

## ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ НА ОСНОВІ КРИТЕРІЮ ВІДНОСНОЇ ЕНТРОПІЇ

© Олешко І.В., 2012

Біометричні системи дають змогу ідентифікувати людину на основі фізіологічних або поведінкових характеристик. Ми визначаємо термін “біометрична інформація” так: зменшення невизначеності ідентичності людини за рахунок вимірювання набору біометричних характеристик. Показано, що особливості біометричної інформації можуть бути розраховані за допомогою відносної ентропії. Виконано порівняльний аналіз джерел біометричної інформації.

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

Biometric systems allow identification of human person based on physiological or behavioral characteristics. We define biometric information as the decrease in uncertainty about the identity of a person due to a set of biometric measurements. It is shown that the characteristics of biometric data can be calculated using the relative entropy. Comparative analysis of biometric information sources was performed.

**Key words:** biometric information, relative entropy, biometric identification, covariance matrix, principal component analysis, Gaussian distribution.

### Вступ

Ідентифікація на основі біометричних даних – це засіб автоматичного розпізнавання особистості на основі унікальних фізичних або поведінкових параметрів. Оскільки біометричні характеристики унікальні для кожної людини, вони не можуть бути вкрадені, підроблені або забуті, на відміну від пароля або персонального ідентифікаційного номера (PIN). Для біометричної ідентифікації досить фізичних параметрів людини і не потрібні жодні файли, які можна скопіювати, або паролі, які можна зламати. Сьогодні біометричні технології ідентифікації особи отримали широке розповсюдження в різних областях забезпечення безпеки: від контролю і управління доступом в офісні приміщення до громадянської ідентифікації на правоохоронних об’єктах. Одним із головних показників розвитку біометрії є стандартизація цієї області [1–4].

Сьогодні активно використовуються такі біометричні ознаки: відбиток пальця, геометрична форма кисті руки, форма і розміри обличчя, особливості голосу, візерунок райдужної оболонки і сітківки ока [5]. Найчастіше порівняльний аналіз методів біометричної автентифікації виконують на основі помилок першого і другого роду. Використання відносної ентропії як критерію ефективності ідентифікації робить можливим не тільки порівняння біометричних ознак між собою, але й з персональним ідентифікаційним номером (ПІН), паролем та іншими методами автентифікації. Для порівняння систем біометричної автентифікації на основі критерію відносної ентропії необхідно визначити кількість інформації кожної системи окремо. Визначимо термін біометрична інформація так: зменшення невизначеності ідентичності людини за рахунок вимірювання набору біометричних характеристик. Спираючись на це визначення, у роботі розвинено підхід до вимірювання біометричної інформації за допомогою відносної ентропії та наведений її порівняльний аналіз для методів автентифікації по обличчю та райдужній оболонці ока.

### Алгоритм для обчислення біометричної інформації

Алгоритм для обчислення біометричної інформації з використанням відносної ентропії полягає у таких кроках: 1) висування вимог; 2) відносна ентропія біометричних характеристик; 3)

гауссова модель для обчислення біометричних характеристик і відносної ентропії, 4) методи регуляризації для вироджених характеристик; 5) методи регуляризації для неповних даних.

Основні вимоги до особливостей біометричної інформації:

1. Якщо розподіл характеристик людини  $p$  дорівнює розподілу характеристик між людьми  $q$ , тоді немає інформації, яка відрізняє людину, і тоді інформація про її біометричні характеристики дорівнює 0.
2. Якщо вимірювання особливості стає точнішим, легше визначити когось у популяції, і біометрична інформація збільшується.
3. Якщо у людини незвичайне значення характеристики (далеке від значення популяції), інформація про її біометричну характеристику збільшується.
4. Біометрична інформація про некорельовані характеристики дорівнює сумі біометричних інформацій цих характеристик.
5. Характеристики, які не пов'язані із визначенням ідентичності, не підвищують біометричну інформацію.
6. Біометричні характеристики, такі як вага і зріст, менш інформативні.

Найкраща міра для характеристики біометричної інформації – відносна ентропія  $D(p||q)$ , де  $p(x)$  і  $q(x)$  – розподіли біометричних характеристик людини і населення відповідно.  $D(p||q)$ , або відстань Кульбака–Лейблера, визначається як “додаткові біти” інформації, необхідні для подання  $p(x)$  стосовно  $q(x)$ .  $D(p||q)$  визначається як [6]:

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

У загальній біометричній системі біометричні характеристики  $S$  вимірюються для створення вектора біометричних характеристик  $x(S \times 1)$  для кожної людини. Для людини  $p$  ми маємо  $N_p$  зразків характеристик, в той час, як для населення –  $N_q$  зразків характеристик. Визначивши  $x$  як значення випадкової величини  $X$ , ми обчислюємо середнє зображення населення  $m_q$  [6]:

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

Середнє зображення людини визначається аналогічно, замінюючи  $q$  на  $p$ . Матриця коваріації населення  $\Sigma_q$  визначається так [6]:

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

Матриця коваріації людини обчислюється аналогічно. Однією з важливих проблем під час прямих вимірювань теоретичної інформації є придатність даних. Розподіл важко оцінити точно, особливо на кінцях; для невеликих значень  $p(x)$  та  $q(x)$ ,  $\log_2(p(x)/q(x))$  матиме велике абсолютне значення. Зазвичай у таких випадках виконується перехід до моделі з невеликою кількістю параметрів. Найзагальнішою моделлю є гауссовий розподіл. Допускається використання такої моделі для отримання верхньої оцінки ентропії. Ґрунтуючись на гауссовій моделі і відповідних  $p$  і  $q$  [6], обчислимо розподіли біометричних характеристик людини та населення:

$$p(x) = \frac{1}{\sqrt{|2p\Sigma_p|}} \exp\left(-\frac{1}{2}(x - m_p)^T \Sigma_p^{-1}(x - m_p)\right); \quad (4)$$

$$q(x) = \frac{1}{\sqrt{|2p\Sigma_q|}} \exp\left(-\frac{1}{2}(x - m_q)^T \Sigma_q^{-1}(x - m_q)\right). \quad (5)$$

З наведених формул обчислимо  $D(p||q)$ :

$$D(p||q) = \int p(x)(\log_2 p(x) - \log_2 q(x))dx = -k(\ln|2\pi\Sigma_p| - \ln|2\pi\Sigma_q|) + 1 - \mathbb{E}[(x - \mu_q)^T \Sigma_q^{-1} (x - \mu_q)] = k(\ln \frac{|2\pi\Sigma_q|}{|2\pi\Sigma_p|} + \text{trace}((\Sigma_p + T)\Sigma_q^{-1} - I)), \quad (6)$$

де  $T = (\mathbf{m}_p - \mathbf{m}_q)(\mathbf{m}_p - \mathbf{m}_q)^t$ ,  $k = \log_2 \sqrt{e}$ .

Цей вираз обчислює відносну ентропію для гауссового розподілу  $p(x)$  і  $q(x)$ . Він відповідає більшості вимог до особливостей біометричної інформації, наведеної вище.

Розглянемо регуляризаційний метод для вироджених характеристик. Для захисту від чисельної нестабільності в наших вимірах ми прагнемо отримати взаємно незалежний набір  $W$  «важливих» характеристик ( $W \leq S$ ). Для генерації відображення ( $U^T : X \rightarrow Y$ ) з початкових біометричних характеристик  $x(S \times 1)$  у новий простір характеристик  $Y$  розміром  $W \times 1$  ми використовуємо метод головних компонент (PCA) [7]. PCA може бути обчислено на основі декомпозиції єдиного значення (SVD) [8] матриці коваріантних характеристик:

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

де  $\Sigma_q$  – позитивна матриця;  $U$  – ортогональна матриця власних векторів;  $S_q$  – діагональна матриця власних значень. Значення  $S_q$  визначають значущість кожної характеристики у просторі PCA. Характеристика  $j$  з невеликим значенням  $[S_q]_{j,j}$  чинитиме незначний ефект на усю біометричну інформацію. Беручи за основу  $U$ , ми розкладаємо на складові частини особистісну коваріантність у просторі характеристик  $Y$ :

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

де  $S_p$  – необов'язкова діагональна матриця. Тим не менше, оскільки  $p$  і  $q$  описують подібні дані, ми очікуємо, що  $S_p$  матиме сильну діагональну компоненту.

Ґрунтуючись на вищенаведених твердженнях, перепишемо  $D(p||q)$  в PCA-просторі:

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

де  $\mathbf{b} = \ln \frac{|S_q|}{|S_p|}$  та  $S_t = U^T T U$ .

Розглянемо регуляризаційний метод для неповних даних. Матриця коваріації для людини може залишатися єдиною у загальному випадку, коли тільки невелика кількість зразків кожної особистості доступна. Нехай дано  $N_p$  зображень особистості, з яких обчислено  $G$  характеристик. Матриця коваріації буде єдиною, якщо  $G \geq N_p$ . На практиці цей випадок є загальним, оскільки більшість біометричних систем обчислюють багато сотень характеристик і рідко буває більше десяти зразків кожної людини. Щоб позбутись цієї проблеми, ми розробили оцінку, яка може бути нижньою границею під час розрахунку відносної ентропії. Щоб зробити це, ми робимо такі припущення:

1. Оцінки дисперсії характеристик матриці  $[S_p]_{i,i}$ , дійсні для усіх  $i$ .
2. Оцінки коваріації характеристик матриці  $[S_p]_{i,j}$  для  $i \neq j$  дійсні тільки для найважливіших  $L$  характеристик, де  $L < N_p$ .

Характеристики, які не вважаються правильними, ґрунтуючись на цих припущеннях, встановлюються в 0 множенням  $S_q$  на маску  $M$ :

$$M = \begin{cases} 1, \text{ якщо } i = j \\ 0, \text{ в інших випадках.} \end{cases} \quad (10)$$

Вибір  $L$  являє собою компроміс між використанням усіх наявних вимірювань (з використанням великого  $L$ ) і униканням чисельної нестабільності, коли  $S_p$  близьке до сингулярного (з використанням невеликого  $L$ ). Дослідним шляхом ми обрали  $L=3/4 N_p$ .

### Обчислення біометричної інформації обличчя

Обчислення біометричної інформації обличчя виконувалося на основі вищеописаного методу. Використовуючи базу даних по обличчю Aberdeen, ми обираємо 18 зображень обличчя кожного з 16 осіб, для яких ми обчислюємо PCA компоненти обличчя та лінійні дискримінанти Фішера (FLD) [9]. Усі зображення бази Aberdeen належать до типу “indoor”, мають розмір 150 на 200 пікселів та представлені в шкалі “GRAY” (256 рівнів яскравості). Для обличчя бази Aberdeen характерна зміна: повороту голови у просторі 3D в межах не менше  $\pm 20^\circ$  по осі X,  $\pm 30^\circ$  – по осі Y та не менше  $\pm 10^\circ$  – по осі Z; міміки обличчя, промальовування деталей обличчя (їх згладжування). Для знаходження компонент PCA та FLD було обчислено 288 окремих векторів і 100 найвагоміших характеристик використовувались для подальшого аналізу. На рис. 1 проілюстровано головні компоненти обличчя. Зліва направо показані компоненти під номерами 3, 15, 35, 55.



Рис. 1. PCA компоненти обличчя

Використовуючи FLD- і PCA-компоненти для кожного з 16 осіб розраховано відносну ентропію  $D(p||q)$ , використовуючи вираз (9). Після приведення розподілів  $p(x)$  та  $q(x)$  до гауссової моделі, ми аналізуємо характеристики біометричної інформації для FLD- та PCA-методів. PCA- та FLD-компоненти показані на рис. 2. З графіка бачимо поступове зменшення біометричної інформації після другої головної компоненти для PCA. Таку форму графіка можна пояснити природою PCA-декомпозиції: чим вищий номер характеристики, тим вищі частоти деталей. Оскільки шум збільшується з частотою, біометрична інформація за вищих номерів PCA-характеристик буде меншою. Для FLD-характеристик найбільша біометрична інформація характерна для домінуючих “облич Фішера”.

Використовуючи алгоритм, викладений вище, ми обчислюємо загальну біометричну інформацію для різних алгоритмів розпізнавання по обличчю. Для PCA-компонент середня біометрична інформація  $D(p||q) = 45$  біт, а для FLD дискримінант –  $D(p||q) = 37$  біт. Сумарна середня біометрична інформація для PCA- та FLD-компонент –  $D(p||q) = 55,6$  біт. Біометрична інформація для FLD-дискримінант менша, ніж для PCA-компонент. Це можна пояснити тим, що PCA головні компоненти містять інформацію про вираз обличчя і освітлення. Більше значення біометричної інформації вказує на те, що набір характеристик, що використовується в біометричній системі, містить більше розрізнявальної інформації, що повинно привести в результаті до зниження помилок першого і другого роду.

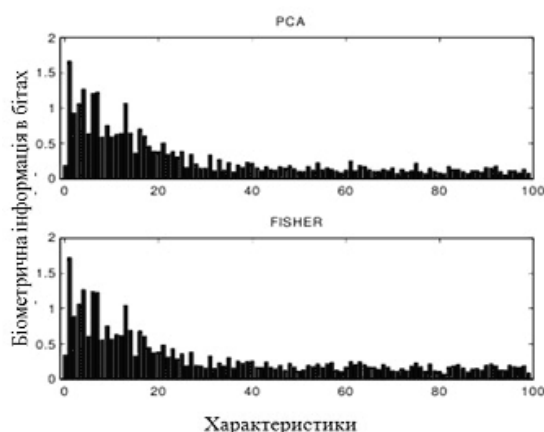


Рис. 2. Залежність біометричної інформації від номера характеристики для PCA та FLD

### Обчислення біометричної інформації райдужної оболонки ока

Для обчислення біометричної інформації райдужної оболонки ока використовувалась база даних CASIA, яка містить 689 зображень райдужних оболонок ока, взятих у 109 людей (шість–сім зображень кожної людини). Зображення райдужної оболонки було попередньо оброблене. Далі були обчислені PCA- та ICA- (Independent Component Analysis) компоненти. Для знаходження головних компонент використовувались 327 векторів характеристик. На рис. 3 проілюстрована біометрична інформація, обчислена для 327 PCA- та ICA-компонент райдужної оболонки ока. Стандартне відхилення показано внизу кожного графіка.

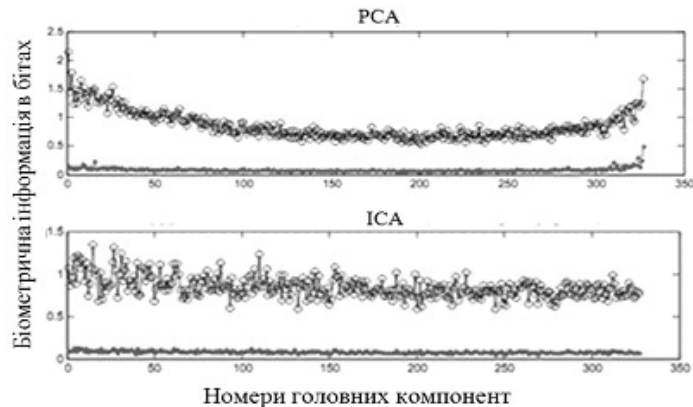


Рис. 3. Залежність біометричної інформації від номера характеристики для PCA та ICA

Використовуючи алгоритм для обчислення біометричної інформації, описаний вище, і базу даних CASIA, отримаємо, що середня біометрична інформація райдужної оболонки ока для PCA головних компонент становить 278 біт, а для ICA – 288 біт. Як бачимо, кількість біометричної інформації для ICA- і PCA-компонент дуже близька. ICA-компоненти містять більше інформації, тому що вони відповідають моделі характеристичних даних райдужки ока краще.

### Порівняльний аналіз методів біометричної автентифікації

Порівняльний аналіз методів біометричної автентифікації проводився на основі критерію відносної ентропії. Порівняння проводилось за такими біометричними джерелами: зображення обличчя, райдужна оболонка ока. Виділення компонент з зображення обличчя відбувалося з використанням алгоритмів PCA, FLD і ICA, а з райдужної оболонки ока – за допомогою ICA і PCA. Результати порівняння наведені у таблиці. Враховуючи дані з таблиці, можна зробити висновок про те, що метод розпізнавання на основі райдужної оболонки ока має більше біометричної інформації і тому дає можливість надійніше виконувати автентифікацію особи.

### Порівняльний аналіз методів розпізнавання по обличчю та райдужній оболонці ока

Алгоритм	Біометрична інформація (біти)	
	обличчя	райдужка
PCA	45	278
ICA	39	288
FLD	37	

### Висновки

Сьогодні біометричні технології ідентифікації особи отримали поширення. Актуальним є завдання вибору того чи іншого методу. Порівняльний аналіз джерел біометричної інформації можна проводити на основі критерію відносної ентропії. На основі розглянутого алгоритму, була розрахована відносна ентропія обличчя: для методу PCA  $D(p||q) = 45$  біт, для FLD характеристик –

$D(p||q) = 37$  біт, а для ІСА – 39 біт. Для райдужної оболонки ока середня біометрична інформація для РСА головних компонент становить 278 біт, а для ІСА – 288 біт. Такі результати сумісні з попередніми дослідженнями. Так, Daugman заявляв, що комбінаторна складність фазової інформації райдужної оболонки ока людини становить приблизно 249 ступенів свободи. Cover та Thomas, використовуючи райдужну оболонку ока діаметром 11 мм, розраховали, що її біометрична інформація становить 241 біт. Різницю наших даних зі значеннями Daugman та Thomas можна пояснити тим, що діаметр райдужної оболонки ока в нашій системі набував значення 11,5 мм. Різниця у 0,5 мм збільшує біометричну інформацію на 28,52 біта.

На основі проведеного аналізу, можна зробити висновок про те, що метод розпізнавання на основі райдужної оболонки ока має більше біометричної інформації, тобто має найбільшу ентропію, і тому дає змогу надійніше виконувати автентифікацію особи. Використання відносної ентропії як критерію ефективності ідентифікації робить можливим не тільки порівняння біометричних ознак між собою, але і з ПІНом, паролем та іншими методами автентифікації.

1. ГОСТ Р ИСО/МЭК 19794-6-2006. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. – Ч. 6: Данные изображения радужной оболочки глаза. – М.: ИПК Изд-во стандартов, 2006. – 27 с. 2. ГОСТ Р ИСО/МЭК 19794-5-2006. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. – Ч. 5: Данные изображения лица. – М.: ИПК Изд-во стандартов, 2006. – 42 с. 3. ISO/IEC 19794-5 Information technology – Biometric data interchange formats – Part 5: Face image data. – S.: ISO copyright office, 2005. – 132 p. 4. BS ISO/IEC 19794-6:2011 Information technology. Biometric data interchange formats. Iris image data. – S.: BSI, 2011. – 30 p. 5. Технологии аутентификации. – [Электронный ресурс]. – 2009. – Режим доступа: <http://ypn.ru/category/data-protection-technologies/authentication-technologies/>. 6. Adler A. Towards a measure of biometric feature information / A. Adler, R. Youmaran, S. Loyka // Pattern Anal. Appl. – 2009. – №12(3). – P. 261–270. 7. Recognizing faces with PCA and ICA / B. Draper, K. Baek, M. Bartlett, J. Beveridge // Computer Vision and Image Understandin. – 2003. – Vol. 91. – P.115–137. 8. Alter O. Singular value decomposition for genome-wide expression data processing and modeling / O. Alter, O. Brown, D. Botstein // Proc Natl. Acad. Sci. – 2000. – Vol. 97. – P.10101–10106. 9. 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.