

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ НА ОСНОВІ КРИТЕРІЮ ВІДНОСНОЇ ЕНТРОПІЇ

© Інна Олешко, 2015

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.

Keywords - biometric information, relative entropy, biometric identification, covariance matrix, principal component analysis, Gaussian distribution.

Біометричні системи дозволяють ідентифікувати людину на основі фізіологічних або поведінкових характеристик. Ми визначаємо термін біометрична інформація наступним чином: зменшення невизначеності ідентичності людини за рахунок вимірювання набору біометричних характеристик. 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.

Вступ

Ідентифікація на основі біометричних даних – це засіб автоматичного розпізнавання особистості на базі унікальних фізичних або поведінкових параметрів. Так як 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 ніякі файли, які можна скопіювати, або паролі, які можна зламати. На сьогоднішній день 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].

В даний час 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 виконується на основі помилок першого і другого роду. Використання відносної ентропії як критерію ефективності ідентифікації робить можливим не тільки порівняння біометричних ознак між собою, але і 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 наступним чином: зменшення невизначеності ідентичності людини за рахунок вимірювання набору біометричних характеристик. Спираючись на це визначення, у статті розвинуто 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 в наступних кроках: 1) висування вимог; 2) відносна ентропія біометричних характеристик; 3) Гаусова 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 $p(x)$ відносно $q(x)$. $D(p||q)$ визначається як [6]:

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

В 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 μ_q [6]:

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

Середнє зображення людини визначається аналогічно, замінюючи q на p . Матриця коваріації населення Σ_q визначається наступним чином [6]:

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

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 $q(x)$, $\log_2(p(x)/q(x))$ буде мати велике абсолютне значення. Зазвичай у таких випадках виконується перехід до моделі з невеликою кількістю параметрів. Найбільш загальною моделлю є Гаусовий розподіл. Допускається використання такої моделі для отримання верхньої оцінки ентропії. Грунтуючись на Гаусовій моделі і відповідних p і q [6], обчислимо розподіли біометричних характеристик людини та населення:

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

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

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

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

де $T = (\mu_p - \mu_q)(\mu_p - \mu_q)^T$, $k = \log_2 \sqrt{e}$.

Цей вираз обчислює відносну ентропію для Гаусового розподілу $p(x)$ і $q(x)$. Він відповідає більшості вимог до особливостей біометричної інформації, що наведені 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 $x(S \times 1)$ в 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 Σ_q – позитивна матриця, U – ортогональна матриця власних векторів, S_q – діагональна матриця власних значень. Значення S_q визначають значимість кожної характеристики в просторі PCA. Характеристика j з невеликим значенням $[S_q]_{j,j}$ буде чинити незначний ефект на всю біометричну інформацію. Використовуючи за основу U , ми розкладаємо на складові частини особистісну коваріантність в 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.

Грунтуючись на наведених вище твердженнях, перепишемо $D(p||q)$ в PCA просторі:

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

де $\beta = \ln \frac{|S_q|}{|S_p|}$ 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 побороти цю проблему, ми розробили оцінку, яка може служити нижньою границею при розрахунку відносної ентропії. Щоб зробити це, ми робимо наступні припущення:

1. Оцінки дисперсії характеристик матриці $[S_p]_{i,i}$ дійсні для всіх i .

2. Оцінки коваріації характеристик матриці $[S_p]_{i,j}$ для $i \neq j$ дійсні тільки для найбільш важливих 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 всіх 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 описаного вище 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]. Всі зображення бази Aberdeen відносяться до типу «indoog», мають розмір 150 на 200 пікселів та представлені в шкалі «GRAY» (256 рівнів яскравості). Для обличчя бази Aberdeen характерна зміна: повороту голови в 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 найбільш вагомих 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 гаусової моделі, ми аналізуємо характеристики біометричної інформації для FLD та PCA методів. PCA та FLD компоненти показані на рис. 2. З графіка видно поступове зменшення біометричної інформації після другої головної компоненти для PCA. Таку форму графіка можна пояснити природою PCA декомпозиції: чим вище номер характеристики, тим більш високі частоти деталей. Так як шум збільшується з частотою, біометрична інформація при більш високих номерах PCA характеристик буде меншою. Для FLD характеристик найбільша біометрична інформація характерна для домінуючих «обличчя 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.

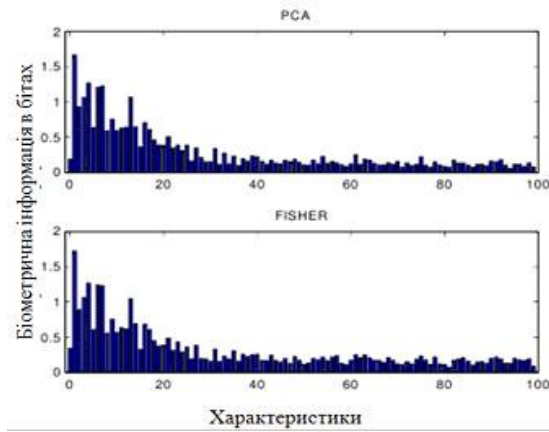


Рис. 2. Залежність біометричної інформації від номера характеристики для PCA та FLD

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 (6-7 зображень кожної людини). Зображення райдужної оболонки було попередньо оброблено. 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. Стандартне відхилення зображено внизу кожного графіка.

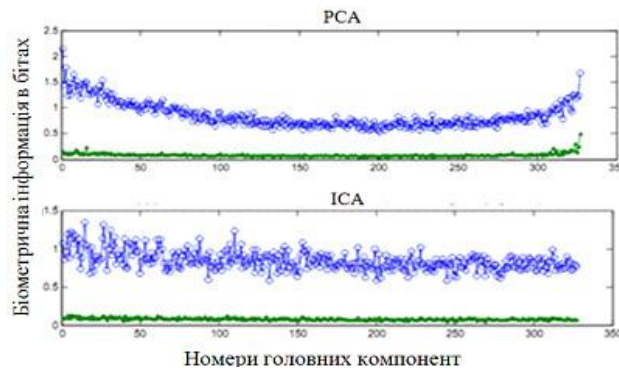


Рис. 3. Залежність біометричної інформації від номера характеристики для PCA та ICA

Використовуючи алгоритм для обчислення біометричної інформації, описаний вище, і базу даних CASIA, отримаємо, що середня біометрична інформація райдужної оболонки ока для PCA головних компонент складає 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 краще.

Порівняльний аналіз методів біометричної автентифікації

Порівняльний аналіз методів біометричної автентифікації проводився на основі критерію відносної ентропії. Порівняння проводилося 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 в таблиці 1. Виходячи з таблиці, можна зробити висновок про те, що метод розпізнавання на основі райдужної оболонки ока має більше біометричної інформації і тому дозволяє більш надійно 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

В даний час біометричні технології ідентифікації особи отримали широке поширення. Актуальною є задача 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 ICA – 39 біт. Для райдужної оболонки ока середня біометрична інформація для PCA головних компонент складає 278 біт, а для ICA – 288 біт. Такі результати сумісні з попередніми дослідженнями. Так, Daugman заявляв, що комбінаторна складність фазової інформації райдужної оболонки ока людини складає приблизно 249 ступенів свободи. Cover та Thomas, використовуючи райдужну оболонку діаметром в 11мм, розрахували, що її біометрична інформація складає 241 біт. Різниця наших даних зі значеннями Daugman та Thomas можливо пояснити тим, що діаметр райдужної оболонки в нашій системі приймав значення 11,5мм. Різниця в 0,5мм 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 дозволяє більш надійно 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.

Література

- ГОСТ Р ИСО/МЭК 19794-6-2006. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Часть 6 Данные изображения радужной оболочки глаза. – М.: ИПК Изд-во стандартов, 2006. – 27 с.
- ГОСТ Р ИСО/МЭК 19794-5 — 2006. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Часть 5 Данные изображения лица. – М.: ИПК Изд-во стандартов, 2006. – 42 с.
- ISO/IEC 19794-5 Information technology — Biometric data interchange formats — Part 5: Face image data. – S.: ISO copyright office, 2005. – 132 p.
- BS ISO/IEC 19794-6:2011 Information technology. Biometric data interchange formats. Iris image data. – S.: BSI, 2011. – 30 p.
- Технологии аутентификации [Электронный ресурс]. – 2009. – Режим доступа: <http://ypr.ru/category/data-protection-technologies/authentication-technologies/>.
- Adler A. Towards a measure of biometric feature information / A. Adler, R. Youmaran, S. Loyka // Pattern Anal. Appl. – 2009. – №12(3). – P. 261-270.
- 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.
- 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.
- 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.