

УДК 528.854; 519.237

ТОЧНОСТЬ ОПРЕДЕЛЕНИЯ АВТОРСТВА СКАНИРОВАННЫХ РУКОПИСНЫХ ТЕКСТОВ С ПОМОЩЬЮ СТАТИСТИКИ ГРАФЕМ

О. А. Тодорико, Г. А. Добровольский

Запорожский национальный университет

ул. Жуковского, 66, г. Запорожье, 69600, Украина. E-mail: o-sun@rambler.ru; gen@znu.edu.ua

М. Г. Добровольская

Санкт-Петербургский государственный университет

Университетская набережная, 7–9, Санкт-Петербург, 199034, Россия. E-mail: dobrovolska.marina@gmail.com

Исследовался способ решения задачи аутентификации личности по особенностям почерка на основе статических изображений рукописного текста с помощью автоматического поиска и анализа графем, представленных криволинейными отрезками с границами в местах пересечения, слияния, ветвления и окончания линий. Набор найденных графем разбивался на группы с помощью иерархической кластеризации. Признаками почерка служили набор кластеров и статистика распределения в них графем. Расстояние между почерками определялось метрикой хи-квадрат. Для оценки качества метода использовались известные тестовые наборы данных IAM и Firemaker. По результатам вычислительного эксперимента строилась зависимость меры отличия от количества графем в образце, составлялась эмпирическая функция распределения, вычислялась 90 % квантиль. Опытным путем было показано, что для сравнения почерков достаточно иметь не более 2000 графем, что составляет около четырех страниц текста.

Ключевые слова: оффлайнная аутентификация почерка, точность аутентификации, выделение графем, сравнение почерка, кластеризация.

ТОЧНІСТЬ ВИЗНАЧЕННЯ АВТОРСТВА СКАНОВАНИХ РУКОПИСНИХ ТЕКСТІВ ЗА ДОПОМОГОЮ СТАТИСТИКИ ГРАФЕМ

О. О. Тодоріко, Г. А. Добровольський

Запорізький національний університет

вул. Жуковського, 66, м. Запоріжжя, 69600, Україна. E-mail: o-sun@rambler.ru; gen@znu.edu.ua

М. Г. Добровольська

Санкт-Петербурзький державний університет

Університетська набережна, 7–9, Санкт-Петербург, 199034, Росія. E-mail: dobrovolska.marina@gmail.com

Досліджується спосіб вирішення задачі аутентифікації особи за особливостями почерку на основі статичних зображень рукописного тексту за допомогою автоматичного пошуку та аналізу графем, представлених криволинійними відрізками з границями в місцях перетину, злиття, розгалуження й закінчення ліній. Набор знайдених графем розбивався на групи за допомогою ієрархічної кластеризації. Ознаками почерку служили набір кластерів та статистика розподілу в них графем. Відстань між почерками визначалася метрикою хи-квадрат. Для оцінки якості методу розрізнення почерків використовувалися відомі тестові набори даних IAM і Firemaker. За результатами обчислювального експерименту будувалася залежність міри відмінності від кількості графем у зразку, знаходилася емпірична функція розподілу, обчислювалася 90 % квантиль. Дослідним шляхом було показано, що для порівняння почерків досить мати не більше 2000 графем, що складає близько чотирьох сторінок тексту.

Ключові слова: оффлайнна аутентифікація почерку, точність аутентифікації, виділення графем, порівняння почерку, кластеризація.

АКТУАЛЬНОСТЬ РАБОТЫ. Несмотря на развитие информационных технологий рукописные документы все еще важны в юридической, медицинской сферах, исследованиях исторических документов и других областях. Это стимулирует научное исследование проверки авторства рукописного текста. Исследователи опираются на тот факт, что почерк обычно медленно меняется со временем, позволяя надёжно определить человека, написавшего текст.

Традиционно задачи аутентификации почерка подразделяются на несколько классов [1]. Динамические методы отличаются от статических доступностью информации о процессе написания. Зависимость от времени координат пера, силы его давления на поверхность даёт дополнительную информацию и тем самым упрощает аутентификацию. Часть методов, ориентированных на работу с подписью, предполагает, что человек всегда пишет один и тот же текст. Более сложным направлением считается

определения авторства рукописей произвольного содержания. При этом заранее известное содержание документа упрощает задачу аутентификации, но требует предварительного распознавания рукописного текста – довольно сложной операции, которая до сих пор не решена для статического случая.

Предложенный ранее авторами статьи метод [2] ориентирован на работу со статическими изображениями произвольных рукописных текстов неизвестного содержания. Он учитывает, что слова могут быть написаны практически без отрыва пера от бумаги. Рукописный текст такого рода является сложной, никогда не повторяющейся системой пересекающихся кривых, что требует дополнительный усилий для выделения устойчивых признаков почерка.

Аналіз публікації. Существоют разные подходы к выделению графем. Традиционным является разделение слова в точках минимума его верхнего контура [3, 4].

В другом методе [5] следы пера покрывались равными прямоугольниками, и каждый из полученных таким образом фрагментов изображения считался отдельной графемой.

В "теневом" способе [6] для получения графем на бинаризованном изображении места перехода с черного в белый цвет вдоль заранее выбранного направления окрашиваются черным, а остальные пиксели получают белый цвет. Полученное в результате множество коротких линий напоминает тень от выпуклых символов. На втором шаге каждая из линий равномерно расширяется во все стороны, превращаясь в область неправильной формы, которая и считается графемой.

В некоторых случаях из слов выделялись символы, которые и были элементарными графемами [7]. Но такой подход возможен в редких случаях.

Иногда текст разделяли на слова, изображение каждого слова делилось на $n \times m$ частей с одинаковым количеством черных пикселей в каждой из n строк и в каждом из m столбцов. После для каждой ячейки вычислялся набор признаков [8].

Графемам, в зависимости от способа построения, сопоставлялись численные признаки, которые использовались в алгоритмах машинного обучения. Например, для тонких линий вычислялись координаты равномерно распределенных вдоль них точек, количество черных пикселей, координаты особых точек [7]. Для графем, которые являлись закрашенными областями, выбирались: отношение площади доминирующего цвета к площади описанного выпуклого многоугольника, отношение полуосей описанного эллипса, свойства кривизны, моменты распределения пикселей по площади описанного многоугольника, интегральный отзыв детектора контуров, гистограмма градиентов, свойства градиента, структуры, выпуклости и т.д. [1].

Кроме выделения графем для описания рукописного текста вычисляются глобальные признаки, например, распределение градиентов и его интегральные характеристики [8], вероятности появления оттенков серого и вычисленные на её основе "энергия", корреляция, обратный разностный момент (inverse difference moment), энтропия [9].

Для учета текстуры оценивались распределения направлений контуров [9, 10] статистика геометрических и топологических характеристик: количество внутренних контуров и внешних кривых, средняя высота и средний наклон символов строки, толщина штриха, среднее расстояние между словами [9], количество вертикальных, горизонтальных, положительных, отрицательных штрихов, средняя высота и средний наклон символов строки.

Иногда для полностью автоматического извлечения признаков использовалось скользящее окно [11].

В масштабах отдельной строки исследовались проекции строк на горизонтальную линию [12], применялись фильтры Габор [8], вейвлет-преобразование [13], интегральное преобразование контуров [14].

Широко использовалась гистограмма угловой координаты [15], измерение наклона на основе рас-

пределения градиентов, хотя существуют работы, ставящие под сомнение ценность данного признака [16].

Для выделения устойчивых дискриминативных признаков почерка и проверки неизвестных образцов применяются методы машинного обучения. Классификаторы на основе дистанции относят неизвестный образец к одному из классов, вычисляя расстояние от него до каждого класса. Для вычисления расстояния используются разные метрики: Евклидова метрика, расстояние городских кварталов, расстояние хи-квадрат, расстояния Чебышева, Хемминга, Минковского, Махаланобиса, Бхаттачария, Хаусдорфа и другие [17]. Успешность применения метрики для классификации зависит от природы вектора признаков, поэтому трудно сделать однозначный вывод о преимуществе какой-либо формулы расстояния. Каждая из метрик обладает определенными преимуществами и недостатками, и в зависимости от задачи, выбирается наиболее подходящая. Например, Евклидова метрика, в отличие от метрики Чебышева, требует больше вычислительных ресурсов, а быстро вычисляемая метрика Хемминга не всегда дает нужную точность. Многие исследователи считают, что к более точным результатам приводит метрика хи-квадрат [1], хотя она вычисляется медленно. Если вектора признаков бинарные, то, согласно экспериментам, лучшие результаты обеспечивает расстояние корреляции [18].

Классификаторы, основанные на дистанции, не обнаруживают кластеры неправильной формы. В таких случаях лучше использовать байесовские сети, метод опорных векторов, скрытые Марковские модели [19], классификаторы, основанные на плотности точек в пространстве признаков [20, 21], нейронные сети Хемминга, Кохонена, нечеткие классификаторы.

Для оценки метода определения авторства рукописи необходимо иметь достаточное количество образцов почерка. Существует множество различных тестовых наборов рукописных документов на разных языках [1].

Например, база Firemaker содержит 1008 отсканированных страниц написанных на датском языке 252 студентами по 4 страницы каждый [22]. База IAM включает 1539 страниц английского текста, написанных 657 разными людьми [23].

Оценка качества методов сравнения почерков проводится разными способами, например, определяется точность результатов, скорость работы метода, количество требуемых вычислительных ресурсов. Показателем качества системы аутентификации чаще всего выбирают процент правильно сопоставленных почерков. В расширенных исследованиях дополнительно находятся доверительные интервалы для показателя схожести почерков или уровня значимости для статистической гипотезы равенства.

Для исследования разработанного авторами способа [2] сравнения почерков необходимо решить такие задачи: предварительная обработка изображений, выделение пригодных для анализа почерка фрагментов (графем), подбор числовых характеристик для описания каждой графемы, составление

вектора признаков почерка с помощью методов машинного обучения, экспериментальная проверка разработанного способа на известных тестовых наборах Firemaker и IAM, статистический анализ полученных результатов.

Целью данной статьи является подробное исследование данного метода.

МАТЕРИАЛ И РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ. Исходными данными для программной системы сравнения почерков являются изображения рукописного текста, отвечающие определенным требованиям: одинаковое разрешение рисунков, достаточное количество страниц, отсутствие помарок, зачеркиваний и т.д. Перед выделением и анализом графем изображения преобразуются к оттенкам серого, проводится бинаризация и скелетизация.

Разделение изображения на графемы начинается с поиска особых точек, в которых заканчивается 1, 3 или более кривых. Для этого исследуются ближайшая окрестность каждого черного пикселя – квадрат 3×3 . Чтобы точка считалась особой, в её окрестности должны находиться одна, три или четыре раздельные группы черных пикселей.

Случай, когда в окрестности точки расположены две раздельные группы, означает, что это просто часть кривой. После нахождения особых точек сначала находятся соединяющие их отрезки кривых – графемы, и потом петли, которые не содержат особых точек. Таким образом, исходный текст преобразуется в набор кривых (рис. 1).

Полученный таким образом набор графем достаточно большой, что позволяет использовать статистические методы, но затрудняет непосредственное сравнение двух наборов. Чтобы упростить сравнение, размерность описания почерка следует уменьшить, например, за счет кластеризации.

Для разбивки набора графем на кластеры каждому криволинейному отрезку сопоставляется набор числовых признаков и задается функция расстояния между ними. Выбирая признаки, следует помнить, что кластеризация в дискретных пространствах большой размерности затруднена.

Признаками графемы были выбраны: количество черных пикселей, координаты концов и приближенная форма кривой.

Приближенная форма вычисляется в несколько этапов, сначала выбирается начальная точка – конец кривой, находящийся наиболее близко к началу координат. Для замкнутых кривых начальная точка выбирается искусственно, например самая нижняя точка на левой вертикальной касательной.

Далее составляется таблица значений, описывающая кривую в параметрическом виде, например, $x(s)$ и $y(s)$, где s – длина пути вдоль кривой от начальной до текущей точки.

Исследуя зависимости $x(s)$ и $y(s)$, можно найти количество и длину участков возрастания и убывания, количество и длину участков со значениями выше и ниже среднего.

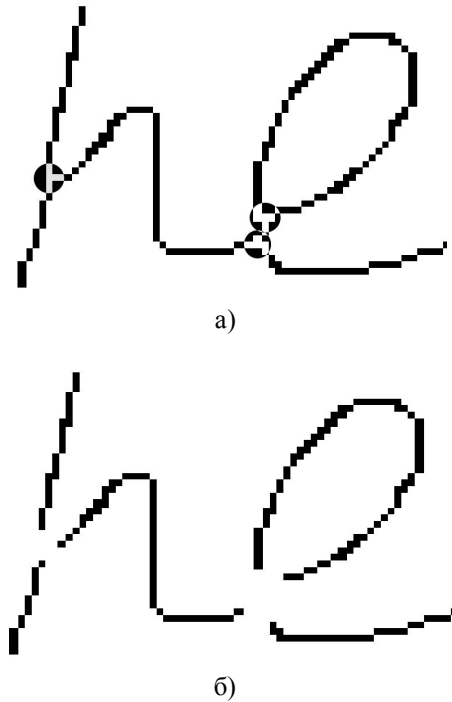


Рисунок 1 – Набор кривых: а) определение особых точек (обозначены кругами); б) полученный после удаления особых точек набор графем

В процессе кластеризации расстояние между наборами признаков вычисляется часто, поэтому функция расстояния должна быть максимально простой. В данной работе функцией расстояния была выбрана метрика Чебышева.

Для ускорения, упрощения и улучшения качества кластеризация выполнялась в несколько итераций, каждая следующая проводилась на результатах предыдущей.

Результатом кластеризации является разбиение пространства признаков на области, содержащие некоторое количество графем.

Такой подход позволяет использовать для дальнейшего сравнения почерка аппарат математической статистики, когда для каждого образца почерка строится распределение вероятности появления графемы в зависимости от номера кластера. Незвестный образец распределяется по тем же кластерам и полученные частоты с помощью критерия Пирсона [17] сравниваются с уже известным распределением.

Экспериментальная проверка. Описанный способ сравнения почерков проверялся на известных тестовых наборах IAM и Firemaker.

В качестве неизвестного образца случайным образом выбирался один из образцов тестовой выборки и вычислялось расстояние χ^2 от этого образца до каждого из оставшихся. По результатам сравнения была построена диаграмма иллюстрирующая зависимость меры отличия от количества графем в образце (рис. 2, 3).

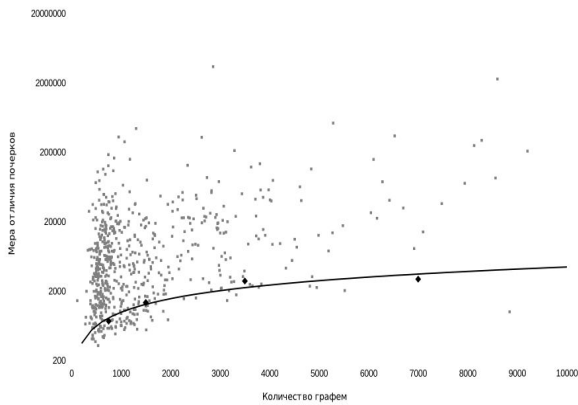


Рисунок 2 – Зависимость меры отличия почерков от количества графем на тестовом наборе IAM

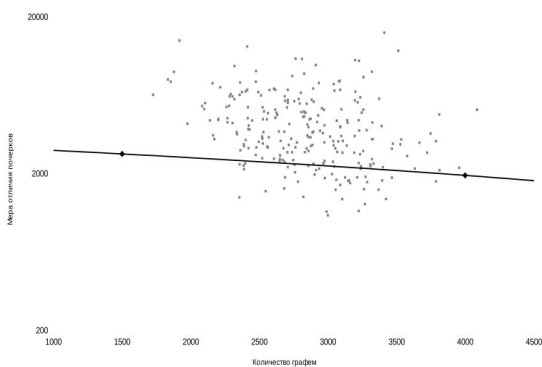


Рисунок 3 – Зависимость меры отличия почерков от количества графем на тестовом наборе Firemaker

На рис. 2 и 3 каждая точка изображает пару (количество графем, мера отличия), и линия показывает оценку для 90 % квантиля эмпирического распределения, которое оценивалось методом Парзена-Розенблатта [24]. Всё, что ниже линии может считаться похожим на неизвестный почерк с вероятностью 90 %. Как и следовало ожидать, увеличение количества признаков положительно сказывается на точности, но, даже имея 2000 и более графем, можно с 90 % вероятностью найти похожие почерки.

ВЫВОДЫ. В случае определения авторства произвольного рукописного текста, в котором символы часто пишутся без отрыва пера от бумаги, используются методы, основанные на выделении графем. В данной работе был предложен способ выделения графем как криволинейных отрезков с границами в местах пересечения, слияния, ветвления и окончания линий. Данный способ больше соответствует интуитивному представлению об элементарных навыках написания слов, и концы кривой вероятнее всего будут местами начала или окончания движения пера. Это позволяет рассматривать графемы, как линии, написанные одним непрерывным подсознательным движением. В результате полученный набор графем анализируется методами машинного обучения, основанные на обработке числовых признаков каждой графемы. В данной работе использовалось меньшее количество числовых характеристик

каждой графемы, чем в аналогичных работах, что позволило сократить время на машинное обучение.

Оценка качества методов различия почерков проводилась на известных тестовых наборах данных IAM и Firemaker. Экспериментальная проверка подтвердила качество работы метода. По результатам эксперимента была построена зависимость меры отличия от количества графем в образце, составлена эмпирическая функция распределения, вычислена 90 % квантиль. Опытным путем было показано, что для сравнения почерков достаточно иметь не более 2000 графем, что составляет около четырех страниц текста.

В дальнейшем планируется модификация предложенного метода за счет улучшенной трассировки линий, более тщательного подбора вектора признаков и способа классификации.

ЛИТЕРАТУРА

1. Awaida. State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text [Electronic resource] / Sameh M. Awaida, Sabri A. Mahmoud // Educational Research and Reviews. – 2012. – Vol. 7 (20). – PP. 445–463. – Access mode: http://www.academicjournals.org/article/article1379684852_Awaida%20and%20Mahmoud.pdf
2. Метод выделения графем для сравнения почерка в сканированных рукописях / О.О. Тодорико, Г.А. Добровольский, М.Г. Добровольская // Вісник ХНТУ. – Херсон: ХНТУ, 2014. – № 3 (50). – С. 174–178.
3. Writer identification by writer's invariants [Electronic resource] / A. Nosary, T. Paquet, L. Heutte // Eighth International Workshop on Frontiers in Handwriting Recognition // IEEE Computing Society, Ontario, Canada. – 2002. – PP. 274–279. – Access mode: <http://lheutte.free.fr/download/iwfh02bensefia.pdf>
4. Automatic Writer Identification Using Connected-Component Contours and Edge-Based Features of Uppercase Western Script / L. Schomaker, M. Bulacu // IEEE Trans. Pattern Anal. Mach. Intel. – 2004. – Vol. 26 (6). – PP. 787–798.
5. A Robust Authentication System Handwritten Documents using Local Features for Writer Identification / Parves Kamal, Faisal Rahman, Saad Mustafiz // Journal of Computing Science and Engineering. – 2014. – Vol. 8, no. 1, March 2014. – PP. 11–16.
6. Static signature verification employing a Kosko-Neuro-Fuzzy approach / K. Franke, Y.-N. Zhang, M. Koppen // In: Pal, N., Sugeno, M. (Eds.), Advances in Soft Computing-AFSS LNAI, 2275. Springer-Verlag, 2002. – PP. 185–190.
7. Schomaker, Lambert. Automatic writer identification using connected-component contours and edge-based features of uppercase Western script [Electronic resource] / Lambert Schomaker, Marius Bulacu // Pattern Analysis and Machine Intelligence, IEEE Transactions on. – Vol. 26, Iss. 6, June 2004. – pp. 787–798 DOI: 10.1109/TPAMI.2004.18. – Access mode: <http://www.ai.rug.nl/~lambert/papers/TPAMI-0101-0503-Schomaker+Bulacu.pdf>
8. Combining Global and Local Features for Writer Identification [Electronic resource] / I. Siddiqi, N. Vincent // Proceedings of the Eleventh International Conference on Frontiers in Handwriting Recognition. –

Montreal, Canada, 2008. – PP. 48–53. – Access mode: <http://www.cenparmi.concordia.ca/ICFHR2008/Proceedings/papers/cr1007.pdf>

9. Comparison of Statistical Models for Writer Verification [Electronic resource] / S. Srihari, G. Ball // *Proceedings on Document Recognition and Retrieval XVI San Jose, CA, USA. – 2009. – PP. 1–8. – Access mode: <http://www.cedar.buffalo.edu/~srihari/papers/SPIE2009-Stat.pdf>*

10. Automatic writer identification using fragmented connected-component contours [Electronic resource] / L. Schomaker, M. Bulacu, K. Franke // *9th International Workshop on Frontiers Handwriting Recognition. – 2004. – pp. 185–190. – Access mode: <http://www.ai.rug.nl/~mbulacu/iwfh2004-schomaker-bulacu-franke.pdf>*

11. Off-Line Handwriting Identification Using HMM Based Recognizers / A. Schlapbach, H. Bunke // *Proc. Int. Conf. Pattern Recognit. 2. – 2004. – PP. 654–658.*

12. Morphological Waveform Coding for Writer Identification [Electronic resource] / E. Zois, V. Anastassopoulos // *Pattern Recognit. – 2000. – Vol. 33 (3). – PP. 385–398. – Access mode: <http://visgraph.cs.ust.hk/biometrics/Papers/Signature/pr2000-03-02.pdf>*

13. Gazzah S, Ben Amara N. Neural Networks and Support Vector Machines Classifiers for Writer Identification Using Arabic Script. // *Int. Arab J. Inform. Technol. 1. – 2008. – PP. 13–75.*

14. A contourlet-based method for writer identification [Electronic resource] / Z. He, Y. Tang, X. You // *IEEE International Conference on System Manufacturing and Cybernetics. – 2005. – Vol. 1. – PP. 364–368. – Access mode: <http://www.comp.hkbu.edu.hk/~pgday/2005/proceedings-jul.pdf#page=64>*

15. Srihari, S. Distance between histograms of angular measurements and its application to handwritten character similarity [Electronic resource] // *Proceedings of the 15th International Conference on Pattern Recognition. IEEE Computing Society. – 2000. – PP. 21–24. – Access mode: http://pdf.aminer.org/000/346/252/distance_between_histograms_of_angular_measurements_and_its_application_to.pdf*

16. Towards Robust Writer Verification by Correcting Unnatural Slant [Electronic resource] / A. Brink, R. Niels, R. van Batenburg, C. van Den Heuvel, L. Schomaker // *Pattern Recognit. Lett. – 2010. – Vol. 32 (3). – PP. 449–457. – Access mode: http://www.researchgate.net/publication/220644231_Towards_robust_writer_verification_by_correcting_unnatural_slant/file/9fcfd5077ab189fe56.pdf*

17. Кластерный анализ / Дюран Б. и Оделл П.; пер. с англ. Е.З. Демиденко, под ред. А.Я. Боярского. – М.: Статистика, 1977. – 128 с.

18. Zhang B. Handwriting Pattern Matching and Retrieval with Binary Features // Ph.D. dissertation, Department of Computer Science and Engineering, State University of New York, Buffalo, NY, 2003. – 172 p.

19. Comparison of Statistical Models for Writer Verification [Electronic resource] / Srihari S., Ball G. // *Proceedings on Document Recognition and Retrieval XVI San Jose, CA, USA, 2009. – Access mode: <http://www.cedar.buffalo.edu/~srihari/papers/SPIE2009-Stat.pdf>*

20. OPTICS: Ordering Points To Identify the Clustering Structure / Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jurg Sander // *ACM SIGMOD international conference on Management of data. ACM Press. – 1999. – PP. 49–60.*

21. A density-based algorithm for discovering clusters in large spatial databases with noise / Martin Ester, Hans-Peter Kriegel, Jurg Sander, Xiaowei Xu // *In Evangelos Simoudis, Jiawei Han, Usama M. Fayyad. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. – 1996. – PP. 226–231.*

22. Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems (research report) / L. Schomaker, L. Vuurpijl // *Nijmegen. – 2000. – 22 p.*

23. The IAM-Database: an English Sentence Database for Offline Handwriting Recognition / U. Marti, H. Bunke // *Int. J. Doc. Anal. Recognit. – 2002. – Vol. 5 (1). – PP. 39–46.*

24. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function // *The Annals of Mathematical Statistics. – 1956. – 27 (3): 832. – Access mode: doi:10.1214/aoms/1177728190.*

ACCURACY OF WRITER IDENTIFICATION OF SCANNED MANUSCRIPTS USING GRAPHEME STATISTICS

O. Todoriko, G. Dobrovolsky

Zaporozhye National University,

vul. Zhukovskiy, 66, Zaporizhzhya, 69600, Ukraine. E-mail: o-sun@rambler.ru; gen@znu.edu.ua

M. Dobrovolska

Saint Petersburg State University,

Universitetskaya naberezhnaya., 7–9, St. Petersburg, 199034, Russia. E-mail: dobrovolska.marina@gmail.com

In this paper is elaborated the suggested before method of writer identification and verification of handwritten text. The method utilizes extracting graphemes from static manuscript images. In the suggested method, graphemes are curvilinear segments, which connect crossing points, branching points and end points on the images. The set of the found graphemes is divided into groups by means of an hierarchical clustering, then the set of clusters and the statistical distribution of graphemes over them are considered as numerical description of handwriting. The distance between handwritings was calculated by using chi-square metrics. To assess the quality of the suggested method of handwriting comparison the known data sets IAM and Firemaker were used. The empirical probability distribution was estimated over number of graphemes and computed distances between a unknown handwriting sample and each sample in dataset, and a 90 % quantile was calculated. It was shown that reliable handwritings comparison requires no more than 2000 graphemes or about 4 pages of the text.

Key words: offline handwriting authentication, authentication accuracy, allocation of graphemes, handwriting comparison, clustering.

REFERENCES

1. Sameh M. Awaida, Sabri A. Mahmoud (2012) "State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text", *Research and Reviews*, vol. 7 (20), pp. 445–463. http://www.academicjournals.org/article/1379684852_Awaida%20and%20Mahmoud.pdf
2. Todoriko, O.A., Dobrovolsky, G.A., Dobrovolska, M.G. (2014) "Graphemes extraction method for handwriting comparison", *Transaction Kherson National Technical University*, vol. 3 (50), pp. 174–178.
3. Nosary, A., Paquet, T., Heutte, L. (2002) "Writer identification by writer's invariants. Eighth International Workshop on Frontiers in Handwriting Recognition", *IEEE Computing Society*, pp. 274–279, Ontario, Canada. <http://lheutte.free.fr/download/iwfh02bensefia.pdf>
4. Schomaker, L., Bulacu, M. (2004) "Automatic Writer Identification Using Connected-Component Contours and Edge-Based Features of Uppercase Western Script", *IEEE Trans. Pattern Anal. Mach. Intell.*, 26 (6), pp. 787–798.
5. Parves, Kamal. (2014) "A Robust Authentication System Handwritten Documents using Local Features for Writer Identification", *Journal of Computing Science and Engineering*, vol. 8, no. 1, March 2014, pp. 11–16.
6. Franke, K., Zhang, Y.-N., Koppen, M. (2002) "Static signature verification employing a Kosko-Neuro-Fuzzy approach", In: Pal, N., Sugeno, M. (Eds.), *Advances in Soft Computing—AFSS 2002, LNAI, 2275, Springer-Verlag*, pp. 185–190.
7. Lambert Schomaker, Marius Bulacu (2004) "Automatic writer identification using connected-component contours and edge-based features of uppercase Western script. Pattern Analysis and Machine Intelligence", *IEEE Transactions on*, vol. 26, iss. 6, June 2004, pp. 787–798 DOI: 10.1109/TPAMI.2004.18 <http://www.ai.rug.nl/~lambert/papers/TPAMI-0101-0503-Schomaker+Bulacu.pdf>
8. Siddiqi I., Vincent N. (2008) "Combining Global and Local Features for Writer Identification", *Proceedings of the Eleventh International Conference on Frontiers in Handwriting Recognition*, pp. 48–53, Montreal, Canada. <http://www.cenparmi.concordia.ca/ICFHR2008/Proceedings/papers/cr1007.pdf>
9. Srihari, S., Ball, G. (2009) "Comparison of Statistical Models for Writer Verification", *Proceedings on Document Recognition and Retrieval XVI San Jose, CA, USA*, pp. 1–8, 7247OE <http://www.cedar.buffalo.edu/~srihari/papers/SPIE2009-Stat.pdf>
10. Schomaker, L., Bulacu, M., Franke, K. (2004) "Automatic writer identification using fragmented connected-component contours", *9th International Workshop on Frontiers Handwriting Recognition*, pp.185–190, <http://www.ai.rug.nl/~mbulacu/iwfh2004-schomaker-bulacu-franke.pdf>
11. Schlapbach, A., Bunke, H. (2004b) "Off-Line Handwriting Identification Using HMM Based Recognizers", *Proc. Int.Conf. Pattern Recognit*, no. 2, pp. 654–658.
12. Zois, E., Anastassopoulos, V. (2000) "Morphological Waveform Coding for Writer Identification", *Pattern Recognit*, vol. 33 (3), pp. 385–398. <http://visgraph.cs.ust.hk/biometrics/Papers/Signature/pr2000-03-02.pdf>
13. Gazzah, S., Ben Amara, N. (2008) "Neural Networks and Support Vector Machines Classifiers for Writer Identification Using Arabic Script", *Int. Arab J. Inform. Technol.*, vol. 1, pp. 13–75.
14. He, Z., Tang, Y., You, X. (2005) "A contourlet-based method for writer identification", *IEEE International Conference on System Manufacturing and Cybernetics*, vol. 1, pp. 364–368. <http://www.comp.hkbu.edu.hk/~pgday/2005/proceedings-jul.pdf#page=64>.
15. Srihari, S. (2000) "Distance between histograms of angular measurements and its application to handwritten character similarity", *Proceedings of the 15th International Conference on Pattern Recognition. IEEE Computing Society*, pp. 21–24. http://pdf.aminer.org/000/346/252/distance_between_histograms_of_angular_measurements_and_its_application_to.pdf.
16. Brink, A., Niels, R., van Batenburg, R., van Den Heuvel, C., Schomaker, L. (2010) "Towards Robust Writer Verification by Correcting Unnatural Slant", *Pattern Recognit. Lett.*, vol. 32 (3), pp. 449–457. http://www.researchgate.net/publication/220644231_Towards_robust_writer_verification_by_correcting_unnatural_slant/file/9fcfd5077ab189fe56.pdf.
17. Duran, B., Odell P. (1974) "Cluster Analysis. A Survey", *Springer-Verlag*, Berlin – Heidelberg – New York, p. 137.
18. Zhang, B. (2003) "Handwriting Pattern Matching and Retrieval with Binary Features", Ph.D. dissertation, *Department of Computer Science and Engineering*, State University of New York, Buffalo, NY, p. 172.
19. Srihari, S., Ball, G. (2009) "Comparison of Statistical Models for Writer Verification", *Proceedings on Document Recognition and Retrieval XVI San Jose, CA, USA*. <http://www.cedar.buffalo.edu/~srihari/papers/SPIE2009-Stat.pdf>
20. Ankerst, Mihael, Breunig, M. Markus, Kriegel, Hans-Peter, Sander, Jürg (1999) "OPTICS: Ordering Points To Identify the Clustering Structure", *ACM SIGMOD international conference on Management of data. ACM Press*, pp. 49–60.
21. Ester, Martin, Kriegel, Hans-Peter, Sander, Jürg, Xu, Xiaowei (1996) "A density-based algorithm for discovering clusters in large spatial databases with noise", In Evangelos Simoudis, Jiawei Han, Usama M. Fayyad, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press, pp. 226–231.
22. Schomaker, L., Vuurpijl, L. (2000) "Forensic Writer Identification: A Benchmark Data Set and a Comparison of Two Systems (research report). Nijmegen.
23. Marti, U., Bunke, H. (2002) "The IAM-Database: an English Sentence Database for Offline Handwriting Recognition", *Int. J. Doc. Anal. Recognit.* 5(1):39-46.
24. Rosenblatt, M. (1956) "Remarks on Some Nonparametric Estimates of a Density Function". *The Annals of Mathematical Statistics* 27 (3): 832. doi:10.1214/aoms/1177728190.

Стаття надійшла 27.09.2014.