

ЗНАЧЕНИЕ И РИСКИ ПРИ ИСПОЛЬЗОВАНИИ АВТОМАТИЗИРОВАННЫХ СИСТЕМ ВЕРИФИКАЦИИ ПОДПИСИ В ЭКСПЕРТНОЙ И ПРЕДЭКСПЕРТНОЙ ДЕЯТЕЛЬНОСТИ

Бахтеев Д.В.

Уральский государственный юридический университет
имени В.Ф. Яковлева,
Екатеринбург, Россия

Статья посвящена авторскому междисциплинарному проекту по разработке интеллектуальной системы верификации рукописных подписей на основе технологии компьютерного зрения. Основная гипотеза исследования состоит в том, что за счёт ограничения вариантов выполнения оригинальной подписи возможно с помощью системы на основе машинного обучения выявить подложную подпись, которая будет выходить за пределы вариационной оригинала. В работе описывается хронометрируемое анкетирование, благодаря которому была определена точность верификации подписи человеком: 70,1 % для носителей специальных знаний в области почерковедения и 60,7 % для иных респондентов. Рассматривается собранный датасет изображений рукописных подписей, приводится его сравнение с иными датасетами в этой области. Дается краткая характеристика технологии сравнения подписей и структуры экспериментального прототипа, включающего мобильное приложение и серверную часть. Описывается круг технических и организационных областей рисков, связанных с разработкой прототипа и потенциальным прикладным использованием предлагаемого технологического решения. Указывается на возможное несоответствие содержимого датасета реальным объектам, проблему оценки результатов в количественной форме, ориентированность на предэкспертное, а не экспертное применение таких систем поддержки принятия решений, возможные методы противодействия в виде использования роботизированных плоттеров.

Ключевые слова: судебная почерковедческая экспертиза, машинная верификация рукописных подписей, искусственный интеллект, цифровизация криминалистики, предэкспертная верификация.

Процессы цифровизации в настоящее время вполне можно назвать всеобъемлющим. Они не только создают новые векторы в правоохранительной деятельности, среди которых особого внимания заслуживают новые системы наблюдения в оперативно-розыскной деятельности, исследования цифровых следов, применение дронов и транскайберов при производстве следственных действий и пр., но и позволяют применить отдельные технологические решения к классическим задачам, таким как исследование рукописных подписей.

Эта задача стоит не только перед следователем или судьёй, как первичными инициаторами исследования, либо судебным экспертом, как лицом, которое разрешает

Адрес для корреспонденции: Бахтеев Дмитрий Валерьевич, профессор кафедры криминалистики Уральского государственного юридического университета имени В. Ф. Яковлева, доктор юридических наук, доцент, Россия г. Екатеринбург, ae@crimlib.info

спорную ситуацию, но и перед любым должностным лицом, чьи функции предполагают проверку документов, в которых одним из ключевых реквизитов является рукописная подпись. Несмотря на распространение цифровых документов, в которых единственным защитным реквизитом является электронная подпись, классические бумажные акты вряд ли полностью исчезнут в ближайшие десятилетия, в связи с чем разработка методов автоматизации их исследования не утрачивает своей актуальности.

На примере проекта, реализуемого под руководством автора настоящей работы, рассмотрим технологию машинной верификации подписей, способы её применения и риски при использовании.

Первоначальная гипотеза для эксперимента звучит следующим образом:

- подпись человека обладает некоторым набором вариантов её исполнения;
- подложная подпись, выполненная другим человеком, будет полностью или отдельными фрагментами выходить за поле нормального распределения таких вариантов.

Если автоматизированная система, функционирующая на основе технологии компьютерного зрения, изучит в процессе обучения достаточное количество подписей, то она сможет с достаточной правильностью верифицировать спорную подпись – то есть определять её оригинальность либо подложность.

Такая достаточность может быть определена экспериментально. Так, 257 респондентам, 14 из которых являются действующими судебными экспертами-почерковедами, была предложена специальная анкета, в которой ставилась задача определить, какие из четырёх подписей, размещённых рядом с достоверно оригинальной, являются подложными. Средняя правильность ответов составила 70,1 % для лиц, обладающих специальными знаниями в области почерковедения, и 60,7 % – для остальных. Полагаем, что именно первое значение следует считать минимально пороговым для оценки эффективности прототипа информационной системы по верификации подписей.

Данное анкетирование хронометрировалось, при этом время не было ограничено. В среднем респонденты затрачивали 12 секунд на одно сравнение (оригинальная подпись – спорная подпись). Даже этот достаточно скромный показатель всё равно существенно больше того времени, что тратит следователь, судья, прокурор или иное лицо на визуальную верификацию подписи в документе. Соответственно, для лиц, не обладающих специальными знаниями, следует предположить более пессимистичные в аспекте правильности верификации результаты на практике.

В связи с этим, представляется необходимым создать систему предэкспертной верификации рукописной подписи, ориентированной на лиц, не обладающих специальными познаниями в сфере почерковедения, которая сможет формировать результат с точностью выше 70,1 % за время, не превышающее нескольких секунд.

В рассматриваемом проекте реализуется так называемое офлайновое распознавание подлога подписей, для реализации которого требуется предварительное обучение искусственной интеллектуальной системы на специализированном массиве данных.

В структуре датасета, названного «NSP», нами реализовано следующее соотношение для одного человека: не менее 56 оригинальных подписей и не менее 112

поддельных подписей. Такие количественные показатели сформированы на основе базового постулата почерковедческой экспертизы: чем больший объём почеркового материала представлен на исследование, тем выше вероятность правильной идентификации и формулирования категорического вывода. Особенно это важно в экспертизе подписей, поскольку данные реквизиты отличаются высокой степенью вариационности. Соответственно для обучения системы была избрана модель воспроизведения условий судебной экспертизы, тогда как на практике происходит сопоставление одной оригинальной подписи с одной спорной.

Всего в сборе данных для описываемого проекта приняло участие 673 человека, общий объём датасета составляет 13,3 ГБ, или 116 193 отдельных изображений подписей. Исходя из оценки работы конкурирующих научных коллективов (см. Таблицу 1) 1-5, очевидно, что уже сейчас настоящий датасет является крупнейшим в мире по числу изображений и количеству реальных людей, задействованных в получении данных, и при этом он продолжает расширяться.

Таблица 1.

Датасеты подписей

Название датасета	Система письма	Количество респондентов подписей	Тип используемых подписей	Количество подписей (подлинные / подложные)	Страна
NSP	Кириллица	673	естественные	116 193 (38 730 / 77 463)	Россия
CEDAR [1]	Латиница	55	естественные	2640 (1320 / 1320)	США
MCYT [2]	Латиница	330	естественные	16 500 (8250 / 8250)	Испания
GPDS-960 [3]	Латиница	960	сгенерированные	51 849 (23 049 / 28 800)	Испания
BHSig260 [4]	Индийское письмо (хинди и бенгальский языки)	260	естественные	14 040 (6240 / 7800)	Индия
The Uyghur signatures database [5]	Уйгурское арабское письмо	500	естественные	24 800 (17400 / 7400)	Китай

Разумеется, размер датасета не является единственным ключевым критерием, но без него создание современных систем на базе машинного обучения попросту невозможно. Кроме того, благодаря таким проектам возможно расширение знаний сугубо криминалистических, к примеру, сбор датасета позволил выявить отдельные особенности как формирования, так и подлога рукописных подписей [6].

Не вдаваясь глубоко в технические подробности работы моделей, опишем общий принцип их функционирования. Во-первых, происходит предпроцессорная обработка изображения подписи: удаляются лишние шумы, подпись центрируется, фон по возможности стирается. Сам процесс сравнения подписей (оригинальной и спорной) реализуется в двух формах. Первая из них предполагает одновременное перемещение окна прямоугольной формы по изображениям подписей с целью выявления дистанции отклонения спорной подписи от оригинала, то есть, по сути, поиск похожих элементов. Во-вторых, производится сравнение направлений штрихов в подписи относительно друг друга.

В настоящее время мы видим прикладную реализацию описываемого проекта в форме мобильного приложения и связанного внешнего сервера. В мобильном приложении пользователь фотографирует оригинальную и спорную подписи, изображения которых затем отправляются на сервер, где и происходит их обработка методами машинного обучения. Внешний вид приложения представлен на рисунке 1.

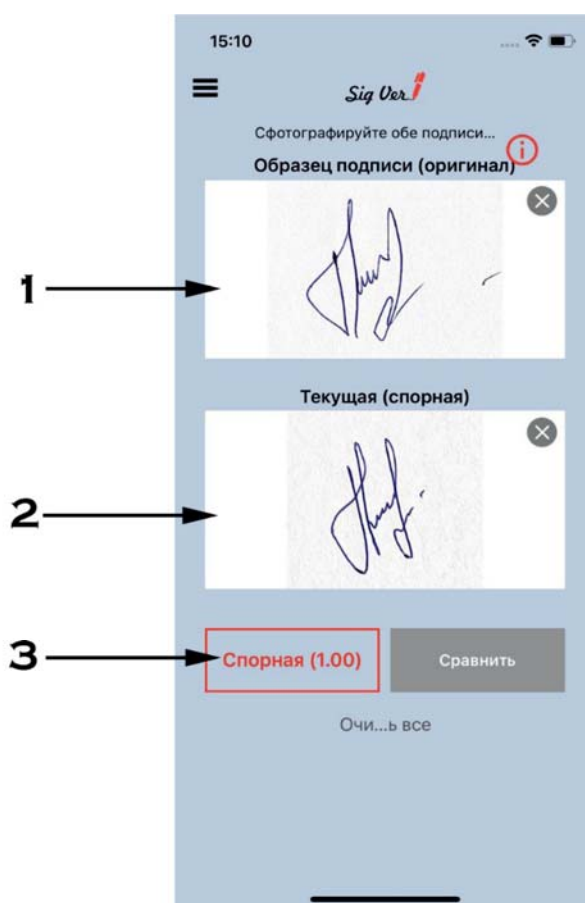


Рис. 1. 1 – место для загрузки фотографии достоверно оригинальной подписи;
2 – место для загрузки фотографии спорной подписи;
3 – результат сравнения изображений подписей.

Учитывая, что модели обучения предполагали сравнение оригинальной и спорной подписей, следует говорить о том, что данный прототип предназначен для решения идентификационных задач. В будущем планируется добавление функций, нацеленных на диагностику отдельных состояний исполнителя рукописной подписи.

В экспертном исследовании диагностический этап является, как правило, более простым, хотя зачастую и более объёмным. В случае с автоматизированными системами идентификация является достаточно простой задачей: на этапе обучения искусственные нейросети получают пары подписей с разметкой, то есть известно, какой статус соответствует каждой из них, на этапе верификации результатов обучения известен только статус оригинальной подписи, а статус второй нужно определить, что в целом линейно воспроизводит операции, осуществлявшиеся в процессе обучения. Диагностические же задачи требуют дополнительной ручной разметки датасета. В настоящее время это возможно осуществить с наименьшими усилиями в отношении признаков возраста, пола и ведущей руки.

Однако, несмотря на описанную актуальность и важность разрабатываемого нами инструмента, как и любое технологическое решение, данная система не свободна от рисков, связанных с её использованием.

Во-первых, возможно радикальное несоответствие содержимого датасета тем подписям, которые будут исследоваться на практике. Речь может идти про цветовые схемы: в датасете подписи отсканированы, цвета как штрихов, так и фона достаточно корректные. На практике документы часто фотографируют без вспышки, соответственно, белая бумага выглядит на изображении серой или даже коричневой. Эта проблема может быть решена добавлением в датасет блока сфотографированных подписей в противовес сканированным.

Во-вторых, если судебный эксперт может по результатам исследования представить ограниченное количество выводов, то результат работы интеллектуальной системы выражается в процентном соотношении (степени уверенности системы в правильности результатов). Какое решение должен принять правоприменитель, если при сопоставлении подписей он увидит результат, к примеру, «вероятность того, что подпись выполнена другим человеком составляет 71 %»? Для одного человека этот показатель достаточен, чтобы инициировать дополнительные процедуры проверки, например, в виде назначения судебной экспертизы, для другого 29 % сомнений достаточно, чтобы считать подпись оригинальной.

В-третьих, результаты анкетирования демонстрируют разницу между восприятием спорных подписей у экспертов-почерковедов и прочих лиц. В качестве получателя рассматриваемой системы в первую очередь мы видим лиц, не обладающих специальными познаниями в области почерковедения. Для них автоматизированная система верификации подписей будет не более чем механизмом поддержки принятия решения, тогда как для экспертов такая «поддержка» слабо отличима от навязывания решения. Как указывалось ранее, для каждой спорной подписи решение эксперта сводится к тому, чтобы определить, является ли её исполнителем лицо, от имени которого она выполнена. Представим ситуацию, в которой судебный эксперт при производстве почерковедческой экспертизы пользуется подобной системой. Если он соглашается с результатом машинной обработки, возможны два варианта: либо система навязывает эксперту решение, либо решения эксперта и системы параллельны. В

первом случае происходит вытеснение человека из процесса экспертного исследования, что недопустимо с позиций современной юридической науки. Во втором случае не ясно, зачем квалифицированному специалисту такой инструмент, если его компетентность позволяет принимать решения самостоятельно. Аналогичный вывод возможен и для ситуации, когда выводы эксперта и информационной системы противоречат друг другу.

В-четвёртых, нельзя отрицать существование такой технологии, как роботизированные плоттеры, способные выполнять рукописный текст ручкой по модели движения руки человека, зачастую в четырёх плоскостях: по высоте, длине, ширине и с произвольным наклоном [7]. На дальнейших этапах реализации проекта мы планируем добавить в датасет подписи, выполненные таким образом.

Подводя итог, отметим, что научное сообщество криминалистов и представителей технических отраслей науки должны объединить свои усилия в деле цифровизации криминалистики и судебной экспертизы. Представленный проект демонстрирует возможности такого междисциплинарного взаимодействия.

Исследование выполнено за счет гранта Российского научного фонда № 23-78-10011, <https://rscf.ru/project/23-78-10011/>.

Список литературы:

1. Srinivasan H. Machine learning for signature verification // Computer Vision, Graphics and Image Processing. Berlin, Heidelberg, 2006.
2. Ortega-Garcia J. [et al]. MCYT baseline corpus: a bimodal biometric database // IEE Proceedings. Vision Image and Signal Processing. 2003. DOI: 10.1049/ip-vis:20031078.
3. Vargas F., Ferrer M., Travieso C., Alonso J. Off-line Handwritten Signature GPDs-960 Corpus // Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). Curitiba, 2007.
4. Jain A., Singh S. K., Singh K. P. Handwritten signature verification using shallow convolutional neural network // Multimedia Tools and Applications. 2020. № 79 (14). DOI: 10.1007/s11042-020-08728-6.
5. Ubul K., Yimin A., Gheni Z. et al. Creation of Uyghur Off-Line Handwritten Signature Database // Proceedings of the 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016). Beijing, 2016.
6. Bakhteev D. V. Criminalistic conclusions on signature forgery process during building an offline signature verification intellectual system // Nowa Kodyfikacja Prawa Karnego. 2021. Vol. 60. DOI: 10.19195/2084-5065.60.2.
7. Рыбалкин Н. А. Актуальные проблемы судебной экспертизы при выявлении документов, выполненных с помощью современных технических средств // Технологии XXI века в юриспруденции: Материалы Пятой международной научно-практической конференции, Екатеринбург, 19 мая 2023 года. Екатеринбург: АНО «Центр содействия развитию криминалистики «КримЛиБ»», 2023. С. 176–183.

ՓՈՐՁԱԳԻՏԱԿԱՆ և ՆԱԽԱՓՈՐՁԱԳԻՏԱԿԱՆ ԳՈՐԾՈՒՆԵՈՒԹՅԱՆ ՄԵՋ ՍՏՈՐԱԳՐՈՒԹՅԱՆ ՍՏՈՒԳՄԱՆ ԱՎՏՈՄԱՏԱՑՎԱԾ ՀԱՄԱԿԱՐԳԵՐԻ ՕԳՏԱԳՈՐԾՄԱՆ ՆՇԱՆԱԿՈՒԹՅՈՒՆԸ և ՌԻՍԿԵՐԸ

Բախտյեև Դ.Վ .

Սույն հոդվածը նվիրված է հեղինակային միջմասնագիտական նախագծին, որը վերաբերում է համակարգչային տեսողության տեխնոլոգիայի հիման վրա ձեռագիր ստորագրությունների ստուգման խելացի համակարգի մշակմանը: Հետազոտության հիմնական վարկածը կայանում է նրանում, որ բնօրինակ ստորագրության կատարման տարբերակների սահմանափակման հաշվին հնարավոր է մեքենայական ուսուցման վրա հիմնված համակարգի միջոցով հայտնաբերել կեղծ ստորագրությունը, որը դուրս կգա բնօրինակի փոփոխականության սահմաններից: Աշխատանքում նկարագրվում է ժամանակային հարցում, որի շնորհիվ որոշվել է անձի կողմից ստորագրության ստուգման ճշգրտությունը՝ 70,1% ձեռագրագիտության ոլորտում հատուկ գիտելիքներ կրողների համար և 60,7% այլ հարցվածների համար: Դիտարկվում է ձեռագիր ստորագրությունների տվյալների հավաքածուն, և տրվում է դրա համեմատությունը այլ տվյալների հավաքածուների հետ այս ոլորտում: Տրվում է ստորագրությունների համեմատության տեխնոլոգիայի համառոտ նկարագրություն և փորձարարական նախադրյալի կառուցվածք, որը ներառում է բջջային հավելվածը և պահոցային մասը: Նկարագրվում է տեխնիկական և կազմակերպչական ռիսկային ոլորտների շրջանակ, որը կապված է նախադրյալի մշակման և առաջարկվող տեխնոլոգիական լուծման հնարավոր կիրառման հետ: Ցույց է տրվում տվյալների հավաքածուի բովանդակության հնարավոր անհամապատասխանությունը իրական օբյեկտների հետ, քանական արդյունքների կերպով գնահատման խնդիրը, որոշումների կայացման, համակարգերի նախափորձնական, այլ ոչ թե փորձագիտական կիրառման կողմնորոշումը, ինչպես նաև հակազդեցության հնարավոր մեթոդները՝ ռոբոտային պլուրերների օգտագործման տեսքով:

Բանալի բառեր. դատաձեռագրաբանական փորձաքննություն, ձեռագիր ստորագրությունների մեքենայական ստուգում, արհեստական բանականություն, դատաբժշկական փորձաքննության թվայնացում, նախափորձագիտական ստուգում :

SIGNIFICANCE AND RISKS OF USING AUTOMATED SIGNATURE VERIFICATION SYSTEMS IN EXPERT AND PRE-EXPERT ACTIVITIES

Bakhteyev D.V.

The article is devoted to the author's interdisciplinary project on the development of an intelligent system for verifying handwritten signatures based on computer vision

technology. The main hypothesis of the research is that by limiting the variants of the original signature, it is possible to use a machine learning-based system to identify a forged signature that will go beyond the variability of the original. The paper describes a timed survey, thanks to which the accuracy of signature verification by a person was determined: 70.1% for holders of special knowledge in the field of handwriting analysis and 60.7% for other respondents. The collected dataset of handwritten signature images is considered, and its comparison with other datasets in this field is given. A brief description of the signature comparison technology and the structure of the experimental prototype, including a mobile application and a server part, is given. A range of technical and organizational risk areas associated with the development of the prototype and potential applied use of the proposed technological solution is described. The possible discrepancy between the dataset content and real objects, the problem of evaluating the results in quantitative form, the focus on pre-expert rather than expert use of such decision support systems, possible countermeasures in the form of using robotic plotters are indicated.

Keywords: forensic handwriting expertise, machine verification of handwritten signatures, artificial intelligence, digitalization of forensics, pre-expert verification.

Статья поступила: 15.08.2024
Принята к печати: 25.03.2025