

УДК 004.93'1

ОГЛЯД НАЯВНИХ ПІДХОДІВ ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ

П.В. Степашко

*Інститут прикладного системного аналізу, НТУУ «Київський політехнічний інститут»,
просп. Перемоги, 37, корп.35
pstepashko@gmail.com*

Стаття містить короткий огляд задачі розпізнавання рукописного тексту і підходів до її розв'язку. В ній розглянуто дві основні категорії розпізнавачів рукописного тексту – онлайн і офлайн, та визначено різницю між ними. Також показано продуктивність декількох комерційних реалізацій систем розпізнавання рукописного тексту і намічено напрямки подальшого розвитку.

Ключові слова: розпізнавання рукописного тексту, онлайн, офлайн, штрих, співставлення шаблону, статистичні методи, структурні методи, нейронні мережі.

This paper presents a brief overview of the handwriting recognition problem and the approaches to its resolution. It overviews the two main categories of handwriting recognizers – online and offline, and defines the distinction between them. Also, the paper shows the performance of some commercial implementations of handwriting recognition software and outlines the directions of further research.

Keywords: handwriting recognition, online, offline, stroke, template matching, statistical techniques, structural techniques, neural networks.

Стаття содержит краткий обзор задачи распознавания рукописного текста и подходов к ее решению. В ней рассмотрены две основные категории распознавателей рукописного текста - онлайн и офлайн, и определено различие между ними. Также показано производительность нескольких коммерческих реализаций систем распознавания рукописного текста и намечены направления дальнейшего развития.

Ключевые слова: распознавание рукописного текста, онлайн, оффлайн, штрих, сопоставление шаблона, статистические методы, структурные методы, нейронные сети.

Вступ

Рукописне введення було відмінним засобом комунікації та документації протягом тисяч років, але останнім часом воно все частіше замінюється введенням з клавіатури. Тим не менше, воно залишається найпростішим і найбільш ефективним за часом способом перетворення людської думки в передавану форму для більшості людей в усьому світі. У часи підйому комп'ютерної ери, використання почерку в обчислювальних машинах було менш зручно застосовувати у великих масштабах, але в пізніші дні, з новітнім розвитком в технологіях сенсорних пристроїв та швидких і точних камер і сканерів, перетворення тексту, написаного людиною, в дані, зрозумілі комп'ютеру, є актуальним завданням. В даний час системи розпізнавання рукописного тексту мають нові горизонти для досягнень в таких областях як машинний переклад, поштові пересилки, перевірки підпису, ведення повсякденних нотаток і т.п. Кінцева мета розпізнавання рукописного введення

– це існування машин, які можуть читати будь-який текст з тією ж точністю розпізнавання, що і люди, але з більшою швидкістю.

1. Властивості почерку і спричинені ними проблеми у розпізнаванні

Письмова мова має алфавіт символів (або букв), розділові знаки і т.д. Фундаментальна властивість письма, яка робить можливим спілкування, полягає в тому, що відмінності між різними символами більш значні, ніж відмінності між різними написаннями такого ж символу.

Почерк складається з часової послідовності штрихів, тобто рухів пера від дотику до паперу до його підняття. Символи письма, як правило, формуються послідовно, один символ завершується перед початком наступного, а літери слідує певному просторовому порядку, наприклад, зліва направо. Є й винятки – наприклад, в англійській та мові точки над і та перетинки на t, як правило, затримуються. Спершу пишеться основна частина слова, а потім слово завершується написанням точок та перетинок.

Всі літери розрізняються їх статичними і динамічними властивостями. Статичні відмінності можуть полягати, наприклад, в розмірі або формі символу. Відмінності в динаміці можуть полягати в кількості штрихів і їх порядку. Англійська може мати більш значні зміни напрямку штриха, ніж китайська. В англійській також більше варіативності в присутності або за відсутності повторних штрихів. Повторний штрих – це штрих в місці, де щось вже написано, зазвичай він робиться, щоб уникнути піднімання пера. Ступінь відмінностей залежить від стилю і швидкості написання, поспішне письмо, як правило, проявляє більшу варіацію [1].

Є багато задач розпізнавання образів для письма і малювання на планшетах. Вони охоплюють відрізнення мовних символів, формул, креслення і жестових символів, наприклад таких, що використовуються в редагуванні. Проблеми розпізнавання символів мови включають в себе, наприклад, великий алфавіт китайських ієрогліфів, японські хірагану і катакану, корейський хангиль, арабську та рукописні алфавіти і шрифти в західних мовах. Проте найбільші проблеми в розпізнавання рукописного тексту – це ті, які викликають труднощі у людей при спробі прочитати навіть власний почерк.

По-перше, це той факт, що більшість символів можуть бути записані по-різному. Наприклад, на рис. 1 наведено різні можливі стилі письма англійською мовою [2], які передова система повинна бути в змозі розрізнити і розпізнавати. Також рідко можна знайти двох людей з дійсно схожим почерком. Ця задача пов'язана з розрізненням шрифтів в класичній задачі розпізнавання тексту. Проте, на відміну від шрифтів, кожна літера в письмі однієї людини може мати інший стиль залежно від контексту, в якому відбувається написання (рис. 1), навколишніх літер, та інших факторів. Щоб справитися з цією проблемою, багато систем включають в себе компонент, який навчається отримуваному почерку, відрізняє користувачів, і застосовує ці дані при прийнятті рішення.

По-друге, декілька символів часто виглядають однаково, або навіть не відрізняються, в почерку однієї людини. Крім очевидних прикладів, таких як O і 0 або I, L і 1, є також випадки f і t, g і q, l і e, і так далі. Більш того, деякі літери можуть бути написані недбало, і об'єктивно виглядати більш схожими на інші. Ця складність зазвичай долається шляхом додавання системи розпізнавання контексту і використання словника, в якому програма може знайти сумнівне слово, і таким чином усунути неоднозначність.

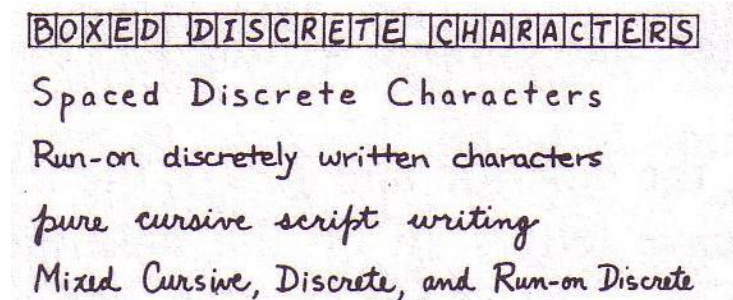


Рис. 1. Різні стилі письма англійською мовою

2. Онлайн-розпізнавання

Зазначимо одразу, що цей термін не має нічого спільного з Інтернетом. Онлайн-розпізнавання рукописного тексту означає, що машина розпізнає написане під час письма користувача. Замість «онлайн» також використовувались терміни «ріал-тайм» та «динамічне» розпізнавання. Залежно від техніки розпізнавання і швидкості комп'ютера, розпізнавання відстає від написання більшою чи меншою мірою. Більшість комерційних систем відстають тільки на один або два символи, більш сучасні реалізації зовсім не мають відставання. Системи онлайн-розпізнавання мають бути швидкими лише настільки, щоб встигати за письмом користувача. Середня швидкість письма дорівнює 1,5-2,5 символи на секунду для англійських букв і цифр або 0,2-2,5 символів/сек для китайських ієрогліфів [1]. Пікова швидкість письма англійською мовою може сягати 5-10 символів/с, наприклад, послідовність 1 та e можна записати дуже швидко. Для більшості алгоритмів розпізнавання вимоги продуктивності легко виконуються сучасними мікропроцесорами.

Для онлайн-розпізнавання рукописного введення вимагається перетворювач, який на ходу зберігає письмо [2, 3, 5]. Найбільш поширеним з цих пристроїв до останніх днів був електронний планшет, або ж діджитайзер, який має роздільну здатність не менше 200 точок на дюйм, частоту дискретизації 100 точок/с, та здатність сприйняти "маркування", або опускання ручки. В даний час найбільш зручними пристроями для розпізнавання рукописного тексту є планшетні комп'ютери, оскільки вони забезпечують найкращий зворотний зв'язок – відразу показують символи, записані пером користувача, і в основному мають достатньо обчислювальної потужності, щоб встигати за введенням.

В більшості випадків, онлайн-дані розпізнаються негайно, але можуть бути застосовані, для яких розпізнавання більш доречно здійснити пізніше.

На відміну від офлайн-систем, онлайн-розпізнавачі здатні захопити часову та динамічну інформацію про письмо: *кількість* штрихів, *порядок* штрихів, *напрямок* написання штриха, і *швидкість* написання. *Штрих* – це рух пера від дотику до паперу до його підняття. В англійській мові, великі літери в середньому пишуться приблизно двома штрихами, а малі одним штрихом, при цьому швидке рукописання може в середньому мати менше одного штриха на літеру [3]. Через те, що онлайн-розпізнавання використовує як динамічну, так і статичну інформацію, воно може бути більш точним, ніж офлайн-розпізнавання.

У такій системі, рух кінчика стилуса (пера) оцифровується через рівні проміжки часу з використанням цифрового планшета і він передається в комп'ютер, який виконує алгоритм розпізнавання рукописного тексту. У більшості систем, сигнал даних проходить певний процес фільтрації. Потім сигнал нормалізується до стандартного розміру і коригується його нахил і крутизна. Після нормалізації написане зазвичай сегментується на базові блоки, і кожен сегмент проходить класифікацію і маркування. Використовуючи алгоритм пошуку в контекстній мовній моделі, найбільш імовірний шлях повертається користувачеві у вигляді очікуваного рядка (див. рис. 2) [6].

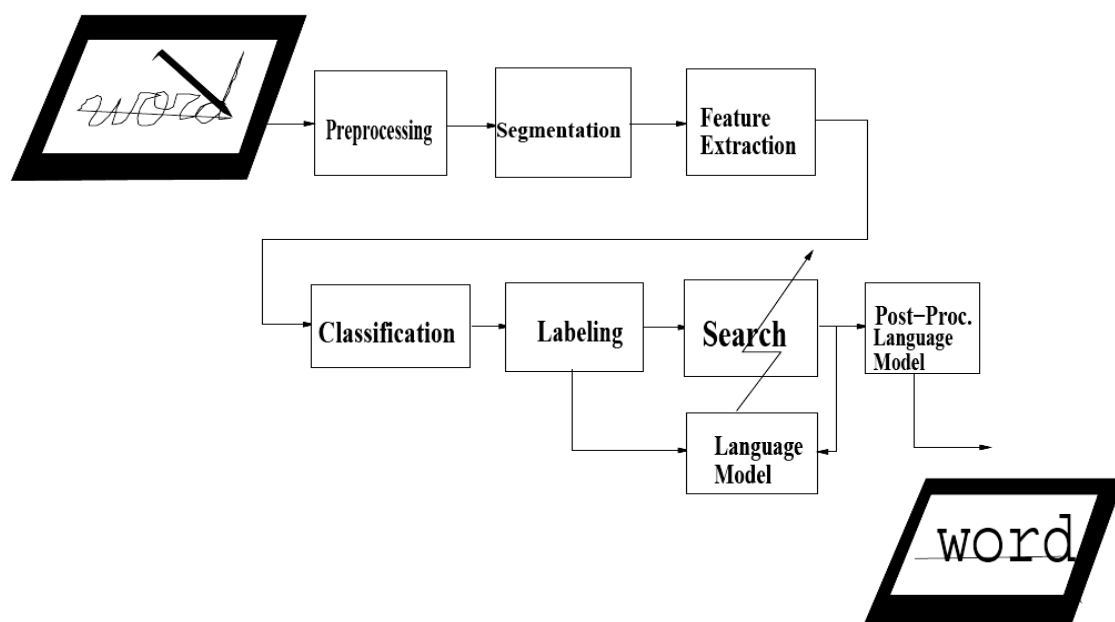


Рис. 2. Загальний процес розпізнавання

Ця динамічна інформація може бути корисна в розрізненні між символами однакової форми, наприклад, 5 та S. Якщо символ пишеться двома штрихами – це, як правило, 5. Крім того, 5, написана одним штрихом часто може бути вирізнена за динамічною інформацією – паузи або сповільнення в кутах, навіть якщо вони різко не виділяються.

Проте динамічна інформація також ускладнює процес розпізнавання. Необхідність обробляти штрихи змушує машину вивчати значно більше різних способів написання символів, додаючи це до і без цього великої кількості зображень, які має пам'ятати система. Наприклад, букву Н можна записати одним, двома або трьома штрихами, з різним порядком і напрямком штрихів, але отримане зображення буде практично те ж саме. Якщо штрихи нечітко розділені, система повинна ігнорувати можливі випадкові штрихи. Таким чином, складність проблеми штрихів і двох напрямків, в яких штрихи можуть бути написані, множить на кількість способів, якими кожен набір штрихів можуть бути накреслений і поєднаний в реальному рукописі.

3. Офлайн-розпізнавання

Офлайн-розпізнавання рукописного тексту, навпаки, виконується після завершення запису [3]. Воно може бути виконано дні, місяці або роки по тому. Оптичний сканер перетворює зображення письма в набір бітів. Сканери зазвичай мають розділення 300-400 точок/дюйм по осях x та y. Системи оптичного розпізнавання символів здатні обробляти сотні символів за секунду.

Офлайн-розпізнавання рукописного тексту є підмножиною задачі оптичного розпізнавання символів (OCR). Хоча більшість робіт з OCR зосереджувались на розпізнаванні друкованих символів, чимало зусиль було витрачено і на рукописний текст.

Інша відмінність між онлайн і офлайн розпізнаванням – в захопленні даних почерку [4]. Онлайн збір даних означає, що машинні дані збираються під час написання символів людиною. Офлайн запис даних означає, що машинні дані виділяються через деякий час після того, як запис створено. Після збору розпізнавач може обробити онлайн- чи офлайн-дані в будь-який час. При цьому офлайн-методи можуть застосовуватись до даних, отриманих динамічно, наприклад, починаючи розпізнавання при завершенні написання слова.

4. Методи і підходи

Є чимало методів для розпізнавання рукописного тексту. Загалом, більшість з них можна застосувати і до онлайн, і до офлайн підходів до розпізнавання. Основна відмінність між ними полягає в наборі характеристик, що розпізнаються. В [7] передбачено поділ на чотири загальних підходи розпізнавання образів, а саме співставлення шаблонів, статистичні методи, структурні методи і нейронні мережі. Такі підходи не обов'язково є незалежними чи відокремленими один від одного. Іноді метод в одному підході також можна віднести і до інших.

Операції співставлення шаблонів визначають ступінь подібності між двома векторами (групами пікселів, фігур, вигинів і т.д.) у просторі ознак. Відповідні методи можуть бути згруповані в три класи: пряме співставлення [8], пружне співставлення [9], і релаксаційне співставлення [10].

Статистичні методи стосуються статистичних функцій прийняття рішень і набір оптимальних критеріїв, що визначають ймовірність належності картини, що спостерігається, до певного класу. Кілька популярних підходів розпізнавання рукописного належать до цієї групи:

- Правило k -найближчих сусідів (k -NN) є популярним непараметричним методом розпізнавання, де апостеріорна ймовірність приблизно визначається з частоти найближчих сусідів невідомого шаблону. Такий підхід, як повідомляється, дає досить хороші результати, наприклад, [11, 12], хоча слід зазначити, що він вимагає значних обчислювальних потужностей в процесі класифікації.

- Прихована модель Маркова (ПММ) є ще одним популярним способом вирішення задачі. Модель є двічі стохастичним процесом: вона включає неспостережуваний процес (звідси "прихована"), що, однак, може спостерігатися через інший випадковий процес, який продукує послідовність спостережень. ПММ широко визнана одним з найпотужніших інструментів для моделювання мови та широкого кола інших реальних сигналів. Ці ймовірнісні моделі мають багато властивостей, бажаних для моделювання символів або слів. Однією з найбільш важливих властивостей є наявність ефективних алгоритмів для автоматичного навчання моделі без необхідності маркування попередньо сегментованих даних. ПММ широко застосовуються для розпізнавання рукописних слів ([13, 14]), в тому числі в поєднанні з іншими підходами, такими як стохастичні граматики [15] і нейронні мережі [16]. Є два основних підходи, які можна виділити в зазначених роботах: модель-дискримінантні ПММ і шлях-дискримінантні ПММ. У першому випадку модель будується для кожного класу (слова, літери, або одиниці сегментації) у фазі підготовки, в другому одна ПММ будується для всієї мови або контексту. Продуктивність обох в основному порівнянна [14].

- Метод опорних векторів (Support Vector Machine – SVM) заснований на статистичній теорії навчання і квадратичній оптимізації. SVM – це, по суті, двійковий класифікатор, і кілька SVM можуть бути об'єднані, щоб сформувати систему для класифікації декількох класів. В останні роки SVM отримав більшу вагу в середовищі машинного навчання через його відмінну продуктивність щодо узагальнення. Було розроблено кілька систем класифікації, заснованих на SVM, для розпізнавання рукописних цифр, і деякі обнадійливі результати були зафіксовані, наприклад, в [17, 18, 19].

Існують також декілька інших статистичних підходів, такі як байєсівський або поліноміальний дискримінантний класифікатори, та вони менш популярні для задачі, що розглядається, і в більшості випадків дають гірші результати.

У структурних методах символи представлені у вигляді об'єднань структурних символів-примітивів, що отримані з почерку, піддаються кількісній оцінці, і можна знайти взаємозв'язок між ними. Структурні методи можна розділити на два класи: граматичні [20] та графічні методи [21].

Найбільш універсальний підхід до задачі – нейромережевий. Нейронна мережа – це обчислювальна структура, яка складається з штучних нейронів – абстракції нервових клітин людини. Створені при спробах імітувати людський мозок, ці структури широко використовуються в розпізнаванні образів, обробці даних та задачах апроксимації функцій. Основні переваги нейронних мереж полягають в здатності навчатися автоматично на основі вибірок, бути продуктивною на зашумлених даних, можливістю паралельної реалізації та бути ефективними інструментами для обробки великих баз даних. Нейронні мережі широко використовуються в області, що розглядається, і були досягнуті перспективні результати, особливо в розпізнаванні рукописних цифр.

Найбільш широко досліджуваною і використовуваною нейронною мережею є багат шаровий перцептрон (Multi-Level Perceptron – MLP). Така структура, що навчається за допомогою зворотного поширення помилки, є однією з найбільш популярних і універсальних форм нейронних мереж-класифікаторів і одним з найбільш часто використовуваних для розпізнавання рукописного тексту [22]. У цьому підході є багато різних методів. Найбільш успішними можна назвати нечіткі нейронні мережі [23], мережу Хеммінга [24], мережу Хопфілда [25], самоорганізувальні карти [26] та ін.

Очевидно, всі перелічені методи мають свої переваги і недоліки. Щоб обійти недоліки, можуть використовуватись комбінації підходів. Розроблялись різні схеми поєднання класифікаторів і експериментально було показано, що деякі з них стабільно випереджають кращі з окремих класифікаторів [13, 27].

5. Деякі комерційні реалізації

Найбільш очевидне застосування розпізнавання рукописного тексту нині знаходиться в портативних пристроях із сенсорними екранами. Наприклад, лінійка планшетів і великих смартфонів Samsung Galaxy Note, що базуються на застосуванні високоточного стилуса, включає цю можливість з метою ведення нотаток. Схоже, що дана система використовує офлайн-підхід, оскільки вона починає розпізнавати текст тільки після того, як нічого не було написано протягом певного часу. Рис. 3 показує процес розпізнавання.

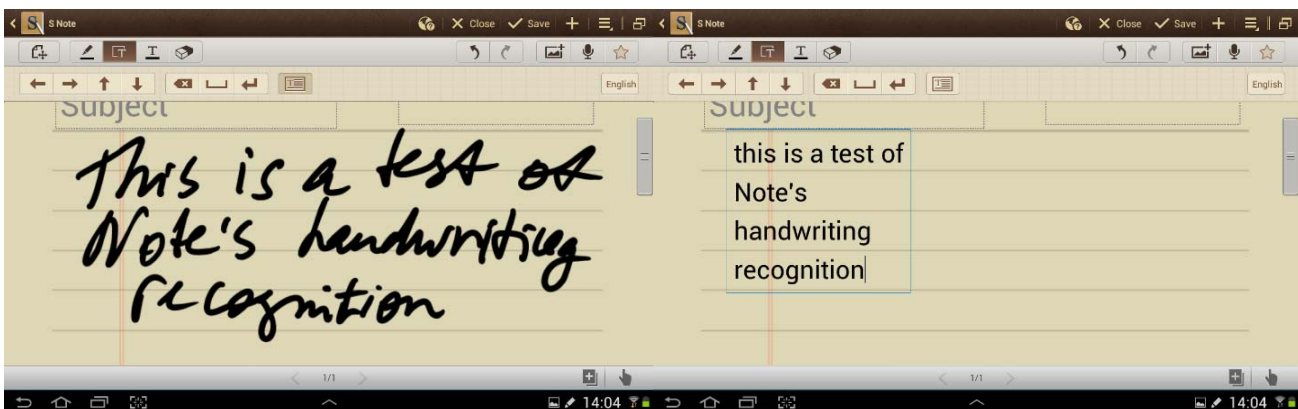


Рис. 3. Процес розпізнавання рукописного тексту на планшеті Samsung Galaxy Note 10.1

Інший приклад використання розпізнавання рукописного тексту в комерційному пристрої є мобільний додаток Google Translate, що дозволяє писати текст однією мовою і миттєво отримувати його переклад іншою. На відміну від попереднього прикладу, використовується онлайн-техніка, про що свідчить те, що ще нерозпізнані штрихи підсвічуються, а текст коригується при додаванні нових штрихів. Це показано на рис. 4.

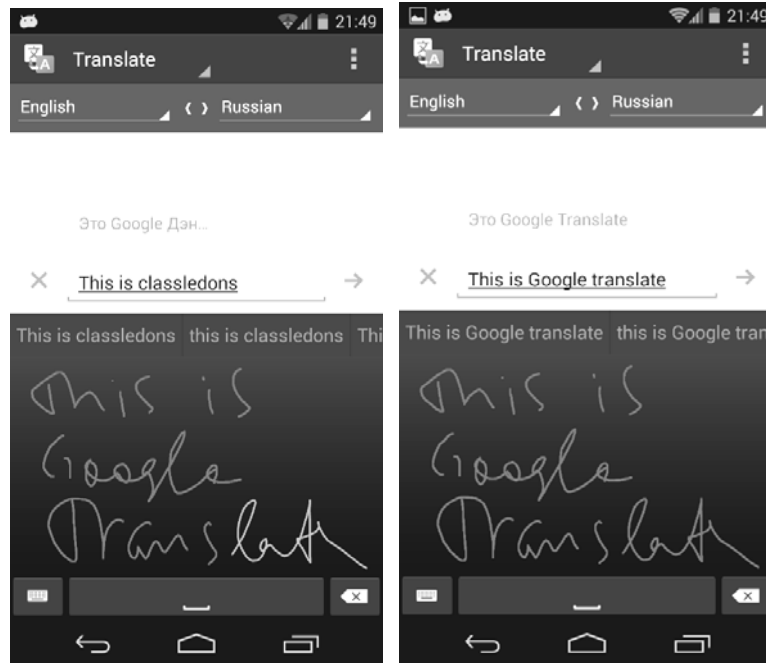


Рис. 4. Процес розпізнавання рукописного тексту в програмі Google Translate для Android.

6. Висновок

У даній роботі розглянуто наявні підходи до задачі розпізнавання рукописного тексту. Дані для розпізнавання можуть бути отримані двома різними способами – офлайн та онлайн. Основна відмінність між ними полягає в наборі ознак, що розпізнаються. Якщо перший спосіб є простим і може працювати з даними, що були отримані без використання будь-якого спеціалізованого обладнання, другий є більш точним, але вимагає значно більш складного набору даних. Поєднання різних методів розпізнавання ознак слід вважати основним напрямком розвитку в цій галузі.

Важливою частиною будь-якої системи розпізнавання символів є підсистема сегментації. Розрізнення написаних слів у зображенні та виділення літер у слова є складним завданням, яке вимагає не менше уваги, ніж власне процес розпізнавання. Ще більш важливою є система виділення ознак, яка має знайти значущі властивості виділених букв і відкинути неважливі.

Сучасна машина розпізнавання рукописного тексту не може існувати без словника і підсистеми розпізнавання контексту. Вони дозволяють машині використовувати зовнішні дані для вирішення конфліктних ситуацій,

наприклад, визначення відмінності між малими і великими літерами або розумінням неясно написаного знаку.

Враховуючи досягнення в цій галузі, не дивно, що останнім часом швидкі і точні програмні системи розпізнавання рукописного тексту були застосовані в комерційних продуктах, таких як мобільний додаток Google Translate, серія портативних пристроїв Samsung Galaxy Note та Microsoft Surface. В майбутньому слід очікувати на ще більш широке впровадження таких засобів.

Література

1. Tappert C.C., Suen C.Y., Wakahara T.: The State of the Art in On-Line Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 12, No. 8, August 1990.
2. Charles C. Tappert, Sung-Hyuk Cha: English Language Handwriting Recognition Interfaces. *Text Entry Systems*, ed. MacKenzie and Tanaka-Ishii, Morgan Kaufman, 2007.
3. Drissman Avi: Handwriting Recognition Systems: An Overview. <http://www.drissman.com/avi/school/HandwritingRecognition.pdf>
4. Arica N. and Yarman-Vural F.T.: An Overview of Character Recognition Focused on Off-Line Handwriting. *Machine Intelligence*, 22(1): 4-37. *IEEE Trans. on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 31, No. 2, May 2001
5. Nathan K.S., Beigi H.S.M., Subrahmonia J., Clary G.J., Maruyama H.: Real-Time On-line Unconstrained Handwriting Recognition Using Statistical Methods. *Int. Conf. on Acoustics, Speech, and Signal Processing*, 1995. ICASSP-95.
6. Homayoon S., Beigi M.: An overview of handwriting recognition. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.7795&rep=rep1&type=pdf>
7. Jain K., Duin R.P.W., Mao J.: Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22 (1): 4-37.
8. Gader P.D., Forester B., Ganzberger M., Billies A., Mitchell B., Whalen M., Youcum T.: Recognition of handwritten digits using template and model matching. *Pattern Recognition*, 1991, 5 (24): 421-431.
9. Dimauro G., Impedovo S., Pirlo G., Salzo A.: Automatic bankcheck processing: A new-engineered system. In S.Impedovo et al, editor, *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, 1997.
10. Xie S.L., Suk M.: On machine recognition of hand-printed chinese character by feature relaxation. *Pattern Recognition*, 1988, 21 (1): 1-7.
11. Perez-Cortes J.C., Llobet R., Arlandis J.: Fast and Accurate Handwritten Character Recognition using Approximate Nearest Neighbours Search on Large Databases. https://prhlt.iti.upv.es/demos/demo_forms/Perez-cortes00b.pdf
12. Mico L., Oncina J.: Comparison of fast nearest neighbour classifier for handwritten character recognition. *Pattern Recognition Letters*, 1999, 19 (3-4): 351-356.

13. Yacoubi E., Gilloux M., Sabourin R., Suen C.Y.: An hmm-based approach for offline unconstrained handwritten word modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, 21(8):752-760.
14. Kundu, Y. He, M. Chen. Alternatives to variable duration hmm in handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.
15. Makhoul J., Starner T., Schwartz R., Chou G.: On-Line Cursive Handwriting Recognition Using Hidden Markov Models and Statistical Grammars. <http://acl.ldc.upenn.edu/H/H94/H94-1086.pdf>.
16. Bengio Y., LeCun Y., Nohl C., Burges C.: A NN/HMM Hybrid for On-Line Handwriting Recognition. *Neural Computation*, Volume 7, Number 5, 1997.
17. Ayat N.E., Cheriet M., Suen C.Y.. Optimization of the SVM kernels using an empirical error minimization scheme. In *Proc. of the International Workshop on Pattern Recognition with Support Vector Machine*, Niagara Falls-Canada, August 2002, pp. 354-369.
18. Oliveira L.S., Sabourin R.: Support Vector Machines for Handwritten Numerical String Recognition. *9th International Workshop on Frontiers in Handwriting Recognition*, October 2004.
19. Ahmad A.R., Viard-Gaudin C., Khalid M., Yusof R.: Online Handwriting Recognition using Support Vector Machine. *Proceedings of the Second International Conference on Artificial Intelligence in Engineering & Technology*, August 3-5 2004, pp. 250-256.
20. Shridhar M., Badreldin A.: Recognition of isolated and simply connected handwritten numerals. *Pattern Recognition*, 1986, 19(1):1-12.
21. Kim H.Y., Kim J.H.: Handwritten korean character recognition based on hierarchical random graph modeling. In *Proc. 6th International Workshop on Frontiers of Handwriting Recognition*, Taegon-Korea, August 1998, pp. 557-586.
22. Zhang G.P.: Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 2000 - Part C: Applications and Reviews, 30(4):451-462.
23. Gómez Sánchez E., Dimitriadis Y.A., Más M.S.-R., García P.S., Cano Izquierdo J.M., Coronado J. L.: On-Line Character Analysis and Recognition with Fuzzy Neural Networks. *Intelligent Automation and Soft Computing*, Vol. 7, No. 3, pp. 161-162, 1998.
24. Verma B., Blumenstein M., Kulkarni S.: Recent Achievements in Off-Line Handwriting Recognition Systems. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.8866&rep=rep1&type=pdf>.
25. Ling, M. Lizaraga, N. Gomes, A. Koerich: A prototype for brazilian bankcheck recognition. In S.Impedovo et al, editor, *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific 1997, pp. 549-569.
26. Kittler J., Hatef M., Duin R., Matas J.: On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(3):226-239.