Analysis of Synthetic Signature Generated From Single Offline Handwritten Signature Specimen Using Different Operations

Sameera Khan¹, Dr. Megha Mishra², Dr. Vishnu Kumar Mishra³

PhD Scholar, Department of Computer Engineering, SSTC, Bhilai, Chhattisgarh, India¹
Associate Professor, Department of Computer Engineering, SSTC, Bhilai, Chhattisgarh, India²
Associate Professor, Department of Computer Engineering, SSTC, Bhilai, Chhattisgarh, India³

ABSTRACT: In biometric technology, data collection is a huge problem. As a consequence of security concerns reported by individuals or unobtainable data, the efficiency of biometric systems is considerably decreased. Synthetic Biometric databases can be used as a solution to this problem. In this paper, a novel framework with diverse algorithms to generate synthetic offline signatures is proposed. The proposed method is designed while considering the intrapersonal variability of the handwritten signatures. Algorithms to incorporate geometric and bio-inspired variability are integrated into the framework, to generatesynthetic signature samples generating voluminous signature databases. The offline signature sample is inspected based on connected components, pixel density, number of components, and other standards. In this paper, four different operations are applied individually and as a combination of all. The difference in general features of the signature is then analyzed concerning the original specimen. The operations applied are morphing, random rotation (component-wise), random displacement (Component wise), and sine transform.

KEYWORDS: Biometric technology, Synthetic biometric, offline signature verification

I. INTRODUCTION

Due to escalating need for the application of automatic systems, the requirement of authentication and verification systems is also shooting up simultaneously. Biometric technology finds its application in such areas. Biometric technology works in the field of authentication and verification of a person based on particular behavioral [1,32] and physical characteristics. This comprises any inimitable behavioral or physical trait of a person like iris [2], fingerprint [3], signature [4], gait [5], voice [6], and face.

This type of biometric systems continuously require a dataset for training the classifier. The interminable necessity of new data for training and evaluation is the major problem in the field of biometric technology. The primary reason which significantly affects the performance of the biometric system is the biometric datasets. In contemporary times, various attempts have been made to ease the process of large biometric database acquisition. The procedure of acquiring large datasets in biometric technology for a huge population with proper variability concerning multi-session, different modes of signals, and varied physical conditions make it high priced concerning workforce, time, and capital. Furthermore, it is a multifaceted procedure as well. Moreover, citing security concerns, many individuals are unwilling to offer their biometric data easily. As a result of the inadequacy of the databases, the performance of biometric systems drops significantly.

Numerous open-source datasets are accessible online to facilitate evaluation procedures unbiased and comparable. Various competitive evaluations are motivated too in this domain. Some of the evaluations are-NISTFacialRecognitionTechnology Evaluations [7], theNISTS SpeakerRecognition Evaluations (SRE) [8], the Fingerprint Verification Competitions (FVC) [9], the Bio Secure Multimodal Evaluation Campaign held in 2007[10], or different signature verification competitions [11,12].

To find a solution to this problem, the biometric scientific group has put a lot of hard work into creating synthetic databases. Synthetic data can be identified as the reorganized data with non-natural characteristics that are not experimentally calculated rather they are drawn randomly from a population with certain features. Synthetic databases thus created for training purposes must-have properties like flexibility, parsimony, and consistency [13]. No restriction on the size of the dataset is imposed in synthetically generated databases.
Generation of the synthetic database is a composite problem which deals with the special case of pattern recognition and image restructuring. Implementing variability (both inter-class and intra-class) amongst the samples of same and different users while keeping the personal traits of a person intact, is a difficult task. Synthetic data generation has numerous benefits like the comfort of data collection, large volume, no legal and security problems, less time consuming, and many more. Even then, to attain efficiency nearly equal to that of the actual dataset with a synthetic dataset using numerous numeric features is a challenge.

In this paper, a novel framework for creating a synthetic handwritten signature dataset is proposed. A signature is considered to be a pattern that constitutes a collection of alphabets or shapes, interconnected with each other and occasionally illegible. In addition to these patterns, sometimes the signer makes some add-on patterns known as flourishes too. Combined creating a unique pattern for an individual, making it apt to be used as a behavioral biometric. In this framework, signatures done in English or western signatures are considered under study. Usually, such signatures are comprised of English alphabets written especially with some flourishes.

Generation of the database with synthetic signatures comprises various phases. After image acquisition, some preprocessing is required, in which the image is resized to a standard size, color images are converted into gray images, and noise removal. After preprocessing, firstly, an imitation of the seed signature is created. After that, components are extracted and labeled distinctly in different colors, it detects flourishes and slant of the seed signature too. In the next phase, intrapersonal variability is induced in form of displacement and distortion to the components of duplicated seed signature. Variability is induced using different types of operations on the components of the signature like-morphing, random rotation(component-wise), random displacement (Component wise), elastic distortion, and sine transform. To analyze the degree of distortion induced by each operator separately and unitedly some general signature features are compared.

II. RELATED WORK

Synthetic biometric has continuously been a problem for forensic experts as they can cause a potential threat to biometric security systems. Alternatively, thanks to recent researches in biometric technology in which the researchers have pleaded the use of synthetic data generation for biometric systems. Such databases can save lots of time and resources that were wasted in collecting data for creating a large training set. Patterns that can be used under biometric technology can be any physical or behavioral characteristic that can be used to recognize an individual exclusively.

According to the literature, the generation of biometric can take one of the following forms-

A. Duplicated Samples- This algorithm uses multiple samples from a user. Various affine and elastic transformations are applied to the original samples. Filters can also be used. All these functions are successively applied to generate a duplicated synthetic sample. Such algorithms are not very fruitful for the creation of new synthetic datasets rather they can be applied to make the dataset voluminous. This approach can successfully increase the number of samples per user by keeping the number of users constant[14,15]. A similar approach has been applied to the handwritten signatures in [16,17] and for face synthesis in [18,19]. Such algorithms can prove their utility in applications where efficiency is proportional to the size of the dataset.

B. The merger of unlike real samples- This technique uses numerous actual samples of a biometric and subsequently applying selected transformations to it produces a new synthetic biometric. Usually, it is advantageous in the text to speech converters[20], captchas like automatic Turing test[21], typewriting, and handwriting analysis [22]. The necessity of real samples for producing synthetic biometric is the shortcoming of this technique, for this reason, their effectiveness in biometric systems is also restricted. It does not increase efficiency in any way. It helps produce numerous samples for one user but not for the completely synthetic database.

C. Synthetic database- This technique is apriori in nature. Due to which prior information of the dataset is required. Certain apt representation of the biometric trait is recognized from the actual dataset, which is further utilized to generate a model that emulates the real dataset generation. Fresh synthetic individuals are created based on this information and the model generated. This technique can be used without even raising the slightest of legal issues to generate voluminous training sets. It also proved to be cheaper in terms of resources and manpower and typically accomplishes higher rates of efficiency as compared to the earlier two techniques. It enables the voluminous dataset and provides the solution for the undying problem of the inadequate dataset. Various mathematical models were used in creating synthetic datasets for several biometric databases including iris, gait, fingerprint, speech, signature, and many others.
Handwritten signatures are widely acknowledged as a way of validation and confirmation around the world. When a handwritten signature is done, a neuro motor mechanism is triggered. Recurrent movements are learned and complex fine motor control creates a signature. Motor equivalence theory explains the same. As per the motor equivalence theory [25,28-31], the brain saves the muscle movement for the recurring task. Handwritten signatures also comprise similar stored movements. This is executed into two steps- (i) Effector independent- it establishes a relationship between the spatial coordinates of every trajectory and its relative positions to eachother. A stroke is defined as the stroke between subsequent pen-ups. (ii) Effector Dependent- Series of motor instructions generated by learned actions to get a specific muscle movement.

Even though, signatures are expected to be the same whenever it is done however they are not the same. They resemble eachother with varying degrees of intrapersonal variability. Such variability can be due to ambiance, psychological obstructions, oldness, disease, etc. The learned muscular pathway involved in the signing process is vulnerable to getting old, climate-related situations, certain neurodegenerative diseases[26] like Parkinson’s, etc, or any other associated health concerns.

Thus, the procedure established is motivated by motor equivalence theory. Effectors independent method takes a signature as the series of points in a cognitive grid plot. The inter-stroke variability of the handwritten signature can be achieved by creating spatial transposition of flourishes or signature components alteration in skew or slant angles. However, interstroke variability can be achieved by inducing spatial deformation which extends over a few control points of the sample. Effector-dependent effects are entirely dependent on the cognitive grid plot. Two of them are separate but the complete intrapersonal variability attained in the handwritten signature is inclusive of both the effects. In this paper, four different operations are applied to a signature and its effect is analyzed based on some general features of handwritten signatures.

**III. COGNITIVE THEORY FOR SYNTHETIC DATA GENERATION**

III. COGNITIVE THEORY FOR SYNTHETIC DATA GENERATION

The proposed system is based on the acquisition of a single sample of offline handwritten signatures from an individual. The sample of handwritten signature thus acquired is used for populating the database for a specific user. Interstroke and intrastroke changeability is attained for a specific specimen by examining strokes, flourishes, skew angle, slant angles, density, etc. The basic methodology is explained in Fig 1.

Samples of actual handwritten signatures are taken from available databases of offline signatures like GPDS corpus 960, CEDAR, MCYT, and SVC 2000[27]. The samples shown in this paper are used from the dataset available in Kaggle. The proposed system is implemented in stages. a) Acquisition stage, b) component identification stage, c) stroke segmentation stage, d) control point selection phase, e) application of selected operation to induce variability and f) generation of new sample phase. In the acquisition stage, one of the real samples is acquired from the database in the system. Preprocessing is done to make images uniform in size and color and noise removal is also performed. Flourishes and key components are recognized individually. The distance and shape of flourishes concerning the key components are analyzed. The nature of flourishes, number of components, the relativity of components with each other, slant angle, skew angle of each component concerning the baseline, etc. are calculated in the second phase. A stroke is defined by the pixels amid one pen-down and pen-up. Every component of a signature is recognized individually and categorized with diverse colors. Further, control points are selected based on the variation in direction vector of the signature stroke. After identification of segments, flourishes, and control points, the image is subjected to the selected operation to get a morphed image with induced variability. The operations used in this method are morphing, random rotation (component-wise), random displacement (Component wise), and elastic transform, to achieve intrapersonal variability in the duplicated samples.

All the images, thus acquired from the database are first preprocessed to make them uniform. In preprocessing, colored samples are converted into black and white logical images and samples are resized to have a general pixel area under consideration. 8- connected approach is used for segmentation and identification of several components. In the stroke segmentation phase, one segment is considered as a set of all the plotted pixels between two pen-ups. Next, control points are identified which are accountable for the generation of the signature curve. Also, some more perceptual points are selected to help the curve fitting. All the curves are concatenated to form the final signature. This signature will be identical to the original signature. Since the handwritten signature of a person is not always the same.
Doing a signature is a complex process that involves neuro-motor movement. As a result, some inconsistencies are always there in the two specimens of a single person. That is why, in this new synthetic sample, some distortions are induced to achieve variability. Variability thus introduced will be in form of displacement of components of a signature and a small amount of face distortion of each component. Adding such distortion will result in both interpersonal and intrapersonal variability.

![Flowchart of the synthetic signature generator](image)

### V. EXPERIMENTAL RESULTS

The proposed method implements a mathematical model that performs stroke segmentation, flourish identification, and generation of duplicated signatures. For the implementation of the proposed system, publically available databases are used. In Fig 2(a& c) actual offline handwritten signature is shown. Whereas in fig 2 (b& d) segmented images with separation of components by using different colors have been shown. To compute the intrapersonal variability signature is divided into distinct components extracted using the 8 connected technique. The actual signature and the result after application of the algorithm are shown in fig 3.

![Fig.2 a. & c. original signature b. & d. segmented signature with different colors for different segments.](image)
Further to introduce variability in the signature specimen, various operations are applied to the original signature specimen. The first operation applied is random displacement in components of a signature. Generalized extreme value random numbers have been used to induce random displacement in the components. To show the displacement the image result has been displayed with the grids. The placement of every component of a signature usually varies every time with a random displacement operation. Each time the operation is applied a random displacement is generated among the components to create a new synthetic signature. While signing, a person usually makes slightly displaced moves concerning components and flourishes. This effect can be mimicked by using a random displacement operation. One component is considered to be the pattern made by the writer between two pen-ups. The following figure shows different results after applying random displacement operations. Displacement can be calculated as:

\[ x = x + d \]

where \( d \) is the random translation constant calculated in a specific range using GEV random numbers.

Additionally, Edge points have been identified to produce variability in the components themselves. Edge points thus identified have been shown in Fig.5.
Fig. 1. The second operation applied is the random rotation. This operation is also applied component-wise. While the process of doing a signature takes place, the relative orientation of the components does not remain the same in every signature done by the same signer. This property is achieved by applying random rotation component-wise. The result of the application of random rotation operation is shown in Fig. 6. Rotation is implemented by using the following rotation matrix:

\[
R = \begin{bmatrix}
\cos(\theta) & \sin(\theta) & 0 \\
-\sin(\theta) & \cos(\theta) & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

where \( \theta \) is angle of rotation in radians. Also the value of angle of rotation is randomly generated to induce change in orientation of the components.

The third operation applied is the morphing operation. Usually, when a signer starts signing, the first component suffers the most degree of change. When compared with a signature specimen of a single user taken at different times, it was observed that the first component suffers maximum distortion. So to achieve this distortion function is applied on the first component only. All other components are left untouched. The result of this operation on the original signature specimen is shown in Fig. 7.

When all the three previously discussed operations are performed simultaneously in an image, the resultant gives the distorted synthetic signature. So the fourth operation applied to the image specimen under consideration is an amalgamation of all the three: random displacement, random rotation, and morphing of prominent components, applied...
in the same order. As a result, following synthetic signature is generated which is shown in Fig-8 along with the original signature specimen.

Elastic distortion operation is also applied to the specimen image. This operation induces a small number of jaggies in the strokes to give a realistic stroke.

Analysis of the effect of these operations can be analyzed based on certain general features of the handwritten signature. Features like total image area in terms of number of pixels, pure height, and pure width in terms of a maximum number of pixels in a particular scan line horizontally and vertically, aspect ratio, vertical and horizontal projection, number of edge points calculated by using Sobel operator, co-ordinates of the centroid are used to identify the percentage of distortion in the synthetic image thus created. These features are extracted individually for all the three operations applied on a particular signature specimen and the fourth operation is the composite transformation formed by applying random displacement, random rotation, morphing respectively. The database used comprises five signature specimens for a single writer. So the features are extracted separately for all five specimens and the average of the values is taken as a benchmark to compare the values. One of these five signatures is taken as the seed to perform all the operations to test the performance of all the operations.

VI. CONCLUSION

A new approach for generating synthetic handwritten signatures has been proposed based on motor equivalence theory and a cognitively inspired model for signature generation. This paper concentrates on generating synthetic samples to populate the training dataset for higher accuracy. A single handwritten signature is acquired and it is segmented into parts by analyzing strokes. Intrapersonal variability is achieved by examining the segments and flourishes. Since variability and consistency are very important concerning signatures, variability is achieved by keeping consistency intact. To incorporate a permissible degree of displacement variability, generalized extreme value random numbers have been used. Additionally, edge points have been identified to incorporate variability concerning the shape and size of the component. In this paper, a signature is studied based on no of components, flourishes, skew, slant, and density.

Five operations have been applied to the seed signature with random values to generate different synthetic signatures. The operations used are random displacement, random rotation, the morphing and composite transformation of all, and elastic distortion. When a signer does a signature repeatedly, the components are not always in the same distance to each other. To implement this property random displacement has been applied. The variation from the original signature due to the application of this operation is approx. 19%. The second operation applied is random
rotation. This operation is used to realize the property of handwritten signature due to which various components are not in the same orientation with each other in each successive rotation. Another important property of a handwritten signature is usually, the first component is distorted more as compared to others while successive signing. So the third operation of morphing has been applied to the first component. The fourth operation is the combination of the previous three. To mimic human-like variability these four operations have been tested against the average features of five signatures of a particular signer available in the database. It is found to be that displacement induces the least variability whereas morphing induces the most. Each of the operations hampers a different feature, based on the property they exploit. The maximum distortion in %age induced by these four operations is about 25%. This percentage can be further tuned down by changing parameters of distortion in each of the components while keeping the consistency intact.

Such algorithms can be used to generate synthetic individuals too. Synthetically generated biometric data can be useful in performance estimation, security evaluation against synthetic individuals, populating a database for training security systems, forensic investigations, and the creation of large datasets for research purposes. Such synthetic databases have the advantage of less cost and it does not have any legal issues too.

Table-1 Feature matrix of five specimen images taken from the database

<table>
<thead>
<tr>
<th>Name of Feature</th>
<th>DB_image_1</th>
<th>DB_image_2</th>
<th>DB_image_3</th>
<th>DB_image_4</th>
<th>DB_image_5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count_area</td>
<td>1091</td>
<td>1027</td>
<td>1092</td>
<td>1124</td>
<td>1100</td>
<td>1086.8</td>
</tr>
<tr>
<td>Pure_width</td>
<td>22</td>
<td>19</td>
<td>21</td>
<td>24</td>
<td>17</td>
<td>20.6</td>
</tr>
<tr>
<td>Pure_height</td>
<td>46</td>
<td>29</td>
<td>40</td>
<td>29</td>
<td>30</td>
<td>34.8</td>
</tr>
<tr>
<td>Aspect_ratio</td>
<td>0.478261</td>
<td>0.655172</td>
<td>0.525</td>
<td>0.827586</td>
<td>0.566667</td>
<td>0.610537</td>
</tr>
<tr>
<td>Vertical Projection</td>
<td>81</td>
<td>64</td>
<td>67</td>
<td>71</td>
<td>63</td>
<td>69.2</td>
</tr>
<tr>
<td>Horizontal Projection</td>
<td>79</td>
<td>103</td>
<td>117</td>
<td>90</td>
<td>110</td>
<td>99.8</td>
</tr>
<tr>
<td>Number of edge points</td>
<td>2133</td>
<td>2021</td>
<td>2143</td>
<td>2203</td>
<td>2144</td>
<td>2128.8</td>
</tr>
<tr>
<td>Cetroid(x)</td>
<td>125.6062</td>
<td>125.6611</td>
<td>125.6301</td>
<td>125.7282</td>
<td>125.6275</td>
<td>125.6506</td>
</tr>
<tr>
<td>Centroid(y)</td>
<td>125.3669</td>
<td>125.4612</td>
<td>125.4547</td>
<td>125.4964</td>
<td>125.4961</td>
<td>125.4551</td>
</tr>
</tbody>
</table>

Table-2 Comparison of effects of each operation on the handwritten signature specimen.

<table>
<thead>
<tr>
<th>Name of Feature Extracted</th>
<th>Average</th>
<th>After displacement</th>
<th>After Rotation</th>
<th>After Morphing</th>
<th>After composite transformation</th>
<th>After Elastic Distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count_area</td>
<td>1086.8</td>
<td>891</td>
<td>1038</td>
<td>860</td>
<td>830</td>
<td>986</td>
</tr>
<tr>
<td>Pure_width</td>
<td>20.6</td>
<td>18</td>
<td>21</td>
<td>19</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>Pure_height</td>
<td>34.8</td>
<td>28</td>
<td>27</td>
<td>30</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>Aspect_ratio</td>
<td>0.61</td>
<td>0.64</td>
<td>0.78</td>
<td>0.63</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>Vertical Projection</td>
<td>69.2</td>
<td>67</td>
<td>83</td>
<td>49</td>
<td>62</td>
<td>61</td>
</tr>
<tr>
<td>Horizontal Projection</td>
<td>99.8</td>
<td>97</td>
<td>99</td>
<td>74</td>
<td>86</td>
<td>77</td>
</tr>
<tr>
<td>No. of edge points</td>
<td>2128.8</td>
<td>1779</td>
<td>2038</td>
<td>1735</td>
<td>1673</td>
<td>1960</td>
</tr>
<tr>
<td>Cetroid(x)</td>
<td>125.65</td>
<td>125.67</td>
<td>125.62</td>
<td>125.48</td>
<td>125.34</td>
<td>125.6376922</td>
</tr>
<tr>
<td>Centroid(y)</td>
<td>125.46</td>
<td>125.44</td>
<td>125.49</td>
<td>125.57</td>
<td>125.72</td>
<td>125.4972527</td>
</tr>
</tbody>
</table>
Table-3 Comparison of the degree of distortion by each operation on the handwritten signature specimen.

<table>
<thead>
<tr>
<th>Name of Feature Extracted</th>
<th>%age Change After displacement</th>
<th>%age Change After Rotation</th>
<th>%age Change After Morphing</th>
<th>%age Change After composite transformation</th>
<th>%age Change After Elastic transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count_area</td>
<td>18.02</td>
<td>4.49</td>
<td>20.87</td>
<td>23.63</td>
<td>9.27</td>
</tr>
<tr>
<td>Pure_width</td>
<td>12.62</td>
<td>1.94</td>
<td>7.77</td>
<td>1.94</td>
<td>1.94</td>
</tr>
<tr>
<td>Pure_height</td>
<td>19.54</td>
<td>22.41</td>
<td>13.79</td>
<td>13.79</td>
<td>22.41</td>
</tr>
<tr>
<td>Aspect_ratio</td>
<td>5.29</td>
<td>27.39</td>
<td>3.73</td>
<td>14.65</td>
<td>27.39</td>
</tr>
<tr>
<td>Vertical Projection</td>
<td>3.18</td>
<td>19.94</td>
<td>29.19</td>
<td>10.40</td>
<td>11.85</td>
</tr>
<tr>
<td>Horizontal Projection</td>
<td>2.81</td>
<td>0.80</td>
<td>25.85</td>
<td>13.83</td>
<td>22.85</td>
</tr>
<tr>
<td>No. of edge points</td>
<td>16.43</td>
<td>4.27</td>
<td>18.50</td>
<td>21.41</td>
<td>7.93</td>
</tr>
<tr>
<td>Cetroid(x)</td>
<td>0.01</td>
<td>0.03</td>
<td>0.14</td>
<td>0.25</td>
<td>0.01</td>
</tr>
<tr>
<td>Centroid(y)</td>
<td>0.01</td>
<td>0.03</td>
<td>0.09</td>
<td>0.21</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Fig.9 Graphical representation of a comparison of the degree of distortion by each operation on the handwritten signature specimen.

REFERENCES


INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH
IN SCIENCE | ENGINEERING | TECHNOLOGY

Impact Factor:
7.512

9940 572 462  6381 907 438  ijirset@gmail.com

www.ijirset.com