

# ОЦЕНКА ТОЧНОСТИ АВТОМАТИЧЕСКОГО КОДИРОВАНИЯ ЧАСТНЫХ ПРИЗНАКОВ ИЗОБРАЖЕНИЙ ОТПЕЧАТКОВ ПАЛЬЦЕВ

**В.Ю. Гудков**

## ACCURACY RATING OF AUTOMATIC MINUTIAE RECOGNITION OF FINGERPRINTS

**V.U. Gudkov**

В статье рассматривается оценка точности автоматического распознавания частных признаков изображений отпечатков пальцев. Предлагается способ, учитывающий неравномерное распределение ошибок автоматического распознавания частных признаков по изображению. Данная оценка служит критерием для обучения алгоритмов классификационного анализа дактилоскопических изображений и выбора их функционального наполнения.

*Ключевые слова: отпечаток пальца, точность, рейтинг, неравномерное, распределение, частные признаки, распознавание.*

This paper presents a method of accuracy rating of automatic minutiae recognition. The method takes into account no uniform distribution of minutiae extraction errors. This accuracy rating serves as criterion to train the classification algorithms and to select their functionality.

*Keywords: fingerprint, accuracy, rating, no uniform, distribution, minutiae, recognition.*

### Введение

В постановке большинства задач распознавания результаты распознавания должны быть отнесены к определенным классам объектов [1, 3]. Формально это отношение задается пятеркой

$$\langle \Omega, C, W, W_t, f_c \rangle,$$

где  $\Omega$  – множество образцов;  $C = \{c_k, c^*\}$  – множество классов, причем  $c^*$  – дополнение до  $\{c_k\}$  и обозначает неизвестный класс;  $W = \{\omega_k \mid \omega_k \in \Omega\}$  – обучающая выборка образцов;  $W_t = \{\omega_k \mid \omega_k \in \Omega\}$  – тестовая выборка образцов;  $f_c$  – экспертный классификатор. Требуется найти функцию классификации  $f: \Omega \rightarrow P^{|C|}$ , которая наилучшим образом приближает  $f_c$  на тестовой выборке  $W_t$ , где вероятность

$$P^n = \{p \in R^n \mid p = (p_1, \dots, p_n), p_k \in R, \sum_{k=1}^n p_k = 1\}.$$

Оценить степень приближения можно при помощи функции ошибки, например, числа ошибочных классификаций на тестовой выборке [5, 10]. Оптимальным классификатором является тот, у которого функция ошибки минимальна [1, 8, 10]. Для стабилизации решений образцы, которые экспертно не классифицируются однозначно, удаляются из  $\Omega$ .

В задачах классификации опираются на понятие признака образца [3-5]. В дактилоскопии образцом является физический объект, например палец. Но компьютерные технологии имеют дело не с пальцами, а с их изображениями, которые получают с устройств ввода [6, 10-12]. На таких дактилоскопических изображениях (ДИ) гребни кожи обычно отображаются темными линиями с контрастно выделяющимися просветами между ними. Линии могут начинаться или оканчиваться, соединяться или расщепляться, образуя частные признаки ДИ. В компьютерных технологиях частные признаки приравниваются к признакам образца [4, 9-12].

Понятие частного признака хорошо проработано и при ручном кодировании оператор просто указывает его местоположение и ориентацию [6, 11]. Это классический случай однозначной классификации и именно так формируются выборки  $W$  и  $W_t$ . Небольшое смещение частного признака от его истинного местоположения (до одной линии) не считается за грубую ошибку, поскольку не оказывает существенного влияния на результаты идентификации [4, 12]. Однако влияние человеческого фактора при ручном кодировании частных признаков затрудняет оценку точности их автоматического распознавания. В работе [4] эта проблема частично решена. Но метод, предложенный в [4], не позволя-

ет акцентировать внимание на компактно расположенных ложных частных признаках. Именно они значительно ухудшают результаты идентификации. Действительно, след «родного» отпечатка пальца, приходящийся на фрагмент ДИ с этими ложными частными признаками, не идентифицируется [5, 6]. Поэтому необходимо обеспечить чувствительность метода оценки точности автоматического распознавания частных признаков к компактности расположения ложных частных признаков.

### 1. Метрика для сопоставления частных признаков

В суде при доказательстве идентичности двух отпечатков пальцев частные признаки являются основными признаками образца. Пусть одному и тому же образцу  $\omega$  функциями поставлены в соответствие шаблоны  $T_1 = f_1(\omega)$  и  $T_2 = f_2(\omega)$  в виде  $T_1 = \{m_i | i \in I = 1..n_1\}$ ,  $m_i = \{x_i, y_i, \theta_i\}$ , (1)  
 $T_2 = \{m_j | j \in J = 1..n_2\}$ ,  $m_j = \{x_j, y_j, \theta_j\}$ , (2)  
 где  $n_1$  – количество частных признаков на  $\omega$  при ручной кодировке ДИ функцией  $f_1$ ;  $n_2$  – количество частных признаков на  $\omega$  при автоматической кодировке ДИ функцией  $f_2$ ;  $(x, y)$  – координаты;  $\theta$  – направление как угол;  $I$  – множество индексов, причем для любого  $i \in I$  элемент  $m_i$  известен;  $J$  – множество индексов, причем для любого  $j \in J$  элемент  $m_j$  известен. Шаблоны (1) и (2) есть способ простого описания ДИ [4, 10, 12].

Введем функцию метрики

$$\mu(r, \alpha, m_i, m_j) \in \{0, 1\}, \quad (3)$$

где  $r$  – радиус как допуск на величину отклонения координат  $m_i$  и  $m_j$ ;  $\alpha$  – угол как допуск на величину отклонения углов направлений  $m_i$  и  $m_j$ ;  $m_i$  – частный признак, закодированный вручную,  $m_i \in T_1$ ;  $m_j$  – частный признак, закодированный автоматом,  $m_j \in T_2$ .

Функция метрики  $\mu$  определяет сопоставимость двух частных признаков, полученных разным образом, по формуле

$$\mu(r, \alpha, m_i, m_j) = \begin{cases} 1, & \text{если } \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r \wedge \text{turn}(\theta_i, \theta_j) \leq \alpha, \\ 0, & \text{иначе,} \end{cases}$$

где функция наименьшего угла поворота  $\text{turn}(\alpha, \beta) =$

$$= \begin{cases} \alpha - \beta, & \text{если } |\alpha - \beta| \leq 180, \\ 360 \text{sign}(\beta - \alpha) + \alpha - \beta, & \text{если } |\alpha - \beta| > 180 \end{cases}$$

и знаковая функция

$$\text{sign}(a) = \begin{cases} +1, & \text{если } a \geq 0, \\ -1, & \text{если } a < 0. \end{cases}$$

Если  $m_i$  сопоставляется с  $m_j$ , то отклонение координат  $m_i$  и  $m_j$  в допуске  $r$  и отклонение направлений  $m_i$  и  $m_j$  в допуске  $\alpha$  и  $\mu = 1$ , иначе  $\mu = 0$ .

Задавая  $r = 0$ , требуем, чтобы частные признаки ручной и автоматической кодировки располагались точка в точке. Это идеальный случай, недостижимый в действительности. Частные признаки не только смещаются, но появляются ложные и исчезают истинные [4, 6, 12]. Вопрос оценки качества признакового пространства объекта вообще актуален [2, 10], а в криминалистике в особенности из-за сильно загрязненных изображений следов отпечатков пальцев с мест преступлений. Если материал, на котором оставлен след, текстурирован, количество ложных частных признаков может возрастать и группироваться. Функция метрики  $\mu$  это не учитывает. Она выносит булево решение о сопоставимости только двух частных признаков, выбранных из шаблонов  $T_1$  и  $T_2$ , на основе указанных допусков, причем шаблон  $T_1$  считается эталонным. Однако это позволяет построить оценку, учитывающую как точность распознавания частных признаков [4], так и компактность расположения ложных частных признаков.

### 2. Оценка точности

Оценка точности автоматического распознавания частных признаков основана на функции массовой ранговой вероятности. Зададим допуск  $\alpha$  константой и ранжируем по допуску  $r$  сопоставляемые по (3) частные признаки  $(m_i, m_j)$ , которые назовем парой. Для допуски  $r$  количество сопоставляемых пар  $(m_i, m_j)$  находится как

$$n(r, \alpha) = \sum_{\substack{m_i \in T_1 \\ m_j \in T_2}} \mu(r, \alpha, m_i, m_j), \quad (4)$$

где  $T_1$  и  $T_2$  – шаблоны по (1) и (2).

**Определение 1.** Рангом пары  $r(m_i, m_j)$  называется минимальная величина допуски  $r \in 1..r_{\max}$ , при которой эта пара сопоставляется.

Количество пар ранга  $r(m_i, m_j)$  находится в виде  $\delta(r) = n(r, \alpha) - n(r - 1, \alpha)$ .

**Определение 2.** Функцией массовой ранговой вероятности  $P(r)$  (МРВ) называется вероятность того, что ранг пары  $r(m_i, m_j)$ .

Шаблоны  $T_1$  и  $T_2$  могут образовать не более  $\min(n_1, n_2)$  сопоставляемых пар. С позиций шаблона  $T_i$ ,  $i \in \{1, 2\}$ , функция МРВ находится в виде

$$P_i(r) = \frac{\delta(r)}{n_i}. \quad (5)$$

Обычно в шаблонах  $n_1 \neq n_2$ . Тогда первичный классификатор на основе функции МРВ (5) строится в виде

$$\hat{f}_c(r) = P_1(r)P_2(r) = \frac{\delta^2(r)}{n_1 n_2}, \quad (6)$$

где произведение в знаменателе (6) максимизирует  $\hat{f}_c(r)$  при  $n_1 \equiv n_2$ . Фактически величина  $\hat{f}_c(r)$  уменьшается при неравенстве количества частных признаков в шаблонах.

Оценку точности автоматического распознавания частных признаков как алгоритмический метод [7] построим следующим образом.

Зададим допуск  $r=1$ . Рассчитаем (6), а сопоставляемые по (3) частные признаки  $(m_i, m_j)$  удалим из шаблонов:  $T_1 = T_1 \setminus \{m_i \mid i \in I\}$  и  $T_2 = T_2 \setminus \{m_j \mid j \in J\}$ . Выберем частные признаки из шаблона  $T_2$  в качестве вершин полного графа  $G_r$ , количество вершин которого  $p(G_r) = |T_2|$ . В графе все вершины связаны. Рассчитаем сумму длин ребер в графе  $G_r$  от частного признака  $m_s \in T_2 \mid s \in J$  в виде

$$L_r(s) = \sum_{j \in J} l_{sj}, \quad (7)$$

где

$$l_{sj} = \sqrt{(x_s - x_j)^2 + (y_s - y_j)^2}.$$

Для исходного графа  $|T_2| = n_2$  и  $L_r(s) = L_0(s)$ .

Инкрементируем допуск  $r$  и рассчитаем  $\hat{f}_c(r)$  по (6), удаляя сопоставляемые частные признаки из шаблонов  $T_1$  и  $T_2$ . Выберем частные признаки из шаблона  $T_2$  в качестве вершин полного графа  $G_r$  и рассчитаем  $L_r(s)$  по (7). Последовательно применяя инкрементацию допуска  $r$ , оценим точность автоматического распознавания частных признаков на образце  $\omega_k$ , являющимся  $k$ -м по счету в тестовой выборке  $W_t$ , при помощи ошибки вероятностной массы по формуле

$$e_k = \max_s \sum_{r=1}^{r_{\max}} \exp\left(-\frac{L_r(s)}{L_0(s)}\right) \hat{f}_c(r) \frac{r}{r_{\max}}, \quad (8)$$

где  $r_{\max}$  – максимальный допуск на величину отклонения координат (до 15). В ошибке вероятностной массы  $e_k$  допуск  $\alpha$  задан константой, а величина ошибки отражает взвешенную вероятность несовпадения шаблонов  $T_1$  и  $T_2$  для образца  $\omega_k$ . Метрика  $\mu$  по (3) устанавливает критерий, по которому частные признаки  $m_i$  и  $m_j$  определяются сопоставляемыми. Чем ближе они располагаются, тем меньше величина  $e_k$ , так как уменьшается величина  $\hat{f}_c(r)$ . Заметим, что (6) не зависит от того, какое множество  $T_1$  или  $T_2$  является действительно полученным автоматом. Множитель  $r/r_{\max}$  в формуле (8) дополнительно увеличивает

чувствительность  $e_k$  к удаленным друг от друга частным признакам из  $T_1$  и  $T_2$ , а отношение сумм длин ребер по (7) обеспечивает чувствительность к компактному расположению ложных частных признаков. Действительно, для малых допусков  $r$  выполняется  $L_r(s) \approx L_0(s)$ , а с увеличением  $r$  сумма длин ребер  $L_r(s)$  снижается. Однако компактно расположенные ложные частные признаки не сопоставляются и можно найти такой частный признак  $m_s$ , который минимизирует  $L_r(s)$  (как центр тяжести). Это позволяет выбором наиболее «неудобного» частного признака  $m_s \mid s \in J$  максимизировать ошибку вероятностной массы  $e_k$ .

Точность автоматического распознавания частных признаков на тестовой выборке образцов  $W_t = \{\omega_k \mid \omega_k \in \Omega\}$ , ранжированной по величине ошибки вероятностной массы  $e_k$ , можно оценить при помощи суммарной ошибки

$$e = \frac{1}{|W_t|} \sum_{\omega_k \in W_t} e_k, \quad (9)$$

где  $|W_t|$  – мощность тестовой выборки образцов. Величина допуска  $\alpha$  фиксирована (30–45). На рис. 1 показана кривая ошибки вероятностной массы (8). Образцы  $\omega_k$  ранжированы по величине  $e_k$ . Старшим индексам  $k$  соответствуют ДИ с успешно распознанными частными признаками. При удалении хороших изображений выделяется множество «плохих лидеров». Ограничивая тестовую выборку образцов  $W_t$  множеством «плохих лидеров», можно определить индекс  $k'$ , начиная с которого оценка (9) наиболее критична, а изображения представляют наибольший интерес для развития методов классификационного анализа.

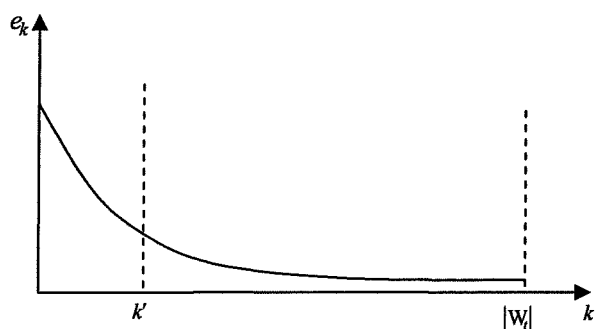


Рис. 1. Кривая ошибки вероятностной массы

#### Выводы

Оценка точности автоматического распознавания частных признаков ДИ на основе МРВ (5) и суммарной ошибки (9) реализована в рамках автоматизированной дактилоскопической информационной системы СОНДА и использована для подготовки к тестам, проводимым независимыми международными организациями. Обучающая и тестовая

вая выборки образцов сформированы вручную на основе изображений, полученных от этих организаций в качестве образцов для пробных предварительных тестов.

Оценки (8) и (9) показали высокую эффективность при подготовке к международным тестам. Они служат удобным средством для предварительного обучения алгоритмов распознавания частных признаков, по которым выполняется идентификация изображений отпечатков пальцев. Выявлена зависимость качества обучения алгоритмов и выбора их функционального наполнения от мощности обучающей выборки образцов.

### Литература

1. Аркадьев, А.Г. Обучение машины классификации объектов /А.Г. Аркадьев, Э.М. Браверман. — М.: Наука, 1971. -192 с.

2. А.с. 1652984 СССР, МКИ G 06 K 9/00. Способ формирования признаков при распознавании изображений объектов /Г.Е. Баскин, В.И. Гордиенко, Л.С. Королюк, Б.П. Русын. - № 4468868/24; заявл. 01.08.88; опубл. 30.05.91, Бюл. №20.-9 с.

3. Боннер, Р.Э. Некоторые методы классификации / Р.Э. Боннер; под ред. Э.М. Бравермана// Автоматический анализ сложных изображений. — М.: Мир, 1969. - С. 209-234.

4. Гудков, В.Ю. Оценка точности автоматического кодирования особенностей отпечатков пальцев / В.Ю. Гудков; под ред. член-корр. РАН В.Л. Арлазорова и д.т.н. проф. Н.Е. Емельянова// Обработка изображений и анализ данных: Труды ИСА РАН. - М.: Книжный дом «ЛИБРОКОМ», 2008. - Т. 38. - С. 357-363.

5. Гуревич, И.Б. Проблемы распознавания изображений: распознавание, классификация, прогноз: математические методы и их применение / КБ. Гуревич. -М.: Наука, 1982. -Вып. 1.-237 с.

6. Дактилоскопическая экспертиза: современное состояние и перспективы развития / В.Е. Корноухов, В.К. Анциферов, Г.П. Морозов и др.; под ред. ГЛ. Грановского. - Красноярск: Изд-во Красноярского университета, 1990. -416 с.

7. Журавлев, Ю.И. Об алгоритмических методах в задачах распознавания и классификации: распознавание, классификация, прогноз: математические методы и их применение / Ю.И. Журавлев. -М: Наука, 1989. - Вып. 1. - С. 9-16.

8. Кирилов, С.Н. Оптимизация признакового пространства в задачах распознавания элементов речи/С.Н. Кирилов, А.С. Шелудяков//Автоматика и телемеханика. -1998. -№5.-С. 157-162.

9. Орлов, П.Г. Статистический метод оценки идентификационных признаков в дактилоскопической экспертизе: применение теории вероятностей и математической статистики в судебной экспертизе: материалы к теоретической конференции/П.Г. Орлов. —М., 1963. -С. 22—28.

10. Bolle, KM. Guide to biometrics /KM. Bolle, J. Y. Connel, S. Pankanti, N.K. Ratha. - New York: Springer-Verlag, 2004. - 368p.

11. Komarinski, P. Automated fingerprint identification systems / P. Komarinski. - New York: Academic Press, 2005. -295 p.

12. Maltoni, D. Handbook of fingerprint recognition / Davide Maltoni, Dario Maio, Anil K. Jain. - New York: Springer-Verlag, 2003. - 348p.

Поступила в редакцию 27 апреля 2009 г.