

## МЕТОД ОЦЕНКИ ОТНОСИТЕЛЬНОЙ ЭНТРОПИИ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИСТОЧНИКОВ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ

*И.Д. ГОРБЕНКО, И.В. ОЛЕШКО*

В работе развивается подход к измерению информации, содержащейся в биометрической характеристике. Показывается, что особенности биометрической информации могут быть рассчитаны с помощью относительной энтропии. Производится сравнительный анализ источников биометрической информации.

*Ключевые слова:* биометрическая информация, относительная энтропия, биометрическая идентификация, матрица ковариации, метод главных компонент, Гауссово распределение.

### ВВЕДЕНИЕ

Идентификация на основе биометрических данных – это средство автоматического опознавания личности на базе уникальных физических или поведенческих параметров. На сегодняшний день биометрические технологии идентификации личности получили широкое распространение в различных областях обеспечения безопасности: криминалистика; системы контроля доступа; системы идентификации личности; системы электронной коммерции; информационная безопасность (доступ в систему, авторизация на ПК); системы голосования, проведения электронных платежей; проекты государственной идентификации (пересечение границ, выдача виз) и т.д. Существует огромное количество методов биометрической идентификации. Актуальной остается проблема выбора того или иного метода. Сравнительный анализ методов биометрической аутентификации чаще всего производится на основании ошибок первого и второго рода. Использование относительной энтропии как критерия эффективности идентификации делает возможным не только сравнение биометрических признаков между собой, но и с персональным идентификационным номером (ПИН), паролем и другими методами аутентификации. Для сравнения систем биометрической аутентификации на основании критерия относительной энтропии необходимо определить количество информации каждой системы отдельно. Определим термин биометрическая информация следующим образом: уменьшение неопределенности идентичности человека за счет измерения набора биометрических характеристик. Основываясь на этом определении, в статье рассматривается алгоритм для измерения биометрической информации с помощью критерия относительной энтропии. Затем мы используем этот алгоритм для сравнительного анализа биометрической информации, содержащейся в различных алгоритмах распознавания по лицу и по радужной оболочке глаза.

### 1. АЛГОРИТМ ДЛЯ ВЫЧИСЛЕНИЯ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ

Алгоритм для вычисления биометрической информации с использованием относительной

энтропии заключается в следующих шагах [1]: 1) выдвижение требований; 2) вычисление относительной энтропии биометрических характеристик; 3) Гауссова модель для вычисления биометрических характеристик и относительной энтропии; 4) методы регуляризации для вырожденных характеристик; 5) методы регуляризации для неполных данных.

Требования к особенностям биометрической информации:

1. Если распределение характеристик человека  $p$  равно распределению характеристик между людьми  $q$ , тогда нет информации, которая отличает человека и тогда биометрическая характеристика информации равна 0.

2. Если измерение особенности становится более точным, легче определить кого-то в популяции и биометрическая информация увеличивается.

3. Если у человека необычное значение характеристики (далекое от значения популяции), он становится более различимым, и информация о его биометрической характеристике увеличивается.

4. Биометрическая информация о некоррелированных характеристиках равна сумме биометрических информационных этих характеристик.

5. Характеристики, которые не связаны с определением идентичности, не повышают биометрическую информацию.

6. Биометрические характеристики, такие как вес и рост менее информативны.

Наиболее подходящая мера для характеристики биометрической информации – относительная энтропия  $D(p\|q)$ , где  $p(x)$  и  $q(x)$  распределения биометрических характеристик человека и населения соответственно.  $D(p\|q)$  или расстояние Кульбака–Лейблера определяется как “дополнительные биты” информации необходимые для представления  $p(x)$  относительно  $q(x)$ .  $D(p\|q)$  определяется как:

$$D(p\|q) = \int_x p(x) \log_2 \frac{p(x)}{q(x)} dx. \quad (1)$$

Эту формулу можно объяснить следующим образом: относительная энтропия  $D(p\|q)$

является дополнительной информацией, необходимой для описания распределения  $p(x)$  на основании предполагаемого распределения  $q(x)$ .  $D(p\|q)$  отличается от энтропии  $H(p)$ , которая является информацией, необходимой, в среднем, для описания особенности  $x$  с распределением  $p(x)$ .  $H$  является не подходящей величиной для измерения особенностей биометрической информации, т.к. она не учитывает, в какой степени каждый признак может идентифицировать человека. К примеру, характеристика, не относящаяся к идентичности – направление человеческого лица.

В общей биометрической системе, биометрические характеристики  $F$  измеряются для создания вектора биометрических характеристик  $(F \times 1)$  для каждого человека. Для человека  $p$  мы имеем  $N_p$  образцов характеристик, в то время как для населения –  $N_q$  образцов характеристик. Для удобства обозначения, мы сортируем измерения человека  $p$  так, чтобы они были первой группировкой населения. Определив  $x$  как значение случайной величины  $X$ , мы вычисляем среднее значение характеристик населения  $\mu_q$ :

$$\mu_q = E[X] = \frac{1}{N_q} \sum_{i=1}^{N_q} x_i. \quad (2)$$

Среднее значение характеристик человека  $p$ ,  $\mu_p$ , определяется аналогично, заменяя  $q$  на  $p$ .

Матрица ковариации характеристик населения  $\Sigma_q$  определяется следующим образом:

$$\Sigma_q = \frac{1}{N_q - 1} \sum_{i=1}^{N_q} (x_i - \mu_q)^t (x_i - \mu_q). \quad (3)$$

Матрица ковариации характеристик человека  $\Sigma_p$  определяется аналогично. Одной из важных общих сложностей с прямым измерением теоретической информацией является пригодность данных. Распределение трудно оценить точно, особенно на концах; для небольших  $p(x)$  и  $q(x)$ ,  $\log_2(p(x)/q(x))$  будет иметь большое абсолютное значение. Обычно в таких случаях выполняется переход к модели с небольшим числом параметров. Гауссово распределение является наиболее общей моделью. На основании Гауссовой модели и соответствующих  $p$  и  $q$ , вычислим распределения биометрических характеристик человека и населения:

$$p(x) = \frac{1}{\sqrt{|2\pi\Sigma_p|}} \exp\left(-\frac{1}{2}(x - \mu_p)^t \Sigma_p^{-1} (x - \mu_p)\right), \quad (4)$$

$$q(x) = \frac{1}{\sqrt{|2\pi\Sigma_q|}} \exp\left(-\frac{1}{2}(x - \mu_q)^t \Sigma_q^{-1} (x - \mu_q)\right). \quad (5)$$

Из приведенных выше формул вычислим  $D(p\|q)$ :

$$\begin{aligned} D(p\|q) &= \int p(x)(\log_2 p(x) - \log_2 q(x))dx = \\ &= -k(\ln |2\pi\Sigma_p| - \ln |2\pi\Sigma_q| + 1 - \end{aligned}$$

$$\begin{aligned} &-E[(x - \mu_q)^t \Sigma_q^{-1} (x - \mu_q)]) = \\ &= k(\ln \frac{|2\pi\Sigma_q|}{|2\pi\Sigma_p|} + \text{trace}((\Sigma_p + T)\Sigma_q^{-1} - I)) \quad (6) \end{aligned}$$

где  $T = (\mu_p - \mu_q)^t (\mu_p - \mu_q)$  и  $k = \log_2 \sqrt{e}$ . Это выражение вычисляет относительную энтропию для Гауссового распределения биометрических характеристик человека и населения. Оно соответствует большинству требований для измерения особенностей биометрической информации, представленных выше:

1. Если распределение характеристик человека соответствует распределению характеристик населения ( $p=q$ ), то  $D(p\|q)=0$ , как и требовалось.

2. По мере улучшения измерения особенностей, значения матрицы ковариации  $\Sigma_p$  будут уменьшаться, приводя к уменьшению  $|\Sigma_p|$  и увеличению  $D(p\|q)$ .

3. Если человек имеет значения характеристик далекие от средних по населению,  $T$  будет больше, и как результат  $D(p\|q)$  будет больше.

4. Комбинация некоррелированных векторов характеристик дает сумму измерений человека  $D(p\|q)$ .

5. Добавление особенностей, не относящихся к идентичности, не будет изменять  $D(p\|q)$ .

6. Коррелированные характеристики менее информативны, чем некоррелированные. Они будут уменьшать  $\Sigma_p$  и  $\Sigma_q$ . Это приведет к понижению точности измерения  $D(p\|q)$ . В предельном случае для коррелированных характеристик,  $\Sigma_p$  станет единственной с нулевым детерминантом и  $D(p\|q)$  будет неопределенно. Таким образом, наше измерение недостаточно в этом случае. Далее предлагается алгоритм для того, чтобы иметь дело с этим эффектом.

Рассмотрим методы регуляризации для вырожденных характеристик. Для защиты от численной нестабильности в наших измерениях мы хотим извлечь взаимно независимый набор  $G$  «важных» характеристик ( $G \leq F$ ). Для генерации отображения ( $U^t: X \rightarrow Y$ ) из первоначальных биометрических характеристик  $(F \times 1)$  в новое пространство характеристик  $Y$  размером  $G \times 1$  мы используем метод главных компонент (PCA) [2]. PCA может быть вычислен на основании декомпозиции единственного значения (SVD) [3] матрицы ковариационных характеристик следующим образом:

$$US_q U^t = \text{svd}(\text{cov}(X)) = \text{svd}(\Sigma_q), \quad (7)$$

где  $\Sigma_q$  – положительная матрица,  $U$  – ортонормированная матрица,  $S_q$  – диагональная матрица. Мы выбираем для выполнения PCA распределение характеристик населения, а не человека, так как оно основано на гораздо большем количестве данных, и поэтому, вероятно, будет более надежной оценкой. Значения  $S_q$  определяют значимость

каждой характеристики в пространстве PCA. Характеристика  $j$  с небольшим  $[S_q]_{i,j}$  будет оказывать незначительный эффект на всю характеристику биометрической информации. Мы используем этот анализ для упорядочивания  $\Sigma_q$  и для отказа от вырожденных характеристик при помощи исключения SVD. Выбирается округленный порог  $j$ , где  $[S_q]_{i,j} < 10^{-10}[S_q]_{1,1}$ . Основываясь на этом пороге,  $S_q$  усекается до  $G \times G$ , а  $U$  усекается до  $F \times G$ . Используя за основу  $U$ , мы раскладываем на составные части личностную ковариантность в пространстве характеристик  $Y$ :

$$S_p = U^T \Sigma_p U, \quad (8)$$

где  $S_p$  – необязательная диагональная матрица. Основываясь на этих регуляризационных схемах, перепишем  $D(p||q)$  в PCA пространстве:

$$D(p||q) = k(\beta + \text{trace}U((S_p + S_i)S_q^{-1} - I)U^T), \quad (9)$$

где  $\beta = \ln \frac{|S_q|}{|S_p|}$  и  $S_i = U^T T U$ .

Рассмотрим регуляризационный метод для неполных данных. Выражение, разработанное выше, решает проблему некорректности  $\Sigma_q$ . Тем не менее,  $\Sigma_q$  может оставаться единственным в общем случае, когда только небольшое число образцов каждой личности доступно. Пусть даны  $N_p$  изображений личности, из которых вычислены  $G$  характеристик,  $S_q$  будет единственной, если  $G \geq N_p$ . На практике этот случай является общим, поскольку большинство биометрических систем вычисляют много сотен характеристик и редко бывает более 10-ти образцов каждого человека. Чтобы побороть эту проблему, мы разработали оценку, которая может служить нижней границей. Чтобы сделать это, мы делаем следующие предположения:

1. Оценки дисперсии характеристик матрицы  $[S_p]_{i,j}$  верны для всех  $i$ .
2. Оценки ковариации характеристик матрицы  $[S_p]_{i,j}$  для  $ij$  верны только для наиболее важных  $L$  характеристик, где  $L < N_p$ .

Характеристики, которые не считаются правильными, основываясь на этих предположениях, устанавливаются в 0 умножением  $S_q$  на маску  $M$ , где

$$M = \begin{cases} 1, & \text{если } i = j \text{ или } (i < L \text{ и } j < L) \\ 0, & \text{в остальных случаях} \end{cases} \quad (10)$$

Это выражение регуляризирует матрицу ковариации характеристик человека  $\Sigma_p$  и гарантирует, что  $D(p||q)$  не отклоняется. Чтобы выяснить влияние регуляризации на  $D(p||q)$ , мы отмечаем, что ковариация характеристик человека  $|\Sigma_p|$  будет уменьшаться до 0, что приведет к тому, что оценка дифференциальной энтропии уйдет в  $\infty$ . Мы рассматриваем эту регуляризационную стратегию для создания нижней границы биометрической характеристики информации. Выбор  $L$  представляет собой компромисс между

использованием всех имеющихся измерений (с использованием большого  $L$ ) и избеганием численной нестабильности, когда  $S_p$  близко к сингулярному (с использованием небольшого  $L$ ).

### 3. ВЫЧИСЛЕНИЕ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ ЛИЦА

Вычисление биометрической информации лица осуществляется с использованием описанного метода. Чтобы протестировать алгоритм необходимо иметь несколько изображений одной личности. Используя базу данных по лицу Aberdeen [4] мы выбираем 18 фронтальных изображений каждого из 16 человек, для которых мы вычисляем PCA компоненты лица, используя алгоритм [5] и характеристики линейной дискриминанты Фишера (FLD), используя алгоритм, описанный в [6]. Первоначально все изображения лиц были зарегистрированы во вращении и определен их масштаб для того, чтобы знать позицию глаз в точках (50,90) и (100,90). Далее изображения были обрезаны до размера 150x200 пикселей, выровнены их гистограммы для покрытия диапазон интенсивности 0-255. Биометрические характеристики были вычислены из набора  $N_q$  изображений, используя различные методы компонентного анализа такие как PCA и FLD.  $\mu_p$  и  $\mu_q$  –  $F \times 1$  вектора среднего распределения характеристик населения и человека,  $\Sigma_p$  и  $\Sigma_q$  – матрицы ковариации характеристик человека и населения соответственно.

Процесс декомпозиции характеристик был проведен на 18 изображениях каждого из 16 человек (всего 288 изображений). Для декомпозиции характеристик PCA и FLD было вычислено 288 отдельных векторов и 100 наиболее значимых компонент использовались для дальнейших анализов. На рис. 1 проиллюстрированы главные компоненты лица PCA. Слева направо показаны компоненты под номерами 3, 15, 35, 55.



Рис. 1. Главные компоненты лица

На рис. 2 проиллюстрированы базисные дискриминанты лица FLD. Слева направо показаны дискриминанты под номерами 7, 10, 30, 50.



Рис. 2. Базисные дискриминанты лица



Используя FLD и PCA компоненты, для каждого из 16 человек рассчитано  $D(p||q)$ , используя выражение (9). Принимается, что  $p$  и  $q$  имеют Гауссово распределение. Чтобы проверить достоверность Гауссовой модели для наших данных мы используем следующие тесты нормальности:

- тест Колмогорова-Смирнова: сравнение распределений значений в 2-х векторах данных  $X_1$  и  $X_2$ , где  $X_1$  представляет собой случайную выборку из исходного распределения, а  $X_2$  – следующее идеальное гауссово с нулевым средним и дисперсией. Нулевая гипотеза состоит в том, что  $X_1$  и  $X_2$  взяты из того же непрерывного нормального распределения. Мы отвергаем нулевую гипотезу при  $p < 0.01$ .

- тест Лилифора [7]: оценивает гипотезу, что  $x$  имеет нормальное распределение с непостоянным средним и дисперсией, против альтернативы, что  $X$  не имеет нормального распределения. Этот тест сравнивает эмпирическое распределение  $X$  с нормальным распределением, имеющим такое же среднее и дисперсию как  $X$ . Мы отвергаем нулевую гипотезу при  $p < 0,01$ .

Используя эти тесты, получаем в среднем 88% и 89% для FLD и PCA компонент предельного распределения распределены нормально.

После подгонки распределений  $p(x)$  и  $q(x)$  к гауссовой модели, мы анализируем биометрическую информацию для FLD и PCA методов. PCA компоненты показаны на рис. 3. Из него видно постепенное снижение биометрической информации после 2-ой главной компоненты.

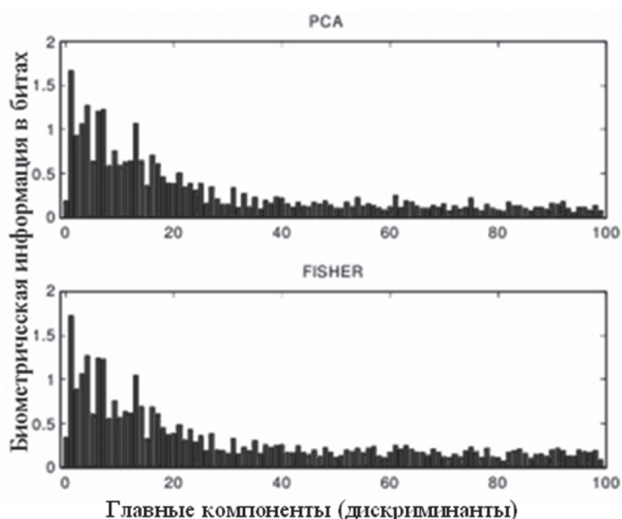


Рис. 3. Зависимость биометрической информации от номера компоненты для PCA и FLD

Такую форму графика можно объяснить природой PCA декомпозиции: чем выше номер главной компоненты, тем более высокие частоты деталей. Так как шум увеличивается с частотой, биометрическая информация при более высоких номерах PCA компонент будет меньше. Сумма биометрической информации за первые 100 PCA главных компонент для одного человека равна 40,5 бит. Биометрическая информация,

вычисленная с использованием FLD дискриминант, кажется похожей на PCA. Для FLD дискриминант наибольшая биометрическая информация характерна для доминирующих ‘лиц Фишера’. Поскольку 18 изображений одного человека используются для вычисления матрицы ковариации, попытка вычислить  $D(p||q)$  для более чем 17 характеристик потерпит неудачу, т.к.  $\Sigma_p$  единственная. Относительно небольшое значение  $S_q$  показывает, что нет вырожденных характеристик для PCA и FLD алгоритмов. Как бы то ни было,  $S_p$  плохо обусловлено. Для преодоления этого мы предлагаем регуляризационную схему (формула 10), основанную на маске с точкой отсечения  $L$ . Эта схема основана на диагональной структуре  $S_p$ , как показано на рис. 4.



Рис. 4. Регуляризованная ковариационная матрица  $S_p$  с доминирующими компонентами вдоль диагонали

Для гарантирования сходимости размер маски  $L$  устанавливается значением меньшим, чем  $N_p$ . Мы решаем эту сингулярность уравнения (9), используя маску для  $S_p$ , основанную на параметре  $L$ . Для дальнейшего исследования воздействия параметров  $L$  и  $N_p$  мы искусственно уменьшаем  $N_p$  с помощью случайного удаления некоторых изображений личности. Результат для PCA компонент каждого человека как функция от  $L$  показан на рис.5.

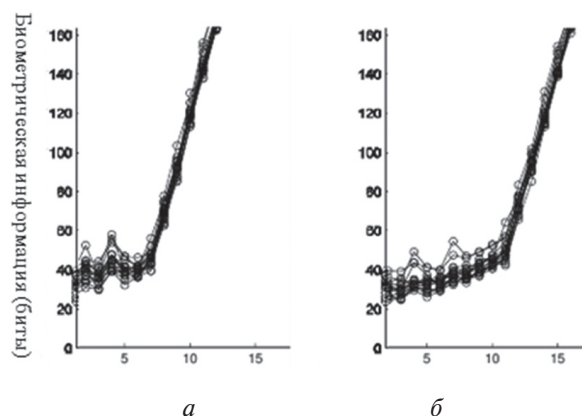


Рис. 5. Биометрическая информация (в битах) относительно размера маски  $L$

Кривая напоминает форму клюшки. На рис. а и б представлены значения для различных  $N_p$  (изображений одного человека): на рис. а –  $N_p=8$ , на рис. б –  $N_p=12$ . Кривая показывает, что

$D(p||q)$  необоснованно увеличивается, когда  $\Sigma_p$  становится сингулярным ( $L \geq N_p$ ). Очевидно, что значения  $D(p||q)$ , находящиеся выше “колена клюшки” не являются достоверными. Относительная энтропия увеличивается с размером маски. При приближении значений  $L$  к  $N_p$  происходит переоценивание  $D(p||q)$ . С другой стороны малые значения  $L$  будут приводить к недооцениванию  $D(p||q)$ . Чтобы получить точную оценку  $D(p||q)$  необходимо идти на компромисс между этими результатами. Мы выбрали  $L=3/4 N_p$ . Используя алгоритм, изложенный выше, и значение  $L$  мы вычисляем общую биометрическую информацию для различных алгоритмов распознавания по лицу. Для PCA компонент средняя биометрическая информация  $D(p||q) = 45$  бит, а для FLD дискриминант  $D(p||q) = 37$  бит. Если PCA и FLD компоненты объединить (сделать 200 характеристик) средняя биометрическая информация  $D(p||q) = 55,6$  бит. Биометрическая информация для FLD дискриминант меньше, чем для PCA компонент. Это можно объяснить тем, что PCA главные компоненты содержат информацию о выражении лица и освещенности. Большее значение биометрической информации указывает на то, что набор характеристик, использующийся в биометрической системе, содержит больше различающей информации, что должно привести в итоге к снижению ошибок первого и второго рода.

#### 4. ВЫЧИСЛЕНИЕ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ РАДУЖНОЙ ОБОЛОЧКИ ГЛАЗА

Методы идентификации личности по радужной оболочке построены по одному и тому же принципу – выделение частотной или какой-либо другой информации о текстуре радужки из изображения и сохранение этой информации в виде специального кода (для системы Daugman этот код получил специальное название – IrisCode (радужковый код)). Построение кода производится в три этапа:

1. Выделение <баранки> радужки из общего изображения
2. Предобработка полученного изображения – например убиение шума (denoising), улучшение изображения (enhancing), в том числе выравнивание гистограммы, убиение блика. Некоторые методы “разворачивают” круглый зрачок в прямоугольное изображение – происходит переход из полярных координат в декартовы. Иногда после такой “развертки” часть изображения отрезается, чтобы накопленная на данном этапе ошибка не повлияла на качество распознавания.
3. Составление кода. Предварительно обработанное изображение фильтруется способом, зависящим от конкретного метода. По результатам фильтрации составляется представление в виде кода.

Для кодов необходимо выработать критерий сравнения. Часто код записывается в виде последовательности битов и критерием сравнения служит код Хэмминга. В частности, код Хэмминга используется в системах Daugman, Tisse [8]. Наглядно система аутентификации по радужной оболочке приведена на рис. 6.

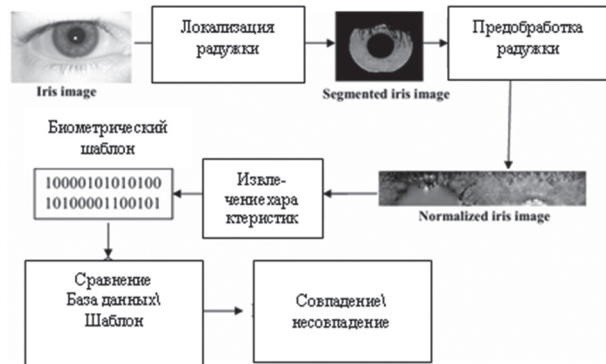


Рис. 6. Различные этапы в системе распознавания по радужке

Для вычисления биометрической информации радужной оболочки глаза использовалась база данных CASIA, которая содержит 689 изображений радужных оболочек глаз, взятых у 108 людей (6-7 изображений каждого человека). Изображение радужной оболочки было предобработано. Далее были вычислены PCA главные компоненты, используя алгоритм, описанный в [5] и ICA (Independent Component Analysis) компоненты, используя алгоритм, описанный в [6]. Для PCA метода и метода независимых компонент (ICA) было вычислено 327 векторов характеристик, которые были использованы для дальнейшего анализа. На рисунках 7 и 8 проиллюстрирована биометрическая информация, вычисленная для PCA и ICA компонент радужки. На рисунке 7 показана биометрическая информация, вычисленная для 327 PCA компонент. Верхний график вычислен для алгоритма Masek, а нижний – с использованием расширенной техники. Стандартное отклонение показано внизу каждого графика.

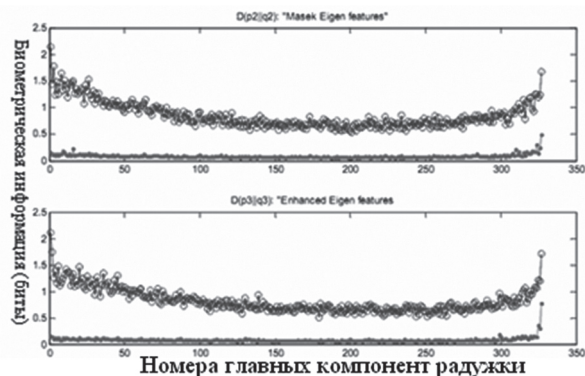


Рис. 7. Зависимость биометрической информации радужки от номера PCA компоненты

На рис. 8 показана биометрическая информация, вычисленная для 327 ICA компонент.

Верхний график вычислен для алгоритма Masek, а нижний – с использованием расширенной техники. Стандартное отклонение изображено внизу каждого графика.

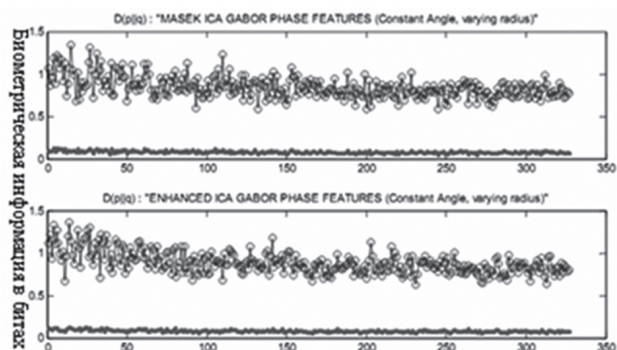


Рис. 8. Зависимость биометрической информации радужки от номера ICA компоненты

Используя алгоритм для вычисления биометрической информации, описанный выше, и базу данных CASIA, получаем 278 бит биометрической информации для PCA главных компонент радужки вычисленных по алгоритму Masek и 267 бит биометрической информации при использовании расширенной техники. Разницу в биометрической информации можно объяснить тем фактом, что алгоритм Masek не полностью устраняет ресницы из области радужки и эти пиксели ложно используются как информация о радужке. Это повышает биометрическую информацию, т.к. алгоритм принимает ресницы за текстуру радужки, делая ее тем самым более отличимой в наборе. Для ICA компонент средняя биометрическая информация составила 288 бит, используя сегментационный алгоритм Masek и 277 бит, используя расширенную сегментационную технику. Как видим, количество биометрической информации для ICA и PCA компонент, очень близко. ICA компоненты содержат больше информации, т.к. они соответствуют модели характеристических данных радужки лучше.

### 5. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИСТОЧНИКОВ БИОМЕТРИЧЕСКОЙ ИНФОРМАЦИИ

Сравнительный анализ источников биометрической информации производился на основании критерия относительной энтропии. Сравнение производилось по следующим биометрическим источникам: изображение лица, радужная оболочка глаза. Выделение компонент из изображения лица происходило с использованием алгоритмов PCA, FLD и ICA, а из радужной оболочки глаза с помощью ICA и PCA. Результаты сравнений приведены в табл. 1.

Исходя из таблицы, можно сделать вывод о том, радужная оболочка глаза содержит больше биометрической (различающей) информации, чем изображение лица и, соответственно, аутентификация на ее основе будет более надежной.

Таблица 1

Сравнительный анализ источников биометрической информации

Алгоритм	Биометрическая информация (биты)	
	Лицо	Радужка
PCA	45	278
ICA	39	288
FLD	37	

### ВЫВОДЫ

В данное время биометрические технологии идентификации личности получили широкое распространение. Актуальной является задача выбора того или иного метода. Чаще всего производится сравнение биометрических методов аутентификации на основании ошибок первого и второго рода. В работе предлагается производить сравнительный анализ источников биометрической информации на основании критерия относительной энтропии. Сравнительный анализ производился по 2-м источникам биометрической информации: радужная оболочка глаза и изображение лица. Выделение компонент лица осуществлялось с использованием методов компонентного анализа PCA, FLD и ICA. Выделение компонент радужки производилось с помощью метода PCA и метода независимых компонент.

Проблемным вопросом в нашей работе было: как определить распределение биометрических характеристик для населения. Мы использовали типичный подход: приняли нашу базу данных за адекватное представление населения.

С использованием описанного нами метода, была вычислена относительная энтропия для лица: для метода PCA она составила 45 бит, для FLD – 37 бит, а для ICA – 39 бит. Количество биометрической информации радужной оболочки глаза составило: для PCA метода 278 бит, а для ICA – 288 бит. Такие результаты совместимы с предыдущими исследованиями радужной оболочки глаза. Так, Daugman заявлял [9], что комбинаторная сложность фазовой информации радужной оболочки глаза различных людей составляет около 249 степеней свободы. Cover и Thomas [10], используя радужку диаметром в 11мм, просчитали, что ее биометрическая информация составляет 241 бит. Различие наших данных со значениями Daugman и Thomas можно объяснить тем, что диаметр радужной оболочки в нашей системе принимал значения от 11 до 11,5мм. А разница в 0,5мм дает увеличение биометрической информации на 28,52 бита.

На основании проведенного нами анализа, можно сделать вывод о том, что радужная оболочка глаза содержит больше биометрической информации, чем изображение лица. А это означает, что набор характеристик, использующийся в системе распознавания по радужке, содержит больше различающей информации, что должно



привести в итоге к снижению ошибок первого и второго рода.

Использование относительной энтропии как критерия сравнительного анализа позволяет сравнивать не только методы биометрической аутентификации между собой, но и с ПИНом, паролем и с другими методами.

#### Литература

- [1] *Adler A.* Towards a measure of biometric feature information / A. Adler, R. Youmaran, S. Loyka // *Pattern Anal. Appl.* – 2009. – №12(3). – P. 261-270.
- [2] *Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.*, “Recognizing faces with PCA and ICA”, *Computer Vision and Image Understanding*, 91:115-137, 2003.
- [3] *Alter O, Brown PO, Botstein D.*, “Singular value decomposition for genome-wide expression data processing and modeling”, *Proc Natl. Acad. Sci.*, 97:10101–10106, 2000.
- [4] *Craw, I., Costen, N.P., Kato, T., Akamatsu, S.*, “How should we represent faces for automatic recognition?”, *IEEE Trans. Pat. Anal. Mach. Intel.* 21725–736, 1999.
- [5] *Grother, P.*, “Software Tools for an Eigenface Implementation” National Institute of Standards and Technology, (2000) <http://www.nist.gov/humanid/feret/>
- [6] *Xiang C.* Face recognition using recursive Fisher linear discriminant / C. Xiang, X.A. Fan, T.H. Lee // *Communications, Circuits and Systems.* – 2004. – Vol.2. – P. 27-29.
- [7] *Conover, W.J.*, *Practical Nonparametric Statistics*, Wiley, 1980.
- [8] *Christel-loic Tisse, Lionel Martin, Lionel Torres, Michel Robert.* Person identification technique using human iris recognition. *Proc. of Vision Interface*, pp.294-299, 2002.
- [9] *Daugman, J.* (2004). How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14 (1): 21-304.
- [10] *Cover, T.M., Thomas, J.A.* (1991). *Elements of Information Theory* New York:Wiley.

Поступила в редколлегию 10.04.2012

**Горбенко Иван Дмитриевич**, фото и сведения об авторе см. на с. 190.



**Олешко Инна Викторовна**, аспирант каф. БИТ ХНУРЭ. Область научных интересов: электронная паспортная система, биометрическая аутентификация

УДК 621.391:519.2:519.7

**Метод оцінки відносної ентропії та порівняльний аналіз джерел біометричної інформації**/ І.Д. Горбенко, І.В. Олешко // *Прикладна радіоелектроніка: наук.-техн. журнал.* – 2012. – Том 11. № 2. – С. 255–261.

У роботі розвивається підхід до вимірювання інформації, що міститься в біометричній характеристиці. Показується, що особливості біометричної інформації можуть бути розраховані за допомогою відносної ентропії. Виконується порівняльний аналіз джерел біометричної інформації.

*Ключові слова:* біометрична інформація, відносна ентропія, біометрична ідентифікація, матриця коваріації, метод головних компонент, Гауссовий розподіл.

Табл. 1. Лл. 8. Бібліогр.: 10 найм.

UDC 621.391:519.2:519.7

**Method of assessing of relative entropy and comparative analysis of biometric information sources**/ I.D. Gorbenco, I.V. Oleshko // *Applied Radio Electronics: Sci. Journ.* – 2012. Vol. 11. № 2. – P. 255–261.

This paper develops the approach to the measurement of the information contained in a biometric characteristic. It is shown that the characteristics of biometric data can be calculated using relative entropy. Comparative analysis of biometric information sources was performed.

*Keywords:* biometric information, relative entropy, biometric identification, covariance matrix, principal components method, Gaussian distribution.

Tab. 1. Fig. 8. Ref.: 10 items.