

Методы обработки биометрических данных рукописного почерка

Б.И. Баянов

*Казанский национальный исследовательский технический университет
имени А. Н. Туполева-КАИ*

Аннотация: В работе представлены методы обработки биометрических данных рукописного почерка. Исходными данными являются координаты положения пера на графическом планшете. В качестве обрабатываемого рукописного текста использовалась собственная база и открытая база данных подписей SVC 2004. В процессе обработки использовались следующие методы – «интерполяция рукописной записи», метод гистограмм, «скорость перемещения пера», направленные на устранение сдвигов исходных значений по оси времени и по оси значений. В результате обработки получены рекомендации по стандартизации, дискретизации исходных данных, количественные показатели оценок качества найденных биометрических признаков. Полученные экспериментальные результаты будут полезны исследователям для усовершенствования собственных разработок в области биометрических систем защиты.

Ключевые слова: рукописный почерк, биометрические данные, графический планшет, верификация, аутентификация, стандартизация, дискретизация, интерполяция, защита информации

Введение

Общая структура алгоритмов биометрических систем защиты описывается следующими укрупненными шагами: считывание биометрических данных, обработка биометрических данных или формирование качественных биометрических признаков (далее БП), распознавание биометрических образов на «Свой» и «Чужой». На первом этапе помимо разработки программ считывания биометрических данных и проведения экспериментов можно воспользоваться публичными базами данных [1]. Второй этап предполагает разработку методов, улучшающих качество БП [2], попадающих под рассмотрение методов распознавания биометрических систем защиты на третьем этапе. Выбор метода распознавания биометрических образов на «Свой» и «Чужой» зависит от поставленной цели разрабатываемой биометрической системы защиты: аутентификации [3], идентификации [4, 5] или формирования надежного пароля на основе БП пользователя [6].

Нами рассматривается публичная база данных SVC 2004 и собственная база данных [7, 8]. Публичная база данных включает в себя 1600 подписей 40 пользователей, где на каждого пользователя приходится 20 подлинных подписей и 20 умелых подделок. Собственная база данных включает всего 1508 рукописных записей 29 пользователей, где на каждого пользователя приходится 26 подлинных записей (личные имена испытуемых), 26 простых подделок (рукописное написание фразы «пароль река»). Существует 3 типа подделок: случайная подделка (злоумышленник ничего не знает о рукописной записи), простая подделка (злоумышленник знает имя испытуемого или рукописную фразу в печатном виде), умелая подделка (злоумышленник наблюдал за авторизованным пользователем во время ввода фразы или подписи). В экспериментах в качестве случайной подделки рассматривались различные друг от друга подлинные рукописные записи.

Методы предварительной обработки исходных данных

Все БП характеризуются непостоянностью своих значений, которые идентифицируют биометрический образ. Биометрическими данными являются координаты положения пера на графическом планшете. Здесь наблюдается проблема со сдвигами по оси значений и по оси времени. Для начала мы предлагаем частично устранить такие сдвиги при помощи Z-стандартизации и дискретизации на основе одномерной кусочно-линейной интерполяции. При одномерной кусочно-линейной интерполяции последовательности значений БП разной длины приводятся в последовательности единой длины. В разработанном нами приложении для этой цели использована функция «interp» пакета данных «numpy» языка программирования Python [9].

Перечислим рассматриваемые типы БП с указанием условных обозначений: значения координат положения пера (X , Y); расстояние между действующим положением пера и центром рукописной записи (DC);

значения углов наклона между прямой, соединяющей соседние точки, и горизонталью (A); значения БП, вычисленные методом «скорости перемещения пера» (P); статические признаки (SF). В качестве статических признаков выбраны следующие: отношение высоты к ширине рукописной записи (SF_1); ордината центра рукописной записи (SF_2); абсцисса центра рукописной записи (SF_3); угол наклона между прямой, соединяющей центры двух половин рукописной записи, и горизонталью (SF_4). Приведем примеры условных обозначений БП: гистограммы H_Y , группы БП с использованием метода «интерполяции рукописной записи» Y_I [10].

Мы предлагаем использовать метод с условным названием «интерполяция рукописной записи», где рукописная запись состоит из равноудаленных друг от друга точек. Такая обработка позволит скорректировать влияние ускорения движения пера при считывании координат положения, что позволит частично устранить проблему со сдвигом значений по оси времени.

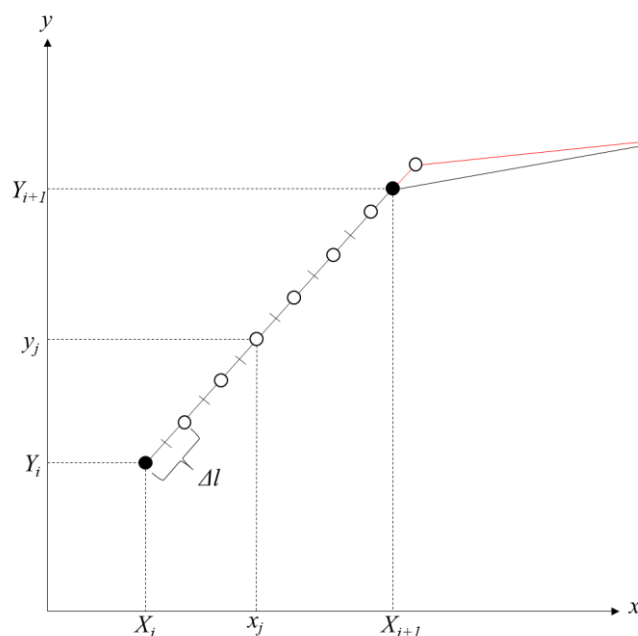


Рис. 1. – Реализация метода «интерполяция рукописной записи»

В представленном примере (рис. 1) имеются исходные точки (X_i, Y_i) и (X_{i+1}, Y_{i+1}) , $i=0, 1, \dots, N$, обозначенные черным цветом. Необходимо найти

координаты точек (x_j, y_j) , $j=0, 1, \dots, M_i$ (точки белого цвета) на прямой $Y=AX+B$, соединяющей исходные точки с соблюдением условия равноудаленности точек (x_j, y_j) друг от друга на шаг Δl . В этом условии могут быть исключения в случае выхода точки (x_j, y_j) за пределы отрезка. Для того, чтобы не нарушать условия равноудаленности точек, точка (X_{i+1}, Y_{i+1}) перемещается в точку (x_{M_j}, y_{M_j}) на расстояние меньше, чем шаг Δl .

Для вычисления координат точек (x_{j+1}, y_{j+1}) преобразованной рукописной записи используются следующие формулы:

$$\begin{cases} x_{j+1} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ y_{j+1} = Ax_{j+1} + B \end{cases}, \quad (1)$$

$$a = 1 + A^2, \quad (2)$$

$$b = 2(AB - Ay_j - x_j), \quad (3)$$

$$c = x_j^2 + (b - y_j)^2 - \Delta l^2, \quad (4)$$

где x_j, y_j – координаты j -ой точки преобразованной рукописной записи; Δl – расстояние между точками (x_j, y_j) и (x_{j+1}, y_{j+1}) .

Значения параметров интерполирующей прямой $Y=AX+B$, соединяющей точки (X_i, Y_i) и (X_{i+1}, Y_{i+1}) исходной рукописной записи, вычисляются по формулам:

$$A = \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i},$$

$$B = Y_i - AX_i.$$

Нами была рассмотрена характеристика скорости перемещения пера, описанная значениями, которые были получены методом с условным названием «скорость перемещения пера». Суть метода заключается в том, чтобы разделить последовательность исходных значений координат X, Y на равные части и рассчитать пройденный пером путь в каждой части.

Оценка качества биометрических признаков

Оценка качества i -го БП по выборке значений B_{Fi} проводится в соответствии с ГОСТ Р 52633.5-2011:

$$Q = q(B_{Fi}) = \frac{|M_{\text{чужой}}(B_{Fi}) - M_{\text{свой}}(B_{Fi})|}{S_{\text{чужой}}(B_{Fi}) + S_{\text{свой}}(B_{Fi})}, \quad (5)$$

где $M_{\text{чужой}}(B_{Fi})$, $M_{\text{свой}}(B_{Fi})$, $S_{\text{чужой}}(B_{Fi})$, $S_{\text{свой}}(B_{Fi})$ – выборочные оценки математических ожиданий и среднеквадратических отклонений БП.

Ниже представлена таблица результатов средних оценок качества БП. Для последовательности значений БП, такого типа как X , Y , DC , A , рассматривались средние значения качества при длине последовательности 20, 25 и 30. Для гистограмм рассматривались количество карманов 3, 4, 5, 6, 7. Для типа БП P рассматривались последовательности значений длины 3, 5 и 7.

В таблице №1 результаты экспериментов наглядно демонстрируют, что метод «интерполяция рукописной записи» значительно увеличивает качество БП. Например, качество последовательностей БП таких типов, как X , Y и DC в среднем по всем базам данных увеличилось не менее, чем на 14%. Также представленный метод улучшил качество БП гистограмм. Например, при публичной базе данных SVC 2004 качество БП таких типов гистограмм как H_X , H_Y и H_{DC} в среднем увеличилось не менее, чем на 5%. На собственной базе данных некоторые показатели качества гистограмм оказались не очень удовлетворительными (например, снижение качества БП H_Y на 1.2%), поэтому не стоит исключать из рассмотрения БП исходных последовательностей X , Y , DC , H_Y , A и т.д.

Таблица № 1

Количественные показатели оценок качества Q_{MEAN}

№ п/п	Тип БП	Q_{MEAN} , случайные подделки (SVC 2004)	Q_{MEAN} , умелые подделки (SVC 2004)	Q_{MEAN} , случайные подделки (собственная база данных)	Q_{MEAN} , простые подделки (собственная база данных)
1	X	1,26	0,96	0,62	0,42
2	X_I	1,5	1,23	0,73	0,47
3	Y	1,18	0,88	0,58	0,3
4	Y_I	1,34	1,07	0,63	0,34
5	DC	1,08	0,83	0,63	0,4
6	DC_I	1,23	1,01	0,73	0,45
7	A	0,64	0,48	0,38	0,23
8	A_I	0,84	0,64	0,37	0,24
9	H_X	1,42	1,11	0,72	0,57
10	H_{XI}	1,55	1,26	0,82	0,58
11	H_Y	1,43	1,06	1,31	0,61
12	H_{YI}	1,49	1,13	1,3	0,6
13	H_{DC}	0,89	0,72	0,62	0,44
14	H_{DCI}	1,07	0,91	0,68	0,47
15	H_A	1,42	1,08	0,91	0,84
16	H_{AI}	1,72	1,27	0,9	0,87
17	P	1,51	1,08	0,97	0,46
18	SF_1	2,43	1,9	1,68	1,19
19	SF_2	1,61	1,23	1,48	0,7
20	SF_3	1,69	1,31	0,84	0,66
21	SF_4	1,56	1,12	3,91	2,39

Представленные статические признаки (SF), описанные единичным значением, обладают относительно высокой степенью качества и хорошо идентифицируют биометрический образ. Несмотря на это, разработчик биометрических систем защиты, учитывая всевозможные условия и внешние факторы, самостоятельно принимает решение включать такие БП в собственную систему защиты или нет.

Метод передискретизации биометрических данных

Экспериментальные данные показали, что при приведении последовательностей значений БП, например таких типов БП как X , Y и DC , к единой длине качество БП в начале и в конце рукописной записи выше, чем в центральной части. Для этого мы предлагаем увеличить частоту точек на краях рукописной записи и сделать передискретизацию последовательностей БП на основе арифметической прогрессии (далее АП).

Пусть N_T – размерность исходной рукописной записи при равномерной дискретизации (условно будем считать период дискретизации равным 1). Необходимо произвести неравномерную передискретизацию с целью получения выборки размерности n . Реализация этой процедуры проводится на основе следующих формул (приведены для двух случаев размерности выборки n):

$$S_{A1} = s_1 + (s_1 + \tau_1) + (s_1 + 2\tau_1) + \dots + s_1 k = \frac{(s_1 + s_1 k)(n_1 + 1)}{2}, \quad (6)$$

$$S_{A2} = s_2 + (s_2 + \tau_2) + (s_2 + 2\tau_2) + \dots + s_2 k = \frac{(s_2 + s_2 k)n_2}{2}, \quad (7)$$

$$2S_{A1} - s_1 k = N_T, \quad (8)$$

$$2S_{A2} = N_T, \quad (9)$$

$$\tau_1 = \frac{2s_1(k-1)}{n_1 - 1}, \quad (10)$$

$$\tau_2 = \frac{2s_2(k-1)}{n_2 - 2}, \quad (11)$$

$$s_1 = \frac{2N_T}{(k+1)(n_1 + 1) - 2k}, \quad (12)$$

$$s_2 = \frac{2N_T}{(k+1)n_2}, \quad (13)$$

где S_{A1} , S_{A2} – суммы АП; s_1 , s_2 – первые члены АП; τ_1 , τ_2 – разности АП; n_1 , n_2 – размерность выборки при нечетном и четном количестве точек

соответственно; k – коэффициент увеличения периода дискретизации в центральной части выборки.

Приведем некоторые примеры. При $N_T=30$, $n_2=8$, $k=2$ получаем следующий массив периодов дискретизации $[s_2, s_2+\tau_2, s_2+2\tau_2, s_2k, s_2k, s_2+2\tau_2, s_2+\tau_2, s_2]$ и преобразованный массив точек передискретизации $[0, 2.5, 5.833, 10, 15, 20, 24.167, 27.5, 30]$. При этом центральные периоды дискретизации в k раз больше, чем первый и последний периоды. При $n_2=7$ получаем следующие периоды дискретизации $[3, 4, 5, 6, 5, 4, 3]$ и преобразованный массив точек передискретизации $[0, 3, 7, 12, 18, 23, 27, 30]$.

Например, при случайных подделках базы данных SVC 2004 среднее значение качества последовательности БП Y равно 1.18, а при неравномерной передискретизации значение увеличилось до 1.36 с коэффициентом $k=5$. Рассматривая Y_t , этот показатель увеличился с 1.34 до 1.51, что в очередной раз доказывает, что при приведении последовательности значений БП к единой длине, качества БП на концах рукописной записи выше, чем в центральной части.

Заключение

В статье предложены методы по обработке исходных биометрических данных рукописного почерка в виде координат положения пера X , Y на графическом планшете. Для этого были рассмотрены случайные, простые, умелые подделки рукописных записей собственной базы данных и известной в литературе открытой базы данных SVC 2004. Нами найден список биометрических признаков, позволяющий идентифицировать биометрический образ. В работе экспериментально доказана целесообразность и эффективность использования таких методов, как: «интерполяция рукописной записи», метод гистограмм, «скорость перемещения пера», неравномерная передискретизация. Хорошие результаты

количественного показателя качества продемонстрировали: статические признаки SF , гистограммы H_{AI} , H_{YI} , H_Y , H_A , H_{XI} , H_X , последовательности X_I , DC_I , Y_I , P . Полученные результаты будут полезны исследователям для разработки и усовершенствования биометрических систем защиты.

Литература

1. Diaz M., Ferrer M.A., Impedovo D., Malik M.I., Pirlo G., Plamondon R. A Perspective Analysis of Handwritten Signature Technology // ACM Computing Surveys. 2019. Vol. 51. № 6. pp. 117-155.
 2. Ponce-Hernandez W., Blanco-Gonzalo R., Liu-Jimenez J., Sanchez-Reillo R. Fuzzy Vault Scheme Based on Fixed-Length Templates Applied to Dynamic Signature Verification // IEEE ACCESS. 2020. Vol. 8. pp. 11152-11164.
 3. Ходашинский И.А., Костюченко Е.Ю., Сарин К.С., Анфилофьев А.Е., Бардамова М.Б., Самсонов С.С., Филимоненко И.В. Аутентификация пользователя по динамике подписи на основе нечёткого классификатора // Компьютерная оптика. 2018. Т. 42. № 4. С. 657-666.
 4. Самойлов А.Н., Сергеев Н.Е., Дайебал Д.Б., Кучерова М.С. Модель интеллектуальной информационной системы для распознавания пользователей социальной сети с использованием биоинспирированных методов // Инженерный вестник Дона, 2018, №4. URL: ivdon.ru/magazine/archive/n4y2018/5411.
 5. Явна Д.В., Бабенко В.В. Метод извлечения из цифровой фотографии лица информации, достаточной для его идентификации // Инженерный вестник Дона, 2014, №3, URL: ivdon.ru/magazine/archive/n3y2014/2494.
 6. Фатхи Д.В., Галушка В.В. Повышение сложности пароля пользователя на основе комплексирования символов пароля и временных интервалов между ними // Инженерный вестник Дона, 2019, №1. URL: ivdon.ru/magazine/archive/n1y2019/5594.
-



7. Yeung D.Y., Chang H., Xiong Y., George S., Kashi R., Matsumoto T., Rigoll G. SVC2004: First International signature verification competition. In Biometric Authentication // Lecture Notes in Computer Science. 2004. Vol. 3072. pp. 16-22.

8. Баянов Б.И. Исследование средств и методов считывания биометрических данных рукописного почерка // Двадцать пятые тулолевские чтения (школа молодых ученых): материалы Международной молодёжной научной конференции. Казань: ИП Сагиева А.Р., 2021. С. 74-77.

9. NumPy Documentation. One-dimensional linear interpolation. URL: numpy.org/doc/stable/reference/generated/numpy.interp.html.

10. Баянов Б.И., Исмагилов И.И. Сравнительный анализ биометрических параметров в задаче формирования криптографического ключа на основе рукописного почерка // Математические методы в технологиях и технике. 2021. № 6. С. 55-58.

References

1. Diaz M., Ferrer M.A., Impedovo D., Malik M.I., Pirlo G., Plamondon R. A Perspective Analysis of Handwritten Signature Technology. ACM Computing Surveys. 2019. Vol. 51. № 6. pp. 117-155.

2. Ponce-Hernandez W., Blanco-Gonzalo R., Liu-Jimenez J., Sanchez-Reillo R. Fuzzy Vault Scheme Based on Fixed-Length Templates Applied to Dynamic Signature Verification. IEEE ACCESS. 2020. Vol. 8. pp. 11152-11164.

3. Khodashinskiy I.A., Kostyuchenko E.Yu., Sarin K.S., Anfilof'ev A.E., Bardamova M.B., Samsonov S.S., Filimonenko I.V. Komp'yuternaya optika. 2018. Vol. 42. № 4. pp. 657-666.

4. Samoylov A.N., Sergeev N.E., Dayebal D.B., Kucherova M.S. Inzhenernyy vestnik Dona. 2018. №4. URL: ivdon.ru/magazine/archive/n4y2018/5411.

5. Yavna D.V., Babenko V.V. Inzhenernyy vestnik Dona. 2014. №3. URL: ivdon.ru/magazine/archive/n3y2014/2494.



6. Fatkhi D.V., Galushka V.V. Inzhenernyy vestnik Dona. 2019. №1. URL: ivdon.ru/magazine/archive/n1y2019/5594.
7. Yeung D.Y., Chang H., Xiong Y., George S., Kashi R., Matsumoto T., Rigoll G. SVC2004: First International signature verification competition. In Biometric Authentication. Lecture Notes in Computer Science. 2004. Vol. 3072. pp. 16-22.
8. Bayanov B.I. Dvadtsat' pyatye tupolevskie chteniya (shkola molodykh uchennykh): materialy Mezhdunarodnoy molodezhnoy nauchnoy konferentsii (Twenty-fifth Tupolev Readings (school of young scientists): Proc. of the International Youth Scientific Symp.). Kazan: IP Sagieva A.R., 2021. pp. 74-77.
9. NumPy Documentation. One-dimensional linear interpolation. URL: numpy.org/doc/stable/reference/generated/numpy.interp.html.
10. Bayanov B.I., Ismagilov I.I. Matematicheskie metody v tekhnologiyakh i tekhnike. 2021. № 6. pp. 55-58.