CrossMark

A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation

Raashid Hussain[1], Ahsen Raza[2], Imran Siddiqi[3], Khurram Khurshid[4*] and Chawki Djeddi[5]

Abstract

Handwriting has remained one of the most frequently occurring patterns that we come across in everyday life. Handwriting offers a number of interesting pattern classification problems including handwriting recognition, writer identification, signature verification, writer demographics classification and script recognition, etc. Research in these and similar related problems requires the availability of handwritten samples for validation of the developed techniques and algorithms. Like any other scientific domain, the handwriting recognition community has developed a large number of standard databases allowing development, evaluation and comparison of different techniques developed for a variety of recognition tasks. This paper is intended to provide a comprehensive survey of the handwriting databases developed during the last two decades. In addition to the statistics of the discussed databases, we also present a comparison of these databases on a number of dimensions. The ground truth information of the databases along with the supported tasks is also discussed. It is expected that this paper would not only allow researchers in handwriting recognition to objectively compare different databases but will also provide them the opportunity to select the most appropriate database(s) for evaluation of their developed systems.

Keywords: Standard databases, Handwritten samples, Handwriting recognition

1 Review

The availability of databases is a fundamental requirement for development and evaluation in all scientific research domains. Standard datasets provide a platform for comparison and evaluation of different techniques on the same grid thus abstracting any possible bias [1, 2]. The task of collecting samples for database development is naturally cumbersome and tedious as it involves getting a maximum possible variety of samples from sundry participants. Having standard databases not only prevents the researchers from compiling the databases but also provides them with an opportunity to have an objective as well as comparative performance evaluation of their developed systems. Benchmark construction is not just the accumulation of samples, but an organized process of cull and abnegation of samples to be included

in the database. Like any other scientific domain, document analysis and recognition community (DAR) has also developed a large number of document databases. The most researched and significant task in document analysis and recognition is handwriting recognition. Naturally, most of the standard databases developed by the document recognition community are handwritten databases.

The process of development of handwritten databases is as old as the problem of document analysis and recognition itself. This development of standard databases started to receive a notable attention in the early 1990s and the process still continues. Most important and widely used handwritten databases include IAM [2–4], RIMES [5], NIST [6], MNIST [7], CENPARMI [8–13], CEDAR [14], UNIPEN [15], ETL9 [16] and PE92 databases [17]. Although most of these databases have been developed using text in languages based on the Latin alphabet, development of databases in Chinese [18], Korean [17], Arabic [13, 19–22], Farsi [10, 12, 23, 24] and Indian scripts [25] is also on the rise. The trend of multi-script handwritten databases [26, 27] has also been observed in the last

*Correspondence: khurram.khurshid@ist.edu.pk
[4]Institute of Space Technology, Islamabad, Pakistan
Full list of author information is available at the end of the article

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 2 of 24

few years. These handwritten databases comprise a variety of samples including handwritten digits [6, 7, 13, 21, 28], characters [14, 17, 29–31], words [13, 14, 19, 21, 23, 24, 28], or complete sentences [3, 5, 18, 26, 27]. A step ahead to benchmark construction is the organization of evaluation campaigns and competitions allowing researchers to compare their systems under the same experimental setups.

This paper is intended to provide a comprehensive survey of the handwritten databases developed during the last two decades. We not only discuss the statistics of these databases but also present a comparative analysis on different dimensions including the size of database, number of contributors, textual content of the database, data acquisition mode (online or offline), writing script and the tasks which could be evaluated on a given database. This study is likely to be helpful for researchers in selecting the most appropriate databases for evaluation of their developed systems. We first discuss the basics of handwriting benchmarks in Section 2 followed by a detailed review of the well-known handwritten databases, their structure and usage in Section 3. Section 4 provides an overview of the evaluation campaigns and competitions organized using these databases while the last section concludes the paper with a discussion on future trends on the subject.

2 Handwriting benchmarks: basics

Research in handwriting recognition and related problems has been carried out in online as well as offline domains. Benchmarks have, therefore, been developed both for offline and online analysis of handwriting. Offline samples of handwriting are collected by making individuals write on paper with a typical writing instrument (pen or a pencil) and digitizing the paper documents using a scanner. Online databases of handwriting are produced by requiring the subjects to directly write on a digitizing tablet or similar devices. Writing is produced using a stylus or directly through finger. In addition to the writing strokes in terms of x-y coordinates of the pen position, online handwriting also contains additional information including pen pressure, writing speed, stroke order, etc. Offline datasets of handwritten text may comprise alphanumeric characters, isolated words, or complete paragraphs. Generally, these databases are produced by requiring the subjects to fill standardized forms with already specified or an arbitrary text. These forms are then scanned into a digital format. Online handwriting databases also comprise isolated characters, words, or sentences. Since the collection of online data requires the subjects to directly produce their samples on digitizing devices, online data collection is generally considered relatively easier but naturally requires specialized hardware for acquisition of samples.

The next step after data collection is the labeling of data to produce the ground truth. The ground truth associated with a database determines the tasks that could be evaluated using the database. Labeling is generally carried out at character, word, or line levels to support the traditional preprocessing, segmentation and recognition tasks. In addition, some databases also support evaluation of tasks like document layout analysis, word spotting, writer demographics classification, writer identification and writer verification.

The next section presents a detailed discussion on the handwritten databases developed during the last two decades.

3 Handwriting benchmarks survey: structure and usage

A large number of handwritten benchmark datasets supporting the evaluation of a variety of preprocessing, segmentation and recognition tasks have been developed over the years. These database could be categorized on different dimensions including the data acquisition method (online or offline), script, size, or the types of tasks supported. In our discussion, we have grouped the databases as a function of the script of the writing samples. These include the databases of Roman/Latin script, Chinese, Japanese and Korean (CJK) writings covering East Asian languages, Arabic and Arabic-like scripts and different Indian scripts. The handwritten databases developed in each of these scripts are discussed in the following.

3.1 Databases in the Roman script

The Roman or Latin script is the most widely used writing system based on the letters of classical Latin alphabet. With minor variations, Roman script covers English, French, German, Spanish, Portuguese, Swedish and Dutch languages. Some other languages have also migrated to this script, Malaysian and Indonesian being the most notable of these. Consequently, a significant proportion of the handwritten databases comprise text in the Roman script. The following sections discuss in detail the well-known handwriting databases in the Roman script.

3.1.1 IAM databases

The IAM databases are easily the most widely used collections of handwritten samples employed for a variety of segmentation and recognition tasks. A number of offline and online databases have been developed under the IAM umbrella as discussed in the following.

IAM-DB:IAM handwriting database The IAM Handwriting Database [2, 3] comprises handwritten samples in English which can be used to evaluate systems like text

segmentation, handwriting recognition, writer identification and writer verification. The database is developed on Lancaster-Oslo/Bergen Corpus and comprises forms where the contributors copied a given text in their natural unconstrained handwriting. Each form was subsequently scanned at 300 dpi and saved as gray level (8-bit) PNG image. A complete filled form, sample lines of text and some words extracted from a sample form in the database are illustrated in Fig. 1. The IAM Handwriting Database 3.0 includes contributions from 657 writers making a total of 1539 handwritten pages comprising 5685 sentences, 13,353 text lines and 115,320 words. The database is labeled at sentence, line and word levels. The database has been widely used in word spotting [32–35], writer identification [36–40], handwritten text segmentation [41–43] and offline handwriting recognition [44–47].
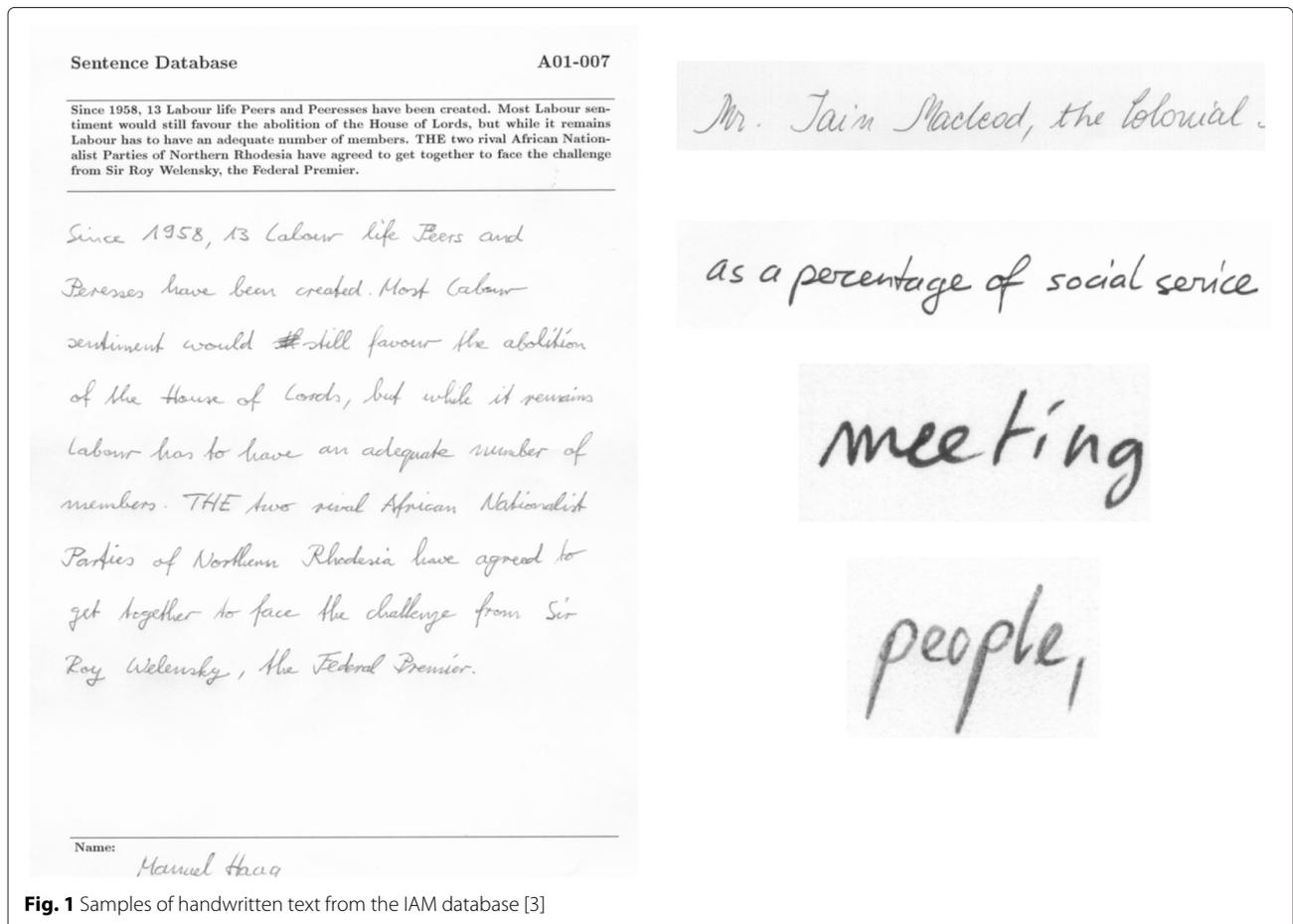
IAM On-Line Handwriting Database (IAM-OnDB) IAM-OnDB [4] is a collection of online handwritten samples on a white board acquired with the E-Beam System. Data is stored in xml-format which, in addition to the transcription of text, also contains information on writers and writer demographics. The database comprises 221 writers contributing a total of more than 1700 forms with 13,049 labeled text lines and 86,272 word instances from a dictionary of 11,059 words. In addition to recognition of online handwriting [48, 49], the database has also been employed for online writer identification [50] and gender classification from handwriting [51].

IAM Online Document Database (IAM on-Do) The IAM on-Do [52] is a relatively new database of online handwritten documents containing text, drawings, diagrams, formulas, tables, lists and markings as indicated in Fig. 2. The database can be employed for document layout analysis and different segmentation and recognition tasks. The database consists of 1000 documents produced by approximately 200 writers. Few constraints were imposed on the writers while creating the documents. Nonetheless, the database has a stable distribution of the different content types and presents a collection of samples close to those encountered in real-world scenarios. The database has been employed for mode (content type) detection [53], keyword spotting [54] and classification of text/non-text objections [55].

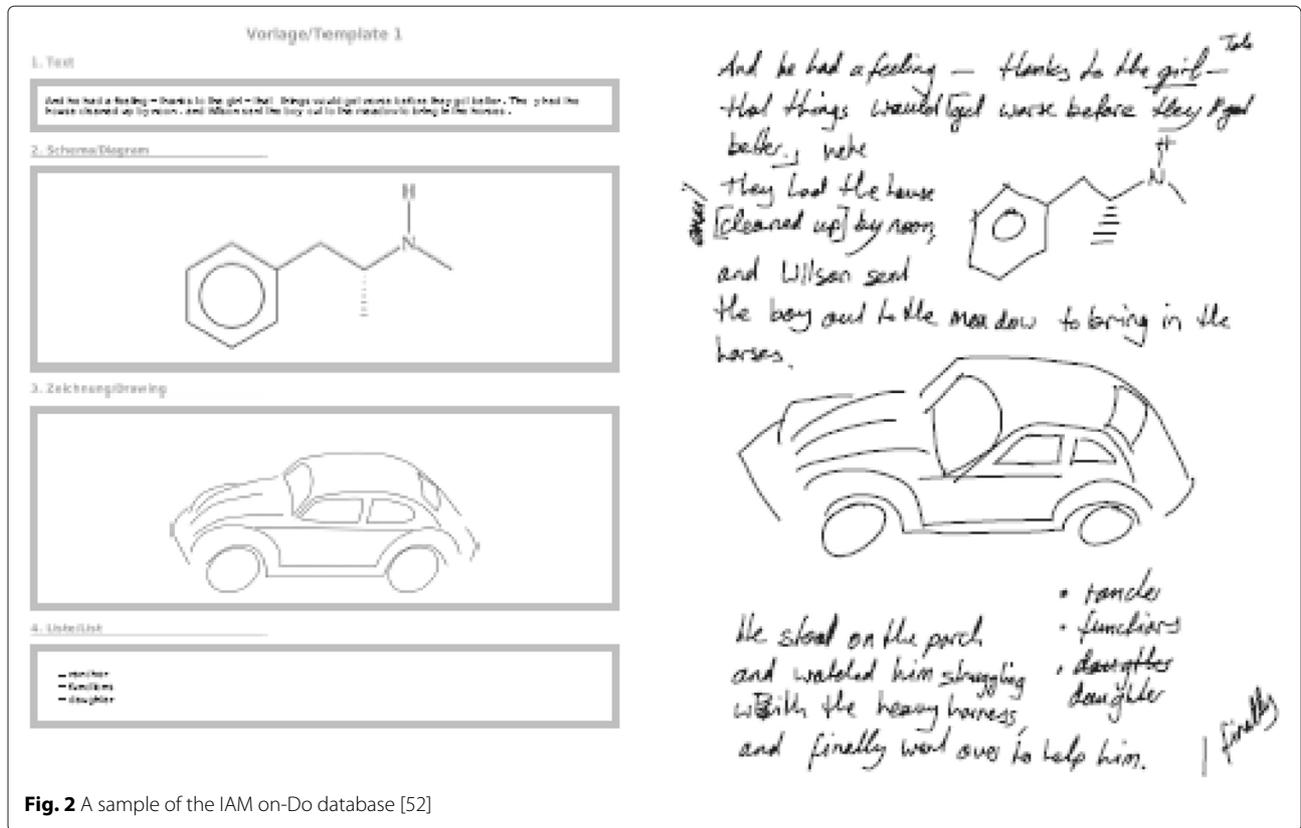**Fig. 1** Samples of handwritten text from the IAM database [3]

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 4 of 24

Fig. 2 A sample of the IAM on-Do database [52]

IAM Historical Document Database (IAM-HistDB)
The IAM-HistDB is a repository comprising handwritten historical manuscript images together with ground truth data. The IAM-HistDB currently includes Saint Gall Database [56] of ninth century containing manuscripts written by a single writer in Carolingian script. The original manuscript is housed at the Abbey Library of Saint Gall, Switzerland. The manuscript images are made available online by the E-codices (Virtual Manuscript Library of Switzerland) project and a text edition was attached at page-level by the Monumenta project. IAM additionally added binarized and normalized text line images to the manuscript data. Altogether, the manuscript data contains page images (jpeg, 300 dpi), binarized and normalized text line images and text edition at page-level (word spelling, capitalization, punctuations, etc.). These images have been employed for text-line segmentation [57, 58], binarization [59, 60], keyword spotting [34, 61] and handwriting recognition [62].

3.1.2 RIMES
RIMES [5, 63] is a representative database of an industrial application. The main idea of developing this database was to collect handwritten samples similar to those that are sent to different companies by individuals. Each contributor was assigned a fictitious identity and a maximum of up

to five different scenarios from a set of nine themes. These themes included real-world scenarios like 'damage declaration' or 'modification of contract'. The subjects were required to compose a letter for a given scenario using their own words and layout on a white paper using black ink. A total of 1300 volunteers contributed to data collection providing 12,723 pages corresponding to 5605 mails. Each mail contains two to three pages including the letter written by the contributor, a form with information about the letter and an optional fax sheet. The pages were scanned, and the complete database was annotated to support evaluation of tasks like document layout analysis [64, 65], mail classification [66], handwriting recognition [67–72] and writer recognition [38, 73].

3.1.3 NIST: handwriting sample image databases
The National Institute of Standards and Technology, NIST, developed a series of databases [6] of handwritten characters and digits supporting tasks like isolation of fields, detection and removal of boxes in forms, character segmentation and recognition. A sample form from the database is illustrated in Fig. 3. The form comprises boxes containing writer information, 28 boxes for numbers and 2 for alphabets while 1 box for a paragraph of text. The NIST Special Database 1 comprised samples contributed by 2100 writers. The latest version of the database, the
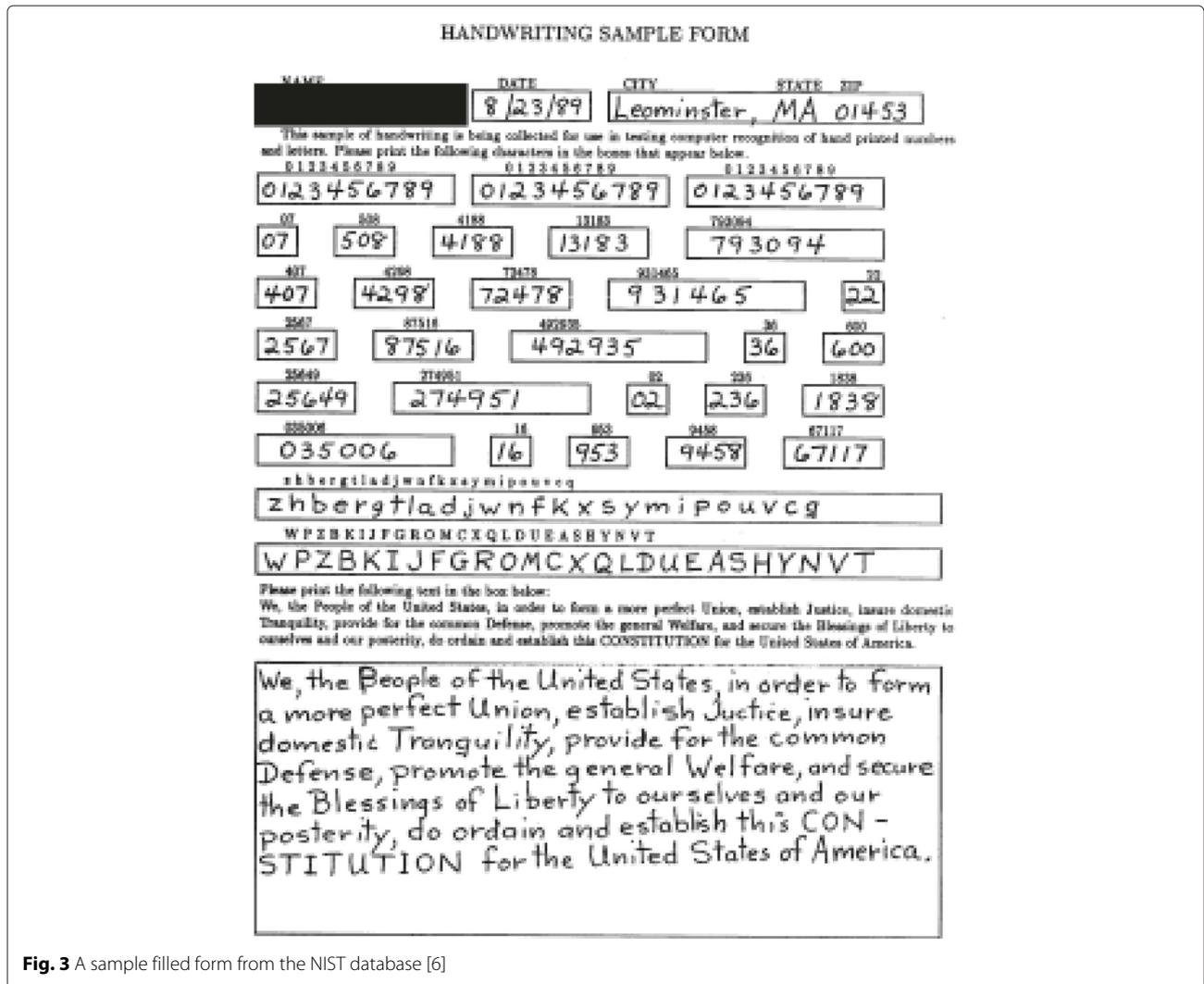
Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 5 of 24

Fig. 3 A sample filled form from the NIST database [6]

Special Database 19, comprises handwritten forms of 3600 writers with 810,000 isolated character images along with ground truth information. This database has been widely employed in a variety of handwritten digit [74–77] and character recognition systems [78–81].

3.1.4 MNIST: a database of handwritten digits

MNIST is a large collection of handwritten digits [7] with a training set of 60,000 and a test of 10,000 samples. MNIST is a subset of the NIST database discussed earlier and is composed of samples from the NIST Special Database 3 (SD-3) and Special Database 1 (SD-1). Initially, SD-3 was proposed to be employed as training and SD-1 as test set. However, samples in SD-3 were contributed by the employees of the Census Bureau while those of SD-1 were written by high school students. As a result, SD-1 offered more challenges in terms of recognition as opposed to SD-3. To ensure uniform distribution of samples from SD-1 and SD-3 in training and test sets, the MNIST database was compiled with a training set of 30,000 images from SD-1 and 30,000 from SD-3. In a similar fashion, the test set comprised 5000 samples each from SD-1 and SD-3 databases. The database has been extensively employed in a number of digit recognition systems [82–87].

3.1.5 CEDAR databases

The Center of Excellence for Document Analysis and Recognition (CEDAR), at the State University of New York at Buffalo, has developed a number of handwritten databases [14] including handwritten words, ZIP codes, digits and alphanumeric characters. These databases were mainly intended to support research in automatic processing of postal addresses on the envelopes. The samples

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 6 of 24

contain 5632 city words, 4938 state words and 9454 ZIP codes. This makes a total of 27,835 alphanumeric characters segmented from address blocks and 21,179 digits segmented from ZIP codes. The words in the database are divided into separate subsets for training and test. This database has been used for the evaluation of a number of systems including handwriting segmentation [88, 89], cursive digit recognition [74, 90–92] character recognition [90, 93, 94] and word segmentation [95] and recognition [96–98].

3.1.6 IRONOFF: the IRESTE on/off dual handwriting database

The Institute de Recherche et d'Enseignement Supérieur aux Techniques de l'Electronique, IRESTE, developed a dual on/off database [99], named IRONOFF. The database comprises handwritten samples of French writers including characters, digits and words. The contributors were required to fill forms having predefined boxes, and the ground truth information and the filled forms were later inspected by human operators. Each contributor filled three types of forms which have been named as B, C and D. The information on form B includes the lower- and uppercase letters of the alphabet, digits, the Euro symbol and the frequently occurring strings in French checks. Forms C and D comprise cursive words in French. The database contains a total of 1000 forms with 32,000 isolated characters and 50,000 cursive words. The online collection of these images is stored in UNIPEN format at a sampling rate of 100 points/second. The database has been employed in a variety of recognition tasks [100–103] as well as online writer identification [104].

3.1.7 The RODRIGO database

RODRIGO [105] is one of the very large databases containing diverse samples of historical manuscripts in Spanish. The RODRIGO database was generated from an old manuscript (of 1545) written in old Castilian (Spanish) by a single author. The writing style is mainly influenced from the Gothic style. The database is spread on 853 pages and is further divided into 307 chapters describing chronicles from the Spanish history. Each page contains a well-separated single text block of calligrapher handwriting. The complete manuscript was digitized by the experts of Spanish ministry of culture in 300 dpi with true colors. The database can be employed for research on historical manuscripts [106–108].

3.1.8 Indonesian handwritten text database

A database of Indonesian handwritten text [109] was compiled to support recognition and segmentation tasks. The database was developed by writing samples contributed by college students making a total of 200 scanned forms. These forms comprise isolated and cursive digits, isolated

upper- and lowercase characters and words and can be employed for evaluation of a number of recognition tasks.

3.1.9 Database for bank-check processing

A recent database for evaluation of check processing and word recognition systems is presented in [110]. The database comprises cursive words in English, courtesy amounts and signatures. The ground truth of the database is developed in XML and, in addition to transcription of text, also contains the identity of the contributing writers. The database can be employed for recognition of words as well as verification of signatures.

3.1.10 IBM UB database

The IBM UB database [111] developed at the Center for Unified Biometrics and Sensors (CUBS) at the University at Buffalo is a multi-lingual online/offline database of handwritten samples. The writing samples include paragraphs of free text, filled forms, words, isolated characters and symbols. Writing samples were collected on IBM's CrossPad with an electronic pen which simultaneously produced ink on the paper and captured the trajectory of the pen. The online data is available in the InkML format while the offline images are scanned as 'png' files. The database is divided into two parts, IBM UB 1 and IBM UB 2. The IBM UB 1 comprises cursive handwritten texts in English with more than 6500 online pages collected from 43 writers and around 6000 offline pages contributed by 41 writers. The IBM UB 2 contains short phrases, digits and isolated characters in French produced by 200 writers. The database has been used for online handwriting recognition [112] and writer identification tasks [113].

3.1.11 CVL Database

CVL [27] is a database of handwritten samples supporting handwriting recognition, word spotting and writer recognition. The database comprises seven different handwritten texts, one in German and six in English. A total of 310 volunteers contributed to data collection with 27 authors producing 7 and 283 writers providing 5 pages each. The ground truth data is available in XML format which includes transcription of text, the bounding box of each word and the identity of writer. The database has been used for writer recognition and retrieval [114] and can also be employed for other recognition tasks. A sample image from the database is shown in Fig. 4.

In addition to text, a database of handwritten digit strings contributed by 303 students has also been compiled [115]. Each writer provided 26 different digit strings of different lengths making a total of 7800 samples. Isolated digits were extracted from the database to form a separate dataset—the CVL Single Digit Dataset. The Single Digit Dataset comprises 3578 samples for each of the digit classes (0-9). A subset of this database has also been
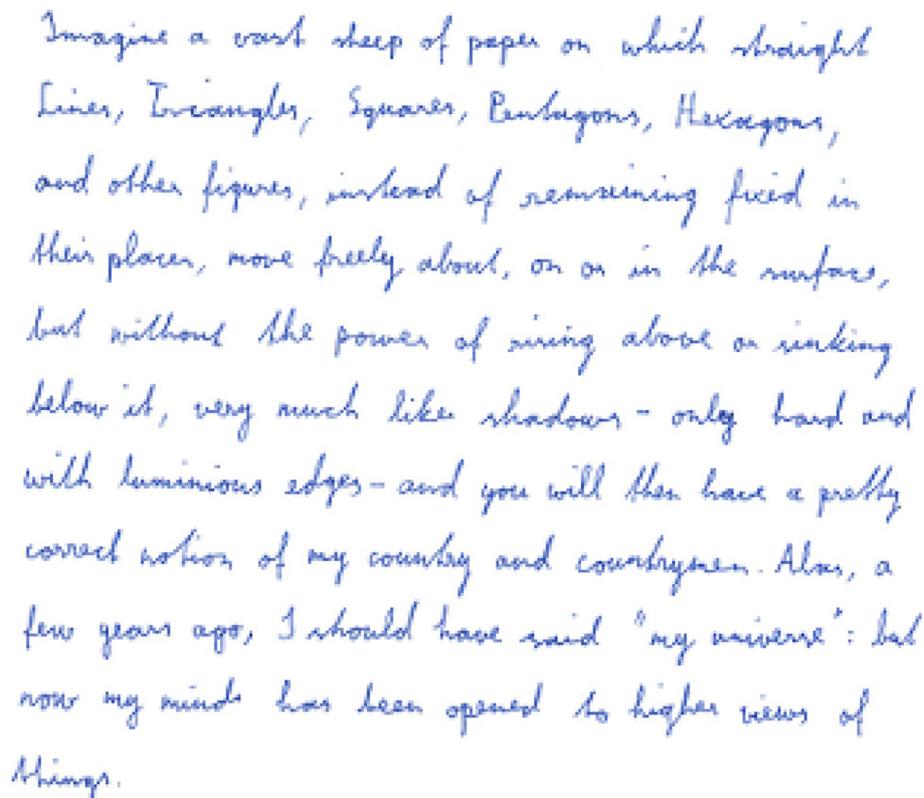
Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 7 of 24

Fig. 4 Sample handwritten text from CVL Database

used in the ICDAR 2013 digit recognition competition [115].

3.1.12 Firemaker Database

The Firemaker database [116] comprises handwritten samples of 250 Dutch individuals with 4 samples per writer making a total of 1000 writing samples. Each writer copied a given text in the normal writing style on page 1 while on page 2 the writers copied the given text using only uppercase letters. Page 3 of each writer comprised 'forged' text whereas on page 4 the writers provided their own text describing the contents of a given cartoon. All pages were scanned at 300 dpi as gray-scale images. The database has been mainly employed for evaluation of writer identification and verification systems [37, 117].

3.2 Databases in the Arabic and Arabic-like scripts

Arabic is the second most widely used script after Roman and supports languages like Arabic, Urdu, Pashto and Farsi. The initial research in document analysis and recognition mainly focused on text in Roman scripts only and it was relatively late that Arabic and other Arabic-like scripts started to receive notable research interest. During the last decade, however, significant research has been carried out on Arabic handwriting recognition and other related

tasks. Consequently, a significant number of Arabic and similar databases have been developed in the recent years. We present an overview of well-known handwritten Arabic and Arabic-like databases in the following sections.

3.2.1 IFN/ENIT

The IFN/ENIT database [19] comprises handwritten Arabic words representing names of towns and villages in Tunisia along with the postal code of each. The database has been developed by contributions from 411 volunteers each filling a specified form. The total words (city/town names) in the database sum up to 26,400 corresponding to 210,000 characters. The ground truth data with the database includes information on the sequence of character shapes, baseline and the writer. All filled forms were digitized at 300 dpi and stored as binary images. The database mainly targets preprocessing [118–120] and recognition of Arabic handwritten words [121–126] but has also been employed to evaluate writer identification systems [127–129]. Figure 5 illustrates a town name from the database written by 12 different writers.

3.2.2 The Arabic Database: ADAB

The online Arabic database ADAB was jointly developed by the Institut Fuer Nachrichtentechnik (IFN) and the

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 8 of 24

Fig. 5 Examples from the IFN/ENIT-database: a town/village name written by 12 different writers [19]

Research Group on Intelligent Machines (REGIM) aiming to support research in online Arabic handwriting recognition. Writing samples were collected from 170 writers making a total of more than 20,000 Arabic words. The database is also accompanied with a tool which allows not only online data collection but also data verification and correction of erroneous data. The database has been employed in a number of segmentation and recognition studies [130–133] as well as for online writer identification [133, 134].

3.2.3 Arabic Handwriting Database: AHDB

The AHDB [20, 135] is an offline database of Arabic handwriting together with several pre-processing procedures. It contains Arabic handwritten paragraphs, words and the words used to represent numbers on checks produced by 100 different writers. The database was mainly intended to support automatic processing of bank checks, but it also contains pages of unconstrained text (as indicated in Fig. 6) allowing evaluation of generic Arabic handwriting recognition systems as well. The database can be employed in handwriting recognition [136] and writer identification tasks [137].

3.2.4 Arabic checks database

This database has been developed to advance research in automatic recognition and processing of Arabic checks [13]. The database comprises a collection of 7000 images of checks containing about 30,000 sub-words and more than 15,000 digits. A sample check from the database

is shown in Fig. 7. The database can be employed for evaluation of automatic check processing and recognition systems [138].

3.2.5 The ARABASE

ARABASE [139] is a rich database for online as well as offline handwriting recognition. The database also supports recognition of offline machine printed Arabic text. The database includes complete paragraphs, words, isolated characters, digits and signatures. The database is also accompanied with a tool supporting traditional document analysis tasks on the database. The database can be employed for evaluation of (online/offline) handwriting recognition and signature verification systems.

3.2.6 CENPARMI Arabic handwriting database

The CENPARMI Arabic database [11] for offline Arabic handwriting recognition comprises isolated digits, letters, numerical strings and words. To support data acquisition, a two-page form was designed that was filled by 100 participants from Canada and 228 participants from Saudi Arabia. These forms comprised a sample Arabic date, 2 samples each of 20 digits, 38 numerical strings, 35 isolated letters and 70 Arabic words. The database is split into three sets. The first set comprises the forms of first 100 writers while the second set contains the forms filled by 228 writers. The third set is a combination of samples from set 1 and set 2. The database has been used for recognition of Arabic characters [140] and numerals [141] as well as word spotting [142].

Fig. 6 An example of free handwriting in Arabic from AHDB database

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46
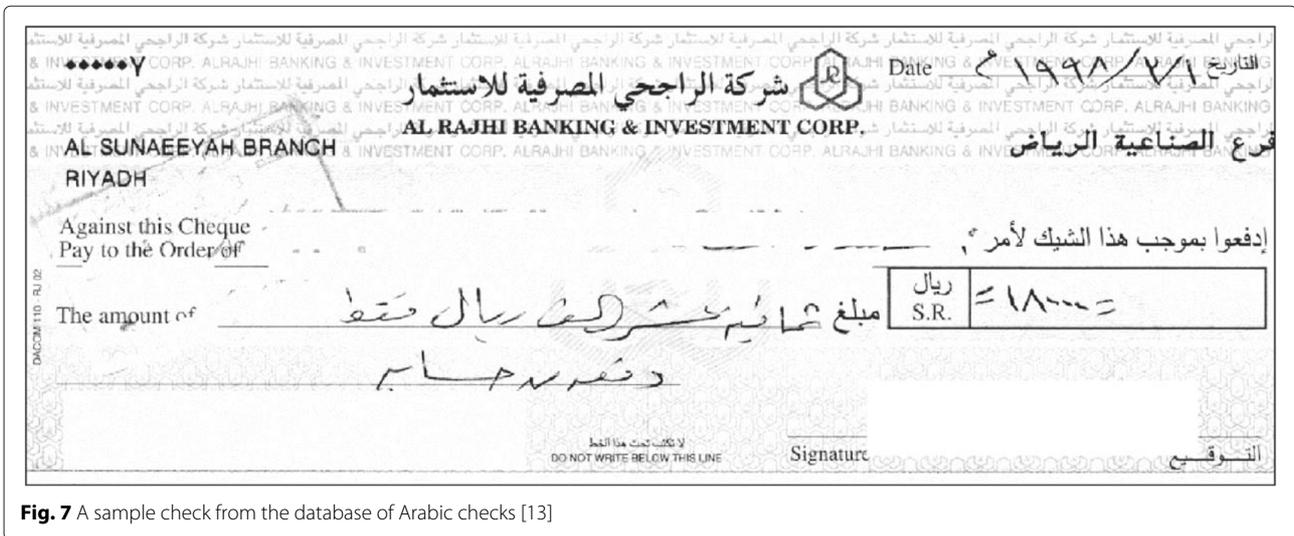
Page 9 of 24

Fig. 7 A sample check from the database of Arabic checks [13]

3.2.7 IBN-E-SINA database

The IBN-E-SINA database [143] is developed on a manuscript provided by the Institute of Islamic Studies (IIS), McGill University, Montreal. The database is a part of the RaSI project which is aimed at creating a large-scale database of Islamic philosophical and scientific manuscripts, mostly written in Arabic with some contributions in Persian and Turkish. The document images were obtained using camera imaging (21 mega-pixels) at a resolution of 300 dpi. The selected dataset consists of 51 folios which correspond to 20,722 connected components (almost 500 CCs on each folio). The database has been used in a variety of interesting research problems on historical manuscripts [144, 145].

3.2.8 Al-Isra Arabic Database

The Al-Isra Database [21] is a large collection of handwritten samples containing words, digits, signatures and sentences compiled by researchers at the University of British Columbia. The samples were gathered from 500 students at Al-Isra University, Jordan. Each student produced a preselected list of words, digits and phrases. The database comprises 500 unconstrained Arabic sentences, 37,000 words, 10,000 digits and 2500 signatures. The database can be employed for handwriting recognition and writer identification tasks.

3.2.9 LMCA database

The On/Off LMCA ("Lettres, Mots et Chiffres Arabe" in French) is a dual Arabic database comprising characters, words and digits [22]. The database includes samples of 55 participants making a total of 500 words and 30,000 digits. The database is compiled in the UNIPEN format [15], the same as that of IRONOFF [99] database. The database can

be used for online as well as offline recognition of Arabic words and digits.

3.2.10 KHATT database

KHATT [146, 147] is a comprehensive database of Arabic handwritten text comprising 1000 forms produced by same number of writers from different countries. Each form is scanned at three different resolutions, 200, 300 and 600 dpi. The textual content of the database comprises 2000 paragraphs randomly picked from multiple sources. The ground truth of the database is provided in xml format and includes the transcription of text at line and paragraph levels. The information about the writer of each sample is also stored. The database is also accompanied with tools that allow segmentation of text images into lines and paragraphs. In addition to recognition of handwriting, the database supports evaluation of a number of pre-processing and segmentation tasks as well as writer identification systems.

3.2.11 QUWI database

The Qatar University Writer Identification (QUWI) [148] database is a comprehensive collection of writing samples of 1017 writers of different cultural and educational backgrounds. A unique feature of this database is that it is a bi-script database where each author contributed four pages, two in English and two in Arabic. This allows using this database in a number of interesting writer identification scenarios. Another feature of this database is that page 1 and page 3 for each writer contains an arbitrary text from the writer's own imagination in Arabic and English, respectively, while page 2 and page 4 of each writer comprises a fixed predefined text (in Arabic and English). This allows the database to be used in text-independent as well as text-dependent evaluation scenarios. The database

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 10 of 24

was mainly developed to support evaluation of writer identification [149] and writer demographic classification systems [150–152] but can also be used for handwriting recognition and similar related tasks.

3.2.12 AHTID/MW

The Arabic Handwritten Text Images Database by Multiple Writers (AHTID/MW) [153] has been developed to support research in the Arabic handwriting segmentation and recognition. In addition, the database can also be employed to evaluate the writer identification systems. The database comprises 3710 text lines and 22,896 words contributed by 53 different native writers of Arabic and is supported by the ground truth annotations. The database has been employed for evaluation of segmentation [154] and writer identification tasks [155].

3.2.13 IAUT/PHCN database

The IAUT/PHCN Database [24] is a collection of handwritten words representing Persian city names. The database was compiled using 1140 forms filled by 380 individuals. The database comprises a total of 200,000 characters and the ground truth includes Unicodes of characters in the city name and baseline information. All forms were scanned at 300 dpi and stored as binary images. The database has been mainly designed to support Farsi word recognition and preprocessing tasks [156–158].

3.2.14 IFN Farsi database

Inspired by the IFN/ENIT Arabic database [19], the IFN Farsi database [23] was developed which comprises more than 7000 images for about 1080 Iranian city/provinces names. A total of 600 individuals contributed to data collection where each writer filled a maximum of two forms with 24 city/province names and their respective postcodes. The ground truth data, in addition to the transcription of text, comprises information on sequence and number of characters, dots and partial words. The database can be used to evaluate Farsi handwritten word and digit recognition systems.

3.2.15 CENPARMI Farsi database

The Center for Pattern Recognition and Machine Intelligence (CENPARMI) Farsi database [10] has been developed to support research in handwriting recognition and word spotting on Farsi text. The database is compiled from 400 native Farsi writers and comprises 432,357 images with dates, words, isolated letters, digits and numeral strings. Each image is provided in gray scale as well as binarized form. The database has been employed in evaluation of symbol/digit recognition [159] as well as Farsi handwriting recognition [160, 161].

3.2.16 FHT: Farsi handwritten text database

FHT database [162] is a repository of unconstrained handwritten texts produced by 250 participants who filled 1000 forms containing Farsi text. The database includes a total of 106,600 handwritten Farsi words, 230,175 subwords and 8050 sentences. Due to its diverse nature, FHT database can be used to evaluate a wide variety of systems including recognition of words and subwords, segmentation of words into characters, baseline detection, machine printed and handwritten textual content discrimination, writer identification and document layout analysis.

3.2.17 HaFT: Farsi text database

HaFT [163] is a large collection of unconstrained Farsi handwritten documents produced by 600 different writers. Each writer contributed three samples at different intervals of time and each sample comprises eight lines of text. This makes a total of 1800 handwritten text images. The database is mainly designed for training and evaluation of Farsi writer identification and writer verification systems but can also be used for different recognition and segmentation tasks.

3.2.18 CENPARMI Urdu database

The CENPARMI Urdu handwritten database [164] comprises Urdu words, characters, digits and numeral strings. A number of native Urdu speakers from different parts of the world contributed to the data collection process. The lexicon of 57 Urdu words and 44 Urdu characters mainly comprises financial terms to support recognition of offline Urdu words, characters and digits. This is the first published database on Urdu handwriting and has been employed in recognition and spotting of Urdu handwritten words [165, 166].

3.2.19 Urdu handwritten sentence database

A relatively new database of unconstrained Urdu handwritten text along with few pre-processing and segmentation algorithms is presented in [167]. The database comprises 400 forms filled by 200 different writers by copying the text given on each form. The forms were generated by taking text from six different categories of news with each category having up to 70 forms. The ground truth of the database includes transcription of text, information on lines and the identity of the writer. The database can be employed for recognition of Urdu text, line segmentation and writer identification.

3.3 CJK databases

CJK, the Chinese, Japanese and Korean, are the main East Asian languages. The writing systems of these languages partially or completely use the Chinese characters Hanzi, Kanji or Hanza. To facilitate research in different areas of handwriting recognition in these languages, a number of standard databases have been developed

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 11 of 24

and distributed. We discuss the notable databases in the following sections.

3.3.1 PE92: handwritten Korean character image database

PE92 [17] is a very large and unique database comprising 100 handwritten image sets of 2350 Hangeul characters (Fig. 8). More than 500 writers contributed to the generation of first 70 sets while the last 30 sets were produced by one person. Writers filled pre-defined forms by writing characters in specified boxes. The database has been used in a variety of recognition tasks [168–170].

3.3.2 Online Japanese character pattern database

A database of online Japanese character patterns [31, 171] was compiled to support research in Japanese character recognition systems. These characters were extracted from unconstrained textual phrases provided by 80 writers. The text was collected from Japanese newspapers and produced 1227 frequently occurring Japanese character categories. The patterns were manually inspected and corrected to remove errors and wrongly written characters.

The database has been used in a number of online character recognition systems [172–175].

3.3.3 HCL-2000 Database

HCL-2000 [176, 177] is a large collection of frequently used Chinese characters produced by 1000 writers. In addition to the ground truth information of 3755 characters, information about the writers, their age and gender is also stored allowing evaluation of writer identification or demographic classification systems. The database has been employed in a number of Chinese character recognition systems [178–180]

3.3.4 SCUT-COUCH2009: online unconstrained Chinese handwriting database

The SCUT online handwriting Chinese character recognition database, SCUT-COUCH2009 [181] is a revision and an enhanced version of SCUT-COUCH2008 [182] database. The database contains 11 datasets of diverse kinds of vocabularies and has been mainly developed to facilitate research in unconstrained online Chinese

Fig. 8 Handwritten samples PE92 Korean character image database [17]

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 12 of 24

handwriting recognition. The database comprises individual Chinese characters in different standards, complete Chinese words and isolated symbols. The total character count in the database is more than 3.6 million. A sample image from the database is shown in Fig. 9. All the samples were gathered using PDAs (Personal Digital Assistants) and smart phone devices with touch screens and a total of 190 different individuals contributed to data collection. This database was the first publicly available large online Chinese handwriting database and has been employed in a number of online handwriting recognition tasks [183–186].

3.3.5 CASIA: online and offline Chinese handwriting databases

CASIA [187, 188] is a widely used Chinese handwritten database comprising handwritten paragraphs as well as isolated characters The data was collected from 1020 individuals who produced writings on paper with a digital pen. This allowed capturing the online trajectory information as well as the offline images of text. The database was divided into six subsets, three comprising isolated characters (DB 1.0–1.2) and three having handwritten paragraphs (DB 2.0–2.2). The datasets of isolated characters comprise a total of about 3.9 million Chinese characters while the datasets of text (paragraphs) contain about 1.35 million characters. This database has been employed in a number of recognition [189–192] word spotting tasks [193].

3.3.6 Touching character database

In order to assess character segmentation algorithms, a database of touching Chinese characters was compiled from the CASIA handwriting database [187, 188]. This database was termed as CASIA-HWDB-T [194]. The database includes more than 56,000 strings with two or more touching characters. More than 1800 strings comprise multiple touching characters. The database is also divided into interesting subsets like strings comprising all Chinese characters, mixed strings and digits. The ground truth data includes information on character classes and

locations of touching points. This database can be used for character segmentation [195, 196] and recognition of broken or touching characters.

3.4 Databases in Indian scripts

Significant research has been carried out on document analysis and recognition problems in different Indian scripts. Several hundred languages are spoken and written in India with Hindi, Tamil, Telugu, Bengali, Kannada and Gujrati being the popular ones. Some of the languages share common scripts while others have unique scripts of their own. Well-known Indian scripts include Devanagari, Telugu, Tamil and Kannada. These diverse scripts offer a variety of interesting and challenging problems to the document recognition community. Despite a rich diversity of scripts and languages, the number of standard databases on Indian scripts is relatively small. We discuss the databases developed on different Indian scripts in the following sections.

3.4.1 Handwritten numeral databases of Indian scripts

A large database of handwritten numerals in two popular Hindi scripts is presented in [197]. The database was compiled by collecting numerals from postal mails and job application forms in Devanagari and Bangla scripts. A total of 22,556 Devanagari numerals were collected from 368 postal mails and 274 job application forms. In a similar fashion, 23,392 Bangla numerals were collected from 465 mails and 268 job applications. All images were digitized at 300 dpi and saved as gray scale 'tif' images. The database has been used to evaluate a number of digit recognition systems [25, 198–200].

3.4.2 Kannada handwritten document dataset

The Kannada Handwritten Text Database (KHTD) [201] comprises 204 writing samples in a popular Indian script Kannada. The database has been developed by collecting writing samples from 51 native speakers of Kannada, and the textual content comes from four different categories. The database has a total of more than 4000 lines of text and 26,000 words. The database can be employed

Fig. 9 Handwritten samples from the SCUT-COUCH 2009 Database [181]

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 13 of 24

in a number of segmentation and recognition tasks at line, word or character levels.

3.4.3 A database of Tamil handwritten city names

A database of handwritten city names in Tamil, a popular script in India and Sri Lanka, is presented in [202]. The database includes a total of 265 different city names with 109 cities from Indian state of Tamil Nadu and 156 cities from Sri Lanka. Each city name has 100 instances in the database and a total of 500 writers with different educational backgrounds contributed to data collection. The database is also accompanied with algorithms to automatically segment city names from the image. Out of the 265 city names, 258 comprise only 1 word, 5 names include 2 words and 2 names contain 3 words with an average of 7 characters per city name. The database can be used for recognition of handwritten Tamil words.

3.4.4 Devanagari numeral and character database

A database comprising Devanagari numerals and characters is presented in [203]. Writing samples of 750 individuals belonging to different educational backgrounds, ages and professions were collected. The database comprises a total of 5137 isolated numerals and 20,305 isolated characters stored as binary 'tif' images. The database has been made available publically and can be used for recognition of Devanagari characters

3.5 Miscellaneous

After having discussed the handwritten databases in Roman, Arabic, CJK and Indian scripts, we now present few other databases in the following.

3.5.1 AMHCD: a database for Amazigh handwritten character recognition research

This database has been developed to support research activities on Amazigh text. Amazigh is spoken by millions of people in Africa mostly for oral communication. The Moroccan government took the initiative to promote Amazigh in mass media as well as the educational system. As a part of these efforts, the IRF-SIC Laboratory at the Ibn Zohr University, Morocco developed the AMHCD database [204] comprising a total of 25,740 isolated characters contributed by 60 different writers (Fig. 10). Each author produced 13 examples of each Amazigh character. The collected documents are scanned at 2400 dpi and are stored as colored 'jpeg' images. The database mainly targets the recognition system for handwritten Amazigh characters [205, 206].

3.5.2 GRUHD: database of Greek unconstrained handwriting

The GRUHD [29] database is a huge collection of unconstrained Greek text. The database includes sentences, characters, digits and other symbols. The writings have
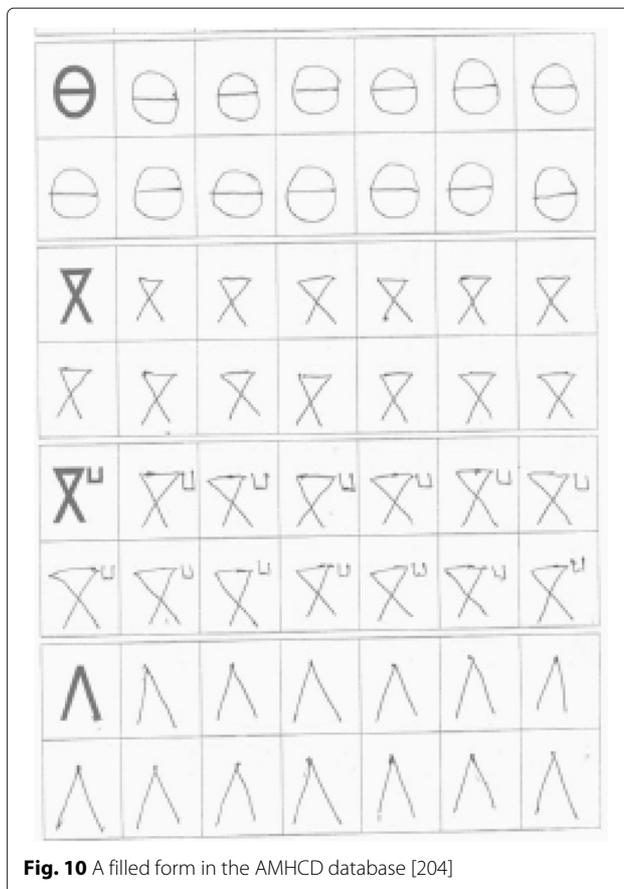
Fig. 10 A filled form in the AMHCD database [204]

been produced by 1000 writers with equal distribution of male and female writers. The database comprises 1760 forms having 667,583 symbols and 102,692 words. The database has been employed for character/symbol recognition [207–209] and discrimination of machine-printed and handwritten texts [210].

3.5.3 MRG-OHTC database

MRG-OHTC [211] is a collection of online Tibetan writings facilitating research in online Tibetan character recognition. A total of 130 Tibetan writers produced the database comprising 910 Tibetan characters from the basic and extended Tibetan character set. The writing samples are collected on a digital tablet using an electronic pen. The database has been employed for evaluation of Tibetan character recognition systems [212, 213].

3.6 Discussion

After having discussed the databases in different scripts, we now present a comparative overview of these databases in Table 1 along with a critical appreciation. The databases are ordered by year of publication and are compared on the basis of the following criteria.

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 14 of 24

Table 1 An overview of the databases discussed in the paper

Database	Year	Language	Content	Mode	Writers	Statistics
PE92 [17]	1993	Korean	Isolated characters	Offline	500+	235,000 characters
CEDAR [14]	1994	English	Words, characters, digits	Offline		10,570 words, 27,835 characters, 21,179 digits
NIST [6]	1995	English	Isolated Digits	Offline	3600	810,000 digits
JPCD [171]	1997	Japanese	Characters	Online	80	1227 character categories
MNIST [7]	1998	English	Isolated digits	Offline		70,000 digits
Al-Isra [21]	1999	Arabic	Sentences	Offline	500	500 sentences, 37,000 words, 10,000 digits
IRONOFF [99]	1999	French	Words, characters, digits	Online		50,000 words, 32,000 characters
Firemaker [116]	2000	English	Paragraphs	Offline	250	4 samples/writers
GRUHD [29]	2001	Greek	Text, symbols	Offline	1000	1760 forms, 667,583 symbols, 102,692 words, 123,256 digits
IAM [2, 3]	2002	English	Sentences	Offline	657	1539 forms, 5685 sentences, 115,320 words
IFN/ENIT [19]	2002	Arabic	Words	Offline	411	2265 forms, 26,449 city names
Checks DB [13]	2003	Arabic	Check amounts	Offline		7000 cheques, 29,498 subwords, 15,000 digits
AHDB [20, 135]	2004	Arabic	Sentences,check amounts	Offline	100	105 Forms
ARABASE [139]	2005	Arabic	Sentences, words, letters	On/off	400	400 forms
IAM-OnDB [4]	2005	English	Sentences	Online	221	1700 forms; 86,272 words
Numerals DB [197]	2005	Bangla Devanagari	Digits	Offline		45,948 numerals
IAUT/PHCN [24]	2008	Farsi	Isolated words	Offline	380	1140 forms, 34,200 words
RIMES [63]	2008	French	Sentences	Offline	1300	12,723 pages
IFN Farsi [23]	2008	Farsi	Words	Offline	600	7271 words, 23,545 subwords
CENPARMI-A [11]	2008	Arabic	Words, characters, digits	Offline	328	13,439 digits, 21,426 characters, 11,375 words
LMCA [22]	2008	Arabic	Words, characters, digits	Online	55	30,000 digits, 100,000 characters, 500 words
CENPARMI-U [164]	2009	Urdu	Words, characters, digits	Offline		18,000 words
FHT [162]	2009	Farsi	Sentences	Offline	250	1000 forms, 106,600 words, 8050 sentences
HCL2000 [177]	2009	Chinese	Characters	Offline	1000	3755 characters
CENPARMI-F [10]	2009	Farsi	Words, letters, digits	Offline	400	432,357 images
IAM-OnDO [52]	2010	English	Text, drawings, tables etc.	Online	200	1000 documents
RODRIGO [105]	2010	Spanish		Offline		1853 pages
ADAB [130]	2011	Arabic	Words	Online	170	20,000+ words
CASIA [188]	2011	Chinese	Text, characters	On/off	1020	3.5M isolated characters, 1.35M characters in text
SCUT-COUCH [181]	2011	Chinese	Characters	Online	190	3.6M characters
Indonesian TDB [109]	2011	Indonesian	Sentences	Offline	200	200 forms
AMHCD [204]	2011	Amazigh	Characters	Offline	60	25,740 characters
MRG-OHTC [211]	2011	Tibetan	Characters	Online	130	910 character classes
KHTD [201]	2011	Kannada	Sentecnes	Offline	51	4000 lines, 26,000 words
KHATT [146]	2012	Arabic	Sentences	Offline	1000	1000 forms
Devanagari DB [203]	2012	Devanagari	Digits, characters	Offline	750	5137 isoloated numerals
UHSD [167]	2012	Urdu	Sentences	Offline	200	400 forms
QUWI [148]	2013	Arabic English	Sentences	Offline	1017	4068 forms
HaFT [163]	2013	Farsi	Sentences	Offline	600	1800 images
CVL [27]	2013	English German	Sentences	Offline	311	2163 forms
Tamil DB [202]	2013	Tamil	Words	Offline	500	265,00 city names
AHTID-MW [153]	2015	Arabic	Text lines	Offline	53	3710 lines

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 15 of 24

- Content of writing (sentences, words, characters or digits)
- Handwriting mode (online or offline)
- Language or script
- Total number of writers
- Total number of samples
- Problems on which databases could be employed

It can be observed from Table 1 that the trend of development of standard databases and their ground truth labeling has witnessed a notable growth in the last few years. Attempts have been made to capture as much variation in writing as possible by considering a large number of writers in the data collection process. In terms of number of writers, RIMES database [5, 63] seems to be the most comprehensive with around 1300 individuals contributing their writing samples. From the view point of number of writing samples, RIMES comprises more than 12,000 pages of handwritten text, one of the largest collection of unconstrained handwritten images. This database, however, is not publically available. In terms of usage, the IAM handwriting database [2, 3] is one of the most widely used databases for a number of recognition tasks. The only major issue with IAM database is the non-uniform distribution of samples per writer which varies from more than 50 for 1 writer to 1 for about 350 writers. This complicates the evaluation protocols for writer identification and verification systems where varied amount of text per writer is available to train and test the systems. Nevertheless, IAM databases remain one of the most popular databases employed by the handwriting recognition community. Likewise, for research on Arabic handwriting, the IFN/ENIT database has been most extensively employed for Arabic handwriting recognition and Arabic writer identification.

Naturally, most of the databases discussed in our study are based on English or Arabic writing samples. This is due to the significant research attention these languages have received over the last three decades. During the last few years, however, research on text in other languages has also gained interest resulting in the development of handwritten databases in many languages like Farsi and Urdu. A trend of having multi-script databases can also be witnessed in the recently developed databases. Such collections provide an opportunity to study the interesting scenarios of finding common writing patterns of individuals across different scripts. QUWI database [148] is an example of such a multi-script database comprising writing samples in Arabic and English. Another interesting aspect in recent databases is that instead of simply keeping the identity of the writer, additional information including the age, gender and background of the writer is also stored allowing development and evaluation of automatic user demographics classification

systems, a relatively less explored area in handwriting analysis.

From the view point of textual content, the preliminary databases in all the scripts mostly comprised isolated characters, digits or words. These databases were mainly employed to evaluate the initial research endeavors in recognition of characters, digits and words. With the advancement in computerized recognition of handwriting, databases comprising unconstrained text (paragraphs) in natural writing styles of contributors were developed. These databases allowed evaluation of unconstrained handwriting recognition rather than simply character or word recognition.

An important parameter in the analysis of different databases is how well they represent the real-world scenarios. Databases where the acquisition is unconstrained and provides writers the flexibility to write in their natural styles are more close to the writing samples encountered in the real-world problems. For applications like handwriting recognition, significant training data is available, but for problems like forensic document analysis (writer identification, writer verification, etc.), the amount of text available to learn the characteristics of an individual is, in general, limited. Same is the case in the test phase where only limited text may be available to find the identity of an individual from a given writing sample. Systems developed for such applications should therefore be evaluated in experimental setups which match the real-world constraints. There is also a need to consolidate the large number of databases at a common platform allowing researchers in document analysis and recognition choose the most appropriate database(s) for development and evaluation of their systems.

As discussed earlier, the problems that could be evaluated using a given database are a function of the ground truth information provided with the database. For all recognition tasks, the database must be accompanied by the corresponding transcription (character, word or paragraph level). Likewise, systems dealing with identification or verification of writers and prediction of user demographics from handwriting require writer information to be stored along with each writing sample. Table 2 groups the databases discussed in this paper as a function of tasks in which they can be employed. Expectedly, most of the handwriting databases have been developed for evaluation of offline handwriting recognition systems. Few recent databases support evaluation of online recognition systems as well. The least explored area seems to be user demographics classification from handwriting and only a few databases contain the required ground truth (writer) information to evaluate such systems.

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 16 of 24

Table 2 Usage of databases

Task	Databases
Offline handwriting recognition	IAM, IAM-HistDB, RIMES, CEDAR, CVL, IFN-ENIT, AHDB, ARABASE, CENPARMI-Arabic, Al-Isra, LMCA, KHATT, QUWI, IAUT-PHCN, IFN-Farsi, CENPARMI-Farsi, HaFt, CENPARMI-Urdu, PE92, HCL2000, CASIA, KHTD, AMHCD, GRUHD, MRG-OHTC
Online handwriting recognition	IAM-onDB, IRONOFF, IBM-UB, ADAB, ARABASE, LMCA, SCUT-COUCH,CASIA
Digit Recognition	NIST, MNIST, CEDAR, CVL, CENPARMI-Arabic, IFN-Farsi, CENPARMI-Farsi
Offline writer identification/verification	IAM, RIMES, CVL, FireMaker, IFN/ENIT, AHDB, Al-Isra, KHATT, QUWI, AHTID/MW, FHT, HaFT, HCL2000
Online writer identification/verification	IAM-onDB, IRONOFF, IBM-UB, ADAB
Word spotting	IAM, IAM-onDo, IAM-HistDB, CENPARMI-Arabic, CENPARMI-Urdu,CASIA
Handwriting segmentation	IAM, IAM-HistDB, CEDAR, AHTID-MW, FHT, HaFT,KHTD
Gender classification	IAM-onDB, QUWI, HCL2000

4 Campaigns, projects, competitions and results

During the recent years, the development of standardized datasets and their labeling has moved a step further to the organization of different evaluation campaigns and competitions. These competitions, related to different classical tasks of document analysis and recognition, not only allow a meaningful comparison of different algorithms under the same experimental conditions but also provide a platform for exchange of ideas and knowledge. This section is dedicated to the discussion of these campaigns and contests, but prior to that we present the UNIPEN project, a major milestone in online handwriting recognition.

4.1 UNIPEN project for online data exchange

The UNIPEN project [15] of data exchange was initiated by International Association of pattern recognition (TC-11) in 1992 with the objective of proposing a uniform format for representation and exchange of online data. The format was developed in collaboration with a group of 14 experts in online handwriting recognition. The participants of the project were asked to submit a minimum of 12,000 characters in any form (sentences, words or individual characters) and the approved data from National Institute of Standards and Technology (NIST) was made publically available. Presently, 11 datasets comprising characters, words and sentences have been compiled and software toolkits to manipulate the UNIPEN files are also provided with the database.

4.2 RIMES evaluation campaign

The RIMES project [5] funded by French ministries of defense and research was initiated to develop and evaluate automatic systems for indexing and recognition of handwritten letters. The project aimed at not only creating a large annotated database but also to organize a set of evaluation campaigns covering a variety of document recognition tasks which could eventually fit in different industrial applications. The first phase of the evaluation campaign [63] comprised tasks including document (letters and fax) layout analysis, handwriting recognition (isolated characters, words and blocks of text), writer identification (on words and paragraphs), writer verification, logo recognition and identification of scenario from letters. The second phase of the campaign [214] focused on three themes, document layout analysis, handwriting recognition and writer identification and a total of seven tasks. Five French research labs participated in this second phase of evaluations. After two successful phases of evaluations, the database was employed in a number of International competitions, discussed later in this paper.

4.3 Organization of competitions

The last few years have seen an increasing trend in the organization of International competitions on different tasks in document analysis and recognition. These contests are mainly advertised and organized in conjunction with the reputed document recognition conferences, International Conference on Document Analysis and Recognition (ICDAR) and International Conference on Frontiers in Handwriting Recognition (ICFHR) being the two most notable platforms. These contests provide training and validation datasets to the participants and require them to submit either the executables of their developed algorithms or the results on the unlabeled test datasets. A major proportion of these competitions are based on handwriting recognition and other related tasks. In most cases, the evaluation is carried out on published and well-known handwritten databases. In Table 3, we present a summary of the competitions based on the databases discussed in Section 3. It can be seen that IFN/ENIT is easily the most widely used database in the regularly organized Arabic handwriting recognition competitions. Recognition of online handwriting in different languages has also

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 17 of 24

Table 3 An overview of databases used in different competitions

Database	Competition	Participants
IFN/ENIT	ICDAR2005 Arabic Handwriting Recognition Competition [221]	6 teams, 6 systems
	ICDAR2007 Arabic Handwriting Recognition Competition [222]	8 teams,14 systems
	ICDAR 2009 Arabic Handwriting Recognition Competition [217]	7 teams,17 systems
	ICFHR2010 Arabic Handwriting Recognition Competition [223]	4 teams, 6 systems
	ICDAR 2011 Arabic Handwriting Recognition Competition [224]	4 teams, 4 systems
ADAB Database	ICDAR 2009 Online Arabic Handwriting Recognition Competition [225]	3 teams,7 systems
	ICDAR 2011 Online Arabic Handwriting Recognition Competition [226]	3 teams,5 systems
Farsi CENPARMI	ICDAR 2009 Handwritten Farsi/Arabic Character Recognition Competitions [227]	4 systems
RIMES	ICDAR 2009 Handwriting Recognition Competition [68]	8 teams,10 systems
	ICDAR 2011-French handwriting recognition competition [69]	9 systems
CASIA Database	ICDAR 2011 Chinese Handwriting Recognition Competition [228]	8 teams, 25 systems
	ICDAR 2013 Chinese Handwriting Recognition Competition [229]	10 teams, 27 systems
QUWI Database	ICFHR 2012 Competition on Writer Identification (Arabic Script) [149]	43 teams
	ICDAR 2013 Competition on Gender Prediction from Handwriting [150]	194 teams
	ICDAR 2015 Multiscript Writer Identification & Gender Classification Competition [220]	8 teams, 13 systems
CVL Digit Database	ICDAR 2013 Competition on Handwritten Digit Recognition [115]	7 teams, 9 systems
	ICFHR 2014 Competition on Handwritten Digit String Recognition [230]	5 teams, 6 systems
KHATT/AHTID-MW	ICFHR 2014 Arabic Writer Identification Contest [231]	3 systems

received an increased research attention. Other than the traditional recognition tasks, competitions on prediction of gender from handwriting have recently gained significant interest. Although a very large number of groups participated in these competitions, a relatively lesser number of groups actually revealed their identities and provided a description of their algorithms [149, 150]. In addition to the contests mentioned in Table 3, a number of other competitions on handwritten databases have also been organized but since they employ non-published or private databases, they are beyond the scope of our discussion.

4.4 Experimental protocols, evaluation metrics and state-of-the-art results

In this section, we discuss the experimental settings and evaluation metrics that are employed by researchers to solve the problems based on analysis of handwriting. As discussed earlier, the most important of these tasks is handwriting recognition which is carried out at character, word and line levels. Consequently, these systems report results in terms of character and word recognition rates. In some cases, the edit distance between the recognized text and ground truth text is used to quantify the recognition performance. Likewise, the handwritten keyword spotting systems are evaluated using the standard precision and recall measures. The two measures are

generally combined into a single f-measure to represent the performance by a single number.

For writer identification systems, the performance is evaluated either using a leave-one-out-approach or by splitting the database into training and test sets, the later being more commonly employed. In most cases, in addition to the identification rate, the Top-K identification rates are also reported where for a given query document, a list of most similar K writers is retrieved which increases the chances of finding the true writer of the query document. Similarly, the performance of writer verification systems is represented through receiver operating characteristic (ROC) curves and is quantified through area under the curve (AUC) or equal error rates (EER). The closely related task of gender (and user demographics) prediction from handwriting is evaluated using the classification rate.

To provide an idea on the performance of state-of-the-art systems on different handwriting recognition tasks, we present a summary of some of the best results reported in the literature on commonly used handwriting databases in Table 4. Few of these results have been taken from the findings of different International competitions while others have been compiled from the literature as reported by the respective researchers (to the best of the authors' knowledge). For handwriting recognition, a high word recognition rate of 94.85 % [69] is reported on the RIMES database. The recognition rates on the IAM database vary

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 18 of 24

Table 4 Overview of state-of-the-art results on commonly used databases

Database	Task	Experimental protocol	Evaluation metric	Best results
IFN/ENIT	Handwriting recognition	Train: 32,492 words, Test: 8671 words	Recognition rate	93.37 % [217]
	Writer identification	411 writers	Top 1 Identification Rate	90 % [219]
QUWI	Writer identification	206 Writers, 2 paragraphs in training 1 in test	Top 1 Identification Rate	95.30 % [149]
	Gender Classification	Train: 282 writers, test: 193 writers	Classification rate	69.25 % [150]
	Multi-script writer identification	Train: 300 writers, test: 100 writers	Top 1 Identification Rate	55 % [220]
	Multi-script gender classification	Train: 300 samples, test: 100 samples	Classification rate	65 % [220]
IAM	Handwriting recognition	Varied	Recognition rate	80-90 % [215, 216]
	Writer identification	657 writers, one sample each in training & test	Top 1 Identification Rate	96.7 % [218]
RIMES	Word recognition	Training: 50,000 words, validation: 7000, test: 5740 words	Recognition rate	94.87 % [69]
KHATT	Writer identification	1000 writers with 4 samples/writers in training, test: 1822 images	Top 1 Identification Rate	73.4 % [231]
CVL	Digit recognition	Train: 7,000, validation: 7,000, test: 21,780	Precision	97.74 % [115]
CASIA	Character recognition	3755 character classes	Recognition rate	Offline:94.77 %, online: 97.39 % [229]

as different studies employ different evaluation protocols and an objective comparison is hard to make. The standard protocol for IAM lines comprises 6161 lines (45,000+ words) for training, 920 lines (7000+ words) for validation and 2781 lines (around 20,000 words) for testing. Word recognition rates in the range 80–90 % are reported by a number of studies [215, 216]. These recognition rates, in general, are lower than those reported on the RIMES database. It should however be noted that the RIMES test set comprises 1600 unique words while the complete IAM database comprises a vocabulary of more than 10,000 words, a major reason for relatively lower recognition rates. Regarding Arabic handwriting recognition, a high word recognition rate of 93.37 % [217] is reported on the IFN/ENIT database.

Writer identification systems have been most evaluated and compared on the IAM database and a highest identification rate of 96.7 % is reported in [218] with one sample of each of the 657 writers in training and one in the test set. Like Arabic handwriting recognition, the writer identification systems targeting Arabic handwritings mostly employ the IFN/ENIT database. The system presented in [219] realizes the highest identification rate of 90 % on the 411 writers of this database. Writer identification rates on the recently developed KHATT database are relatively lower (73.4 %) mainly due to a large number of writers (1000) in the database. The QUWI database which includes writer demographics information has been employed for gender classification in a number of recent studies and a highest classification rate of 69.25 % is realized [150]. Although a two-class problem, gender

prediction from handwriting is a challenging task as the correlation between handwriting and gender is not known to be very strong, a major reason for low classification rates. A step further in evaluation of writer identification and gender classification systems is the multi-script experimental setup where training and test samples come from different scripts. Naturally, the recognition rates (55 % on writer identification and 65 % on gender classification [220]) on these challenging problems are not as high as in case of a single script. Robust systems which exploit the common features of writers across different scripts need to be investigated to enhance the current state-of-the-art on these tasks.

5 Conclusions

Research in handwriting recognition and related areas is a challenging problem. The field has seen more than 30 years of intensive research, and state-of-the-art solutions have been developed for many problems. A number of handwriting recognition problems still remain inviting for the document recognition community and significant research targeting different aspects of handwriting recognition is being carried out presently. During the recent years, there has been an increasing trend of developing standard databases, compiling the ground truth data to support different recognition tasks and exposing the databases to the research community to explore and investigate their algorithms. In general, the statistics and ground truth information of each database is detailed in their respective publications.

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 19 of 24

This paper is an endeavor to provide a comprehensive survey of notable databases of handwritten text developed over the last two decades. For each database, we provided details on its structure, statistics, ground truth information and the tasks supported. Typically, these databases target one or more of the preprocessing, segmentation and recognition tasks. The type of task(s) that can be evaluated with a given database is a function of the ground truth data accompanying the database. In addition to the location and transcription of text, information about contributors is also stored in some cases allowing evaluation of writer recognition and writer demographics classification tasks as well.

We also discussed the evaluation campaigns and competitions organized using these databases. Organization of competitions in conjunction with reputed document and handwriting recognition conferences has become a regular activity for the last few years. The increasing number of participants in these competitions is a clear indication of the kind of research attention different problems of handwritten documents are attracting. In addition to the description of databases, we also summarized the state-of-the-art results on commonly used databases for a number of recognition tasks.

This contribution is likely to provide a summarized review of different databases allowing researchers choose the most appropriate datasets for evaluation of their proposed systems.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]National University of Sciences and Technology, Islamabad, Pakistan. [2]Document Image and Pattern Analysis (DIPA) Group, Islamabad, Pakistan. [3]Bahria University, Islamabad, Pakistan. [4]Institute of Space Technology, Islamabad, Pakistan. [5]Larbi Tebessi University, Tebessa, Algeria.

References

1. SM Lucas, A Panaretos, L Sosa, A Tang, S Wong, R Young, in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*. Icdar 2003 robust reading competitions, (2003), pp. 682–687
2. U-V Marti, H Bunke, in *Proceedings of the Fifth International Conference on Document Analysis and Recognition*. A full english sentence database for off-line handwriting recognition, (1999), pp. 705–708
3. UV Marti, H Bunke, The iam-database: An english sentence database for offline handwriting recognition. Int. J. Doc. Anal. Recognit. **5**(1), 39–46 (2002)
4. M Liwicki, H Bunke, in *Proceedings of the Eighth International Conference on Document Analysis and Recognition*. Iam-ondb - an online english sentence database acquired from handwritten text on a whiteboard, (2005), pp. 956–961
5. E Augustin, J Brodin, M Carre, E Geoffrois, E Grosicki, F Preteux, in *Proceedings of International Workshop on Frontiers in Handwriting Recognition*. Rimes evaluation campaign for handwritten mail processing, (2006), pp. 231–235
6. R Wilkinson, J Geist, S Janet, P Grother, C Burges, R Creecy, B Hammond, J Hull, N Larsen, T Vogl, C Wilson, *The First Census Optical Character Recognition Systems Conference*. (The U.S. Bureau of Census and the National Institute of Standards and Technology, 1992)
7. Y LeCun, L Bottou, Y Bengio, P Haffner, Gradient-based learning applied to document recognition. Proc. IEEE. **86**(11), 2278–2324 (1998)
8. MW Sagheer, CL He, N Nobile, CY Suen, in *Image Analysis and Processing Lecture Notes in Computer Science*. A new large urdu database for off-line handwriting recognition, vol. 5716, (2009), pp. 538–546
9. MI Shah, CL He, N Nobile, CY Suen, in *Proceedings of the 14th Conference of the International Graphonomics Society*. A handwritten pashto database with multi-aspects for handwriting recognition, (2009), pp. 157–161
10. HP Jifroodian, N Nicola, CL He, CY Suen, in *Image Analysis and Recognition Lecture Notes in Computer Science*. A new large-scale multi-purpose handwritten farsi database, vol. 5627, (2009), pp. 278–286
11. H Alamri, J Sadri, CY Suen, N Nobile, in *Proceedings of the 11th Intl. Conference on Frontiers in Handwriting Recognition*. A novel comprehensive database for arabic off-line handwriting recognition, (2008), pp. 664–669
12. F Solimanpour, J Sadri, CY Suen, in *Proceedings of the Tenth International Workshop on Frontiers in Handwriting Recognition*. Standard databases for recognition of handwritten digits, numerical strings, legal amounts, letters and dates in farsi language, (2006), pp. 743–751
13. Y Al-Ohali, M Cheriet, CY Suen, Databases for recognition of handwritten arabic cheques. Pattern Recognit. **36**, 111–121 (2003)
14. JJ Hull, A database for handwritten text recognition research. IEEE Trans. Pattern Anal. Mach. Intell. **16**(5), 550–554 (1994)
15. I Guyon, L Schomaker, R Plamondon, M Liberman, S Janet, in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Conference B: Computer Vision and Image Processing*. Unipen project of on-line data exchange and benchmarks, (1994), pp. 29–33
16. T Saito, H Yamada, K Yamamoto, On the data base etl9 of handprinted characters in jis chinese characters and its analysis. IEICE Trans (1985)
17. D Kim, Y Hwang, S Park, E Kim, S Paek, S Bang, in *Proceedings of the 2nd International Conference on Document Analysis and Recognition*. Handwritten korean character image database pe92, (1993), pp. 470–473
18. T-H Su, T-W Zhang, DJ Guan, Corpus-based hit-mw database for offline recognition of general-purpose chinese handwritten text. Int. J. Doc. Anal. Recognit. **10**(1), 27–38 (2007)
19. M Pechwitz, SS Maddouri, V Maergner, N Ellouze, H Amiri, in *In Proc. of CIFED*. Ifn/enit-database of handwritten arabic words, vol. 2, (2002), pp. 127–136
20. S Al-Maadeed, D Elliman, CA Higgins, A data base for arabic handwritten text recognition research. Int. Arab J. Inf. Technol. **1**, 117–121 (2004)
21. N Kharma, M Ahmed, R Ward, in *IEEE Canadian Conference on Electrical and Computer Engineering*. A new comprehensive database of handwritten arabic words, numbers, and signatures used for ocr testing, vol. 2, (1999), pp. 766–768
22. M Kherallah, A Elbaati, HE Abed, AM Alimi, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. The on/off (lmca) dual arabic handwriting database, (2008)
23. S Mozaffari, HE Abed, V Margner, K Faez, A Amirshahi, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. Ifn/farsi-database: A database of farsi handwritten city names, (2008)
24. AM Bidgoli, M Sarhadi, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. Iaut/phcn: Islamic azad university of tehran/persian handwritten city names, a very large database of handwritten persian words, (2008), pp. 192–197
25. U Bhattacharya, BB Chaudhuri, Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals. IEEE Trans. Pattern Anal. Mach. Intell. **31**(3), 444–457 (2009)
26. SA Maadeed, W Ayouby, A Hassaine, JM Aljaam, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. Quwi: An arabic and english handwriting dataset for offline writer identification, (2012), pp. 746–751
27. F Kleber, S Fiel, M Diem, R Sablatnig, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting, (2013), pp. 560–564
28. G Dimauro, S Impedovo, R Modugno, G Pirlo, in *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*. A new database for research on bank-check processing, (2002), pp. 524–528
29. E Kavallieratou, N Liolios, E Koutsogeorgos, N Fakotakis, G Kokkinakis, in *Proceedings of the 6th International Conference on Document Analysis and*

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 20 of 24

Recognition. The gruhd database of greek unconstrained handwriting, (2001), pp. 561–565

30. D Llorens, F Prat, A Marzal, JM Vilar, MJ Castro, JC Amengual, S Barrachina, A Castellanos, S Espana, JA Gomez, J Gorbe, A Gordo, V Palazon, G Peris, R Ramos-Garijo, F Zamora, in *Proceedings of the Sixth International Conference on Language Resources and Evaluation*. The ujipenchars database: a pen-based database of isolated handwritten characters, (2008)

31. M Nakagawa, K Matsumoto, Collection of on-line handwritten japanese character pattern databases and their analyses. Int. J. Doc. Anal. Recognit. **7**(1), 69–81 (2004)

32. S Wshah, G Kumar, V Govindaraju, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Script independent word spotting in offline handwritten documents based on hidden markov models, (2012), pp. 14–19

33. S Wshah, G Kumar, V Govindaraju, in *Proceedings of 21st International Conference on Pattern Recognition*. Multilingual word spotting in offline handwritten documents, (2012), pp. 310–313

34. A Fischer, A Keller, V Frinken, H Bunke, Lexicon-free handwritten word spotting using character hmms. Pattern Recognit. Lett. **33**(7), 934–942 (2012)

35. V Frinken, A Fischer, R Manmatha, H Bunke, A novel word spotting method based on recurrent neural networks. IEEE Trans. Pattern Anal. Mach. Intell. **34**(2), 211–224 (2012)

36. A Bensefia, T Paquet, L Heutte, A writer identification and verification system. Pattern Recognit. Lett. **26**(13), 2080–2092 (2005)

37. M Bulacu, L Schomaker, Text-independent writer identification and verification using textural and allographic features. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 701–717 (2007)

38. I Siddiqi, N Vincent, Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognit. **43**(11), 3853–3865 (2010)

39. ZA Daniels, HS Baird, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Discriminating features for writer identification, (2013), pp. 1385–1389

40. R Jain, D Doermann, in *Proceedings of International Conference on Document Analysis and Recognition*. Offline writer identification using k-adjacent segments, (2011), pp. 769–773

41. RP dos Santos, GS Clemente, TI Ren, GDC Cavalcanti, in *Proceeding of the 10th International Conference on Document Analysis and Recognition*. Text line segmentation based on morphology and histogram projection, (2009), pp. 651–655

42. M Zimmermann, H Bunke, in *Proceedings of the 16th International Conference on Pattern Recognition*. Automatic segmentation of the iam off-line database for handwritten english text, vol. 4, (2002), pp. 35–39

43. D Salvi, J Zhou, J Waggoner, S Wang, in *Proceedings of IEEE Workshop on Applications of Computer Vision*. Handwritten text segmentation using average longest path algorithm, (2013), pp. 505–512

44. S Gunter, H Bunke, Ensembles of classifiers for handwritten word recognition. Int. J. Doc. Anal. Recognit. **5**(4), 224–232 (2003)

45. H Bunke, S Bengio, A Vinciarelli, Offline recognition of unconstrained handwritten texts using hmms and statistical language models. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 709–720 (2004)

46. P Dreuw, P Doetsch, C Plahl, H Ney, in *Proceedings of the 18th IEEE International Conference on Image Processing*. Hierarchical hybrid mlp/hmm or rather mlp features for a discriminatively trained gaussian hmm: A comparison for offline handwriting recognition, (2011), pp. 3541–3544

47. B Gatos, I Pratikakis, SJ Perantonis, in *Proceedings of 18th International Conference on Pattern Recognition*. Hybrid off-line cursive handwriting word recognition, vol. 2, (2006), pp. 998–1002

48. M Liwicki, H Bunke, in *Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition*. Hmm-based on-line recognition of handwritten whiteboard notes, (2006)

49. M Liwicki, A Schlapbach, H Bunke, in *Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*. Writer-dependent recognition of handwritten whiteboard notes in smart meeting room environments, (2008), pp. 151–157

50. A Schlapbach, M Liwicki, H Bunke, A writer identification system for on-line whiteboard data. Pattern Recognit. **41**(7), 2381–2397 (2008)

51. M Liwicki, A Schlapbach, H Bunke, Automatic gender detection using on-line and off-line information. Pattern Anal. Appl. **14**(1), 87–92 (2011)

52. E Indermuhle, M Liwicki, H Bunke, in *Proceedings of the 9th International Workshop on Document Analysis Systems*. Iamondo-database: An online handwritten document database with non-uniform contents, (2010), pp. 97–104

53. E Indermuhle, V Frinken, H Bunke, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Mode detection in online handwritten documents using blstm neural networks, (2012), pp. 302–307

54. E Indermuhle, V Frinken, A Fischer, H Bunke, in *Proceedings of International Conference on Document Analysis and Recognition*. Keyword spotting in online handwritten documents containing text and non-text using blstm neural networks, (2011), pp. 73–77

55. A Delaye, CL Liu, in *Pattern Recognit Communications in Computer and Information Science*. Text/non-text classification in online handwritten documents with conditional random fields, vol. 321, (2012), pp. 514–521

56. A Fischer, E Indermuhle, H Bunke, G Viehhauser, M Stolz, in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. Ground truth creation for handwriting recognition in historical documents, (2010), pp. 3–10

57. IB Messaoud, H Amiri, HE Abed, V Margner, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. A multilevel text-line segmentation framework for handwritten historical documents, (2012), pp. 515–520

58. R Saabni, A Asi, J El-Sana, Text line extraction for historical document images. Pattern Recognit. Lett. **35**, 23–33 (2014)

59. IB Messaoud, H Amiri, H El Abed, V Margner, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Region based local binarization approach for handwritten ancient documents, (2012), pp. 633–638

60. I Ben Messaoud, H Amiri, H El-Abed, V Margner, in *Proceedings of the 11th International Conference on Information Science, Signal Processing and Their Applications*. Binarization effects on results of text-line segmentation methods applied on historical documents, (2012), pp. 1092–1097

61. EF Can, P Duygulu, A line-based representation for matching words in historical manuscripts. Pattern Recognit. Lett. **32**(8), 1126–1138 (2011)

62. V Frinken, A Fischer, CD Martnez-Hinarejos, in *Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing*. Handwriting recognition in historical documents using very large vocabularies, (2013), pp. 67–72

63. E Grosicki, M Carre, J-M Brodin, E Geoffrois, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. Rimes evaluation campaign for handwritten mail processing, (2008)

64. F Montreuil, E Grosicki, L Heutte, S Nicolas, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Unconstrained handwritten document layout extraction using 2d conditional random fields, (2009), pp. 853–857

65. F Montreuil, S Nicolas, E Grosicki, L Heutte, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. A new hierarchical handwritten document layout extraction based on conditional random field modeling, (2010), pp. 31–36

66. C Kermorvant, J Louradour, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Handwritten mail classification experiments with the rimes database, (2010), pp. 241–246

67. L Guichard, AH Toselli, B Couasnon, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Handwritten word verification by svm-based hypotheses re-scoring and multiple thresholds rejection, (2010), pp. 57–62

68. E Grosicki, H El-Abed, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Icdar 2009 handwriting recognition competition, (2009), pp. 1398–1402

69. E Grosicki, H El-Abed, in *Proceedings of 11th International Conference on Document Analysis and Recognition*. Icdar 2011-french handwriting recognition competition, (2011), pp. 1459–1463

70. O Morillot, L Likforman-Sulem, E Grosicki, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Comparative study of hmm and blstm segmentation-free approaches for the recognition of handwritten text-lines, (2013), pp. 783–787

71. T Bluche, H Ney, C Kermorvant, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Feature extraction with convolutional neural networks for handwritten word recognition, (2013)

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 21 of 24

72. A-L Bianne-Bernard, F Menasri, RA-H Mohamad, C Mokbel, C Kermorvant, L Likforman-Sulem, Dynamic and contextual information in hmm modeling for handwritten word recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(10), 2066–2080 (2011)

73. U Garain, T Paquet, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Off-line multi-script writer identification using ar coefficients, (2009), pp. 991–995

74. TM Ha, H Bunke, Off-line, handwritten numeral recognition by perturbation method. IEEE Trans. Pattern Anal. Mach. Intell. **19**(5), 535–539 (1997)

75. AK Jain, D Zongker, Representation and recognition of handwritten digits using deformable templates. IEEE Trans. Pattern Anal. Mach. Intell. **19**(12), 1386–1390 (1997)

76. M Shi, Y Fujisawa, T Wakabayashi, F Kimura, Handwritten numeral recognition using gradient and curvature of gray scale image. Pattern Recognit. **35**(10), 2051–2059 (2002)

77. L C-L, H Sako, H Fujisawa, Effects of classifier structures and training regimes on integrated segmentation and recognition of handwritten numeral strings. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1395–1407 (2004)

78. SJ Smith, MO Bourgoin, K Sims, HL Voorhees, Handwritten character classification using nearest neighbor in large databases. IEEE Trans. Pattern Anal. Mach. Intell. **16**(9), 915–919 (1994)

79. L S-W, L E-J, in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. Integrated segmentation and recognition of connected handwritten characters with recurrent neural network, vol. 1, (1995), pp. 413–416

80. Z R, D X, Z J, in *Proceedings of the 6th International Conference on Document Analysis and Recognition*. Offline handwritten character recognition based on discriminative training of orthogonal gaussian mixture model, (2001), pp. 221–225

81. E Kavallieratou, N Fakotakis, G Kokkinakis, in *Proceedings of the 16th International Conference on Pattern Recognition*. Handwritten character recognition based on structural characteristics, vol. 3, (2002), pp. 139–142

82. C-L Liu, K Nakashima, H Sako, H Fujisawa, Handwritten digit recognition: benchmarking of state-of-the-art techniques. Pattern Recognit. **36**(10), 2271–2285 (2003)

83. E Kussul, T Baidyk, Improved method of handwritten digit recognition tested on mnist database. Image Vision Comput. **22**(12), 971–981 (2004)

84. L Z, C Z, S W-C, Extraction and optimization of b-spline pbd templates for recognition of connected handwritten digit strings. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 132–139 (2002)

85. F Lauer, CY Suen, G Bloch, A trainable feature extractor for handwritten digit recognition. Pattern Recognit. **40**(6), 1816–1824 (2007)

86. S Benzoubeir, A Hmamed, H Qjidaa, in *Proceedings of International Conference on Multimedia Computing and Systems*. Hypergeometric laguerre moment for handwritten digit recognition, (2009), pp. 449–453

87. W Z, H Y, L S, W L, in *Proceedings of the 18th IEEE International Conference on Image Processing*. A biologically inspired system for fast handwritten digit recognition, (2011), pp. 1749–1752

88. M Cheriet, R Thibault, R Sabourin, in *Proceedings of IEEE International Conference on Image Processing*. A multi-resolution based approach for handwriting segmentation in gray-scale images, vol. 1, (1994), pp. 159–163

89. M Blumenstein, B Verma, in *Proceedings of the 6th International Conference on Document Analysis and Recognition*. Analysis of segmentation performance on the cedar benchmark database, (2001), pp. 1142–1146

90. TM Breul, in *Proceedings of the 2nd International Conference on Document Analysis and Recognition*. Recognition of handprinted digits using optimal bounded error matching, (1993), pp. 493–496

91. S Singh, M Hewitt, in *Proceedings of the 15th International Conference on Pattern Recognition*. Cursive digit and character recognition in cedar database, vol. 2, (2000), pp. 569–572

92. GE Hinton, P Dayan, M Revow, Modeling the manifolds of images of handwritten digits. IEEE Trans. Neural Netw. **8**(1), 65–74 (1997)

93. M Blumenstein, XY Liu, B Verma, in *Proceedings of IEEE International Joint Conference on Neural Networks*. A modified direction feature for cursive character recognition, vol. 4, (2004), pp. 2983–2987

94. F H-C, X Y-Y, Multilinguistic handwritten character recognition by bayesian decision-based neural networks. IEEE Trans. Signal Process. **46**(10), 2781–2789 (1998)

95. M Blumenstein, B Verma, in *Proceedings of International Joint Conference on Neural Networks*. A new segmentation algorithm for handwritten word recognition, vol. 4, (1999), pp. 2893–2898

96. H Yamada, Y Nakano, Cursive handwritten word recognition using multiple segmentation determined by contour analysis. IEICE Trans. Inf. Syst. **79**(5), 464–470 (1996)

97. MA Mohamed, P Gader, Generalized hidden markov models. ii. application to handwritten word recognition. IEEE Trans. Fuzzy Syst. **8**(1), 82–94 (2000)

98. B Verma, P Gader, W Chen, Fusion of multiple handwritten word recognition techniques. Pattern Recognit. Lett. **22**(9), 991–998 (2001)

99. C Viard-Gaudin, PM Lallican, S Knerr, P Binter, in *Proceedings of the 5th International Conference on Document Analysis and Recognition*. The ireste on/off (ironoff) dual handwriting database, (1999), pp. 455–458

100. YH Tay, P-M Lallican, M Khalid, C Viard-Gaudin, S Knerr, in *Proceedings of the 6th International, Symposium on Signal Processing and Its Applications*. Offline handwritten word recognition using a hybrid neural network and hidden markov model, vol. 2, (2001), pp. 382–385

101. E Poisson, C Viard Gaudin, P-M Lallican, in *Proceedings of the 9th International Conference on Neural Information Processing*. Multi-modular architecture based on convolutional neural networks for online handwritten character recognition, vol. 5, (2002), pp. 2444–2448

102. CO Freitas, LS Oliveira, F Bortolozzi, SB Aires, Handwritten character recognition using nonsymmetrical perceptual zoning. Int. J. Pattern Recognit. Artif. Intell. **21**(01), 135–155 (2007)

103. M Dinesh, MK Sridhar, in *Proceedings of the 9th International Conference on Document Analysis and Recognition*. A feature based on encoding the relative position of a point in the character for online handwritten character recognition, vol. 2, (2007), pp. 1014–1017

104. GX Tan, C Viard-Gaudin, AC Kot, Automatic writer identification framework for online handwritten documents using character prototypes. Pattern Recognit. **42**(12), 3313–3323 (2009)

105. N Serrano, F Castro, A Juan, in *Proceedings of the 8th Language Resources and Evaluation Conference*. The rodrigo database, (2010)

106. N Serrano, A Sanchis, A Juan, in *Proceedings of the 15th International Conference on Intelligent User Interfaces*. Balancing error and supervision effort in interactive-predictive handwriting recognition, (2010), pp. 373–376

107. V Romero, N Serrano, AH Toselli, A S, E Vidal, Handwritten text recognition for historical documents. Language Technologies for Digital Humanities and Cultural Heritage. **90** (2011)

108. N Serrano, A Gimenez, J Civera, A Sanchis, A Juan, Interactive handwriting recognition with limited user effort. Int. J. Doc. Anal. Recognit. **17**(1), 47–59 (2014)

109. PR Aryan, I Supriana, A Purwarianti, in *International Conference on Electrical Engineering and Informatics*. Development of indonesian handwritten text database for offline character recognition, (2011), pp. 1–4

110. S Impedovo, G Facchini, FM Mangini, in *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems*. A new cursive basic word database for bank-check processing systems, (2012), pp. 450–454

111. A Shivram, C Ramaiah, S Setlur, V Govindaraju, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Ibm_ub_1: A dual mode unconstrained english handwriting dataset, (2013), pp. 13–17

112. A Shivram, B Zhu, S Setlur, M Nakagawa, V Govindaraju, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Segmentation based online word recognition: A conditional random field driven beam search strategy, (2013), pp. 852–856

113. A Shivram, C Ramaiah, V Govindaraju, A hierarchical bayesian approach to online writer identification. IET Biometrics. **2**(4), 191–198 (2013)

114. S Fiel, R Sablatnig, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Writer identification and writer retrieval using the fisher vector on visual vocabularies, (2013), pp. 545–549

115. M Diem, S Fiel, A Garz, M Keglevic, F Kleber, R Sablatnig, in *Proceedings of 12th International Conference on Document Analysis and Recognition*. Icdar 2013 competition on handwritten digit recognition (hdrc 2013), (2013), pp. 1422–1427

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 22 of 24

116. L Schomaker, L Vuurpijl, L Schomaker, Forensic writer identification: a benchmark data set and a comparison of two systems (2000)

117. M Bulacu, L Schomaker, in *Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition*. Combining multiple features for text-independent writer identification and verification, (2006)

118. F Farooq, V Govindaraju, M Perrone, in *Proceedings of the 8th International Conference on Document Analysis and Recognition*. Pre-processing methods for handwritten arabic documents, (2005), pp. 267–271

119. SS Maddouri, FB Samoud, K Bouriel, N Ellouze, H El-Abed, in *Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition*. Baseline extraction: Comparison of six methods on ifn/enit database, (2008)

120. HM Eraqi, S Abdelazeem, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. A new efficient graphemes segmentation technique for offline arabic handwriting, (2012), pp. 95–100

121. M Pechwitz, V Maergner, in *Proceedings of the 7th International Conference on Document Analysis and Recognition*. Hmm based approach for handwritten arabic word recognition using the ifn/enit-database, (2003)

122. R Al-Hajj, L Likforman-Sulem, C Mokbel, in *Proceedings 8th International Conference on Document Analysis and Recognition*. Arabic handwriting recognition using baseline dependant features and hidden markov modeling, (2005), pp. 893–897

123. H El-Abed, V Margner, in *Proceedings of the 9th International Symposium on Signal Processing and Its Applications*. The ifn/enit-database-a tool to develop arabic handwriting recognition systems, (2007), pp. 1–4

124. F Menasri, N Vincent, E Augustin, M Cheriet, in *Proceedings of the 9th International Conference on Document Analysis and Recognition*. Shape-based alphabet for off-line arabic handwriting recognition, vol. 2, (2007), pp. 969–973

125. A-HM R, L Likforman-Sulem, C Mokbel, Combining slanted-frame classifiers for improved hmm-based arabic handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(7), 1165–1177 (2009)

126. Y Kessentini, T Paquet, A BH, Off-line handwritten word recognition using multi-stream hidden markov models. Pattern Recognit. Lett. **31**(1), 60–70 (2010)

127. M Bulacu, L Schomaker, A Brink, in *Proceedings of the 9th International Conference on Document Analysis and Recognition*. Text-independent writer identification and verification on offline arabic handwriting, vol. 2, (2007), pp. 769–773

128. MN Abdi, M Khemakhem, H Ben-Abdallah, in *Proceedings of the 24th International Symposium on Computer and Information Sciences*. A novel approach for off-line arabic writer identification based on stroke feature combination, (2009), pp. 597–600

129. D Chawki, S-M Labiba, in *Proceedings of International Conference on Machine and Web Intelligence*. A texture based approach for arabic writer identification and verification, (2010), pp. 115–120

130. H El-Abed, M Kherallah, V Margner, AM Alimi, On-line arabic handwriting recognition competition: Adab database and participating systems. Int. J. Doc. Anal. Recognit. **14**(1), 15–23 (2011)

131. I Hosny, S Abdou, A Fahmy, in *Proceedings of the 11th International Symposium on Distributed Computing and Applications to Business, Engineering and Science*. Using advanced hidden markov models for online arabic handwriting recognition, (2011), pp. 565–569

132. SA Azeem, H Ahmed, *Recognition of segmented online arabic handwritten characters of the adab database*, (2011), pp. 204–207

133. HM Eraqi, SA Azeem, in *Proceedings of International Conference on Document Analysis and Recognition*. An on-line arabic handwriting recognition system: Based on a new on-line graphemes segmentation technique, (2011), pp. 409–413

134. A Chaabouni, H Boubaker, M Kherallah, AM Alimi, H El-Abed, in *Proceedings of International Conference on Document Analysis and Recognition*. Multi-fractal modeling for on-line text-independent writer identification, (2011), pp. 623–627

135. S Al-Maadeed, D Elliman, CA Higgins, in *Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition*. A data base for arabic handwritten text recognition research, (2002), pp. 485–489

136. S Al-Madeed, C Higgins, D Elliman, Off-line recognition of handwritten arabic words using multiple hidden markov models. Knowledge-Based Syst. **17**(2), 75–79 (2004)

137. S Al-Maadeed, Text-dependent writer identification for arabic handwriting. J. Electr. Comput. Eng (2012)

138. M Cheriet, Y Al-Ohali, NE Ayat, CY Suen, in *Digital Document Processing*. Arabic cheque processing system: Issues and future trends, (2007), pp. 213–234

139. NB Amara, O Mazhoud, N Bouzrara, N Ellouze, Arabase:a relational database for arabic ocr systems. Int. Arab J. Inf. Technol. **2**(4), 259–266 (2005)

140. AT Sahlol, CY Suen, MR Elbasyouni, A Sallol, A proposed ocr algorithm for the recognition of handwritten arabic characters. J. Pattern Recognit. Intell. Syst, 8–22 (2014)

141. H Alamri, C He, CY Suen, in *Computer Analysis of Images and Patterns Lecture Notes in Computer Science*. A new approach for segmentation and recognition of arabic handwritten touching numeral pairs, (2009), pp. 165–172

142. M Khayyat, L Lam, CY Suen, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Arabic handwritten word spotting using language models, (2012), pp. 43–48

143. F MR, M Cheriet, M Adankon, K Filonenko, R Wisnovsky, in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*. Ibn sina: a database for research on processing and understanding of arabic manuscripts images, (2010), pp. 11–18

144. M Cheriet, RF Moghaddam, in *Guide to OCR for Arabic Scripts*. A robust word spotting system for historical arabic manuscripts, (2012), pp. 453–484

145. HZ Nafchi, SM Ayatollahi, RF Moghaddam, M Cheriet, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. An efficient ground truthing tool for binarization of historical manuscripts, (2013), pp. 807–811

146. SA Mahmoud, I Ahmad, M Alshayeb, WG Al-Khatib, MT Parvez, GA Fink, V Margner, HE Abed, in *Proceedings of the International Conference on Frontiers in Handwriting Recognition*. Khatt: Arabic offline handwritten text database, (2012), pp. 449–454

147. SA Mahmoud, I Ahmad, WG Al-Khatib, M Alshayeb, MT Parvez, V Märgner, GA Fink, Khatt: An open arabic offline handwritten text database. Pattern Recognit. **47**(3), 1096–1112 (2014)

148. S Al Maadeed, W Ayouby, A Hassaine, JM Aljaam, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Quwi: An arabic and english handwriting dataset for offline writer identification, (2012), pp. 746–751

149. A Hassaine, S Maadeed, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. Icfhr 2012 competition on writer identification challenge 2: Arabic scripts, (2012), pp. 835–840

150. A Hassaine, S Al Maadeed, J Aljaam, A Jaoua, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Icdar 2013 competition on gender prediction from handwriting, (2013), pp. 1417–1421

151. I Siddiqi, C Djeddi, A Raza, L Souici-Meslati, Automatic analysis of handwriting for gender classification. Pattern Anal. Appl. (2014)

152. S Al-Maadeed, A Hassaine, Automatic prediction of age, gender, and nationality in offline handwriting. EURASIP J. Image Video Process. (2014)

153. A Mezghani, S Kanoun, M Khemakhem, HE Abed, in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition*. A database for arabic handwritten text image recognition and writer identification, (2012), pp. 399–402

154. F Slimane, V Margner, in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition*. A new text-independent gmm writer identification system applied to arabic handwriting, (2014), pp. 708–713

155. F Khan, A Bouridane, F Khelifi, R Almotaeryi, S Almaadeed, in *Proceedings of Control, Decision and Information Technologies (CoDIT)*. Efficient segmentation of sub-words within handwritten arabic words, (2014), pp. 684–689

156. R Ravani, P Nooralishahi, AS Amani, in *Proceedings of the 3rd European Workshop on Visual Information Processing*. A novel approach for persian/arabic intelligent word recognition (IEEE, 2011), pp. 292–297

157. R Ravani, P Nooralishahi, in *Proceedings of the World Congress in Computer Science, Computer Engineering, and Applied Computing*. Using dynamic time warping for persian handwriting recognition, (2011)

158. F Nadi, J Sadri, A Foroozandeh, in *Proceedings of the First Iranian Conference on Pattern Recognit. and Image Analysis*. A novel method for slant correction of persian handwritten digits and words, (2013), pp. 1–7

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 23 of 24

159. N Nobile, CL He, MW Sagheer, L Lam, CY Suen, in *Proceedings of International Conference on Document Analysis and Recognition*. Digit/symbol pruning and verification for arabic handwritten digit/symbol spotting, (2011), pp. 648–652

160. HP Jifroodian, CY Suen, in *Guide to OCR for Arabic Scripts*. Handwritten farsi word recognition using hidden markov models, (2012), pp. 273–295

161. Z Imani, A Ahmadyfard, A Zohrevand, M Alipour, in *Proceedings of the 8th Iranian Conference on Machine Vision and Image Processing*. Offline handwritten farsi cursive text recognition using hidden markov models, (2013), pp. 75–79

162. M Ziaratban, K Faez, F Bagheri, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Fht: An unconstraint farsi handwritten text database, (2009), pp. 281–285

163. R Safabaksh, AR Ghanbarian, G Ghiasi, in *Proceedings of the 8th Iranian Conference on Machine Vision and Image Processing*. Haft: A handwritten farsi text database, (2013), pp. 89–94

164. MW Sagheer, C He, N Nobile, CY Suen, in *Image Analysis and Processing Lecture Notes in Computer Science*. A new large urdu database for off-line handwriting recognition, (2009), pp. 538–546

165. MW Sagheer, N Nobile, H CL, CY Suen, in *Proceedings of the 20th International Conference on Pattern Recognition*. A novel handwritten urdu word spotting based on connected components analysis, (2010), pp. 2013–2016

166. MW Sagheer, H CL, N Nobile, CY Suen, in *Proceedings of the 20th International Conference on Pattern Recognition*. Holistic urdu handwritten word recognition using support vector machine, (2010), pp. 1900–1903

167. A Raza, I Siddiqi, A Abidi, F Arif, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. An unconstrained benchmark urdu handwritten sentence database with automatic line segmentation, (2012), pp. 491–496

168. SH Kim, D J-I, in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. Off-line recognition of korean scripts using distance matching and neural network classifiers, vol. 1, (1995), pp. 34–37

169. J S-H, N Y-S, K H-K, in *Proceedings of the 7th International Conference on Document Analysis and Recognition*. Non-similar candidate removal method for off-line handwritten korean character recognition, (2003), pp. 323–328

170. K Seo, J Kim, J Yoon, K Chung, Comparison of feature performance and its application to feature combination in off-line handwritten korean alphabet recognition. Int. J. Pattern Recognit. Artif. Intell. **12**(02), 251–261 (1998)

171. M Nakagawa, T Higashiyama, Y Yamanaka, S Sawada, L Higashigawa, K Akiyama, in *Proceedings of the 4th International Conference on Document Analysis and Recognition*. On-line handwritten character pattern database sampled in a sequence of sentences without any writing instructions, (1997)

172. O Velek, S Jaeger, M Nakagawa, in *Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition*. A new warping technique for normalizing likelihood of multiple classifiers and its effectiveness in combined on-line/off-line japanese character recognition, (2002), pp. 177–182

173. M Nakagawa, B Zhu, M Onuma, A model of on-line handwritten japanese text recognition free from line direction and writing format constraints. IEICE Trans. Inf. Syst. **88**(8), 1815–1822 (2005)

174. B Zhu, X-D Zhou, C-L Liu, M Nakagawa, A robust model for on-line handwritten japanese text recognition. Int. J. Doc. Anal. Recognit. **13**(2), 121–131 (2010)

175. B Zhu, M Nakagawa, in *Proceedings of International Conference on Document Analysis and Recognition*. On-line handwritten japanese characters recognition using a mrf model with parameter optimization by crf, (2011), pp. 603–607

176. H Zhang, J Guo, in *Proceedings of Sino-Japan Symposium on Intelligent Information Networks*. Introduction to hcl2000 database, (2000)

177. H Zhang, J Guo, G Chen, C Li, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Hcl2000-a large-scale handwritten chinese character database for handwritten character recognition, (2009), pp. 286–290

178. K Ueki, T Hayashida, T Kobayashi, in *Proceedings of the 18th International Conference on Pattern Recognition*. Improved handwritten character recognition performance by heteroscedastic linear discriminant analysis, vol. 1, (2006), pp. 880–883

179. H Liu, X Ding, Handwritten chinese character recognition based on mirror image learning and the compound mahalanobis function. J.-Tsinghua Univ. **46**(7), 1239 (2006)

180. Z Zhang, L Jin, K Ding, X Gao, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Character-sift: a novel feature for offline handwritten chinese character recognition, (2009), pp. 763–767

181. L Jin, Y Gao, G Liu, Y Li, K Ding, Scut-couch2009 a comprehensive online unconstrained chinese handwriting database and benchmark evaluation. Int. J. Doc. Anal. Recognit. **14**, 53–64 (2011)

182. YY Li, LW Jin, XH Zhu, in *Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition*. A comprehensive online unconstrained chinese handwritingdataset, (2008)

183. S Huang, L Jin, J Lv, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. A novel approach for rotation free online handwritten chinese character recognition, (2009), pp. 1136–1140

184. G Liu, L Jin, K Ding, H Yan, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. A new approach for synthesis and recognition of large scale handwritten chinese words, (2010), pp. 571–575

185. Y Gao, L Jin, C He, G Zhou, in *Proceedings of International Conference on Document Analysis and Recognition*. Handwriting character recognition as a service: A new handwriting recognition system based on cloud computing, (2011), pp. 885–889

186. D Tao, L Liang, L Jin, Y Gao, in *Proceedings of International Conference on Document Analysis and Recognition*. Similar handwritten chinese character recognition using discriminative locality alignment manifold learning, (2011), pp. 1012–1016

187. D-H Wang, C-L Liu, J-L Yu, X-D Zhou. A database of online handwritten chinese characters, (2009), pp. 1206–1210

188. C-L Liu, F Yin, D-H Wang, Q-F Wang, in *Proceedings of International Conference on Document Analysis and Recognition*. Casia online and offline chinese handwriting databases, (2011), pp. 37–41

189. D-H Wang, C-L Liu, X-D Zhou, An approach for real-time recognition of online chinese handwritten sentences. Pattern Recognit. **45**(10), 3661–3675 (2012)

190. Q-F Wang, F Yin, C-L Liu, in *Proceedings of the 10th IAPR International Workshop on Document Analysis Systems*. Improving handwritten chinese text recognition by unsupervised language model adaptation, (2012), pp. 110–114

191. Q-F Wang, E Cambria, C-L Liu, A Hussain, Common sense knowledge for handwritten chinese text recognition. Cognitive Comput. **5**(2), 234–242 (2013)

192. Y Shao, C Wang, B Xiao, Fast self-generation voting for handwritten chinese character recognition. Int. J. Doc. Anal. Recognit. **16**(4), 413–424 (2013)

193. H Zhang, C-L Liu, in *Proceedings of International Conference on Document Analysis and Recognition*. A lattice-based method for keyword spotting in online chinese handwriting, (2011), pp. 1064–1068

194. L Xu, F Yin, Q-F Wang, C-L Liu, in *Proceedings of Intl. Conference on Frontiers in Handwriting Recognition*. A touching character database from chinese handwriting for assessing segmentation algorithms, (2012), pp. 89–94

195. L Xu, F Yin, C-L Liu, in *Proceedings of Chinese Conference on Pattern Recognit*. Touching character splitting of chinese handwriting using contour analysis and dtw, (2010), pp. 1–5

196. L Xu, F Yin, Q-F Wang, C-L Liu, An over-segmentation method for single-touching chinese handwriting with learning-based filtering. Int. J. Doc. Anal. Recognit. **17**(1), 91–104 (2014)

197. U Bhattacharya, BB Chaudhuri, in *Proceedings of 8th International Conference on Document Analysis and Recognition*. Databases for research on recognition of handwritten characters of indian scripts, (2005), pp. 789–793

198. U Bhattacharya, SK Parui, B Shaw, K Bhattacharya, in *Proceedinsgs of the 10th International Workshop on Frontiers in Handwriting Recognition*. Neural combination of ann and hmm for handwritten devanagari numeral recognition, (2006)

199. N Sharma, U Pal, F Kimura, S Pal, in *Computer Vision, Graphics and Image Processing Lecture Notes in Computer Science*. Recognition of off-line

Hussain *et al. EURASIP Journal on Image and Video Processing* (2015) 2015:46

Page 24 of 24

handwritten devnagari characters using quadratic classifier, vol. 4338, (2006), pp. 805–816

200. MJK Singh, R Dhir, R Rani, Performance comparison of devanagari handwritten numerals recognition. Int. J. Comput. Appl. **22** (2011)

201. A Alaei, P Nagabhushan, U Pal, in *Proceedings of International Conference on Document Analysis and Recognition*. A benchmark kannada handwritten document dataset and its segmentation, (2011), pp. 141–145

202. S Thadchanamoorthy, ND Kodikara, HL Premaretne, U Pal, F Kimura, in *Proceedings O the 12th International Conference on Document Analysis and Recognition*. Tamil handwritten city name database development and recognition for postal automation, (2013), pp. 793–797

203. VJ Dongre, VH Mankar, Development of comprehensive devnagari numeral and character database for offline handwritten character recognition. Appl. Comput. Intell. Soft Comput (2012)

204. YE Saady, A Rachidi, M Yassa, Amhcd: A database for amazigh handwritten character recognition research. Int. J. Comput. Appl. **27**(4) (2011)

205. M Amrouch, Y Es-saady, A Rachidi, M El-Yassa, D Mammass, Handwritten amazigh character recognition system based on continuous hmms and directional features. Int. J. Graphics Vis. Image Process (2012)

206. M Amrouch, Y Es-saady, A Rachidi, M El-Yassa, D Mammass, A novel feature set for recognition of printed amazigh text using maximum deviation and hmm. Int. J. Comput. Appl. **44** (2012)

207. E Kavallieratou, N Fakotakis, G Kokkinakis, in *Proceedings of the 16th International Conference on Pattern Recognition*. Handwritten character recognition based on structural characteristics, vol. 3, (2002), pp. 139–142

208. E Kavallieratou, N Fakotakis, G Kokkinakis, An unconstrained handwriting recognition system. Int. J. Doc. Anal. Recognit. **4**(4), 226–242 (2002)

209. G Vamvakas, B Gatos, S Petridis, N Stamatopoulos, in *Proceedings of the 9th International Conference on Document Analysis and Recognition*. An efficient feature extraction and dimensionality reduction scheme for isolated greek handwritten character recognition, vol. 2, (2007), pp. 1073–1077

210. E Kavallieratou, S Stamatatos, H Antonopoulou, in *Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition*. Machine-printed from handwritten text discrimination, (2004), pp. 312–316

211. L Ma, H Liu, J Wu, in *Proceedings of the International Conference on Document Analysis and Recognition*. Mrg-ohtc database for online handwritten tibetan character recognition, (2011), pp. 207–211

212. L Ma, J Wu, in *Proceedings of International Conference on Frontiers in Handwriting Recognition*. A component-based on-line handwritten tibetan character recognition method using conditional random field, (2012), pp. 704–709

213. L Ma, J Wu, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Semi-automatic tibetan component annotation from online handwritten tibetan character database by optimizing segmentation hypotheses, (2013), pp. 1340–1344

214. E Grosicki, M Carree, J-M Brodin, E Geoffrois, in *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Results of the rimes evaluation campaign for handwritten mail processing, (2009), pp. 941–945

215. A Giménez, I Khoury, J Andrés-Ferrer, A Juan, Handwriting word recognition using windowed bernoulli hmms. Pattern Recognit. Lett. **35**, 149–156 (2014)

216. F Zamora-Martinez, V Frinken, S Espana-Boquera, MJ Castro-Bleda, A Fischer, H Bunke, Neural network language models for off-line handwriting recognition. Pattern Recognit. **47**(4), 1642–1652 (2014)

217. H El-Abed, V Margner, Icdar 2009-arabic handwriting recognition competition. Int. J. Doc. Anal. Recognit. **14**(1), 3–13 (2011)

218. D Bertolini, LS Oliveira, E Justino, R Sabourin, Texture-based descriptors for writer identification and verification. Expert Syst. Appl. **40**(6), 2069–2080 (2013)

219. MN Abdi, M Khemakhem, A model-based approach to offline text-independent arabic writer identification and verification. Pattern Recognit. **48**(5), 1890–1903 (2015)

220. C Djeddi, A Tebessa, S Al-Maadeed, A Gattal, I Siddiqi, L Souici-Meslati, A Annaba, H El Abed, in *Proceedings of the 13th International Conference on Document Analysis and Recognition*. Icdar2015 competition on multi-script writer identification and gender classification using quwi database, (2015)

221. V Margner, M Pechwitz, HE Abed, in *Proceedings of 8th International Conference on Document Analysis and Recognition*. Icdar 2005 arabic handwriting recognition competition, (2005), pp. 70–74

222. V Margner, H El-Abed, in *Proceedings of 9th International Conference on Document Analysis and Recognition*. Arabic handwriting recognition competition, vol. 2, (2007), pp. 1274–1278

223. V Margner, HE Abed, in *Proceedings of International Conference on Frontiers in Handwriting Recognition (ICFHR)*. Icfhr 2010 - arabic handwriting recognition competition, (2010), pp. 709–714

224. V Margner, HE Abed, in *Proceedings of International Conference on Document Analysis and Recognition*. Icdar 2011 - arabic handwriting recognition competition, (2011), pp. 1444–1448

225. H El-Abed, V Margner, M Kherallah, A Alimi, in *Proceedings of 10th International Conference on Document Analysis and Recognition*. Icdar 2009 online arabic handwriting recognition competition, (2009), pp. 1388–1392

226. M Kherallah, N Tagougui, A Alimi, H El-Abed, V Margner, in *Proceedings of International Conference on Document Analysis and Recognition*. Online arabic handwriting recognition competition, (2011), pp. 1454–1458

227. S Mozaffari, H Soltanizadeh, in *Proceedings of 10th International Conference on Document Analysis and Recognition*. Icdar 2009 handwritten farsi/arabic character recognition competition, (2009), pp. 1413–1417

228. CL Liu, F Yin, QF Wang, DH Wang, in *Proceedings of International Conference on Document Analysis and Recognition*. Icdar 2011 chinese handwriting recognition competition, (2011), pp. 1464–1469

229. F Yin, QF Wang, XY Zhang, CL Liu, in *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Icdar 2013 chinese handwriting recognition competition, (2013), pp. 1464–1470

230. M Diem, S Fiel, F Kleber, R Sablatnig, JM Saavedra, D Contreras, JM Barrios, LS Oliveira, in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition*. Icfhr 2014 competition on handwritten digit string recognition in challenging datasets (hdsrc 2014), (2014), pp. 779–784

231. F Slimane, S Awaida, A Mezghani, MT Parvez, S Kanoun, S Mahmoud, V Margner, *et al*, in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition*. Icfhr2014 competition on arabic writer identification using ahtid/mw and khatt databases, (2014), pp. 797–802