UTSig: A Persian Offline Signature Dataset

Amir Soleimani 1*, Kazim Fouladi 2, Babak N. Araabi 1,3

1 Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
2 School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
3 School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
* A.Soleimani.B@gmail.com

This paper is a preprint of a paper submitted to IET Biometrics and is subject to Institution of Engineering and Technology Copyright. If accepted, the copy of record will be available at IET Digital Library

Abstract:
The crucial role of datasets in signature verification systems has motivated researchers to collect signature samples. However, with regard to the distinct characteristics of Persian signature, existing offline signature datasets cannot be used in Persian systems. This paper presents a new and public Persian offline signature dataset, UTSig, which consists of 8280 images from 115 classes that each class has 27 genuine, 3 opposite-hand signatures of the genuine signer, and 42 skilled forgeries made by 6 forgers from 230 people. Compared to the other public datasets, UTSig has larger number of samples, classes, and forgers. Meanwhile its samples were collected by considering variables such as signing period, writing instrument, signature box size, and number of observable samples for forgers. Reviewing the main characteristics of offline signature datasets, we statistically show that Persian signatures has fewer number of branch points and end points. We propose and test four different training and testing setups for UTSig. Results of our experiments show that training genuine samples along with opposite-hand signed samples and random forgeries can improve the performance in terms of equal error rate and minimum cost of log likelihood ratio which is an information theoretic criterion.

1. Introduction

Signature is one of the most widespread personal attributes used for authentication. It is simple, cheap and accepted by people, official organizations and courts. However, signature verification suffers from factors that reduce its performance, such as variations caused by writing instrument, paper and physical condition of writer, possibility of being forged and the need to considerable number of samples in authentication process.

Designing automatic signature verification systems, researchers have attempted to help Forensic Handwriting Experts (FHEs) in decision making. In the literature, automatic signature verification systems, based on acquisition approaches, are divided into two categories: offline and online. In offline systems, signatures are known as images while in online systems, position, velocity, pen orientation, and pressure sequences describe samples [1–5]. It is clear that the abundance and uniqueness of information in online mode provide more accurate results [2], but online mode is not completely natural for users [6]. Therefore, whereas offline system has less accuracy, it would be more applicable if it obtains desirable results.
Signatures based on authenticity are divided into three categories: genuine (authentic); forgery and disguised (i.e. when author tries to show his signature as forgery [7]). Moreover, different terms are used to distinguish forgery types, for instance in [8], authors divide forgery into: simple forgery (i.e. when forger has no attempt to mimic a signature); random forgery (i.e. when forger uses his signature instead of genuine signature); and freehand or skilled forgery (i.e. when forger tries to practice and simulate genuine signature as closely as possible). In addition, traced forgery is defined in some papers as following a signature with an instrument [9, 10]. Authors in [11] define simple forgery as forging by ordinary people and skilled forgery as the result of expert’s efforts. However, similar to the most common definition in the literature, we divide forgery into random forgery and skilled forgery. For an authentic person, a sample is random forgery when it is completely dissimilar (i.e. signed without having the genuine sample) or when it is a genuine signature from other authors. Skilled forgery is defined as the samples made by ordinary people with relatively remarkable efforts when they have the genuine sample(s).

Datasets are vital part of signature verification systems. They are prerequisites for training classifiers and for comparing different systems by using common datasets. Therefore, collecting datasets can strongly cause advancement in the filed. However, to provide more realistic results, firstly, datasets must be the samples of the people signatures in the community that the verification system is designed for. The reason is that signatures look different in distinct cultures [12], for instance, English signatures usually consist of reshaped writers’ names while Persian signatures are cursive and independent of their names [13]. Consequently, a Persian signature verification system strongly needs a Persian signature dataset. Secondly, datasets must be rich. To explain, richness for signature datasets refers to the number of samples and participants, and the variables involved in sample collection procedure such as signing period, writing instrument, paper, provided space for signing, samples showed to forgers, forgers’ efforts, meta-data and etc. These variables exist in daily life, for instance, signature of a person can change during the time, or when it is written by different pens or when the provided space is limited. Hence these variables must be considered as much as possible to make datasets more similar to what occurs in real conditions.

In the offline signature verification literature, datasets are mainly collected in Western, Chinese, or Japanese societies. Samples of these datasets are strongly different from the Persian signature styles and consequently cannot be used for Persian signature verification systems. To our knowledge, there is only one small and non-rich Persian offline signature dataset that does not address the researchers’ needs in this area.

In addition to the need of collecting and introducing novel and rich culture-based signature datasets, it is strongly essential to define standard experimental setups to provide standard and fair comparison.
between results of different systems. As these standard setups are not clearly available for all public datasets, many results of existing papers are incomparable because of their different training and testing conditions (e.g. different number of genuine samples in training set, adding or neglecting random forgeries in training set).

This paper introduces a new Persian offline signature dataset, UTSig, which is available to the research community. This rich dataset contains a significant number of classes and samples that several variables have been considered during signature collection procedure. UTSig provides the research community with the opportunity of training, testing, and comparing different Persian offline signature systems, and evaluating different culture-independent classifiers on a rich dataset using its proposed standard experimental setups.

In this paper, we compare the new dataset with other public datasets in terms of considered variables and show the distinct characteristics of Persian signatures (number of branch points and end points). We propose and evaluate four standard experimental setups for UTSig, and to see the performance of a common verification system on the public datasets in similar conditions, we test a same setup on UTSig and other datasets.

For more information and details about signature verification background we address to surveys done in this field. Authors in [8] and [14] survey the literature up to 1988 and 1993, respectively. In [15], signature verification is described as an application of handwritten recognition. Authors in [6], survey the literature up to 2007 and more recent developments can be found in [10], [15] and [16].

The rest of this paper is organized as follows: Section 2 reviews popular offline signature datasets. Section 3 introduces the new Persian offline signature dataset. Section 4 proposes experimental setups. In section 5, offline signature datasets are compared. Experiments and results are presented in section 6 and section 7 concludes the paper.

2. State of the Art in Offline Signature Datasets

There are several offline signature datasets collected in a few communities. This section reviews the main characteristics of popular datasets in the literature.

Spanish dataset MCYT-75 [18], a sub-corpus of MCYT bimodal database [19], has 75 classes containing 15 genuine signatures and 15 forgeries contributed by 3 user-specific forgers. Signatures were acquired by inking pen and paper over a pen tablet. To make forgeries, they gave forgers genuine images and asked them to imitate the shape and natural dynamics.

---

1 UTSig Dataset is freely available at MLCM lab website: http://mlcm.ut.ac.ir/Datasets.html
GPDSsignature [20] has 160 classes with 24 genuine samples gathered in a single day and 30 forged samples imitated by 10 forgers form 10 genuine specimens. To make forgeries, forgers had a random genuine sample and enough time. The collection was built by black or blue ink on white papers with 2 different size boxes.

GPDS-960 [11] contains 960 classes with 24 genuine signatures and 30 forgeries for each one. Genuine samples were signed in a single day on papers with 2 different size boxes. Totally, forgeries were produced by 1920 people apart from genuine persons and each forger had 5 genuine samples from 5 specimens (one sample per specimen).

ICDAR2009 signature competition [2] offline dataset, contains training and evaluations sets. For training, NISDCC signature collection acquired in WANDA project [21] is used. It has 12 classes, for each one 5 genuine and 5 forged samples. Forged signatures were written by 31 forgers. Evaluation set was collected in the Netherlands Forensic Institute (NFI) and contains 100 classes. Each class has 12 genuine samples and 6 forgeries created by 4 forgers from 33 writers.

A Persian signature dataset (FUM) was used in [22]. It has 20 different classes; each class contains 20 genuine and 10 forged signatures. Further information about this dataset is not available.

In 4NSigComp2010 [7] La Trobe signature collection was used. Its training set is composed of 9 reference signatures by one author and 200 questioned signatures, including: 76 genuine; 104 simulated (forged) signatures written by 27 freehand forgers and 20 disguised samples by its own author. Genuine and disguised samples were signed over a week. In addition, the author wrote another 81 genuine signatures to be used in forgery process. To build forgeries, 3 samples from 81 samples were given to forgers and asked them to imitate without tracing in 2 ways: forging 3 times without practice and simulating 3 times after 15 times practice. The test set has 25 signatures of another person written during 5 days and 100 questioned samples, including: 3 genuine; 7 disguise and 90 simulated (forged) signatures written by 34 freehand forgers from lay persons and calligraphers. Both sets were written by ball-point pen and same paper.

SigComp2011 [3] offline dataset contains Chinese and Dutch signature. Chinese and Dutch training set have: 235 and 240 genuine; 340 and 123 forged signatures totally for 10 authors, respectively. Furthermore respectively, 116 and 648 references; 120 and 648 questioned; 367 and 638 forged samples are provided for 10 and 54 different Chinese and Dutch authors in test sets.

In 4NSigComp2012 [23] a new dataset was used. It contains 3 authentic authors (classes) with 15 to 20 signatures as references and 100 to 250 questioned signatures, including: 20 to 50 genuine; 8 to 47 disguised and 42 to 160 forged samples. Genuine and disguised samples were collected during 10 to 15
days. The number of forgers varied from 2 to 31 and they had 3 to 6 authentic samples. Forgers were asked to imitate signature (without tracing) by pen, pencil, with and without practice.

SigWiComp2013 [5] introduced new Dutch and Japanese offline signature dataset. Dutch one has 27 authentic persons who signed 10 times with arbitrary writing instruments during 5 days. For forgeries, 9 persons could use any to all of the supplied specimen signatures as model(s). In average, there are 36 forgeries for each class. Japanese dataset images were converted from online signatures on a tablet PC. It contains 30 classes that each one has 42 genuine samples generated in 4 days and 36 forgeries by 4 forgers.

In current datasets, there is only one small Persian dataset with no description about data collection process. Therefore, a new and rich dataset is required in Persian offline signature verification. Among non-Persian datasets, MCYT-75 and GPDS-960 are more common in the field, but MCYT-75 has relatively small number of genuine and forged samples per class, and GPDS-960 with the largest number of classes, suffers from small period of genuine samples collection, also at the moment it is not available to the research community.

3. UTSig Dataset

UTSig (University of Tehran Persian signature) dataset consists of 8280 images from 115 classes. Each class belongs to one specific authentic person and has 27 genuine and 45 forged samples of his signatures. Fig.1 shows samples from 4 classes. Participants in signature collection procedure were randomly selected from Persian undergraduate and graduate students of University of Tehran and Sharif University of Technology. Their age was between 18 and 31 (average 24.14) and 90% of them were right-hand writers. Participants in genuine procedure were males but those played rule as forgers were 40% females and 60% males. Individuals, whether authentic and forgers, by signing on UTSig genuine and forgery forms, showed their complete consents for using and publishing their signatures while their identities are kept hidden.

Participants were allowed to write with arbitrary pens on A4-sized white forms in predetermined boxes (Fig.2). Scanning procedure was done with 600 dpi resolution and forms were stored as 8 bit grayscale TIF files. Two preprocessing were done on images: manually removing considerably large artifacts (e.g. artifacts from bad printing) and a simple threshold noise removal that converted pixels with intensity higher than a threshold to pure white (i.e. pixels brighter than 237 were assigned to 255). The threshold was estimated by finding the darkest pixel of 5 scanned blank papers.
Fig. 1. Four genuine samples from UTSig, their opposite-hand and forgeries (similarity scores were determined by forgers)

Fig. 2. Data collection forms
a Genuine and opposite-hand form
b Forgery form
3.1. Genuine Signatures

To obtain genuine signatures, 115 participants were asked to sign 10 times on a form and repeat it for 3 days. In each day, first 9 signatures were genuine and the last one was an opposite-hand signed signature of its writer that can be used as forgery or disguise. Therefore, totally 3105 genuine and 345 opposite-hand signatures were collected. We consider opposite-hand signed samples as forgery, since their traits are not as genuine as real genuine samples.

The forms (Fig.2) used for this procedure, contained 10 boxes (6 different sizes): 9 for genuine and 1 for opposite-hand signatures. The reason for using different sizes is to provide natural conditions or constraints that can occur in public service application forms which cause authentic changes and consequently affect the accuracy of verification systems [24]. To keep the nature of signature unchanged; writers were free to sign either vertically or horizontally. Furthermore, the boxes’ sizes were large enough for Persian signatures but if a signature crossed the boxes boundaries another form was given to the author to repeat.

3.2. Forged Signatures

UTSig has 5175 forgeries divided into 3 categories. The first category is 3 opposite-hand signed signatures of the authentic authors. These are 345 samples. Opposite-hand signed signatures can be considered as disguise samples, because their authentic authors who know the genuine traits, signed with a hand that made signatures unnatural. However, they can be used as forgery since they lack all the genuine traits of the genuine signatures. In real conditions (e.g. banks), collecting skilled forged samples produced by forgers is not practical, and consequently training verification systems using such skilled forgeries is not possible. Hence in order to enhance the performance, opposite-hand samples can be useful if they are used to train systems. Note that we consider these samples as skilled forgery and not as random forgery since the overall shape of these samples are relatively similar to genuine samples and they are made by persons who know the genuine traits.

The second category is skilled forged samples obtained from 230 people apart from authentic ones. Each person forged 3 different signatures (from different classes) 6 times in forms with 6 same size boxes (Fig.2). Their observable genuine samples varied from 1 to 3 signatures of same class. In other words, each class or genuine class was forged by 6 persons in a way that 2 skilled forgers saw 1 randomly selected sample of genuine signatures; 2 persons saw 2 and 2 persons saw 3 randomly selected samples.
For each forged sample, the forger was asked to determine its similarity to genuine sample(s) (Fig.1) from very low to very high (1, 2, 3, 4 or 5). The average of scores is 2.72 (2.67, 2.71 and 2.79 when respectively 1, 2 or 3 genuine sample were observable). Totally this category has 4140 skilled forged samples and 91% of them (74% of total forgeries) were ranked by their forgers.

The third category is also a skilled forgery but its samples were forged by a more skilful person than two pervious categories. The form used was such as the second category and the observable sample for this category was only one random genuine sample. Altogether, 690 samples were built in this category.

In all categories writers had enough time to practice and they were ask to be cautious about boxes’ boundaries but we manually refined samples crossed the boundaries (less than 1% of forgeries). Fig.3 shows a sample before and after manual refinement. The refinement procedure similar to [11] was done by manually removing black lines of boxes in a way that the signature curve, in the point that it crossed the box in comparison with previous and subsequent points, remained visually natural. Although this procedure is not completely ideal, as shown in Fig.3 recovered sample provides satisfactory visual results.

![Fig. 3. An example for manual refinement](image)

a Original sample which crossed the boundaries  
b Refined sample

### 4. Proposed Experimental Setups

Writer-Independent (WI) and Writer-dependent (WD) are two major approaches in verification systems. In WI, all genuine samples and all forgeries, regardless of their authors, are compared pair-to-pair in training phase to estimate the distribution of within-writer and between-writer similarities. Next in testing phase to determine the authenticity, questioned signature is checked in the both distributions. In contrast, WD approach uses author-based samples, in other words, training phase is done separately for each authentic person by samples of his class [25].

#### 4.1. Training and Testing Setups

In this paper we focus merely on WD approach and propose 4 different training and their corresponding testing setups for UTSig dataset. Note that to design realistic verification system, using
skilled samples in training set is not recommended, because in real conditions (e.g. banks), it is not easily practical to collect skilled forgery for new users. Therefore, proposed setups merely consist of genuine signatures, random forgeries, and opposite-hand samples.

Genuine vs. random forgery (setup 1): training verification system for each author using 12 randomly selected of his genuine samples and 5 random forgeries from each other classes \((570 = 5 \times (115 − 1))\). This is a common setup in the literature and used in many papers such as [26] and [27].

Genuine vs. random forgery and opposite-hand (setup 2): training verification system for each author using 12 randomly selected of his genuine samples, 5 random forgeries from each other classes, and all his opposite-hand signed samples. This new setup can be applied merely for datasets with opposite-hand signed or disguise samples. As mentioned, it is not recommended to use skilled forgeries in training phase, meanwhile, random forgeries are significantly different from the skilled forgeries that the system encounters in testing phase. Therefore, it is assumed that opposite-hand signed samples can improve the performance of the system on detection skilled forgeries.

Genuine vs. opposite-hand (setup 3): training verification system for each author using 12 randomly selected of his genuine samples and all his opposite-hand signed samples. This new setup may improve the performance, but it should be used with a classifier that is suitable for small sample size problems, because there are only 12 positive and 3 negative training sample.

Genuine alone (setup 4): training verification system for each author using 12 randomly selected of his genuine samples. This is a common setup in the literature and used by many researchers such as [20] and [28].

It is proposed to test the system with 15 remaining genuine samples, along with remaining skilled forgeries and random forgeries. Note that, we consider all the remaining samples of other classes as random forgeries, for instance, for setup 1 it consists of \((7638 = (72 − 5) \times (115 − 1))\) samples. Table 1 shows suggested training and testing setups for UTSig dataset. This table shows proposed setups for each class, therefore to train and test all classes it must be repeated for all 115 classes.

<table>
<thead>
<tr>
<th>Setup 1</th>
<th>12 Genuine + 570 Random</th>
<th>15 Genuine + 45 Skilled + 7638 Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 2</td>
<td>12 Genuine + 570 Random + 3 Opposite-hand</td>
<td>15 Genuine + 42 Skilled + 7638 Random</td>
</tr>
<tr>
<td>Setup 3</td>
<td>12 Genuine + 3 Opposite-hand</td>
<td>15 Genuine + 42 Skilled + 8208 Random</td>
</tr>
<tr>
<td>Setup 4</td>
<td>12 Genuine</td>
<td>15 Genuine + 45 Skilled + 8208 Random</td>
</tr>
</tbody>
</table>

Table 1 Proposed training and testing setups for each class of UTSig.
4.2. Evaluation

In this work, we follow the paradigm shift introduced in SigComp2011 [3] to use both decision and likelihood based criteria in signature verification systems. For the former, we deal with false acceptance rate (FAR) (i.e. the percent of forged samples that are incorrectly accepted) separately for random and skilled forgeries, false rejection rate (FRR) (i.e. the percent of genuine samples that are incorrectly rejected), and equal error rate (EER) (i.e. the rate at which skilled forgery FAR and FRR are equal). For the latter, we refer to two information-theoretic measures, cost of log-likelihood-ratio (\(\hat{c}_{llr}\)) and its minimal possible value (\(\hat{c}^{\text{min}}_{llr}\)). \(\hat{c}_{llr}\) is a positive unbounded measure, that is calculated by normalizing the weighted summation of \(\log = (1 + \exp(y))\) and \(\log = (1 + \exp(-y))\), in which \(y\) is the classifier output score. \(\hat{c}^{\text{min}}_{llr}\), the criteria designed for final evaluation, is the minimum or optimized value of \(\hat{c}_{llr}\) which is bounded between 0 and 1. The value of these two criteria are affected by the probability scores produced by the system for questioned samples. In other words, the better value obtains if the system assigns higher scores for its true and lower scores for its false rejections and false acceptances. Further details and precise mathematical definitions are provided in the original paper [29]. However, to compare the performance of different verification system, we use both EER and genuine versus skilled forgery \(\hat{c}^{\text{min}}_{llr}\), because the former has been a standard criterion in the literature and the latter was proposed in recent signature competitions (4NSigComp2012 [23] and SigWiComp2013 [5]) for evaluating signature verification systems. Therefore, systems that have less value of EER or \(\hat{c}^{\text{min}}_{llr}\) have better performance in terms of EER or \(\hat{c}^{\text{min}}_{llr}\), respectively. In SigWiComp2013 it was shown that good EER does not always result good \(\hat{c}^{\text{min}}_{llr}\) [5]. Results should be the average of 10 independent and new experiments. In each experiment, each criterion is calculated once and not separately for each class.

5. Comparison of datasets

Some datasets with few changes in number of samples are currently publicly available. Table 2 compares UTSig and other public datasets in terms of statistics and variables considered in data collection processes. UTSig dataset in comparison with other datasets, including the only existing Persian offline signature dataset, FUM, is better in terms of number of classes, total number of samples, and number of forgers. Although some datasets surpass UTSig in terms of average number of genuine and forged signatures per class, their very small number of authentic authors reduces usefulness of those dataset. Meanwhile, UTSig surpasses in case of number of forgers, number of different size boxes, number of different observable samples, and meta-data (self-score).
Table 2 Comparison between public offline signature datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Authentic Authors Ave. Genuine/Author</th>
<th>Ave. Skilled Forger/Author</th>
<th>Ave. Disguised Opposite-hand/Author</th>
<th>Total Samples</th>
<th>Number of Forgers</th>
<th>Period of Collecting Genuine Samples (days)</th>
<th>Number of Different Size Boxes</th>
<th>Number of Different Observable Samples for Forgers</th>
<th>Opposite-hand or Disguise Samples</th>
<th>Arbitrary Pen</th>
<th>Meta-data (e.g. Self Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCYT-75 [18] [19]</td>
<td>75</td>
<td>15</td>
<td>15</td>
<td>0</td>
<td>2250</td>
<td>75</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>ICDAR2009 [2]</td>
<td>91</td>
<td>11</td>
<td>27</td>
<td>0</td>
<td>3462</td>
<td>64</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>FUM [22]</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>0</td>
<td>600</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>4NSigComp2010 [7]</td>
<td>2</td>
<td>57</td>
<td>97</td>
<td>14</td>
<td>335</td>
<td>61</td>
<td>5-7</td>
<td>NA</td>
<td>1</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SigComp2011 Dutch [3]</td>
<td>64</td>
<td>24</td>
<td>12</td>
<td>0</td>
<td>2295</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>SigComp2011 Chinese [3]</td>
<td>20</td>
<td>24</td>
<td>34</td>
<td>0</td>
<td>1176</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>4NSigComp2012 [23]</td>
<td>3</td>
<td>55</td>
<td>91</td>
<td>21</td>
<td>501</td>
<td>39</td>
<td>10-15</td>
<td>1</td>
<td>2</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SigWiComp2013 Dutch [5]</td>
<td>27</td>
<td>10</td>
<td>36</td>
<td>0</td>
<td>1241</td>
<td>9</td>
<td>5</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SigWiComp2013 Japanese [5]</td>
<td>20</td>
<td>42</td>
<td>36</td>
<td>0</td>
<td>1566</td>
<td>4</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>UTSig</td>
<td>115</td>
<td>27</td>
<td>42</td>
<td>3</td>
<td>8280</td>
<td>230</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

6. Experiments

6.1. Persian Signatures Characteristic
By observations we see that signatures in distinct cultures have different shapes. Therefore, it seems clear to use different feature extraction and classification methods for the signatures of distinct cultures. For instance in [30] and [31], authors use two-stage approach to improve verification accuracy for multiscrypts signatures. First, they use an identification system to see whether questioned signature is Hindi or English, and then use different verification systems for each type of signature.
To statistically show signatures differences in available datasets, we use morphological operations to count branch points and end points in each sample. A branch point is a point that the signature cross itself and an end point is the beginning and final point of each connected component. First, genuine samples are binarized with a simple threshold, which is approximately the darkest pixel of blank areas of images. Then by connected-component analysis, components with less than 10 pixels, are removed from binary images. Next, to refine samples, after applying horizontal dilation operator, we set a pixel to black if five or more pixels in its 3-by-3 neighbours are black (i.e. majority operation). Standard morphological skeletonization [32] is applied and finally end points and branch point of skeletonized image are extracted by finding pixels which merely have one neighbouring pixel (i.e. morphological end-point operator) and finding pixels which have more than two neighbours (i.e. morphological branch-point operator), respectively. To remove unreal points caused by skeletonization operation, if the Euclidean distance between two or more adjacent branch points or end points is less than a threshold (i.e. 10), only one of them is preserved. Fig. 4 shows two branch points and two end points for a signature.

Fig. 4. A signature with two branch points (circles) and two end points (squares)

Analysis on the number of branch points and end points shows that Persian signature in UTSig and FUM [22] datasets have fewer number of branch points and end points in comparison with other datasets, including: Dutch, Spanish, Chinese and Japanese (see Fig. 5).
6.2. Verification System

6.2.1 Classifier: Support Vector Machine (SVM) is a statistical approach to design a supervised classifier. It was originally developed for two-group classification problems. To separate two classes, SVM uses kernels mapping input vectors to a high-dimension feature space, and then constructs a linear decision surface in the high-dimension space [33]. We use SVM to separate genuine and non-genuine samples of an authentic individual. For the first, second and third setup, linear kernel is used, which multiplies feature dimensions to build high-dimension space \( K(x_1, x_2) = x_1^T x_2 \). For setup 4, which is a one-class problem, we use one-class SVM with radial basis functions (RBF) kernel which is defined as \( K(x_1, x_2) = \exp\left(-\frac{||x_1-x_2||^2}{2\sigma^2}\right) \). SVM produces labels that determine the class of the input vector but in order to calculate EER and construct likelihood ratio, we map SVM outputs into probabilities or scores by the method proposed in [34].

6.2.2 Feature Extraction: We use fixed-point arithmetic which was described in [20] as: “description of the signature envelope and the interior stroke distribution in polar and Cartesian coordinates”. In fixed-point arithmetic feature extraction, using sample’s geometric centre, three parameters are calculated in polar coordinate: derivative of radius of signature envelope; its angle and the number of black pixels that the
radiuses cross when rotate from one point to next point. Moreover, in Cartesian coordinates, height, width and the number of transitions form black to white or white to black pixels of signatures are calculated with respect to their geometric centre [20].

6.2.3 Results: To find numerical results of UTSig dataset, we repeated experiments for each author 10 times and average their results. Results are available in Table 3. In terms of both EER and $\hat{c}^{\text{min}}_{ltr}$, setup 2 has the best performance, after that in descending order setup 1, 4 and 3 have better results. To statistically find the best result, we use t-test that indicates with a 90% confidence interval, setup 2 surpasses setup 1 and strongly the other setups in terms of EER and $\hat{c}^{\text{min}}_{ltr}$. Therefore, the fact that using opposite-hand signed signature along with random forgeries can improve the performance is confirmed with 90% confidence interval. Moreover, this fact shows that collecting opposite-hand signed samples in future datasets and real conditions can be promising.

Note that in the fourth setup due to the nature of one-class SVM, a threshold must be selected. In this paper, one similar ideal threshold is selected for all authors but better results may be obtained using user-based threshold.

Table 3 Sets results for UTSig including 90% confidence interval.

<table>
<thead>
<tr>
<th>Setup</th>
<th>EER</th>
<th>Genuine FRR</th>
<th>Skilled FAR</th>
<th>Random FAR</th>
<th>$\hat{c}_{ltr}$</th>
<th>$\hat{c}_{ltr}^{\text{min}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29.71% ±0.29</td>
<td>39.27%</td>
<td>32.29%</td>
<td>0.08%</td>
<td>0.996</td>
<td>0.819±0.004</td>
</tr>
<tr>
<td>2</td>
<td>29.33%±0.22</td>
<td>41.70%</td>
<td>18.34%</td>
<td>0.07%</td>
<td>0.995</td>
<td>0.813±0.002</td>
</tr>
<tr>
<td>3</td>
<td>34.14%±0.30</td>
<td>0.02%</td>
<td>93.23%</td>
<td>89.53 %</td>
<td>1.43</td>
<td>0.896±0.002</td>
</tr>
<tr>
<td>4</td>
<td>32.46%±0.34</td>
<td>32.50%</td>
<td>32.43%</td>
<td>4.33%</td>
<td>2.65</td>
<td>0.868±0.003</td>
</tr>
</tbody>
</table>

According to the participant self-scores in forgery process, the behaviour of the trained systems has a trend toward distinguishing the low score samples more than the high score samples. The average score of true negative samples is 2.65 while for false positive it is 2.86.

To check results with respect to the number of samples or random forgeries in training phase, we repeated the first setup. Table 6 shows that if the number of genuine samples in training set increases, EER and FRR decrease and skilled forgery FAR increases. Meanwhile, adding more random forgeries in training set increases EER and FRR, but decreases skilled forgery FAR.
To check the performance of the similar system on the other datasets, it seems reasonable to perform only the setup 4, since other setups need a variety of random forgeries (from distinct authors) or opposite-hand samples which are not available in all datasets. Note that some datasets due to unavailability, missed or low genuine samples were not used. Results of the same verification system on other datasets are provided in Table 4. According to EER and $\hat{c}_{\text{ltr}}^{\text{min}}$, it is clear that the performance of the system significantly decreases when it is tested on SigWiComp2013, UTSig, SigComp2011, and 4NSigComp2010.

Table 4 Datasets results for setup 4 including 90% confidence interval.

<table>
<thead>
<tr>
<th>Setup</th>
<th>EER</th>
<th>Genuine FRR</th>
<th>Skilled FAR</th>
<th>Random FAR</th>
<th>$\hat{c}_{\text{ltr}}$</th>
<th>$\hat{c}_{\text{ltr}}^{\text{min}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCYT-75</td>
<td>25.9±0.28%</td>
<td>27.2%</td>
<td>24.5%</td>
<td>4.9%</td>
<td>0.815</td>
<td>0.722±0.002</td>
</tr>
<tr>
<td>FUM</td>
<td>26.2±0.27%</td>
<td>26.2%</td>
<td>26.1%</td>
<td>0.22%</td>
<td>0.810</td>
<td>0.741±0.002</td>
</tr>
<tr>
<td>4NSigComp2010</td>
<td>29.1±0.30%</td>
<td>28.2%</td>
<td>30.1%</td>
<td>0.10%</td>
<td>0.771</td>
<td>0.782±0.004</td>
</tr>
<tr>
<td>SigComp2011 Dutch</td>
<td>32.0±0.31%</td>
<td>31.6%</td>
<td>31.1%</td>
<td>5.9%</td>
<td>1.008</td>
<td>0.851±0.003</td>
</tr>
<tr>
<td>4NSigComp2012</td>
<td>22.3±0.24%</td>
<td>21.4%</td>
<td>23.5%</td>
<td>3.8%</td>
<td>0.825</td>
<td>0.604±0.003</td>
</tr>
<tr>
<td>SigWiComp2013 Japanese</td>
<td>33.1±0.30%</td>
<td>34.9%</td>
<td>33.2%</td>
<td>7.5%</td>
<td>1.372</td>
<td>0.879±0.004</td>
</tr>
<tr>
<td>UTSig</td>
<td>32.46%±0.34</td>
<td>32.50%</td>
<td>32.43%</td>
<td>4.33%</td>
<td>2.65</td>
<td>0.868±0.003</td>
</tr>
</tbody>
</table>

7. Conclusion

In this paper a new and rich Persian offline signature dataset, was introduced. It has 115 classes, each one consists of 27 genuine, 3 opposite-hand and 42 skilled forged samples. UTSig surpasses the other Persian signature dataset in case of size and variables considered in sample collection procedure.
Furthermore, in comparison with the other public datasets collected in other communities, UTSig is richer, because it has higher number of samples, classes, forgers, different size boxes, different observable samples, and available meta-data (self-score). Meanwhile, UTSig has opposite-hand signed samples, and in its sample collection using arbitrary pens was allowed and genuine samples were signed in 3 days period. These traits make UTSig a useful dataset for Persian signature systems and for evaluating other culture-independent systems such as classifiers.

We proposed 4 different standard WD training and testing setups for UTSig dataset and statistically showed that involving opposite-hand signed signatures in the training set, can enhance the performance of the system. We evaluated 4 setups with SVM using fixed-point arithmetic and the best system obtained EER=29.33% and $\hat{c}_{\text{lr}}^{\text{min}}=0.813$.

Counting branch points and end points of signatures in UTSig and other datasets, it was showed that Persian signatures in comparison with other datasets have fewer number of branch points and end points.

8. Acknowledgments
We appreciate participants in sample collection procedure for their cooperation and permission to use and publish their signatures.

9. References


28 Guerbai, Y., Chibani, Y., Hadjadji, B.: ‘The effective use of the one-class SVM classifier for handwritten


