Offline signature verification using locally optimized distance-based classification

Yaseen Moolla*†, Serestina Viriri†, Fulufhelo Nelwamondo*, Jules-Raymond Tapamo†

*Council for Scientific and Industrial Research, South Africa
†University of KwaZulu-Natal, South Africa

ABSTRACT

Although handwritten signature verification has been extensively researched, it has not achieved an optimal classification accuracy rate. Therefore, efficient and accurate signature verification techniques are required since signatures are still widely used as a means of personal verification. This research work presents efficient distance-based classification techniques as an alternative to supervised learning classification techniques (SLTs). The Local Directional Pattern (LDP) feature extraction technique was used to analyze the effect of using several different distance-based classification techniques. The classification techniques tested are the Euclidean, Manhattan, Fractional, weighted Euclidean, weighted Manhattan, weighted fractional distances and individually optimized resampling of feature vector sizes. The best accuracy, of 90.8%, was achieved through applying a combination of the weighted fractional distances and locally optimized resampling classification techniques to the Local Directional Pattern features extracted. These results are compared with results from literature, where the same feature extraction technique was classified with SLTs. The distance-based classification was found to produce greater accuracy than the SLTs by 8.5%.

KEYWORDS: Biometrics, pattern recognition, distance-based classification

CATEGORIES: I.5.3, I.4.7

1 INTRODUCTION

Biometrics is the measure of human characteristics for authentication or identification of individuals [1]. Biometric modalities are regularly becoming an important aspect of automated electronic security systems. For such a system to be successful, it requires methods and techniques that produce high accuracy.

These systems are used either for recognition or verification. Recognition entails the identification of a biometric trait, or set of traits, as belonging to a specific individual from a given set of individuals. Conversely, verification entails authenticating a claim that a biometric trait, or set of traits, belongs to a specific individual [1].

One of the most common and widely accepted biometric modality is the handwritten signature. It has been used for verification since before the advent of electronic machines [2]. Due to its widespread use and acceptance, handwritten signatures are an ideal candidate for automated biometric verification systems. The two methods of capture for signatures are categorizes as offline and online. For offline capture, a static image of a completed signature is recorded. For online

Email: Yaseen Moolla `ymoolla@csir.co.za`, Serestina Viriri `viriris@ukzn.ac.za`, Fulufhelo Nelwamondo `fnelwamondo@csir.co.za`, Jules-Raymond Tapamo `tapamoj@ukzn.ac.za`

capture, the creation of a signature is recorded as a function of time [2].

Offline signature verification systems have much potential for worldwide usage and there are several cases in which only offline signatures can be used. These include automated authentication of bank cheques and legal documents [2]. Additionally, equipment for the capture of offline signatures is cheaper [3], which will allow for greater adoption of automated signature verification, especially for small-to-medium businesses and in developing economies.

A high accuracy is required for the acceptance of such a system, but offline signature verification systems are hindered by low accuracy rates, having not yet reached an acceptable level of accuracy in comparison to online signatures. These low accuracy rates are due to the inherent randomness that is characteristic of behavioural biometrics. Static signatures also have much less information in comparison to online signatures, since the time dimension is not present [3].

There is more current research that attempts various techniques to improve the accuracy of offline signature verification systems. Some of the research focuses on novel feature extraction techniques [4] [5] [6] [7] [8]. These are designed to extract the most relevant and constant features of a signature while also attempting to reduce the extraction of features that are not stable or constant in an individual's signature.

Other research concentrates on using different clas-

sification techniques and finding the optimal classification technique for a particular feature set. Classification can be categorized into supervised learning techniques (SLTs) and distance-based measures. Most offline classification emphasizes the use of different SLTs, such as support vector machines [9] [10] [11] [12], hidden Markov models [11] [13] [14] [15], artificial neural networks [16] and fuzzy logic [17]. There is very little research into using different distance-based classification techniques, except for the Euclidean distance, which is the most commonly known distance measure in geometric space. This is discussed further in Section 1.1.

There are many other distance-based measures that could be used for classification. This work improves the accuracy of offline signature verification systems by investigating the use of some of these techniques, specifically weighted and fractional distances in Lebesque space, also called L^P-space, and applies them to various different feature extraction techniques. The results obtained are compared with those from literature. In particular, the distance-based classification techniques are compared with SLTs with the use of the same feature extraction techniques.

1.1   Literature Survey

Many techniques exist for the classification of signatures and other biometrics. They can be broadly categorized into supervised learning techniques (SLTs) and distance-based classification techniques. SLTs include neural networks, hidden Markov models (HMM) [18], support vector machines (SVM) [19] and fuzzy logic. Linear techniques include Euclidean distance, Mahalanobis, Manhattan distance, weighted Euclidean distances [20] and fractional distances [21]. Some of the notable, recent works are discussed below.

Kovari et al. [16] uses artificial neural networks for classifying a feature vector comprised of both local and global features, and achieves an equal error rate (EER) of just over 20%. Coetzer et al. [13] applies HMM classification to discrete Radon transform, which is a global feature, and achieved an EER of 12.2% using skilled forgeries and 4.5% using random forgeries. Panton and Coetzer further improves the EER to 8.6% by using a fusion of HMM classifiers and adding local features to the feature vector [14]. Yilmaz et al. [9] performs SVM classification on a combination of gradient-based and binary pattern-based features. It is found that a user-dependent, also called user-specific, classifier worked better than a globally applied classifier for all users. They achieved a best average error rate (AER) of 15.41%. Vargas et al. [10] proposes a system with least squares SVM classification on features extracted from the Fourier transform of a signature image. An EER of 6.20% is obtained, which by literature comparison performs better than similar techniques using the same database. Batista et al. [15] uses the statistical fusion of HMM classifiers via a multi-hypothesis approach and user specific codebooks. An AER of 7.79% is acheived. The best recent results for SLTs are obtained by using multi-hypothesis

fusion approaches. Ferrer et al. [11] performs a comparison between the Euclidean distance, HMM and SVM, using geometric features based on Cartesian and polar coordinates. It is found that, for the feature extractions used, the best performance was for HMM, with a false rejection rate (FRR) of 14.1% and false acceptance rate (FAR) of 12.3%, using skilled forgeries. The SVM performed better than the Euclidean distance, but worse than the HMM. Kisku et al. [12] develops another multi-hypothesis approach where Euclidean distance, Mahalanobis distance and Gaussian empirical rule results are fused together using SVM. A private signature database comprising 9 authentic signatures and a single forged signature were collected for each of 180 individuals. Individually, each classification technique performs very well with EER values below 10%, and combined, a best EER of 2.15% is achieved. While multi-hypothesis techniques can greatly improve accuracy, they also require much more processing for training and testing, since multiple classifiers are used in place of a single classifier.

The Mahalanobis distance was first documented in 1936 by Prasanta Mahalanobis [22]. It performs best with multivariate normal data distributions [23]. Fang et al. [24] reports one of the earliest classifications using the Mahalanobis distance. A best EER of 19.1% was achieved. Nguyen et al. [25] also compares squared Mahalanobis distance and Gaussian kernel SVM classification on a local gradient-based feature extraction. The AER for the SVM is 15.02%. The gradient distance with SVM classification performs better than the squared Mahalanobis distance, and better than their previous tests with MDF feature extraction and SVM classification. Sigari et al. [26] performs Mahalanobis distance classification on features extracted using Gabor wavelets. Verification tests are performed on 3 relatively small, yet diverse, signature databases. EER values of 15.0%, 16.8% and 9.0% were obtained. The Mahalanobis distance performs best with multivariate normal data distributions [23]. This limits its applicability to feature extraction techniques.

The most well known distance-based measure is the Euclidean distance. There are many works that have used the Euclidean distance for authentication due to it's simplicity of implementation. Shekar et al. [27] uses the Euclidean distance for the verification of feature vectors created using an Eigenvector-based feature extraction technique. Different sizes of feature vectors were tested, and a single globally applied feature vector size was eventually chosen. An EER of 14.33% is achieved when using 10 signatures for training and 14 for testing. A better EER of 8.78% is achieved when using 15 signatures for training and 9 for testing, but Shekar et al. notes that in related works, the former configuration of training and testing samples were used. Rekik et al. [28] also test global and local feature extractions with Euclidean classification. A fusion of local and global features is found to perform better than using local or global features independently. The best EER with skilled forgeries is 11.0% using signatures from 75 different individuals.

Ramachandra et al. [7] finds the smallest Euclidean distance between cross-validated graphs of signatures, using the Hungarian method [29]. Various feature vector sizes are tested and an EER of 27.78% is achieved using skilled forgeries and 6.25% using random forgeries for two different vector sizes. The Euclidean distance accuracy is limited by the concentration of distances of large feature vectors.

The weighted Euclidean distance is not a commonly used classification technique. Zhu et al. [20] uses the weighted Euclidean distance for iris recognition in 2000. Alizadeh et al. [30] uses the weighted Euclidean distance for online signature verification. This is a promising technique that can improve the accuracy of distance-based classification.

Fractional distances are another uncommon, yet promising, classification technique. They have not been used in offline signature verification before. However, they have been used for online signature recognition in Vivaracho-Pascual et al. [21], and for face recognition in Espinosa-Dur et al. [31].

In this work, the concept of weighting by standard deviation is applied to the fractional distance to produce a novel distance-based classification technique, called the weighted fractional distance. Additionally, locally optimized thresholding is optimized by choosing the optimal feature vector length per individual.

2 TECHNIQUES AND METHODOLOGY
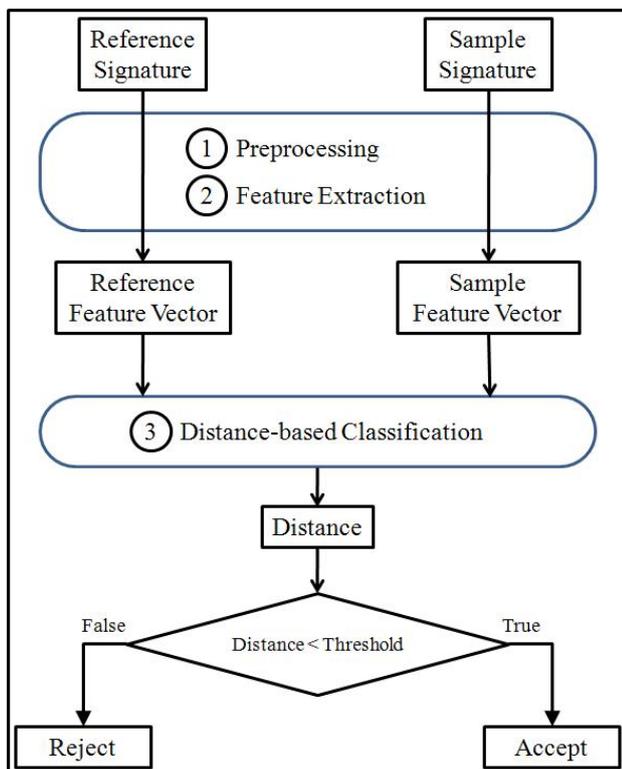
2.1 Design Overview

Figure 1: Overview of the design of a biometric verification system

Fig. 1 shows the overview of the verification system. The three main steps are preprocessing, feature extraction and classification. Preprocessing prepares the signature image for the extraction of features which are then used to classify the signature as authentic or forged. For feature extraction, the Local Directional Pattern (LDP) features are used. Before feature extraction, the required preprocessing steps are binarization, dilation and bounding box extraction. For classification, distance-based classification techniques, such as the Euclidean, Manhattan and fractional distances are tested.

In the training phase, the reference feature vector is determined by averaging a subset of randomly chosen authentic feature vectors. Sample feature vectors are obtained from other authentic and forged signatures individually. The classifier is then trained to determine the threshold that provides the optimal accuracy for the system, using sample feature vectors of known classification as input. In the testing phase, further sample signatures are used, but the classifier will determine their classification independently based on the threshold obtained from the training phase. Accuracy is gauged based on the number of signatures that the classifier correctly accepts as authentic and correctly rejects as forgeries.

2.2 Feature Extraction

2.2.1 Local Directional Pattern

The Local Directional Pattern (LDP) [5] [32] is a grayscale texture based feature method that characterizes the spatial structure of an image. For the purposes of this paper, this is an image of a signature. The LDP utilizes the 8 orientations of Kirsch masks, as shown in Fig. 2, to detect the presence of edges or corners and their orientations. Values of the 8 mask orientations, $m_0, m_1, \ldots, m_7$, are obtained by performing a convolution of Kirsch masks with the image at each pixel, followed by a binarization. In other words, given the source signature image I_{src}, we will compute the transformed image I_{LDP}, using Algorithm 1, where each transformed pixel $I_{LDP}(x, y)$ encodes the LDP spatial information of the source image pixel $I_{src}(x, y)$ and its 8 neighbours.

Fig. 3 shows an example of an LDP transformation using Algorithm 1. A source image pixel $I_{src}(x, y)$ is transformed, using its 8-neighbourhood and Kirsch masks, into a new LDP code pixel value $I_{LDP}(x, y)$. The LDP code is a byte value, where each of the 8 bits is derived from applying one of the 8 convolutions of the Kirsch masks.

A histogram, H_{LDP}, can then be created from the image I_{LDP}. Since each 8-bit pixel has exactly three bits with the value 1, and five bits with the value 0, this allows for only 56 possible permutations. Therefore, the histogram will only account for these 56 possible values. A sample histogram derived from a signature image is shown in Fig. 4. This histogram of 56 values is used as the feature vector.

Algorithm 1 Local Directional Pattern, I_{LDP}, calculation of an image, I_{src}

Require: I_{src},   ▷ Source Image
Ensure: I_{LDP},   ▷ Image Transformed using Local Directional Pattern

1: **for** each pixel (x,y) **do**
2:     **for** i=0 to 7 **do**
3:         **for** k= - 1 to 1 **do**
4:             **for** l = -1 to 1 **do**
5:                 $m_i = m_i + M_i(k+1, l+1) \times I_{src}(x+k, y+l)$
6:             **end for**
7:         **end for**
8:         Transform the three highest values m_i into 1s and the rest into 0s
9:     **end for**
10:    $powerof2 = 1$
11:    $I_{LDP}(x,y) = 0$
12:    **for** i = 0 do 7
13:        $I_{LDP}(x,y) = I_{LDP}(x,y) + m_i \times powerof2$
14:        $powerof2 = 2 \times powerof2$
15:    **end for**
16: **end for**

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

East M_0 North East M_1 North M_2 North West M_3

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

West M_4 South West M_5 South M_6 South East M_7

Figure 2: The 8 orientations of Kirsch Masks [5]. Each orientation is applied to a pixel and its 8 neighbours to calculate 8 mask values.

2.2.2 Feature Vectors and Resampling

Further, it is possible to divide the image $I_{LDP}(x,y)$ into blocks by splitting it a specified number of parts vertically (s_V) and horizontally (s_H) and extract a 56-value histogram from each block. The final feature vector, FV_{LDP}, is then obtained by concatenating all of these histograms, $FV = H_{LDP}^1 + H_{LDP}^2 + ... + H_{LDP}^{s_V \times s_H}$. A sample signature after binarization, dilation, cropping and splitting is shown in Fig. 5.

In this work, multiple feature vector sizes are analyzed so as to understand the effect of resampling on the LDP. Resampling refers to changing the number of splits, and thus the size and depth of information in the feature vector. Feature vectors between 1 and 8 vertical splits, s_V, and horizontal splits, s_H, are tested. Since each histogram has a size of 56 and they are concatenated, the feature vector size for LDP feature extraction is calculated as

$$FV_{LDP} = 56 \times s_V \times s_H \qquad (1)$$

Resampling will be discussed further in the Section 2.3.6 which discusses Individual Optimized Resampling.

Mask index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask value	-300	100	164	540	308	92	-508	-396
Rank	5	7	6	1	4	8	2	3
Code bit	0	0	0	1	0	0	1	1
LDP code	19							

Figure 3: calculation of the LDP code [5] obtained by applying each of the 8 Kirsch masks to the central pixel, with value 50, and its 8 neighbouring pixels
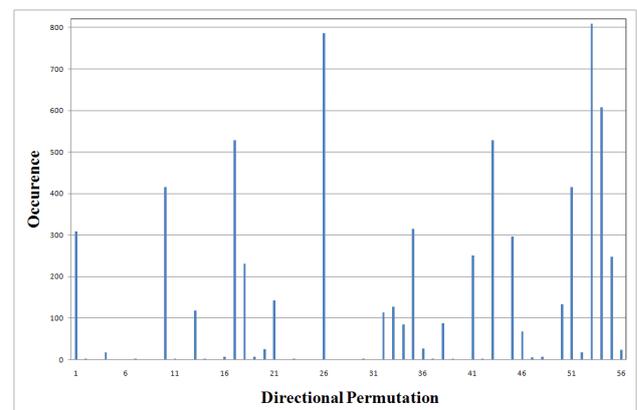

Figure 4: A sample LDP histogram showing the occurrences of each directional permutation from the image in Fig. 5 with no splits

2.3 Classification

2.3.1 Determining the Threshold

Authentic signatures are expected to have distance values below a certain threshold while forged signatures would have values above that threshold. This threshold is determined by finding the optimal Equal Error Rate (EER) during the training phase. Authentic signatures with distances above the threshold are regarded as false negatives and contribute to the False Rejection Rate (FRR) while forged signatures with distances below the threshold are regarded as false positives and contribute to the False Acceptance Rate (FAR). This is further split into the FAR for skilled forgeries (FARS) and for random forgeries (FARR). The threshold is chosen where the distance for the FRR and FARS are equal. This rate is also called the Equal Error Rate (ERR).

2.3.2 Euclidean Distance

One of the most common distance-based classification techniques for determining the accuracy of biometric systems is the calculation of the Euclidean distance [33] between a reference vector (derived as a mean of several authentic signatures of an individual) and other feature vectors. The Euclidean distance is used to calculate the distance between points in a Cartesian plane. It is a distance calculated in L^P-space, also
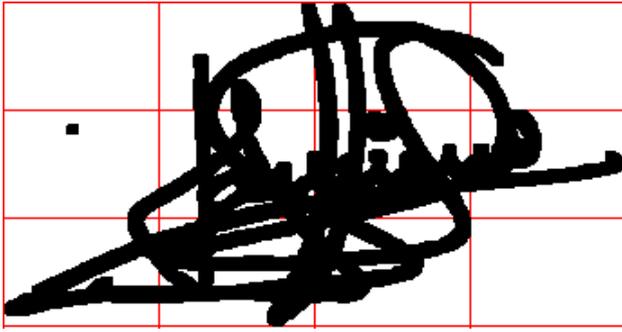
Figure 5: A dilated image with splits segmented by 3 horizontal splits and 4 vertical splits

called Lebesgue space, where the p-norm value is 2, or the L_2 distance.

The equation for determining the Euclidean distance between vectors $x = (x^i)_{i=1,2,...,m}$ and $y = (y^i)_{i=1,2,...,m}$ is defined as

$$||x - y||_p = (\sum_{i=1}^{m} |(x^i - y^i)|^p)^{1/p} \qquad (2)$$

where $p = 2$. This makes the equation

$$||x - y||_2 = (\sum_{i=1}^{m} |(x^i - y^i)|^2)^{1/2} \qquad (3)$$

2.3.3 Manhattan Distance

The Manhattan distance [34], also called the City-block or Taxicab distance, is the distance between two points determined as the sum of the absolute difference of their respective coordinates. The equation for determining the Manhattan distance between vectors $x = (x^i)_{i=1,2,...,m}$ and $y = (y^i)_{i=1,2,...,m}$ is computed as defined in Equation (2) where $p = 1$. Therefore, Equation (2) can be re-written as

$$||x - y||_1 = \sum_{i=1}^{m} |(x^i - y^i)| \qquad (4)$$

2.3.4 Fractional Distance

The fractional distance [33] is another distance in L^P-space where the p-norm value, also called the Minowski norm exponent, is any fractional value less than 1. Francois et al. [35] discusses the use of fractional distance as an alternative to the Euclidean distance to counteract the concentration phenomenon. This phenomenon is when large feature vectors cause the results of the Euclidean distance to concentrate, or cluster. This clustering of values, which is an intrinsic property of L^P-space distances, makes classification difficult for large feature vectors. Fractional distances generally produce less concentrated results than the Euclidean distance, which allows for better classification of data sets.

Francois et al. states, "Fractional norms are not always less concentrated than other norms. They seem, however, to be more relevant as a measure of similarity when the noise affecting the data is strongly

non-Gaussian." [35] Much of the noise generated by behavioral biometrics is due to random variations in human action, that may not follow a normal distribution. This makes fractional distances a viable investigative route of classifying offline handwritten signatures.

The equation for determining fractional p-norm distance between vectors x and y is computed as

$$min(||x - y||_p) = (\Sigma |(x - y)|^p)^{1/p} \qquad (5)$$

where $0.1 \leq p \leq 2.0$.

The optimal value of p is when the distance calculated using Equation (5) is at its minimum for all values of p within the given range.

2.3.5 Weighted Distances in L^P-space

The weighted Euclidean distance measure is a technique adapted from Zhu et al. [20] to improve the classification accuracy by adding weight, or statistical importance, to the most reliable features from the feature vector. Firstly, the standard deviation for the reference signatures is obtained.

Let the n reference signatures be

$$\begin{aligned} x_1 &= (x_1^1, x_1^2, \ldots, x_1^m) \\ x_2 &= (x_2^1, x_2^2, \ldots, x_2^m) \\ &\vdots \\ x_n &= (x_n^1, x_n^2, \ldots, x_n^m) \end{aligned} \qquad (6)$$

Let x_i^j be the jth component of the ith reference signature where $1 \leq i \leq n$ and $1 \leq j \leq m$.

Then the mean of the jth component the reference signatures, μ^j, is computed as

$$\mu^j = \frac{1}{n} \sum_{i=0}^{i<n} x_i^j \qquad (7)$$

and their standard deviation σ^j is defined as

$$\sigma^j = \sqrt{\frac{1}{n} \sum_{i=0}^{i<n} (x_i^j - \mu^j)^2} \qquad (8)$$

The weighted Euclidean distance is then calculated using the standard deviation

$$||x - y||_p = \left(\sum_{j=0}^{j<m} \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \qquad (9)$$

where $p = 2$.

The equation can be rewritten as

$$||x - y||_2 = \left(\sum_{j=0}^{j<m} \frac{|(x^j - y^j)|^2}{\sigma^j} \right)^{1/2} \qquad (10)$$

Further, the Manhattan distance and weighted Euclidean distance are combined to form the weighted Manhattan distance

$$min(||x - y||_1) = \sum_{j=0}^{j<m} \frac{|(x^j - y^j)|}{\sigma^j} \qquad (11)$$

The fractional distances and weighted Euclidean distance are also combined to form the weighted fractional distance

$$min(||x - y||_p) = \left(\sum_{j=0}^{j<m} \frac{|(x^j - y^j)|^p}{\sigma^j} \right)^{1/p} \qquad (12)$$

where $0.1 \leq p \leq 2.0$.

As with Equation 5, the optimal value of p is when the distance calculated using Equation 12 is at its minimum for all values of p within the given range.

2.3.6   Individually Optimized Resampling

Resampling of the feature vector allows it to be resized. This is a form of spatial normalization. Different resampling sizes results in changing accuracies. It is expected that choosing the best resampling size per user will optimize the overall accuracy of the system. When a feature vector is resampled, its size is normalized to produce a uniform feature vector size, either globally for all signature sets or locally per individual set. Vivaracho-Pascual et al. [21] try several resampling sizes for their feature vectors, in research with online signatures. They note that there is no single resampling size that is optimal for all signatures. They further state that local optimization, obtained by choosing the best feature vector size per individual, is a non-trivial approach and interesting for future study. However, for their work, they chose a global resampling size for all individuals.

The resampling method used for LDP is described in Sections 2.2.1 and 2.2.2. In the works of Ferrer et al. [5], the signatures are split into 4 blocks vertically and 3 blocks horizontally, giving a total of 12 blocks. It is unclear if other split combinations were tested. In their work, each block overlapped by 60% and feature vectors were transformed using the discrete cosine transform before classification with a support vector machine. In this work, images are tested with block splits between 1 and 8 in both the vertical and horizontal directions. This provides block numbers between 1 and 64 blocks per image.

3   RESULTS AND DISCUSSION

Tables 1 to 12 show the results of tests on the LDP using different numbers of horizontal and vertical splits (s_H and s_V respectively), with various different classification techniques. The feature vector sizes are determined by the $s_H \times s_V \times H$ where H is the length of the histogram, which is always 56 in the LDP extraction technique. Splits between 1 (no split) and 8 are tested. Further splits were omitted due to the very large vector size's adverse effect on processing speed. The same experiment setup is repeated for each of the Euclidean,

Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances.

3.1   Euclidean Distance

Table 1 shows the EER results using the Euclidean distance calculations as defined in Equation 3. The EER improved with increasing feature vector sizes at first, since more data points allows better classification. The smallest feature vector, from 1×1 splits, provided a high EER of 25.0%. This signifies a poor accuracy. As feature vector size increased, the EER improved to a best of 21.7% at splits 2×3. However, due to the concentration phenomenon, which causes distance values to cluster, the EER worsened for feature vectors larger than this, while some of the largest feature vectors resulted in worse accuracy that the smallest feature vector. The worst EER of 26.0% was obtained with the largest feature vector which had splits of 8×8.

A similar trend is observable with the FARR in Table 2, obtained using the Euclidean distance calculations, with the best FARR in the same feature vector size vicinity as the best EER, and the worst FARR agrees closely with the worst EER. Once again this can be attributed to the concentration phenomenon, which makes the differentiation between authentic and forged classes difficult with large feature vectors. The distance measures discussed below are used to counter the effects of the concentration phenomenon.

3.2   Manhattan Distance

Table 3 shows the EER results using the Manhattan distance calculations as defined in Equation 4. The best EER of 19.2% is with splits 3×3, which gives a feature vector slightly larger than with the Euclidean distance. However, while the largest feature vectors do not provide the best EER, they still provide a better EER than the smallest feature vectors. This is in contrast to the Euclidean distance where the largest feature vectors resulted in a worse EER than the smallest feature vectors. This shows that p-norm distance measures other than the Euclidean distance can provide a better result when feature vectors are larger and information is greater.

The FARR in Table 4 shows a similar trend, where the best FARR of 0.18% is also with splits of 3×3. While the FARR increase for larger feature vector sizes, it is still better than for the smallest feature vector. This further advocates the use of p-norm distances other than the Euclidean distance.

The fractional distances discussed next are used to counter the concentration phenomenon even further.

3.3   Fractional Distance

Fig. 6 shows the effect of globally applied fractional distances on small and large feature vectors. The small feature vector was obtained from images with no split (1 block), while the large feature vector was obtained from images with 7×7 splits (49 blocks). While fractional distances provide an improvement over the

Euclidean distance with the small feature vector, the effect is much more pronounced with the large feature vector. With the small feature vector, the best EER was obtained with $p = 0.4$. This agrees with the observation of Vivaracho-Pascual et al. [21] Accuracy then sharply decreases and $p = 0.1$ gives a worse EER than the Euclidean distance. Conversely, when the feature vector is very large, the Euclidean distance provides a slightly worse EER in comparison to the small feature vector, but the EER greatly improves when $p < 2.0$. The best EER was found to then be with $p = 0.1$, which is 5.8% better than the best globally applied fractional distance for the smaller feature vector. This discrepancy with Vivaracho-Pascual et al. is most likely due to their use of insufficiently large feature vectors. Additionally, the best EER for the fractional distance with the large feature vector was 8.8% better than the associated Euclidean distance and 11.2% better than the worse fractional distance with the small feature vector. This proves that the fractional distances do counter the concentration effect that occurs especially with large feature vectors.
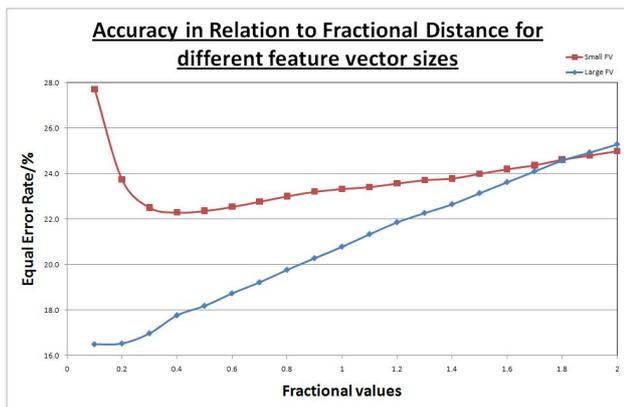

Figure 6: EER in Relation to individual fractional distances for large and small feature vector sizes

The fractional distances were then applied with local optimization. Table 5 shows the EER results using the locally optimized fractional distance calculations as defined in Equation 5. This means that the best fractional distance within the range $0.1 \leq p \leq 2.0$ was chosen per individual, since some fractional distances work better than others for different individuals. The best EER of 14.7% is with one of the largest feature vectors, 7×8. This is smaller than the smallest feature vector using fractional distances by 4.2% and it is better than the best Euclidean distance by 7.0%. This further shows that the fractional distances provide a solution to the concentration phenomenon that occurs with the Euclidean distance. There is a trend of improvement as feature vectors become larger. The fractional distances appears to reach their best improvement in overcoming the concentration phenomenon between splits 7×7 and 8×8, which has associated feature vector sizes between 2744 features and 3584 features. There is also an improvement over globally applied fractional distance, by 1.8%.

The FARR in Table 6 shows a slightly different trend, where the best FARR does not correspond with the best EER. This is most likely due to choosing the best p-norm value locally for each individual, rather than a globally used p-norm as in the case of the Euclidean and Manhattan distances. Due to this localized optimization of the fractional distances, in some cases, larger p-norm values were used for the best EER, resulting in worsening of the concentration phenomenon for larger feature vectors. However, the FARR for fractional distances is still better than that for Euclidean and Manhattan distances, especially in the case of larger feature vectors.

3.4 Weighted Euclidean Distance

Table 7 shows the EER results using the weighted Euclidean distance calculations as defined in Equation 10. The best EER of 14.5% is with splits 4×5, which gives a feature vector larger than with the best unweighted Euclidean or Manhattan distance. This is of a better EER than the Euclidean and Manhattan distances, and of an almost equal EER in comparison with the fractional distance. The low EER is due to extra weight, or importance, being given to the most reliable features in the feature vector, which have the smallest intra-class difference per individual. This weighting also counteracts the clustering of the concentration phenomenon to a small extent. This is emphasized further by the largest feature vectors having a lower EER than the smallest feature vectors.

In contrast, the FARR in Table 8 shows a different trend, where the FARR worsens as the feature vectors become larger. This is because the weighting can not always differentiate between different sets of signatures and some features in random forgeries may be given extra weight if they are similar to authentic signatures. This is particularly visible with the largest feature vectors which will also experience the effects of the concentration phenomenon as well.

3.5 Weighted Manhattan Distance

Table 9 shows the EER results using the weighted Euclidean distance calculations as defined in Equation 11. The best EER of 14.3% is with splits 7×7, which gives a feature vector larger than with the best unweighted Euclidean, Manhattan and weighted Euclidean distances. It is of equal size in comparison to the fractional distance. The larger feature vector size provided a better EER due to a combination of a smaller p-norm distance and the weighting by standard deviation, which counteracted the concentration phenomenon in tandem. Due to this combination, it is also of a better EER than all the distances tested with the LDP before it.

The FARR in Table 8 shows a similar trend to the FARR of the weighted Euclidean distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a smaller p-norm value, the largest feature vectors have a smaller FARR in comparison with the largest feature vector sizes of the weighed Euclidean distance.

Table 1: The effect of different s_H and s_V on EER(%) using the Euclidean distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	25.0	23.9	22.6	23.0	22.6	22.8	23.0	23.0
2	23.8	22.3	**21.7**	22.0	22.0	22.3	22.4	22.6
3	24.0	22.7	21.9	21.9	22.2	22.3	22.6	22.8
4	24.1	23.0	22.5	22.5	22.8	23.0	23.1	23.5
5	24.3	23.7	23.1	23.2	23.5	23.6	23.9	24.3
6	24.8	23.9	23.4	23.7	24.0	24.2	24.5	25.1
7	25.2	24.4	24.2	24.1	24.5	25.0	25.3	25.4
8	26.0	25.2	24.7	25.0	25.2	25.6	25.7	26.0

Table 2: The effect of different s_H and s_V on FARR(%) using Euclidean distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	1.70	0.96	0.60	0.58	0.55	0.55	0.59	1.35
2	1.09	0.48	0.31	0.34	0.34	0.35	0.41	0.45
3	0.85	0.43	0.28	0.33	0.37	0.38	0.44	0.54
4	0.72	0.43	0.35	0.37	0.42	0.54	0.59	0.72
5	0.81	0.52	0.42	0.53	0.61	0.80	0.88	1.09
6	0.89	0.57	0.52	0.68	0.83	1.02	1.28	1.52
7	0.97	0.77	0.73	0.94	1.20	1.53	1.84	2.26
8	1.02	0.90	0.92	1.19	1.50	1.89	2.24	2.87

Table 3: The effect of different s_H and s_V on EER(%) using the Manhattan distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	23.3	21.6	20.4	20.3	20.2	20.1	20.2	20.2
2	21.7	20.2	19.3	19.7	19.4	19.7	19.8	20.0
3	21.5	20.0	**19.2**	19.4	19.3	19.4	19.5	19.7
4	21.4	19.9	19.6	19.5	19.8	19.5	19.6	20.1
5	21.6	20.2	19.6	19.8	19.9	20.1	20.1	20.3
6	21.5	20.6	20.0	20.3	20.2	20.4	20.6	20.8
7	21.9	20.6	19.8	20.1	20.5	20.5	20.8	21.0
8	22.0	21.0	20.2	20.5	20.7	20.9	20.9	21.2

Table 4: The effect of different s_H and s_V on FARR(%) using Manhattan distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	1.39	0.77	0.46	0.43	0.39	0.43	0.37	0.89
2	0.79	0.34	0.24	0.24	0.23	0.23	0.28	0.30
3	0.54	0.22	0.18	0.20	0.23	0.24	0.29	0.32
4	0.45	0.22	0.22	0.23	0.26	0.32	0.36	0.44
5	0.47	0.25	0.24	0.27	0.37	0.45	0.54	0.60
6	0.45	0.25	0.30	0.34	0.44	0.55	0.64	0.79
7	0.47	0.30	0.31	0.43	0.55	0.71	0.93	1.08
8	0.50	0.35	0.38	0.50	0.67	0.87	1.10	1.33

Table 5: The effect of different s_H and s_V on EER(%) using the fractional distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	18.9	17.7	17.0	16.9	16.7	16.6	16.6	16.7
2	17.9	15.0	16.2	16.1	16.1	15.9	16.0	16.0
3	17.5	16.4	15.6	15.7	15.4	15.2	15.2	15.5
4	17.5	16.2	15.9	15.6	15.4	15.3	15.3	15.2
5	17.5	16.3	15.3	15.4	15.3	15.2	15.0	15.2
6	17.7	16.1	15.7	15.4	15.2	15.0	14.9	14.9
7	17.4	16.2	15.4	15.3	15.0	14.8	14.7	**14.7**
8	17.7	16.5	15.5	15.3	15.3	14.8	14.7	14.7

Table 6: The effect of different s_H and s_V on FARR(%) using fractional distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	1.59	0.95	0.53	0.50	0.50	0.48	0.45	0.52
2	1.02	0.48	0.40	0.37	0.38	0.44	0.51	0.52
3	0.70	0.33	0.32	0.30	0.35	0.32	0.42	0.43
4	0.57	0.31	0.37	0.39	0.34	0.40	0.45	0.51
5	0.59	0.34	0.39	0.31	0.45	0.47	0.54	0.57
6	0.54	0.36	0.44	0.44	0.45	0.51	0.54	0.73
7	0.52	0.42	0.41	0.47	0.51	0.66	0.77	0.78
8	0.53	0.54	0.50	0.51	0.60	0.70	0.76	0.89

3.6 Weighted Fractional Distance

Following the success of the weighted Manhattan distance, the weighted fractional distances, which combines the weighted Euclidean distance, are tested. Table 11 shows the EER results using the weighted fractional distance calculations as defined in Equation 12. At splits 7×7, this provides the best EER of 12.2% which is better than all of the classification techniques tested so far. Once again, the use of fractional distances and weighting allow greater accuracy with larger feature vectors. By combining two techniques that are individually better than the Euclidean distance, an overall much better accuracy was achieved. The worst EER for the weighted fractional distance was 16.83% using no splitting of the image. This is nearly 5% better than the best results for the standard Euclidean distance. Conversely, the best EER is 9.5% better than the best recorded Euclidean distance, and 2.5% and 2.7% better than the fractional and weighted Euclidean distances respectively.

The FARR in Table 12, shows a similar trend to the FARR of the weighted Euclidean distance and the weighted Manhattan distance, where the FARR worsens as the feature vectors become larger. However, due to the use of a wide range of p-norm values, the FARR values are better than the FARR for the Euclidean distance, but slightly worse than for the Manhattan distance.

3.7 Individually Optimized Resampling

Tables 13 to 14 show the test results for individually, or locally, optimized resampling performed in conjunction with each of the above six discussed distance-based classification techniques, namely the Euclidean, Manhattan, fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances. Optimizations from 4 split and 9 split combinations were tested. It was found that 9 splits yield better results than 4 splits. These are symbolized in the following tables with the [x:y] notation. For example, [1:3] represents the 9 combinations of 1×1, 1×2, 1×3, 2×1, 2×2, 2×3, 3×1, 3×2 and 3×3.

3.7.1 Individually Optimized Euclidean Distance

Table 13 shows the results of tests with the p-norm distances and individually optimized resampling and the Table 14 shows those for weighted p-norm distances. The best EER for the Euclidean distance with individually optimized resampling was 16.5% from the [1:3] combination. The optimal combinations of feature vector sizes corresponds with the lowest single Euclidean distance EER, of 21.7%, from splits 2×3 when no individual optimization was performed. The combination provides a better EER than non-optimized Euclidean and Manhattan distances. The FARR values also improved through the individually optimized resampling. This suggests an association between individuals with a low EER and a low FARR.

3.7.2 Individually Optimized Manhattan Distance

The best EER for individual optimization on the Manhattan distance was 14.3% from the [1:3] combination. The optimal combination of feature vector sizes corresponds with the lowest single Manhattan distance EER, of 19.2%, from splits 3×3 when no individual optimization was performed. The 9-combination provides an EER better than or equal to all of the non-locally optimized distance-based measures, except for the weighted fractional distance whose best EER is 12.2%. The FARR values also improved through the individually optimized resampling.

3.7.3 Individually Optimized Fractional Distance

The best EER for individually optimized fractional distances was 11.3% from the [1:3] combination and worst was 11.8% in the [6:8] combination. The 9-combination provides an EER better than all measures tested before it, including the weighted fractional distance whose best EER is 12.2%. The FARR values also showed an improvement through the individually optimized resampling.

The optimal combinations of feature vector sizes do not correspond with the lowest single fractional distance, even though the best and worst EER are very close, with only a 0.5% difference between them. This discrepancy may be explained by Fig. 7. In the worse case scenario, with the combination [6:8], the majority occurring p-norm in 0.1. This is larger than the second highest, 0.2, occurrence by over 40%. Conversely, for the best case scenario of [1:3], the 5 highest occurring distances are all within a 10% range of each other, and are the 5 smallest p-norm values. This allows a higher accuracy, since sometimes, among smaller feature vectors, different p-norm values are better for difference individuals. However, with the largest feature vectors, the p-norm value of 0.1 outperforms all others. A combination of smaller feature vectors with more variable p-norm values can therefore perform slightly better than a large feature vector with a single dominant p-norm value.
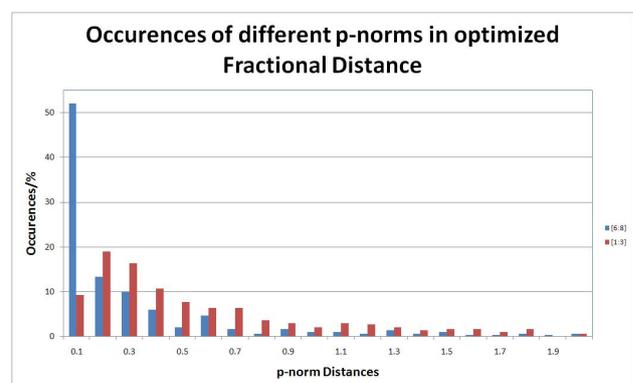

Figure 7: The number of occurrences of each p-norm in the fractional distance with individually optimized resampling combinations [1:3] and [6:8]

Table 7: The effect of different s_H and s_V on EER(%) using the weighted Euclidean distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	18.6	17.9	16.8	16.5	16.1	16.0	16.0	16.0
2	18.4	16.5	15.8	15.7	15.4	15.3	15.2	15.2
3	17.3	16.0	15.5	15.3	14.6	14.8	14.6	14.9
4	17.4	15.7	15.2	15.1	**14.5**	14.8	14.8	15.0
5	17.3	15.6	14.7	14.8	14.6	14.6	14.7	15.1
6	17.0	15.7	14.8	15.0	14.7	14.8	14.7	14.8
7	16.9	15.5	14.8	14.9	15.0	14.7	14.9	15.1
8	17.2	15.8	15.1	15.3	15.2	14.9	15.1	15.3

Table 8: The effect of different s_H and s_V on FARR(%) using weighted Euclidean distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	0.98	0.65	0.50	0.48	0.60	0.64	0.71	1.00
2	0.69	0.51	0.69	0.96	1.08	1.44	1.59	1.85
3	0.55	0.62	0.84	1.04	1.32	1.48	1.53	1.80
4	0.54	0.68	0.93	1.26	1.46	1.69	1.84	1.94
5	0.68	0.87	1.33	1.44	1.79	1.86	2.10	2.22
6	0.77	1.14	1.51	1.66	1.93	2.05	2.17	2.42
7	0.81	1.39	1.54	1.82	1.95	2.19	2.37	2.66
8	0.91	1.35	1.70	1.90	2.08	2.39	2.50	2.92

Table 9: The effect of different s_H and s_V on EER(%) using the weighted Manhattan distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	19.4	18.6	17.3	17.3	17.0	16.6	16.7	16.7
2	19.3	17.2	16.6	16.3	15.9	15.6	15.5	15.6
3	18.3	16.5	15.8	15.5	15.1	15.0	14.7	14.7
4	18.1	16.2	15.4	15.3	14.8	14.8	14.5	14.7
5	18.0	16.2	15.0	14.9	14.8	14.7	14.7	14.7
6	17.7	16.1	15.1	15.1	14.9	14.6	14.5	14.7
7	17.4	16.0	15.0	15.0	14.5	14.3	**14.3**	14.4
8	18.0	16.3	15.2	15.1	14.7	14.6	14.5	14.6

Table 10: The effect of different s_H and s_V on FARR(%) using weighted Manhattan distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	0.91	0.56	0.36	0.35	0.31	0.36	0.38	0.50
2	0.63	0.32	0.37	0.48	0.52	0.59	0.70	0.86
3	0.44	0.32	0.36	0.44	0.50	0.66	0.74	0.85
4	0.36	0.32	0.44	0.53	0.62	0.74	0.81	0.95
5	0.39	0.42	0.60	0.67	0.91	0.96	1.11	1.26
6	0.42	0.46	0.70	0.80	1.01	1.16	1.29	1.50
7	0.43	0.53	0.77	0.96	1.02	1.31	1.52	1.71
8	0.47	0.59	0.82	1.01	1.23	1.44	1.70	1.94

Table 11: The effect of different s_H and s_V on EER(%) using the weighted fractional distance

s_H \ s_V	1	2	3	4	5	6	7	8
1	16.8	16.1	15.1	14.8	14.5	14.4	14.2	14.6
2	16.7	15.0	14.2	14.0	13.7	13.4	13.3	13.4
3	15.7	14.3	13.6	13.2	12.7	12.6	12.5	12.6
4	15.8	14.0	13.3	13.0	12.5	12.7	12.4	12.5
5	15.3	14.1	12.8	12.7	12.5	12.3	12.3	12.5
6	15.1	13.9	12.9	12.7	12.6	12.5	12.3	12.4
7	15.0	13.8	12.9	12.7	12.5	12.3	**12.2**	12.4
8	15.4	13.9	13.0	13.0	12.6	12.5	12.5	12.6

Table 12: The effect of different s_H and s_V on FARR(%) using weighted fractional distances

s_H \ s_V	1	2	3	4	5	6	7	8
1	1.08	0.70	0.47	0.42	0.52	0.57	0.57	0.72
2	0.65	0.48	0.54	0.84	0.81	0.93	1.22	1.55
3	0.54	0.42	0.66	0.75	0.75	0.81	1.09	1.33
4	0.56	0.53	0.23	0.91	1.00	1.11	1.14	1.34
5	0.63	0.79	1.00	1.02	1.20	1.28	1.28	1.61
6	0.68	0.76	1.16	1.08	1.34	1.36	1.45	1.67
7	0.68	0.88	1.18	1.36	1.44	1.63	1.64	1.78
8	0.66	1.04	1.21	1.42	1.53	1.84	1.84	2.11

Table 13: The effect of implementing individually optimized resampling by choosing the best EER from 9 feature vectors for each of the p-norm distances

Split Range	Euclidean		Manhattan		Fractional	
	EER	FARR	EER	FARR	EER	FARR
[1:3]	**16.5**	0.67	**14.3**	0.49	**11.3**	0.57
[2:4]	17.1	0.31	14.9	0.21	11.7	0.31
[3:5]	17.7	0.34	15.3	0.21	11.7	0.29
[4:6]	19.0	0.55	16.3	0.30	11.8	0.36
[5:7]	20.1	0.80	17.1	0.45	11.9	0.45
[6:8]	21.4	1.36	17.8	0.69	11.8	0.62

Table 14: The effect of implementing individually optimized resampling by choosing the best EER from 9 feature vectors for each of the weighted *p*-norm distances

| Split Range | Weighted Euclidean | | Weighted Manhattan | | Weighted Fractional | |
	EER	FARR	EER	FARR	EER	FARR
[1:3]	11.0	0.70	11.5	0.50	9.65	0.60
[2:4]	10.8	0.72	11.8	0.34	9.66	0.49
[3:5]	**10.7**	1.18	**11.4**	0.48	**9.25**	0.80
[4:6]	11.1	1.59	11.6	0.68	9.50	0.97
[5:7]	11.2	1.88	11.9	1.06	9.60	1.18
[6:8]	11.8	2.20	12.0	1.27	10.0	1.36

3.7.4 Individually Optimized Weighted Euclidean Distance

The best EER for individually optimized weighted Euclidean distance was 10.7% from the [3:5] combination. The optimal combination of feature vector sizes correlate with the lowest single weighted Euclidean distance EER, of 14.5%, from splits 4×5 when no individual optimized resampling was performed. This provides better EER than previous distance-based classification measures. This includes better performance than the best weighted fractional distance of 12.2% without individually optimized resampling and all of the previously tested individually optimized resamplings of the LDP. The FARR values correlate closely with those for the weighted Euclidean distance without individually optimized resampling.

3.7.5 Individually Optimized Weighted Manhattan Distance

The best EER for individually optimized weighted Manhattan distance was 11.4% from the [3:5] combination. The optimal combination of feature vector sizes corresponds with the lowest single weighted Manhattan distance EER, of 14.5%, from splits 7×7 when no individual optimized resampling was performed. This provides better EER than previous distance-based classification measures, where no individually optimized resampling was performed. However, the individually optimized weighted Manhattan distance performs worse than the individually optimized weighted Euclidean distance. This may be because the standard deviation used in the weighting equation is calculated with a *p*-norm of 2 while the Manhattan distance is calculated with a *p*-norm of 1. The FARR values closely agree with those for the weighted Manhattan distance without individually optimized resampling.

3.7.6 Individually Optimized Weighted Fractional Distance

The best EER for the individually optimized weighted fractional distance performed better than all other test before. This good performance is due to choosing both the best *p*-norm and best split size per individual, i.e. locally optimized classification. The optimal split sizes for the individually optimized fractional distance do not correlate with the optimal split sizes for the non-locally optimized weighted fractional distance, al-

though there is a mere difference of 0.75% between the best and worst EER. An analysis of Fig. 8, shows that the pattern for the highest EER, from combination [6:8], is similar to that for the highest EER for the individually optimized fractional distance in Fig. 7. In both cases the smallest *p*-norm size of 0.1 has the greatest percentage of occurrences by far. Similarly, the best EER, from combination [3:5], for the individually optimized fractional distance has a comparable pattern with the best EER of [1:3] for the individually optimized fractional distance. In both cases, the first 4 *p*-norm distances are among the largest and within a close range. Additionally, the greatest occurrence is still for the *p*-norm value of 0.1. The smaller feature vectors in combination [1:3] provide an EER in between the highest and lowest. In Fig. 8, is can be seen that the greatest occurrence was for *p*-norm of 0.6 rather than 0.1. The latter provides much lower EER values with larger feature vector sizes. The larger *p*-norm of 0.6 is more effect with smaller feature vectors, in comparison with 0.1. However, 0.1 performs much better than 0.6 in comparison with larger feature vectors.

An analysis of the weighted Euclidean, individually optimized weighted Euclidean, weighted Manhattan and weighted fractional distances shows that all of them provide their best EER in the feature vector range for splits 3×3 and 5×5. This suggests a strong involvement of the standard deviation function that is used in the weighting. In all cases, the standard deviation was calculated in *p*-norm space of 2, whereas the Manhattan and fractional distances were calculates with smaller *p*-norm distances.

3.8 Literature Comparison

Ferrer et al. [5] tested the Local Binary Pattern (LBP) and Local Directional Pattern (LDP) using several data sets, separately. These sets were 75 individuals from the MCYT database [36], and 75, 300 and 960 users from the GPDS database [37]. While EER with the data sets using 75 individuals is low, these results can not be used for comparison, since the small size of the data set bring the precision and accuracy of results into question. A better comparison of the results is obtained with the data sets containing 300 and 960 individuals. Classification was performed with an Least Squares SVM (LS-SVM) with an RBF kernel. The signatures were split into 12 blocks for the feature extraction, consisting of 4 vertical splits and 3 horizontal splits,
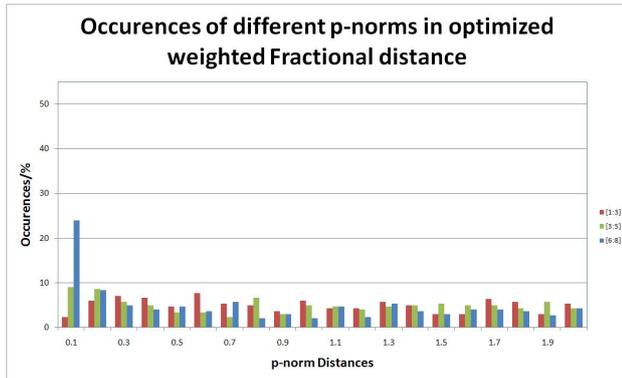
Figure 8: The number of occurrences of each *p*-norm in the weighted fractional distance with individually optimized resampling combinations [1:3] and [6:8]

and an overlap of 60%. From these, the best EER was 17.8% using 300 individuals of the GPDS database, with a corresponding FARR of 0.68%.

In this work, the same feature extraction technique was implemented and tested with multiple distance-based classification techniques, including the Euclidean, fractional, weighted Euclidean and weighted fractional distances. Signatures of 300 GPDS individuals were used, which was also the same database used for the best results in [5]. A larger range of splits sizes were tested. This range was from 1 to 8 splits in both the vertical and horizontal directions. While the best Euclidean and Manhattan distance performances for the LDP was worse than the literature result, at 21.7% and 19.2% respectively, the fractional, weighted Euclidean, weighted Manhattan and weighted fractional distances all performed much better, at 14.7%, 14.5%, 14.3% and 12.2%. It can be seen that the weighted fractional distance performs better than the LS-SVM with an RBF kernel for the classification of the LDP [5]. Further, individually optimized resampling was performed, where the best resampled block sizes per individual were chosen from a possibility of 9. The best individually optimized resampling was for the weighted fractional distance, with an EER of 9.25% and corresponding FARR of 0.80%. The obtained EER is better than the best results in the literature [5] by 8.6%, and both systems obtained an FARR of below 1%.

4   CONCLUSION AND FUTURE WORK

To improve the accuracy of offline signature verification, distance-based classification techniques in L^P-space were tested with application to LDP feature extraction technique. The combination of weighted fractional distance with individually optimized resampling achieved a best EER of 9.25%. This shows that distance-based classification techniques provide a viable alternative to SLTs for the verification of offline signatures.

Future work includes testing the system with foreign signature databases, such as Chinese and Persian; and fusing the signature modality with other biometric modalities to create a multi-modal biometric verification system.

REFERENCES

[1] A. Jain, A. Ross and S. Prabhakar. "An introduction to biometric recognition". *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, pp. 4–20, 2004.

[2] Y. Al-Omari, S. Abdullah and K. Omar. "State-of-the-art in offline signature verification system". In *Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on*, vol. 1, pp. 59–64. IEEE, 2011.

[3] P. Campisi, R. Carter, C. Crooks, V. Govindaraju, W. Hamilton, J. Hurt, A. Ross, C. Tilton and J. D.M. Waymire. *Module 2: Biometric Modalities in Certified Biometrics Professional (CBP)*. IEEE, 2010.

[4] S. Armand, M. Blumenstein and V. Muthukku-marasamy. "Off-line signature verification using the enhanced modified direction feature and neural-based classification". In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pp. 684–691. IEEE, 2006.

[5] M. Ferrer, F. Vargas, C. Traviesto and J. Alonso. "Signature verification using local directional pattern". In *International Carnahan Conference on Security Technology*, pp. 336–340. IEEE, Oct 2010.

[6] B. Majhi, Y. Reddy and D. Babu. "Novel features for off-line signature verification". *International Journal of Computers, Communications & Control*, vol. 1, no. 1, pp. 17–24, 2006.

[7] A. Ramachandra, K. Pavithra, K. Yashasvini, K. Raja, K. Venugopal and L. Patnaik. "Offline signature authentication using cross-validated graph matching". In *Proceedings of the 2nd Bangalore Annual Compute Conference*, p. 7. ACM, 2009.

[8] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla and C. Trullo. "Analysis of stability in static signatures using cosine similarity". In *Proceedings of the International Conference on Frontiers in Handwriting Recognition*, pp. 231–235. IEEE, 2012.

[9] M. Yilmaz, B. Yanikoglu, C. Tirkaz and A. Kholmatov. "Offline signature verification using classifier combination of HOG and LBP features". In *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–7. IEEE, 2011.

[10] J. Vargas, M. Ferrer, C. Travieso and J. Alonso. "Offline signature verification based on pseudo-cepstral coefficients". In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pp. 126–130. IEEE, 2009.

[11] M. Ferrer, J. Alonso and C. Travieso. "Offline geometric parameters for automatic signature verification using fixed-point arithmetic". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 993–997, 2005.

[12] D. Kisku, A. Rattani, P. Gupta and J. Sing. "Offline signature verification using geometric and orientation features with multiple experts fusion". In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, vol. 5, pp. 269–272. IEEE, 2011.

[13] J. Coetzer, B. Herbst and J. Du Preez. "Offline signature verification using the discrete radon transform and a hidden Markov model". *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 559–571, 2004.

[14] M. Panton. *Off-line signature verification using ensembles of local Radon transform-based HMMs*. Ph.D. thesis, Stellenbosch: Stellenbosch University, 2011.

[15] L. Batista, E. Granger and R. Sabourin. "Improving performance of HMM-based off-line signature verification systems through a multi-hypothesis approach". *International Journal on Document Analysis and Recognition*, vol. 13, no. 1, pp. 33–47, 2010.

[16] B. Kovari, A. Horvath, B. Toth, H. Charaf, L. Perlovsky, D. Dionysiou, L. Zadeh, M. Kostic, C. Gonzalez-Concepcion, H. Jaberg et al. "Local feature based off-line signature verification using neural network classifiers". In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*, 11. WSEAS, 2009.

[17] H. Prakash and D. Guru. "Geometric centroids and their relative distances for off-line signature verification". In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pp. 121–125. IEEE, 2009.

[18] L. Baum and T. Petrie. "Statistical inference for probabilistic functions of finite state Markov chains". *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.

[19] C. Cortes and V. Vapnik. "Support-vector networks". *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[20] Y. Zhu, T. Tan and Y. Wang. "Biometric personal identification based on iris patterns". In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, pp. 801–804. IEEE, 2000.

[21] C. Vivaracho-Pascual, M. Faundez-Zanuy and J. Pascual. "An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances". *Pattern Recognition*, vol. 42, no. 1, pp. 183 – 193, 2009.

[22] P. Mahalanobis. "On the generalized distance in statistics". In *Proceedings of the National Institute of Science (India)*, vol. 12, pp. 49–55. National Institute of Science, 1936.

[23] N. Kato, M. Suzuki, S. Omachi, H. Aso and Y. Nemoto. "A handwritten character recognition system using directional element feature and asymmetric Mahalanobis distance". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 258–262, 1999.

[24] B. Fang, C. Leung, Y. Tang, K. Tse, P. Kwok and Y. Wong. "Off-line signature verification by the tracking of feature and stroke positions". *Pattern Recognition*, vol. 36, no. 1, pp. 91–101, 2003.

[25] V. Nguyen, Y. Kawazoe, T. Wakabayashi, U. Pal and M. Blumenstein. "Performance analysis of the gradient feature and the modified direction feature for off-line signature verification". In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, pp. 303–307. IEEE, 2010.

[26] M. Sigari, M. Pourshahabi and H. Pourreza. "Offline handwritten signature identification and verification using multi-resolution Gabor wavelet". *International Journal of Biometric and Bioinformatics*, vol. 5, 2011.

[27] B. Shekar and R. Bharathi. "Eigen-signature: A robust and an efficient offline signature verification algorithm". In *Recent Trends in Information Technology (ICRTIT), 2011 International Conference on*, pp. 134–138. IEEE, 2011.

[28] Y. Rekik, N. Houmani, M. El Yacoubi, S. Garcia-Salicetti and B. Dorizzi. "A comparison of feature extraction approaches for offline signature verification". In *Multimedia Computing and Systems (ICMCS), 2011 International Conference on*, pp. 1–6. IEEE, 2011.

[29] H. Kuhn. "The Hungarian method for the assignment problem". *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 2006.

[30] A. Alizadeh, T. Alizadeh and Z. Daei. "Optimal threshold selection for online verification of signature". In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, pp. 17–19. 2010.

[31] V. Espinosa-Duró, M. Faundez-Zanuy and J. Mekyska. "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums". *Cognitive Computation*, pp. 1–17, 2012.

[32] T. Jabid, H. Kabir and O. Chae. "Gender classification using local directional pattern (LDP)". In I. C. Society (editor), *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR'10, pp. 2162–2165. 2010.

[33] M. Deza and E. Deza. *Encyclopedia of distances*. Springer, 2009.

[34] E. Krause. *Taxicab geometry: An adventure in non-Euclidean geometry*. Courier Dover Publications, 1987.

[35] D. Francois, V. Wertz and M. Verleysen. "The concentration of fractional distances". *Knowledge and Data Engineering, IEEE Transactions on*, vol. 19, no. 7, pp. 873–886, 2007.

[36] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero and Q. I. Moro. "MCYT baseline corpus: A bimodal biometric database". In *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, vol. 150 (6), pp. 395–401. December 2003.

[37] J. Vargas, M. Ferrer, C. Travieso and J. Alonso. "Off-line handwritten signature GPDS-960 corpus". In *Ninth International Conference on Document Analysis and Recognition*, pp. 764 – 768. IEEE Computer Society, 2007.