $\mathbf{x}$ :  $N_l$  — множество номеров концевых вершин на  $l$ -м шаге;  $x'_l = \{P_n, n \in N_l : \cup_n P_n = \mathbf{x}\}$  — слой сегментов;  $x_l = \{Q_n, n \in N_l\}$  — слой аппроксимирующих примитивов на  $l$ -м шаге. Базовыми операциями над сегментом  $P_n \in x'_l$  являются: операция формирования сегментов следующего уровня (segmentation)

$$F_{\text{seg}}(P_n) = \begin{cases} (P_{2n+1}, P_{2n+2}), & \text{если } P_n \text{ — делимый;} \\ P_n, & \text{если } P_n \text{ — неделимый,} \end{cases}$$

и операция аппроксимации

$$F_{\text{app}}(P_n) = Q_n,$$

которая сохраняет примитив, если он выбран для сегмента на предыдущем шаге, либо строит примитив на текущем шаге, если сегмент не имел аппроксимации на предыдущем шаге. Введенные понятия позволяют сформулировать следующий рекурсивный алгоритм представления (representation) объекта  $\mathbf{x} \in \mathbf{X}$ .

**Алгоритм  $a_{\text{rep}}$ .** Начальные условия при  $l = 0$ :

$$P_0 = \mathbf{x}, x'_0 = P_0; \quad Q_0 = F_{\text{app}}(P_0), \quad x_0 = Q_0; \quad L \geq 1.$$

Вычисление слоев сегментов и примитивов на шагах  $l = 1, \dots, L-1$ :

$$x'_l = \{F_{\text{seg}}(P_n), \forall P_n \in x'_{l-1}\}; \quad x_l = \{F_{\text{app}}(P_n), \forall P_n \in x'_l\}.$$

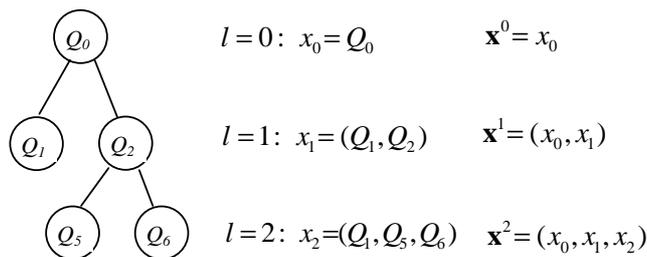


Рис. 1 Пример полного бинарного дерева глубины  $L = 2$

Алгоритм  $a_{\text{тер}}$  строит для объекта  $\mathbf{x} \in \mathbf{X}$ , удовлетворяющего определению 1, многоуровневое представление

$$a_{\text{тер}}(\mathbf{x}) = \mathbf{x}^L = (x_0, \dots, x_l, \dots, x_L) \tag{10}$$

в форме полного бинарного дерева глубины  $L$  (бинарное дерево называется полным, если любая его промежуточная вершина имеет две исходящие ветви). В (10) любое поддерево  $\mathbf{x}^l = (x_0, \dots, x_l)$  глубины  $l = 0, \dots, L$  задано последовательностью слоев примитивов, в которой слой  $x_l$  образован концевыми вершинами этого поддерева. Такое представление обеспечивает растущее с увеличением  $l$  разрешение, определяемое числом примитивов в слое  $x_l$ . Пример структуры вида (10) дан на рис. 1 для случая  $L = 2$ .

Каждый эллиптический примитив определяется набором параметров

$$Q_n = (\mathbf{r}_n, \mathbf{u}_n, \mathbf{v}_n, z_n), \tag{11}$$

где  $n$  — номер вершины дерева, соответствующий сегменту образа;  $\mathbf{r}_n$  — вектор центра эллипса;  $\mathbf{u}_n$  и  $\mathbf{v}_n$  — векторы большой и малой полуосей;  $z_n$  — средний уровень яркости пикселей в аппроксимируемом сегменте. Координаты центра, определяющие вектор  $\mathbf{r}_n$ , и радиусы, определяющие векторы  $\mathbf{u}_n$  и  $\mathbf{v}_n$ , вычисляются как параметры эллипса рассеяния соответствующего сегмента. Векторы  $\mathbf{r}_n$ ,  $\mathbf{u}_n$  и  $\mathbf{v}_n$  задаются в собственной системе координат объекта  $\mathbf{x}$ . Разбиение делимого сегмента с номером  $n \geq 0$  производится осью, проходящей через центр этого сегмента и параллельной одной из собственных осей сегмента с номером  $n = 0$ . Ось разбиения на каждом шаге  $l$  фиксирована и меняется с увеличением номера шага. При указанной нумерации примитив  $Q_n$  находится в полном дереве на уровне  $l = \lfloor \log_2(n + 1) \rfloor$ . Параметры каждого примитива (11) с номером  $n \geq 0$  нормируются относительно корневого примитива с номером  $n = 0$ .

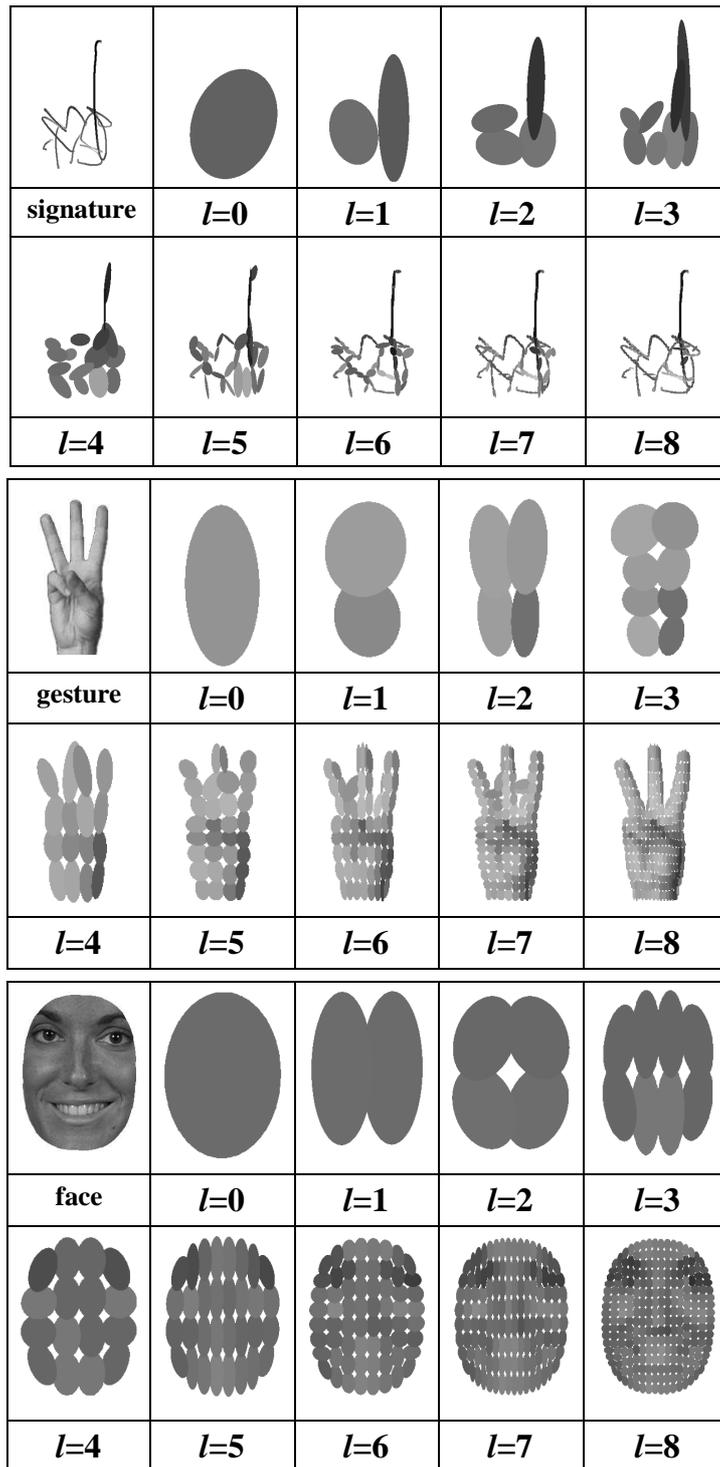
Примеры представлений подписи, жеста руки и лица слоями эллиптических примитивов, которые соответствуют уровням завершеного бинарного дерева глубины  $L = 8$  (с числом вершин  $2^l$ ,  $l = 0, \dots, L$ ) даны на рис. 2.

Построение дерева примитивов в собственных координатах объекта и нормировка их параметров позволяют сформулировать следующее свойство представления (10).

**Утверждение 1.** При достаточно малом размере пикселей и большом числе уровней квантования яркостей бинарное дерево эллиптических примитивов с точностью до размера пикселя и уровня квантования инвариантно к преобразованиям поворота, смещения, изменения масштаба и уровня яркости представляемого объекта.

### 3.2 Мера различия объектов

В этом подразделе вводится семейство мер различия объектов на множестве представлений вида (10). Для определения меры различия объектов  $\mathbf{x} \in \mathbf{X}$ ,  $\hat{\mathbf{x}} \in \hat{\mathbf{X}}$ , представленных поддеревами  $\mathbf{x}^l \in \mathbf{x}^L$ ,  $\hat{\mathbf{x}}^l \in \hat{\mathbf{x}}^L$  глубины  $l$ , потребуются следующие определения.



**Рис. 2** Примеры представлений подписи, жеста руки и лица слоями эллиптических примитивов, образующими завершенные бинарные деревья глубины  $L = 8$

**Определение 2.** Примитивы  $Q_n \in \mathbf{x}^l$  и  $\hat{Q}_n \in \hat{\mathbf{x}}^l$  в  $n$ -х вершинах поддеревьев  $\mathbf{x}^l$  и  $\hat{\mathbf{x}}^l$  называются соответственными. Множество пар соответственных примитивов в поддеревьях  $\mathbf{x}^l$  и  $\hat{\mathbf{x}}^l$  образует пересечение  $\mathbf{x}^l \cap \hat{\mathbf{x}}^l$ .

С учетом определения 2 введем для любой пары поддеревьев  $\mathbf{x}^l, \hat{\mathbf{x}}^l, l = 1, \dots, L$  следующие функции различия порядка  $l$  по параметрам соответственных примитивов:

$$\left. \begin{aligned} \rho_r(\mathbf{x}, \hat{\mathbf{x}}) &= \sum_{n:(Q_n, \hat{Q}_n) \in \mathbf{x}^l \cap \hat{\mathbf{x}}^l} w_n \|\mathbf{r}_n - \hat{\mathbf{r}}_n\|; \\ \rho_{uv}(\mathbf{x}, \hat{\mathbf{x}}) &= \sum_{n:(Q_n, \hat{Q}_n) \in \mathbf{x}^l \cap \hat{\mathbf{x}}^l} w_n (\|\mathbf{u}_n - \hat{\mathbf{u}}_n\| + \|\mathbf{v}_n - \hat{\mathbf{v}}_n\|); \\ \rho_z(\mathbf{x}, \hat{\mathbf{x}}) &= \sum_{n:(Q_n, \hat{Q}_n) \in \mathbf{x}^l \cap \hat{\mathbf{x}}^l} w_n \|z_n - \hat{r}_n\|, \end{aligned} \right\} \quad (12)$$

где  $w_n$  — весовой коэффициент соответственных примитивов с номером  $n$ :

$$w_n = \frac{[\log_2(n+1)] 2^{-\lfloor \log_2(n+1) \rfloor}}{\sum_{n:(Q_n, \hat{Q}_n) \in \mathbf{x}^l \cap \hat{\mathbf{x}}^l} [\log_2(n+1)] 2^{-\lfloor \log_2(n+1) \rfloor}}.$$

Нормы в (12) вычисляются в метрике L1. Функции (12) дают средние отклонения объектов обучающего множества  $\mathbf{X}^{\text{train}}$  относительно объекта  $\hat{\mathbf{x}}$ :

$$\left. \begin{aligned} \sigma_r(\hat{\mathbf{x}}^l) &= \frac{1}{m-1} \sum_{\mathbf{x}^l: \mathbf{x} \in \mathbf{X}^{\text{train}}} \rho_r(\mathbf{x}^l, \hat{\mathbf{x}}^l); \\ \sigma_{uv}(\hat{\mathbf{x}}^l) &= \frac{1}{m-1} \sum_{\mathbf{x}^l: \mathbf{x} \in \mathbf{X}^{\text{train}}} \rho_{uv}(\mathbf{x}^l, \hat{\mathbf{x}}^l); \\ \sigma_z(\hat{\mathbf{x}}^l) &= \frac{1}{m-1} \sum_{\mathbf{x}^l: \mathbf{x} \in \mathbf{X}^{\text{train}}} \rho_z(\mathbf{x}^l, \hat{\mathbf{x}}^l), \end{aligned} \right\} \quad (13)$$

где  $m$  — число объектов в множестве  $\mathbf{X}^{\text{train}}$ . С учетом соотношений (12) и (13) мера различия порядка  $l = 1, \dots, L$  для пары объектов  $\mathbf{x}, \hat{\mathbf{x}}$  определяется функцией:

$$d^l(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\rho_r(\mathbf{x}^l, \hat{\mathbf{x}}^l)}{\sigma_r(\hat{\mathbf{x}}^l)} + \frac{\rho_{uv}(\mathbf{x}^l, \hat{\mathbf{x}}^l)}{\sigma_{uv}(\hat{\mathbf{x}}^l)} + \frac{\rho_z(\mathbf{x}^l, \hat{\mathbf{x}}^l)}{\sigma_z(\hat{\mathbf{x}}^l)}, \quad (14)$$

которая является аналогом махаланобисовой меры [11] для трехмерного пространства признаков с метрикой L1.

### 3.3 Разделяющие функции и решающий алгоритм

Набор мер  $\{d^l(\mathbf{x}, \hat{\mathbf{x}}), l = 1, \dots, L\}$  вида (14) порождает набор разделяющих функций

$$\{g_i^l(\mathbf{x}), l = 1, \dots, L\}_{i=1}^c \quad (15)$$

вида (4), где  $l$  — порядок функции  $g_i^l(\mathbf{x})$ . Функции (15) используются для построения алгоритма поиска решения вида (5) с использованием параметрической стратегии сужения зоны поиска на последовательных уровнях  $l = 1, \dots, L$ . Предполагается, что наборы эталонов по классам в множестве (1) заданы сетью представлений

$$\{\hat{\mathbf{X}}_i^l, l = 1, \dots, L\}_{i=1}^c, \quad (16)$$

в которой  $\hat{\mathbf{X}}_i^l$  — набор представлений поддеревьями глубины  $l$  для набора эталонов  $\hat{\mathbf{X}}_i$ . Стратегия сужения зоны поиска решения в сети (16) задается экспоненциальной функцией

$$c_l = \lfloor c 2^{-\alpha(l-1)} \rfloor, l = 1, \dots, L, \quad (17)$$

где  $\alpha \geq 0$  — свободный параметр, а  $c_l$  — число классов, для которых вычисляются разделяющие функции порядка  $l$  на соответствующем уровне сети (16). Стратегия (17) дает следующий алгоритм, реализующий решение (5).

**Алгоритм  $a_\alpha$ .** На последовательных уровнях  $l = 1, \dots, L - 1$  сети (16) вычисляются разделяющие функции (15) для  $c_l$  классов и среди них отбираются  $c_{l+1}$  классов с наибольшими значениями разделяющих функций; решение (5) принимается на уровне  $l^* = (l < L)[c_{l+1} < 2] + (l = L)[c_L \geq 2]$ , где  $[f]$  — индикатор  $f$ .

В алгоритме  $a_\alpha$  параметр  $\alpha > 0$  соответствует стратегии направленного поиска решения (guided search algorithm  $a_{\alpha>0}^{gs}$ ), а параметр  $\alpha = 0$  — стратегии поиска решения на основе полного перебора (exhaustive search algorithm  $a_{\alpha=0}^{es}$ ).

Вычислительная сложность  $C_\alpha$  алгоритма  $a_\alpha$  определяется количеством элементарных операций, затрачиваемых на вычисление меры различия  $d^l(\mathbf{x}, \hat{\mathbf{x}})$  вида (14) между предъявляемым объектом  $\mathbf{x}$  и всеми эталонными объектами  $\hat{\mathbf{x}} \in \hat{\mathbf{X}}^*$ , на уровнях  $l = 1, \dots, L$  сети (16). Множество эталонов  $\hat{\mathbf{X}}^*$  отбирается согласно (8) на этапе обучения. Вычислительные затраты на вычисление разделяющих функций, их сортировку и отбор наибольших значений, а также затраты на сравнение с порогами малы по сравнению с затратами на вычисление меры и поэтому не учитываются. Элементарной операцией считается сравнение пары соответственных примитивов предъявляемого и эталонного объекта. Поскольку число примитивов в  $l$ -м слое  $x_l$  представления (10) не превосходит  $2^l$ , сложность алгоритма  $a_\alpha$ , реализующего стратегии (17), удовлетворяет оценке

$$C_\alpha \leq \sum_{l=1}^L 2^l \sum_{i=1}^c \hat{m}_i [i \in N_{c_l}] \leq 2c \max_{i=1}^c \hat{m}_i \sum_{l=1}^L 2^{(1-\alpha)(l-1)}, \quad (18)$$

где  $\hat{m}_i$  — число эталонов  $i$ -го класса;  $N_{c_l}$  — множество номеров классов, отобранных в сети (16) на  $l$ -м уровне ( $\|N_{c_l}\| = c_l$ );  $[f]$  — индикатор  $f$ . Модификация оценки (18) получена в работе [9].

При использовании совершенных представляющих деревьев вида (10), в которых уровень с номером  $l = 1, \dots, L$  содержит  $2^l$  вершин-примитивов, оценка (18) позволяет оценить сверху отношение сложностей алгоритмов  $a_{\alpha>0}^{gs}$  и  $a_{\alpha=0}^{es}$ . В случае применения деревьев указанного типа двойная сумма в (18) при  $\alpha = 0$  дает точное значение сложности переборного решающего алгоритма  $a_{\alpha=0}^{es}$ :

$$C_{\alpha=0} = \sum_{l=1}^L 2^l \sum_{i=1}^c \hat{m}_i = 2(2^L - 1)c\hat{m}_{\text{mean}}, \quad (19)$$

где  $\hat{m}_{\text{mean}} = c^{-1} \sum_{i=1}^c \hat{m}_i$ . При  $\alpha > 0$  неравенство (18) дает оценку сверху для сложности иерархического решающего алгоритма  $a_{\alpha>0}^{gs}$ :

$$C_{\alpha>0} \leq 2 \frac{2^{(1-\alpha)L} - 1}{2^{1-\alpha} - 1} c\hat{m}_{\text{max}}, \quad (20)$$

где  $\hat{m}_{\text{max}} = \max_{i=1}^c \hat{m}_i$ . Доля сложности алгоритма  $a_{\alpha>0}^{gs}$  относительно сложности алгоритма  $a_{\alpha=0}^{es}$  определяется отношением  $\gamma = C_{\alpha>0}/C_{\alpha=0}$ , которое с учетом (19) и (20) удовлетворяет оценке

$$\gamma \leq \frac{2^{(1-\alpha)L} - 1}{2^{(1-\alpha)} - 1} \frac{1}{2^L - 1} \frac{\hat{m}_{\text{max}}}{\hat{m}_{\text{mean}}}. \quad (21)$$

**Утверждение 2.** Доля вычислительной сложности решающего алгоритма с параметром  $\alpha \geq 1$  относительно сложности переборного алгоритма с параметром  $\alpha = 0$  удовлетворяет оценке

$$\frac{C_{\alpha \geq 1}}{C_{\alpha = 0}} \leq \frac{L}{2^L - 1} \frac{\hat{n}_{\max}}{\hat{m}_{\text{mean}}}.$$

Сформулированное утверждение следует из неравенства (21), в котором при  $\alpha \geq 1$  точное значение суммы  $L$  членов убывающей геометрической прогрессии заменено оценкой сверху  $(2^{(1-\alpha)L} - 1)/(2^{(1-\alpha)} - 1) \leq L$ . Утверждение 2 демонстрирует экспоненциально растущий вычислительный выигрыш алгоритма  $a_{\alpha \geq 1}^{\text{gs}}$  относительно алгоритма  $a_{\alpha = 0}^{\text{es}}$  с увеличением глубины представляющих деревьев.

## 4 Обучение классификатора

Процедура обучения предполагает отбор наборов эталонов по классам  $\hat{\mathbf{X}}_i^*$ ,  $i = 1, \dots, c$ , и получение оценок порогов  $\Delta_i^*$ ,  $i = 1, \dots, c$ , которые образуют множества  $\hat{\mathbf{X}}^* = \left\{ \hat{\mathbf{X}}_i^* \right\}_{i=1}^c$  и  $\Delta^* = \left\{ \Delta_i^* \right\}_{i=1}^c$ , удовлетворяющие условию (8). Строится оценка функционала ошибок (7) в форме зависимости от наборов эталонов  $\hat{\mathbf{X}}_i$  и порогов  $\Delta_i$  по классам с номерами  $i = 1, \dots, c$ . Предлагается рекурсивная процедура построения наборов эталонов  $\hat{\mathbf{X}}_i$  и параметрическая стратегия выбора порогов  $\Delta_i$  по классам подмножества  $\Omega_c$  для нахождения оптимальных параметров  $\left\{ \hat{\mathbf{X}}_i^*, \Delta_i^* \right\}_{i=1}^c$  путем минимизации оценки скользящего контроля функционала ошибок.

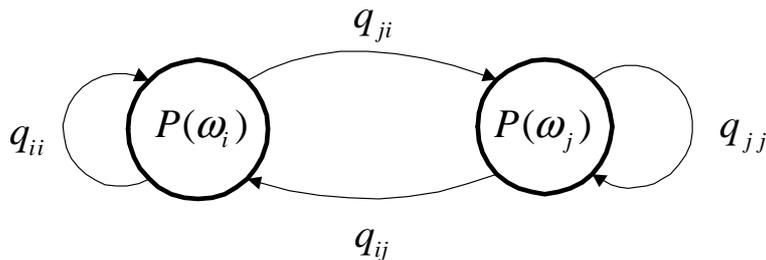
### 4.1 Оценка функционала ошибок

Разделяющие функции (15), вычисляемые согласно (4) по мерам порядка  $l = 1, \dots, L$ , принимают значения на отрезке  $[0, 1]$ . Пусть  $\bar{g}_i^L$  — средние значения разделяющих функций по  $S$  порядковым статистикам:

$$\bar{g}_i^L = \frac{1}{S} \sum_{s=1}^S g_{is}^L, \quad i = 1, \dots, c, \tag{22}$$

где  $g_{is}^L$  —  $s$ -е наибольшее значение (порядковая статистика) для функции  $g_i^L(\mathbf{x})$  на обучающем множестве  $\mathbf{X}^{\text{train}}$ . Средние значения (22) используются в стратегии выбора порогов  $\Delta_i = \delta \bar{g}_i^L$ ,  $i = 1, \dots, c$ , с параметром  $\delta \in [0, 1]$ . Ниже предлагается оценка функционала ошибок (7) в виде зависимости от наборов эталонов  $\left\{ \hat{\mathbf{X}}_i \right\}_{i=1}^c$  и порогового параметра  $\delta$ .

Процедура построения оценок вероятностей ошибок классификации по своим и чужим объектам базируется на схеме принятия решений, заданной графом на рис. 3. Состояния графа соответствуют классам  $\omega_0, \omega_1, \dots, \omega_c$ , а дуги — решениям, принимаемым



**Рис. 3** Схема принятия решений

на классе  $\omega_j$  по объектам класса  $\omega_i$ . Состояния графа помечены безусловными априорными вероятностями  $p(\omega_i)$ ,  $i = 0, \dots, c$ , а дуги — условными вероятностями ошибочных  $q_{ji}$ ,  $j = 0, \dots, c$ ,  $j \neq i$ , или верных  $q_{ii}$  решений по объектам из класса  $\omega_i \in \Omega$ .

С учетом схемы на рис. 3 условные вероятности ошибок в функционале (7) по своим и чужим объектам равны

$$\varepsilon \left( \hat{\mathbf{X}}, \Delta, \mathbf{W} | \Omega_c \right) = \sum_{i=1}^c (1 - q_{ii}) p(\omega_i | \Omega_c); \quad (23)$$

$$\varepsilon \left( \hat{\mathbf{X}}, \Delta, \mathbf{W} | \omega_0 \right) = \sum_{i=1}^c q_{i0}. \quad (23')$$

Полагая, что вероятности  $q_{ii}$  и  $q_{i0}$  являются функциями переменных  $\hat{\mathbf{X}}_i$  и  $\delta$ , зависящими от весов эталонов  $\mathbf{W}_i$ ,  $i = 1, \dots, c$ , в (3) и значений  $\bar{g}_i^L$  вида (22), получаем следующие оценки вероятностей верных и ложных решений на классе  $\omega_i$ ,  $i = 1, \dots, c$ , в терминах true positive rate (TPR) и false positive rate (FPR) [10]:

$$\hat{q}_{ii} = \varepsilon_{\text{tpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right); \quad \hat{q}_{i0} = \varepsilon_{\text{fpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right).$$

Применение этих оценок в формулах (23) дает при равномерном распределении  $p(\omega_i | \Omega_c) = 1/c$ ,  $i = 1, \dots, c$ , следующие оценки условных вероятностей ошибок:

$$\hat{\varepsilon} \left( \hat{\mathbf{X}}, \Delta, \mathbf{W} | \Omega_c \right) = \frac{1}{c} \sum_{i=1}^c \left( 1 - \varepsilon_{\text{tpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right) \right) = 1 - \text{TPR}; \quad (24)$$

$$\hat{\varepsilon} \left( \hat{\mathbf{X}}, \Delta, \mathbf{W} | \omega_0 \right) = \sum_{i=1}^c \varepsilon_{\text{fpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right) = \text{FPR}. \quad (24')$$

**Утверждение 3.** В рамках принятых допущений оценка функционала ошибок (7) с учетом соотношений (24) имеет вид:

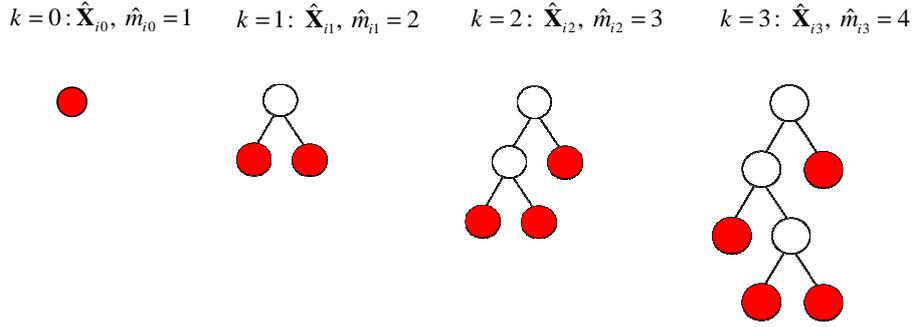
$$\hat{\varepsilon} \left( \hat{\mathbf{X}}, \Delta, \mathbf{W} \right) = \sum_{i=1}^c \varepsilon \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right), \quad (25)$$

где  $\varepsilon \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right)$  — оценка доли ошибок элементарного классификатора, образованного парой классов  $\{\omega_0, \omega_i\}$ :

$$\varepsilon \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right) = \frac{1}{2} \left( \frac{1}{c} \left( 1 - \varepsilon_{\text{tpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right) \right) + \varepsilon_{\text{fpr}} \left( \hat{\mathbf{X}}_i, \delta, \mathbf{W}_i \right) \right). \quad (26)$$

## 4.2 Отбор эталонов

Отбор эталонов в каждом кластере  $\mathbf{X}_i^{\text{train}} \subset \mathbf{X}^{\text{train}}$ ,  $i = 1, \dots, c$ , обучающего множества выполняется независимо с использованием жадного алгоритма, предложенного в работе [9]. Алгоритм отбора эталонов базируется на процедуре дихотомического разбиения кластера  $\mathbf{X}_i^{\text{train}}$  мощности  $m_i$  на фрагменты и выборе в каждом фрагменте эталонного объекта  $\hat{\mathbf{x}}$ , относительно которого объекты этого фрагмента имеют наименьшее рассеяние. Рассеяние вычисляется по мере вида (14) наибольшего порядка  $L$ .



**Рис. 4** Пример наборов эталонов в четырехслойном бинарном дереве

На шаге с номером  $k = 0$  в кластере  $\mathbf{X}_i^{\text{train}}$  выбирается единственный эталон. На каждом последующем шаге дихотомии с номером  $k = 1, \dots, m_i - 1$  делимым фрагментом является фрагмент, объекты которого имеют наибольшее рассеяние относительно своего эталона (на первом шаге делимым сегментом является весь кластер). В результате, на последовательных шагах рекурсивной процедуры формируются наборы эталонов  $\hat{\mathbf{X}}_{ik}$  мощности  $\|\hat{\mathbf{X}}_{ik}\| = \hat{m}_{ik} = k + 1, k = 0, \dots, m_i - 1$ . Набор эталонов  $\hat{\mathbf{X}}_{ik}$ , формируемый на  $k$ -м шаге, соответствует слою концевых вершин в бинарном дереве глубины  $k$ . Пример бинарного дерева, содержащего четыре слоя наборов эталонов, дан на рис. 4.

Формализация построения дерева наборов эталонов состоит в следующем. Пусть  $N_{ik}$  — множество номеров концевых вершин на  $k$ -м шаге дихотомического разбиения кластера  $\mathbf{X}_i^{\text{train}}$ ,  $\mathbf{X}_{ik} = \{X_{in}, n \in N_{ik} : \bigcup_n X_{in} = \mathbf{X}_i^{\text{train}}\}$  — набор фрагментов;  $\hat{\mathbf{X}}_{ik} = \{\hat{\mathbf{x}}_{in} \in X_{in}, n \in N_{ik}\}$  — набор эталонов  $k$ -го шага. Эталон  $\hat{\mathbf{x}}_{in} \in X_{in}$  выбирается из условия:

$$\hat{\mathbf{x}}_{in} = \arg \min_{\hat{\mathbf{x}} \in X_{in}} \max_{\mathbf{x} \in X_{in}} d^L(\mathbf{x}, \hat{\mathbf{x}}), \tag{27}$$

которое обеспечивает минимальное рассеяние  $D(X_{in}) = \max_{\mathbf{x} \in X_{in}} d^L(\mathbf{x}, \hat{\mathbf{x}}_{in})$  объектов фрагмента  $X_{in}$  относительно выбранного эталона  $\hat{\mathbf{x}}_{in}$ . В наборе  $\mathbf{X}_{ik}$  фрагментов  $k$ -го шага дихотомии делимым является фрагмент  $X_{in} \subset \mathbf{X}_{ik}$  с наибольшим рассеянием:

$$D(X_{in}) = \max_{\tilde{X}_{in} \subset \mathbf{X}_{ik}} D(\tilde{X}_{in}).$$

Для разбиения делимого фрагмента  $X_{in} \rightarrow (X_{i2n+1}, X_{i2n+2})$  в нем выбираются опорные объекты  $(\mathbf{x}', \mathbf{x}'') \in X_{in}$ , которые образуют пару наиболее удаленных друг от друга объектов по мере  $d^L(\mathbf{x}', \mathbf{x}'')$ . Фрагменты  $(X_{i2n+1}, X_{i2n+2})$  формируются на наборах, образованных наиболее близкими объектами к соответствующим опорным объектам  $(\mathbf{x}', \mathbf{x}'')$  по указанной мере.

Базовыми операциями над фрагментом  $X_{in} \subset \mathbf{X}_{ik}$  являются: операция формирования фрагментов следующего уровня (fragmentation)

$$f_{\text{frag}}(X_{in}) = \begin{cases} (X_{i2n+1}, X_{i2n+2}), & X_{in} \text{ — делимый;} \\ X_{in}, & X_{in} \text{ — неделимый} \end{cases}$$

и операция отбора (selection) эталона

$$f_{\text{sel}}(X_{in}) = \hat{\mathbf{x}}_{in}$$

в соответствии с условием (27). Указанные операции позволяют сформулировать рекурсивный алгоритм построения наборов эталонов в заданном классе.

**Алгоритм  $a_{\text{sel}}$ .** Начальные условия при  $k = 0$ :

$$X_{i0} = \mathbf{X}_i^{\text{train}}; \quad \mathbf{X}_{i0} = X_{i0}; \quad \hat{\mathbf{x}}_{i0} = f_{\text{sel}}(X_{i0}); \quad \hat{\mathbf{X}}_{i0} = \hat{\mathbf{x}}_{i0}; \quad m_i \geq 1.$$

Построение наборов фрагментов и наборов эталонов на шагах  $k = 1, \dots, m_i - 1$ :

$$\begin{aligned} \mathbf{X}_{ik} &= \{f_{\text{frag}}(X_{in}), \forall X_{in} \subset \mathbf{X}_{i(k-1)}\}; \\ \hat{\mathbf{X}}_{ik} &= \{f_{\text{sel}}(X_{in}), \forall X_{in} \subset \mathbf{X}_{ik}\}. \end{aligned}$$

Алгоритм  $a_{\text{sel}}$  строит на кластере  $\mathbf{X}_i^{\text{train}}$  последовательность наборов эталонов

$$a_{\text{sel}}(\mathbf{X}_i^{\text{train}}) = (\hat{\mathbf{X}}_{i0}, \dots, \hat{\mathbf{X}}_{ik}, \dots, \hat{\mathbf{X}}_{i(m_i-1)}), \quad (28)$$

в которой набор  $\hat{\mathbf{X}}_{ik} = \{\hat{\mathbf{x}}_{in}, \forall n \in N_{ik}\}$  мощности  $\hat{m}_{ik} = k + 1$  образует  $k$ -й слой в бинарном дереве. Выбор из последовательности (28) оптимального набора  $\hat{\mathbf{X}}_{ik^*} = \hat{\mathbf{X}}_i^*$  мощности  $\hat{m}_{ik^*} = \hat{m}_i^*$  осуществляется путем минимизации оценки скользящего контроля для доли ошибок вида (26).

### 4.3 Оптимизация параметров

Аддитивная форма оценки функционала ошибок, полученная в утверждении 3, сводит минимизацию доли ошибок  $\hat{\varepsilon}^{\text{train}}(\hat{\mathbf{X}}, \Delta, \mathbf{W})$  вида (25) на заданном обучающем множестве  $\mathbf{X}^{\text{train}}$  к независимой минимизации долей ошибок  $\varepsilon^{\text{train}}(\hat{\mathbf{X}}_i, \delta, \mathbf{W}_i)$  вида (26) по наборам эталонов  $\hat{\mathbf{X}}_i, i = 1, \dots, c$ , при фиксированных значениях  $\delta \in [0, 1]$ , взятых с шагом  $\Delta\delta > 0$ . Для элементарного классификатора на паре классов  $\{\omega_0, \omega_i\}$  выбор оптимального набора  $\mathbf{X}_i^\delta$  при фиксированном  $\delta$  выполняется на кластере  $\mathbf{X}_i^{\text{train}}$  своих объектов из класса  $\omega_i$ . Оптимизация набора эталонов в классе  $\omega_i$  сводится к нахождению в последовательности (28) набора

$$\hat{\mathbf{X}}_i^\delta = \arg \min_{k=0}^{m_i-1} \varepsilon^{\text{train}}(\hat{\mathbf{X}}_{ik}, \delta, \mathbf{W}_{ik}), \quad (29)$$

где  $\varepsilon^{\text{train}}(\hat{\mathbf{X}}_{ik}, \delta, \mathbf{W}_{ik})$  — доля ошибок вида (26) по всем объектам из кластеров  $\mathbf{X}_0^{\text{train}} \subset \mathbf{X}^{\text{train}}$  и  $\mathbf{X}_i^{\text{train}} \subset \mathbf{X}^{\text{train}}$ , реализуемая на наборе эталонов  $\hat{\mathbf{X}}_{ik}$  с весами  $\mathbf{W}_{ik}$  и пороговым параметром  $\delta$ .

Оптимальный параметр  $\delta^*$ , одинаковый для всех классов долей ошибок, реализуемых на наборах эталонов  $\hat{\mathbf{X}}_i^\delta, i = 1, \dots, c$ , вида (29) и равен

$$\delta^* = \arg \min_{\delta} \sum_{i=1}^c \varepsilon^{\text{train}}(\hat{\mathbf{X}}_i^\delta, \delta, \mathbf{W}_i). \quad (30)$$

Построение древовидно-структурированных наборов эталонов выполняется с помощью алгоритма  $a_{\text{sel}}$  на кластерах  $\mathbf{X}_i^{\text{train}} \subset \mathbf{X}^{\text{train}}, i = 1, \dots, c$ . Вычисление используемых в (29) и (30) долей ошибок  $\varepsilon^{\text{train}}(\hat{\mathbf{X}}_{ik}, \delta, \mathbf{W}_{ik})$  и  $\varepsilon^{\text{train}}(\hat{\mathbf{X}}_i^\delta, \delta, \mathbf{W}_i)$  элементарных классификаторов  $\{\omega_0, \omega_i\}, i = 1, \dots, c$ , выполняется по схеме «класс против всех»: кластер  $\mathbf{X}_i^{\text{train}}$  представляет объекты класса  $\omega_i$ , а множество  $\mathbf{X}^{\text{train}} \setminus \mathbf{X}_i^{\text{train}}$  — объекты класса  $\omega_0$ . При любом фиксированном  $\delta \in [0, 1]$  и значениях  $i = 1, \dots, c$  решение элементарного классификатора  $\{\omega_0, \omega_i\}$  по объектам кластера  $\mathbf{X}_i^{\text{train}}$  принимается согласно правилу

$$i^*(\mathbf{x}) = [g_i^L(\mathbf{x}) \geq \delta \bar{g}_i^L]i.$$

Здесь разделяющая функция  $g_i^L(\mathbf{x})$  вычисляется по модифицированной мере

$$\tilde{d}^L(\mathbf{x}, \hat{\mathbf{x}}) = [\mathbf{x} = \hat{\mathbf{x}}] \operatorname{mean}_{(\mathbf{x}', \mathbf{x}'') \in \mathbf{X}_i^{\text{train}}} d^L(\mathbf{x}', \mathbf{x}'') + [\mathbf{x} \neq \hat{\mathbf{x}}] d^L(\mathbf{x}, \hat{\mathbf{x}}),$$

где  $\operatorname{mean}_{(\mathbf{x}', \mathbf{x}'') \in \mathbf{X}_i^{\text{train}}} d^L(\mathbf{x}', \mathbf{x}'')$  — среднее значение меры различия  $d^L(\mathbf{x}', \mathbf{x}'')$  объектов кластера  $\mathbf{X}_i^{\text{train}}$ , а  $[f]$  — индикатор  $f$ .

Результатом операций (29) и (30) являются оптимальные наборы эталонов  $\{\hat{\mathbf{X}}_i^*\}_{i=1}^c$  с мощностями  $\{\hat{m}_i^* = \|\hat{\mathbf{X}}_i^*\|_{i=1}^c$  и оптимальные пороги  $\{\Delta_i^* = \delta^* \bar{g}_i^L\}_{i=1}^c$ .

## 5 Тестирование классификатора

### 5.1 Состав и схема эксперимента

Разработанный классификатор протестирован на составном источнике полутоновых изображений (8 бит/пиксель) подписей (40 классов по 20 изображений) [12], жестов руки (25 классов по 40 изображений) [13] и лиц (25 классов по 40 изображений) [14]. Множество изображений  $\mathbf{X}$  содержало 2800 объектов; подмножество  $\Omega_c$  включало  $c = 90$  семантически однородных классов; класс  $\omega_0$  эмулирован на объектах подмножества  $\Omega_c$ ; для представления объектов использованы деревья примитивов глубины  $L = 10$ . Результаты тестирования представлены ROC-кривыми в терминах зависимостей TPR от FPR, полученных при различных наборах эталонов по классам, и кривыми зависимости средней доли ошибок  $((1 - \text{TPR}) + \text{FPR})/2$  от вычислительной сложности  $C_\alpha$  решающего алгоритма  $a_\alpha$ , построенными для классификаторов с различными наборами эталонов и оптимизированными порогами.

Реализованы две схемы скользящего контроля, каждая из которых базируется на многократном разбиении множества объектов источника  $\mathbf{X}$  на обучающую  $\mathbf{X}^{\text{train}}$  и тестовую  $\mathbf{X}^{\text{test}}$  выборки. На каждой выборке  $\mathbf{X}^{\text{train}}$  строится классификатор: согласно (29) отбираются наборы эталонов по классам  $\hat{\mathbf{X}}_i^\delta$ ,  $i = 1, \dots, c$ , при различных  $\delta$  и согласно (30) выбирается оптимальный параметр  $\delta^*$ . Построенному классификатору предъявляется выборка  $\mathbf{X}^{\text{test}}$ , на которой регистрируются ошибочные решения по своим или чужим объектам.

Первая схема скользящего контроля использует процедуру leave-one-out для регистрации ошибочных решений по своим объектам. В этой схеме выборки  $\mathbf{X}^{\text{test}}$  образованы всевозможными одиночными объектами из подмножества классов  $\Omega_c$ ; классификаторы строятся на выборках  $\mathbf{X}^{\text{train}} = \mathbf{X} \setminus \mathbf{X}^{\text{test}}$  и содержат наборы эталонов для  $c$  классов; предъявление этим классификаторам выборок  $\mathbf{X}^{\text{test}}$  дает долю ошибок скользящего контроля  $\varepsilon_{\text{loo}}^{\text{test}}(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \Omega_c)$ , которая является оценкой величины  $\hat{\varepsilon}(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \Omega_c)$ , определенной в (24).

Вторая схема скользящего контроля использует процедуру leave-class-out для регистрации ошибочных решений по чужим объектам. В этой схеме выборки  $\mathbf{X}^{\text{test}}$  образованы одиночными кластерами  $\mathbf{X}_i$ ,  $i = 1, \dots, c$ , эмулирующими объектами из класса  $\omega_0$ ; классификаторы строятся на выборках  $\mathbf{X}^{\text{train}} = \mathbf{X} \setminus \mathbf{X}^{\text{test}}$  и содержат наборы эталонов для  $c - 1$  классов; предъявление классификаторам выборок  $\mathbf{X}^{\text{test}}$  дает долю ошибок скользящего контроля  $\varepsilon_{\text{loo}}^{\text{test}}(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \omega_0)$ , которая является оценкой величины  $\hat{\varepsilon}(\hat{\mathbf{X}}, \Delta, \mathbf{W} | \omega_0)$ , определенной в (24').

В рассмотренных схемах скользящего контроля построение классификатора на любой выборке  $\mathbf{X}^{\text{train}}$  выполняется с применением алгоритма отбора наборов эталонов по классам  $a_{\text{sel}}$  и операций оптимизации параметров, заданных соотношениями (29) и (30).

### 5.2 Результаты эксперимента

В ходе эксперимента получено параметрическое множество оценок скользящего контроля

$$\text{TPR}_{\alpha=0}(\delta) = 1 - \varepsilon_{\text{loo},\alpha=0}^{\text{test}}(\hat{\mathbf{X}}^\delta, \Delta^\delta, \mathbf{W}|\Omega_c); \quad \text{FPR}_{\alpha=0}(\delta) = \varepsilon_{\text{loo},\alpha=0}^{\text{test}}(\hat{\mathbf{X}}^\delta, \Delta^\delta, \mathbf{W}|\omega_0) \quad (31)$$

с параметром  $\delta \in [0, 1]$ , где  $\hat{\mathbf{X}}^\delta, \Delta^\delta$  — множества эталонов и порогов при фиксированном значении  $\delta$ . Множество оценок (31) дает ROC-кривую в терминах зависимости TPR от FPR. Весовые коэффициенты эталонов по классам, образующие множество  $\mathbf{W}$ , принимали значения  $\{0, 1\}$ : значение 1 для ближайшего эталона в каждом классе и значение 0 для всех других эталонов. Оценки (31) вычислены на основе решающего правила (5) с применением переборного алгоритма  $a_{\alpha=0}^{\text{es}}$  по всем разделяющим функциям наибольшего порядка  $L = 10$ . Множество порогов  $\Delta^\delta = \{\Delta_i = \delta \bar{g}_i^L\}_{i=1}^c$  строилось для двух случаев: с использованием значений  $\{\bar{g}_i^L = 1\}_{i=1}^c$ , которые порождают одинаковые пороги для всех классов, и для значений  $\{\bar{g}_i^L = g_{i1}^L\}_{i=1}^c$ , которые определяются первыми порядковыми статистиками в формуле (22) и приводят к адаптивным по классам порогам. Семейство ROC-кривых, полученных при различных схемах выбора эталонов по классам, дано на рис. 5, а. На графиках указаны площади AUC (Area Under Curve) под соответствующими кривыми.

«Наихудшую» ROC-кривую (красная кривая,  $\text{AUC}=0,8135$ ) демонстрирует классификатор с наборами эталонов по классам, заданными всеми объектами обучающей выборки (наборами наибольшей мощности), и одинаковыми порогами по классам; «наилучшую» (синяя кривая,  $\text{AUC}=0,9968$ ) — классификатор с наборами эталонов, отобранными с применением алгоритма  $a_{\text{sel}}$ , и адаптивными порогами. Для сравнения приведены две ROC-кривые для классификаторов с адаптивными порогами: с одним эталоном в каждом классе, отобранным алгоритмом  $a_{\text{sel}}$  (черная кривая,  $\text{AUC}=0,8454$ ) и с наборами эталонов наибольшей мощности (зеленая кривая,  $\text{AUC} = 0,9959$ ).

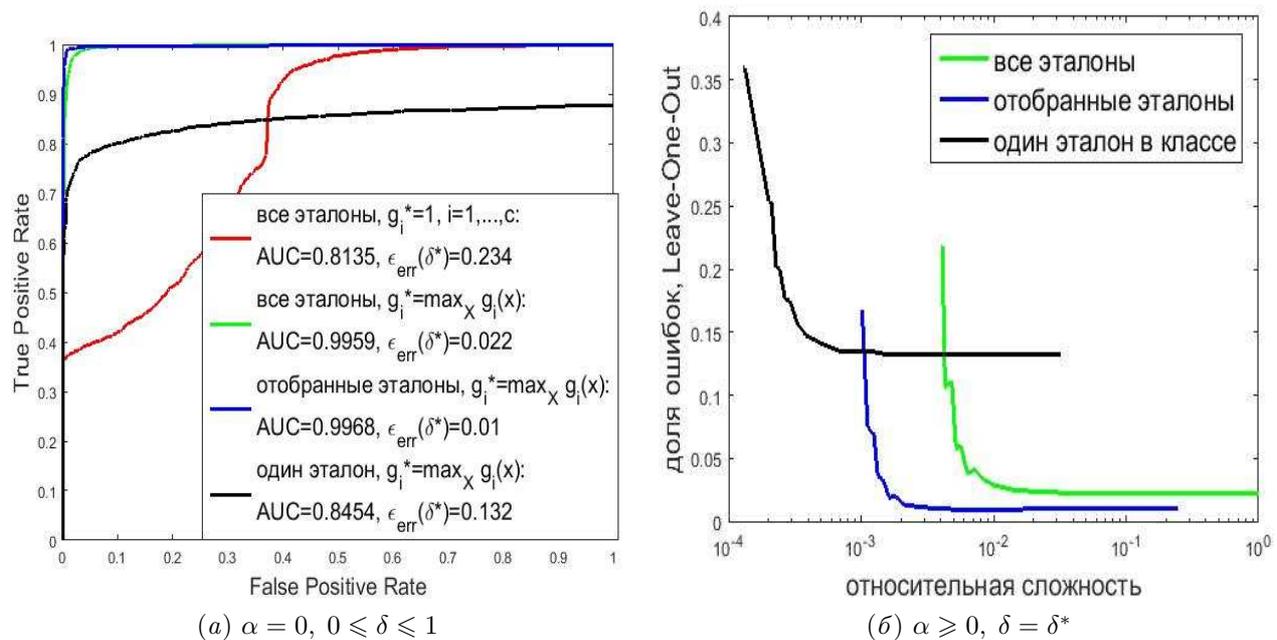


Рис. 5 ROC-кривые (а) и функции «качество–сложность» (б)

Для рассмотренных классификаторов с адаптивными порогами  $\Delta^* = \{\Delta_i = \delta^* \bar{g}_i^L\}_{i=1}^c$  (при оптимальных значениях  $\delta^*$  вида (30)) вычислены оценки долей ошибок скользящего контроля

$$\varepsilon_\alpha^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}) = \frac{1}{2} \left( \varepsilon_{\text{loo},\alpha}^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}|\Omega_c) + \varepsilon_{\text{lco},\alpha}^{\text{test}}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}|\omega_0) \right)$$

с применением решающего алгоритма  $a_\alpha$  с параметром  $0 \leq \alpha \leq 5,5$ . Здесь  $\hat{\mathbf{X}}^*$  — множество эталонов, полученное либо на наборах (29) с оптимальным параметром (30), либо на одиночных эталонах по классам, отобранных алгоритмом  $a_{\text{sel}}$ , либо на наборах наибольшей мощности, заданных соответствующими кластерами обучающей выборки. Используя оценку (18), для классификаторов с указанными параметрами вычислены оценки вычислительной сложности  $C_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})$  и доли сложности

$$\gamma_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W}) = \frac{C_\alpha(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})}{\max_{\hat{\mathbf{X}}^*} C_{\alpha=0}(\hat{\mathbf{X}}^*, \Delta^*, \mathbf{W})}.$$

Для трех классификаторов с «наилучшими» ROC-кривыми зависимости  $\varepsilon_\alpha^{\text{test}}$  от  $\gamma_\alpha$  даны на рис. 5, б. Полученные зависимости демонстрируют наименьшие доли ошибок  $\varepsilon_{\alpha,\min}^{\text{test}}$ , достигаемые на решающих алгоритмах  $a_{\alpha>0}^{\text{gs}}$  и  $a_{\alpha=0}^{\text{es}}$ , и соответствующие этим ошибкам значения вычислительного выигрыша  $1/\gamma_\alpha$  алгоритма  $a_{\alpha>0}^{\text{gs}}$  относительно алгоритма  $a_{\alpha=0}^{\text{es}}$ . Для указанных классификаторов получены следующие характеристики: при наборах эталонов наибольшей мощности (зеленая кривая) наименьшая доля ошибок  $\varepsilon_{\alpha,\min}^{\text{test}} = 0,022$ , вычислительный выигрыш  $1/\gamma_\alpha = 13,7$ ; при оптимальных наборах эталонов (синяя кривая)  $\varepsilon_{\alpha,\min}^{\text{test}} = 0,010$ ,  $1/\gamma_\alpha = 20,9$ ; при одном эталоне в каждом классе (черная кривая)  $\varepsilon_{\alpha,\min}^{\text{test}} = 0,0132$ ,  $1/\gamma_\alpha = 416,7$ . Заметное увеличение достоверности распознавания и вычислительного выигрыша достигнуто на оптимальных наборах эталонов по сравнению с наборами эталонов наибольшей мощности. Использование одиночных эталонов по классам обеспечивает существенное увеличение вычислительного выигрыша по сравнению с оптимальными наборами, но приводит к снижению качества распознавания.

## 6 Заключение

В пространстве древовидных представлений образов с многоуровневым разрешением разработан многоклассовый метрический классификатор с решающим правилом голосования наборов эталонов по классам, которое допускает отказ от классификации. Введена однопараметрическая ROC-функция в терминах зависимости доли верных решений TPR по «своим» объектам, относящимся к положительным классам, от доли ложных решений FPR по «чужим» объектам, которые не относятся к положительным классам и допускают верное решение в виде отказа от классификации. Параметром ROC-функции является параметр пороговых значений, используемых в решающем правиле. Процедура обучения базируется на построении древовидно-структурированных наборов эталонов по классам и оптимизации таких наборов на основе минимизации долей ошибок соответствующих элементарных классификаторов типа «класс против всех». Обучение включает также оценивание порогового параметра на основе минимизации средней доли ошибок  $((1 - \text{TPR}) + \text{FPR})/2$  в многоклассовой схеме. Используя сеть эталонов с многоуровневым разрешением, предложен иерархический решающий алгоритм на основе параметрической стратегии сужения зоны поиска решения на последовательных уровнях разрешения. Выявлена область значений параметра решающего алгоритма, которая обеспечивает экспоненциально растущий вычислительный выигрыш по сравнению с алгоритмом полного перебора.

Эффективность предложенного метода распознавания образов продемонстрирована на множестве изображений лиц, жестов и подписей, содержащих объекты из 90 классов. В режиме скользящего контроля эмулированы классы «своих» и «чужих» объектов. Для классификаторов с различными наборами эталонов и различными стратегиями выбора порогов построены ROC-функции и вычислены характеристики AUC. Для наборов эталонов по классам, заданных одиночными объектами, всеми объектами обучающей выборки и оптимальными наборами, отобранными путем минимизации функционала ошибок, построены соотношения «качество–сложность» в терминах зависимости доли ошибок от вычислительной сложности решающего алгоритма. Наилучшее соотношение «качество–сложность» получено для классификатора с оптимальными наборами эталонов по классам: при наименьшей доле ошибок порядка 1% иерархический решающий алгоритм демонстрирует 20-кратное увеличение быстродействия по сравнению с переборным алгоритмом.

Дальнейшее развитие предложенной модели классификации предполагает разработку и исследование многоклассовых схем с использованием наборов элементарных классификаторов и, в частности, элементарных SVM-классификаторов (Support Vector Machine), обучаемых по схеме «класс против всех». Предполагается также исследовать схемы комплексирования на основе голосования решений классификаторов по отдельным источникам (Majority Voting) и на основе обобщенной меры на ансамбле источников (General Measure).

## Литература

- [1] *Rosenfeld A.* Quadrees and pyramids: Hierarchical representation of images. — University of Maryland, Computer Science, 1982. 14 p.
- [2] *Samet H.* The quadtree and related hierarchical data structures // *ACM Comput. Surv.*, 1984. Vol. 16. No. 2. P. 187–260. doi: 10.1145/356924.356930.
- [3] *Elfiky N. M., Khan F. S., van de Weijer J., Gonzalez J.* Discriminative compact pyramids for object and scene recognition // *Pattern Recogn.*, 2012. Vol. 45. No. 4. P. 1627–1636. <http://dx.doi.org/10.1016/j.patcog.2011.09.020>.
- [4] *Torsello A., Jiang X., Ferrer M.* Editorial for the special issue on graph-based representations in pattern recognition // *Pattern Recogn. Lett.*, 2012. Vol. 33. No. 15. P. 1957. <http://dx.doi.org/10.1016/j.patrec.2012.08.016>.
- [5] *Mestetskiy L., Semenov A.* Binary image skeleton — continuous approach // 3rd Conference (International) on Computer Vision Theory and Applications Proceedings. — INSTI CC, 2008. Vol. 1. P. 251–258.
- [6] *Ganebnykh S. N., Lange M. M., Stepanov D. Y.* Metric classifier using multilevel network of templates // *Pattern Recogn. Image Anal.*, 2012. Vol. 22. No. 2. P. 265–277. doi: 10.1134/S1054661812020034.
- [7] *Pelillo M., Hidovic-Rowe D., Torsello A.* Polynomial-time metrics for attributed trees // *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005. Vol. 27. No. 7. P. 1087–1099. doi: 10.1109/TPAMI.2005.146.
- [8] *Tax D. M. J., Duijn R. P. W.* Growing a multi-class classifier with a reject option // *Pattern Recogn. Lett.*, 2008. Vol. 29. No. . P. 1565–1570. <http://dx.doi.org/10.1016/j.patrec.2008.03.010>.
- [9] *Lange M. M., Ganebnykh S. N.* An efficiency of hierarchical classification in terms of fidelity-complexity ratio // *Machine Learning Data Anal.*, 2014. Vol. 1. No. 8. P. 1126–1136. <http://jmla.org/papers/doc/2014/no8/Lange2014Efficiency.pdf>.
- [10] *Theodoridis S., Koutroumbas K.* Pattern recognition. — 4th ed. — Elsevier, 2009. 978 p. <http://www.sciencedirect.com/science/book/9781597492720>.

- [11] Duda R., Hart P., Stork D. Pattern classification. — 2nd ed. — Wiley, 2001. 680 p. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471056693.html>.
- [12] Database of signatures: First international signature verification competition. 2004. [www.cse.ust.hk/svc2004/download.html](http://www.cse.ust.hk/svc2004/download.html).
- [13] Moeslund T. Gesture recognition database. 2002. [www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html](http://www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html).
- [14] Stepanov D. Yu. Face database. — MIREA, 2013. [http://stepanovd.com/news\\_2015\\_10\\_facedben.html?lang=EN](http://stepanovd.com/news_2015_10_facedben.html?lang=EN).

Поступила в редакцию 29.07.2016

## Multiclass pattern recognition in a space of multiresolution representations\*

M. M. Lange, S. N. Ganebnykh, and A. M. Lange

lange\_mm@ccas.ru, sng@ccas.ru

Federal Research Center “Computer Science and Control” of RAS, 44/2 Vavilova st., Moscow, Russia

For a multiclass source of patterns given by images, a metric classification scheme in a space of tree-structured pattern representations is suggested. At the successive resolution levels, in a set of the pattern representations, both a family of dissimilarity measures at the successive levels and discriminant functions (class likelihoods) by the appropriate measures are defined. A decision of the multiclass classifier is made by voting the values of the discriminant functions. Also, a reject is available. A learning procedure that includes a selection of the template patterns in classes as well as an optimization of the classifier parameters is developed. A parametric decision algorithm that includes both hierarchical and exhaustive search strategies for the decision in the multilevel network of the templates is constructed. Analytical estimation of a computational complexity of the algorithm is obtained. For a composite source of the patterns given by signatures, hand gestures, and faces, an efficiency of the classifiers with different parameters is shown by the appropriate ROC curves as well as by empirical dependencies of the error rate on the computational complexity of the decision algorithm.

**Keywords:** *tree-structured representation; discriminant function; multiclass pattern recognition; multilevel network of templates; ROC-curve; hierarchical search; computational complexity*

DOI: 10.21469/22233792.2.1.06

## References

- [1] Rosenfeld, A. 1982. *Quadtrees and pyramids: Hierarchical representation of images*. University of Maryland, Computer Science. 14 p.
- [2] Samet, H. 1984. The quadtree and related hierarchical data structures. *ACM Comput. Surv.* 16(2):187–260. doi: 10.1145/356924.356930.
- [3] Elfiky, N. M., F. S. Khan, J. van de Weijer, and J. Gonzalez. 2012. Discriminative compact pyramids for object and scene recognition. *Pattern Recogn.* 45(4):1627–1636. Available at: <http://dx.doi.org/10.1016/j.patcog.2011.09.020> (accessed October 15, 2016).

---

\*The research was partially supported by the Russian Foundation for Basic Research (grants 15-01-04671, 15-07-09324, and 15-07-07516).

- [4] Torsello, A., X. Jiang, and M. Ferrer. 2012. Editorial for the special issue on graph-based epresentations in pattern recognition. *Pattern Recogn. Lett.* 33(15):1957. Available at: <http://dx.doi.org/10.1016/j.patrec.2012.08.016> (accessed October 15, 2016).
- [5] Mestetskiy, L., and A. Semenov. 2008. Binary image skeleton — continuous approach. *3rd Conference (International) on Computer Vision Theory and Applications Proceedings*. INSTICC. 1:251–258.
- [6] Ganebnykh, S.N., M.M. Lange, and D.Y. Stepanov. 2012. Metric classifier using multilevel network of templates. *Pattern Recogn. Image Anal.* 22(2):265–277. doi: 10.1134/S1054661812020034.
- [7] Pelillo, M., D.H. Hidovic-Rowe, and A. Torsello. 2005. Polynomial-time metrics for attributed trees. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(7):1087–1099. doi: 10.1109/TPAMI.2005.146.
- [8] Tax, D.M.J., and R.P.W. Duin. 2008. Growing a multi-class classifier with a reject option. *Pattern Recogn. Lett.* 29(10):1565–1570. Available at: <http://dx.doi.org/10.1016/j.patrec.2008.03.010> (accessed October 15, 2016).
- [9] Lange, M.M., and S.N. Ganebnykh. 2014. An efficiency of hierarchical classification in terms of fidelity-complexity ratio. *Machine Learning Data Anal.* 1(8):1126–1136. Available at: <http://jmla.org/papers/doc/2014/no8/Lange2014Efficiency.pdf> (accessed October 15, 2016).
- [10] Theodoridis, S., and K. Koutroumbas. 2009. *Pattern recognition*. 4th ed. Elsevier. Available at: <http://www.sciencedirect.com/science/book/9781597492720> (accessed October 15, 2016).
- [11] Duda, R., P. Hart, and D. Stork. 2001. *Pattern classification*. 2nd ed. Wiley. 680 p. Available at: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471056693.html> (accessed October 15, 2016).
- [12] Database of signatures: First international signature verification competition. 2004. Available at: [www.cse.ust.hk/svc2004/download.html](http://www.cse.ust.hk/svc2004/download.html) (accessed June 14, 2016).
- [13] Moeslund, T. 2002. Gesture recognition database. Available at: [www.prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html](http://www.prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html) (accessed June 14, 2016).
- [14] Stepanov, D. Yu. 2013. *Face database*. MIREA. Available at: [http://stepanovd.com/news\\_2015\\_10\\_facedben.html?lang=EN](http://stepanovd.com/news_2015_10_facedben.html?lang=EN) (accessed June 14, 2016).

Received July 29, 2016