ЗАЩИЩЕННАЯ БИОМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА. II. Качество информационно-математического обеспечения

Protected Biometrical Identification in Accept Control System II. Quality of Information-mathematical Support

Ключевые слова: биометрическая идентификация – biometrical identification; технология – technology; защищенность – security; достоверность – trustworthiness; информационно-математическое обеспечение – information-mathematical support; качество – quality.

Оценивается качество информационно-математического обеспечения новой – защищенной технологии биометрической идентификации в системах контроля доступа. Приведены результаты натурно-имитационного моделирования.

Quality of Information-mathematical support of New – Protected Biometrical Identification in Accept Control System is estimated. Natural-Imitation Modeling Results are presents.

ЛОВЦОВ / LOVTSOV D.
Дмитрий Анатольевич
(dal-1206@mail.ru)
заслуженный деятель науки РФ,
доктор технических наук, профессор,
заместитель по научной работе директора Института точной механики и вычислительной техники им. С.А. Лебедева Российской академии наук,
Москва

КНЯЗЕВ / KNYAZEV K.
Кирилл Владимирович
(dal@ipmce.ru)
аспирант Института точной механики и вычислительной техники им. С.А. Лебедева Российской академии наук,
Москва

Для оценки качества разработанного информационно-математического обеспечения (ИМО) [1] и, в целом, достоверности новой – защищенной технологии биометрической (дактилоскопической) идентификации используются численные показатели, характеризующие наличие ошибок первого рода («не принятие верной гипотезы») и второго рода («принятие неверной гипотезы»).

Ошибка первого рода случается при сравнении «*свой к своему*», когда «*свой*» признается в системе контроля доступа (СКД) «*чужим*», и оценивается вероятностью ошибочного отказа «*своему*» – *FRR* (*False Rejection Rate* – уровень ошибочного отказа). Ошибка второго рода случается при сравнениях «*чужой к чужому*», когда «чужой» признается «*своим*», и оценивается вероятностью пропуска «*чужого*» – *FAR* (*False Acceptance Rate* – уровень ошибочного одобрения). Для комплексной оценки алгоритма сравнения используется параметр *EER* (*Equal Error Rate* – равный уровень ошибок), характеризующий общий уровень неверных решений в биометрической СКД, при котором *FAR* и *FRR* равны.

Получение количественных оценок *достоверности* (вероятности ошибок) и *оперативности* работы ИМО возможно путем его тестирования в СКД и построения графиков так называемой *ROC*-кривой (*Receiver Operator Characteristic* – характеристика работы классификатора) [2].

Для тестирования необходимо заранее подготовить тестовую базу данных и знаний (БДЗ), включающую общее число отпечатков-«образцов», равное $n \times m$, где n – число разных пальцев, m – число вариантов отпечатков каждого пальца, а также правила и операторы взаимных преобразований квадратных и треугольных матриц. От ее мощности и размера зависит точность определения *ROC*-кривых. При создании тестовой БДЗ следует учитывать то, что использование синтезированных отпечатков не позволит получить реальную картину качества работы алгоритмов.

Следовательно, возникает необходимость набора больших тестовых БДЗ с реальными отпечатками разного типа. Для упрощения этой процедуры возможно применение ряда оригинальных алгоритмов, позволяющих значительно сократить объемы БДЗ.

Например, для получения статистики ошибок первого рода необходимо произвести сравнение попарно между отпечатками одного ряда – для обеспечения сравнений типа «*свой к своему*». Первый отпечаток в ряду сравнивается со всеми другими отпечатками ряда, и получается $(m-1)$ сравнений; второй отпечаток в ряду сравнивается со всеми отпечатками, идущими после него, поскольку он уже сравнивался с первым отпечатком, и получается $(m-2)$ сравнений и т.д.; предпоследний отпечаток сравнивается только с последним отпечатком, и получается 1 сравнение. Таким образом, число сравнений в ряду составит:

$$V_i = (m-1) + (m-2) + K + 1 = \frac{m(m-1)}{2} \quad (1)$$

Если число рядов n, тогда возможное число сравнений «*свой к своему*» в БДЗ размера nxm будет равно:

$$V_{FRR} = n\frac{m(m-1)}{2} \quad (2)$$

Для получении статистики ошибок второго рода необходимо произвести сравнения попарно между отпечатками разных рядов – для обеспечения сравнений типа «*чужой к чужому*». Первый отпечаток первого ряда сравнивается со всеми отпечатками всех остальных рядов, и получается $(n-1)$xm сравнений; также сравнивается второй отпечаток первого ряда, и получается еще $(n-1)$xm сравнений. После сравнений m отпечатков первого ряда со всеми отпечатками других рядов получаем $m^2(n-1)$ сравнений. Отпечатки второго ряда сравниваются с отпечатками всех $(n-2)$ рядов после него, поскольку они уже сравнивались с отпечатками первого ряда, и получается еще $m^2(n-2)$ сравнений. Указанная процедура осуществляется до предпоследнего ряда, который сравнивается уже только с единственным, последним, рядом, и получается еще m^2 сравнений. Это значит, что число возможных сравнений «*чужой к чужому*» в БДЗ размера nxm будет равно:

$$V_{FAR} = m^2[(n-1) + (n-2) + K + 1] = m^2\frac{n(n-1)}{2} \quad (3)$$

Для понимания сути ошибок первого и второго рода в ходе биометрической идентификации рассмотрим четырехпольную таблицу сопряженности (табл. 1) [2], которая строится на основе сопоставления экспериментальных результатов классификационной работы ИМО и реальной ситуации и включает:

- T_P (*True Positives* – истинно позитивные случаи) – множество верно классифицированных положительных ситуаций;
- T_N (*True Negatives* – истинно отрицательные случаи) – множество верно классифицированных отрицательных ситуаций;
- F_N (*False Negatives* – ложно отрицательные случаи) – множество положительных ситуаций, классифицированных как отрицательные – ошибка первого рода или так называемый «*ложный пропуск*», когда интересующее нас событие ошибочно не обнаруживается;
- F_P (*False Positives* – ложно позитивные случаи) – множество отрицательных ситуаций, классифицированных как положительные – ошибка второго рода или так называемая «*ложная тревога*» (ложное обнаружение), так как при отсутствии события ошибочно выносится решение о его наличии.

Таблица 1

Классы сопряженности

Результаты ИМО	**Реальная ситуация**	
	Положительно	Отрицательно
Положительно	T_P	F_P
Отрицательно	F_N	T_N

При анализе чаще оперируют не абсолютными показателями, а относительными – долями, выраженными в процентах, такими, как:

1) доля истинно позитивных ситуаций (*True Positives Rate*):

$$T_{PR} = \frac{T_P}{T_P + F_N} \cdot 100\%$$

2) доля ложно позитивных ситуаций (*False Positives Rate*):

$$F_{PR} = \frac{F_P}{T_N + F_P} \cdot 100\%$$

При этом объективная *ценность* ИМО как бинарного классификатора определяется так называемыми чувствительностью и специфичностью.

Чувствительность (*sensitivity*) – это доля истинно позитивных ситуаций:

$$S_e = T_{PR} = \frac{T_P}{T_P + F_N} \cdot 100\% \quad (4)$$

Специфичность (*specificity*) – это доля истинно отрицательных ситуаций, которые были правильно идентифицированы ИМО:

$$S_p = \frac{T_N}{T_N + F_P} \cdot 100\% = 100\% - F_{PR} \quad (5)$$

Достоверность идентификации определяется при этом как:

$$D = \frac{S_e S_p (T_P + T_N)}{S_p T_P + S_e T_N} \cdot 100\%. \quad (6)$$

ИМО с высокой чувствительностью часто дает истинный результат при наличии позитивного исхода (обнаруживает позитивные ситуации). Наоборот, ИМО с высокой специфичностью чаще дает истинный результат при наличии отрицательного исхода (обнаруживает отрицательные ситуации).

Для проведения *натурно-имитационного* эксперимента с разработанным ИМО СКД были использованы реальные отпечатки пальцев 18 человек (1800 отпечатков), распределенные между тремя БДЗ различной мощности, содержащими матрицы отпечатков пальцев: B_1 (300 матриц), B_2 (500), B_3 (1000). При этом с каждого пальца снимались 10 отпечатков (в общем случае, с углом поворота $\chi = \pm 15^{\circ}$).

В результате первичной алгоритмической обработки отпечатков в БДЗ получены топологические изображения на плоскости определенного набора k (как правило, k = 45 - 50) минуций (рис. 1). После их нумерации и соединения «*каждая с каждой*» формируется матрица:

$$B_{[m \times m]} = \left\| d'_{ij} \right\|, i = \overline{1, M}, j = \overline{1, M}$$

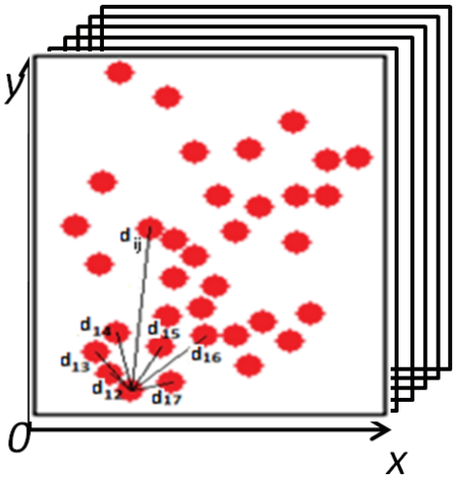


Рис. 1. Топология распределения минуций после первичной обработки изображения отпечатка пальца

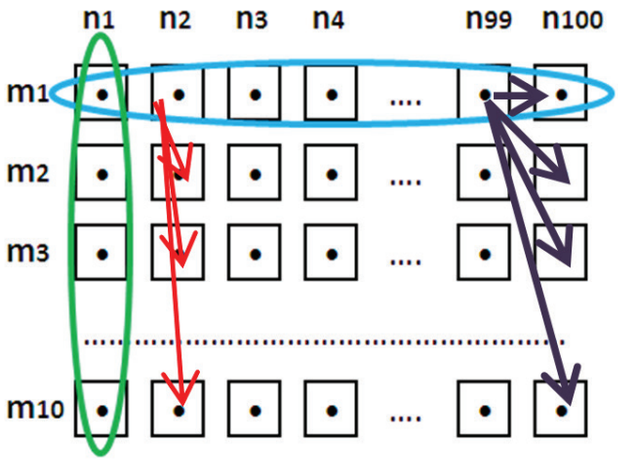


Рис. 2 Принципы сравнения отпечатков

Таблица 2

Количество сравнений по базам данных

БДЗ	*B 1*	*B 2*	*B 3*
Количество отпечатков	300	500	1000
V_{FRR}	1 350	2 250	4 500
V_{FAR}	43 500	122 500	495 000
$V=V_{FRR}+V_{FAR}$	**44 850**	**124 750**	**499 500**

$$B = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1m} \\ d_{21} & d_{22} & \cdots & d_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mm} \end{pmatrix} = \|d_{ij}\|_{i=1,j=1}^{m,m} =$$

$$= \begin{pmatrix} 0 & 12 & \cdots & 10 \\ 12 & 0 & \cdots & 8 \\ \vdots & \vdots & \ddots & \vdots \\ 10 & 8 & \cdots & 0 \end{pmatrix}_{[50\times50]}$$

Для сравнения отпечатков пальцев путем численного сравнения матриц [3] формируются матрицы $A_{i,j}$, $i = 1,...,100$; $j = 1,..,10$ размера $[kxk]$ для каждого j-го отпечатка (m_j) i-го пальца (n_i) по принципам (рис. 2): «*свой к своему*» (выделено красным) и «*чужой к чужому*» (выделено синим).

Расчет количества возможных сравнений (табл. 2) для количественной оценки вероятностей *FAR* и *FRR* проводится следующим образом (для каждой БДЗ):

1. Каждая матрица $A_{i,j}$ сравнивается с оставшимися матрицами отпечатков того же пальца $A_{i,l}$, $l = 1,..,9$; $j \neq l$. Число проведенных сравнений V_{FRR} рассчитывается по формуле (2).

2. Каждая матрица $A_{i,j}$ сравнивается с матрицами отпечатков $A_{i,l}$, $l = 1,..,99$; $i \neq l$. Число проведенных сравнений V_{FAR} рассчитывается по формуле (3).

Использование такого способа позволяет получать достаточно большое количество сравнений, необходимых для оценки вероятностей *FAR* и *FRR* и достоверности идентификации в целом. Имея тестовые БДЗ отпечатков и определив возможное количество сравнений (табл. 3), можно определить количество и вероятности ошибок первого и второго рода.

На основании полученных экспериментальных данных рассчитаны по формулам (4) – (6) соответ-

Таблица 3

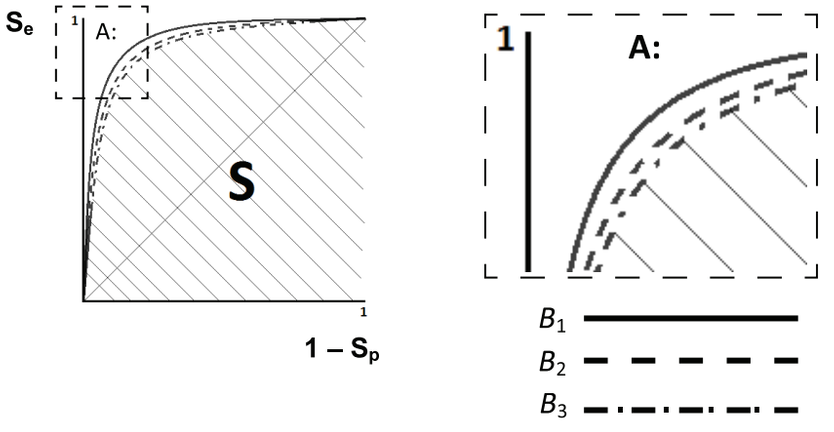
Сопряженность для различных баз данных

Результаты ИМО	Реальная ситуация					
	B 1		*B 2*		*B 3*	
	Положительно	Отрицательно	Положительно	Отрицательно	Положительно	Отрицательно
Положительно	1 247	508	1 986	4 013	3 952	21 685
Отрицательно	103	42 992	264	118 487	548	473 315
Кол-во сравнений	1 350	43 500	2 250	122 500	4 500	495 000
$V=V_{FRR}+V_{FAR}$	**44 850**		**124 750**		**499 500**	

Таблица 4

Результаты эксперимента

Показатели	*B 1*	*B 2*	*B 3*
Dmax (достоверность)	0,986	0,966	**0,955**
Se (чувствительность)	0,924	0,883	**0,878**
1 - Sp (специфичность)	0,988	0,967	**0,956**
S (площадь под кривой)	0,932	0,886	**0,871**



а) графики ROC-кривых *б) фрагмент A:*

Рис. 3. Графическая иллюстрация качества ИМО

Таблица 5

Экспертная шкала

Площадь под кривой	Качество ИМО
0,9 – 1,0	Отличное
0,8 – 0,9	**Очень хорошее**
0,7 – 0,8	Хорошее
0,6 – 0,7	Среднее
0,5 – 0,6	Неудовлетворительное

ствующие чувствительность S_e и специфичность S_p разработанного ИМО и достоверность D_{max} идентификации (табл. 4), а также построены *ROC*-кривые для трех БДЗ $B_{i,i} = 1,2,3$ (рис. 3).

На практике используется специальная экспертная шкала [2] для значений площади S под *ROC*-кривой, по которым можно судить о качестве применяемого ИМО СКД (табл. 5).

Согласно табл. 5 качество разработанного ИМО СКД для заданных имитационных БДЗ оценивается как «*очень хорошее*».

Таким образом, предложенная технология защищенной биометрической (дактилоскопической) идентификации как упорядоченная совокупность элементов ИМО, используемая в системах контроля доступа в условиях информационного соперничества, согласно результатам натурно-имитационного моделирования характеризуется следующими показателями:

– *достоверность* (D) идентификации повышается в среднем на 15-20% по сравнению с традиционной (за счет применения разработанных алгоритмов, обеспечивающих, *во-первых*, более высокое качество обработанных изображений, *во-вторых*, использование одного порога различимости вместо двух традиционных, и *в-третьих*, проведение итогового сравнения после удаления ложных минуций);

– *защищенность* ($I\{A, B\} = 0$) эталонных данных на временном интервале функционирования СКД обеспечивается в результате применения нового способа их хранения в труднообратимом преобразованном виде (в виде матрицы расстояний между минуциями);

– *оперативность* (T) остается в пределах допустимых для СКД значений благодаря компенсации повышенных (за счет применения более сложных алгоритмов) временных затрат на первичную обработку изображений при реализации более оперативного алгоритма итогового сравнения;

– *ресурсоемкость* (Q) также остается в пределах допустимых для компьютерных рабочих станций значений при незначительном увеличении (за счет применения более сложных алгоритмов) объемов хранящихся данных.

Литература

1. *Ловцов Д.А., Князев К.В.* Защищенная биометрическая идентификация в системах контроля доступа. I. Математические модели и алгоритмы // «Информация и Космос». – 2013. – №1. – С. 100-103.

2. Http://www.basegroup.ru/library/analysis/regression/logistic/ – *Н. Паклин*. Логистическая регрессия и *ROC*-анализ – математический аппарат.

3. *Рыканов А.С.* Анализ методов распознавания отпечатков пальца // Системи обробки інформації. – 2010. – Вип. 6 (87). – С. 167-170.