The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Development of Synthetically Generated LEA Signatures to Generalize Probability of False Positive Identification Estimates

Grant Number: 2009-DN-BX-K236

Final Report

Period Covered: January 01, 2010 – July 31, 2012

Submitted by:

Benjamin Bachrach, Pan Gao, Roger Xu, Wei Wang, Ajay Mishra, Kaizhi Tang & Guangfan Zhang

Intelligent Automation, Inc.
15400 Calhoun Drive, Suite 400
Rockville, MD 20855

Abstract

Automated ballistics identification (ABI) systems require a large amount of data to evaluate its performance. It is impractical to fire such a large number of bullets for this purpose. Therefore, it is significantly important to develop a bullet data synthesis methodology for a class of guns based on a set of fired sample bullets, which can be used to generalize the results achieved based on the limited data to a much larger population of firearms.

Land Engraved Area (LEA) signature, a 1D signal computed from a digitized 3D bullet surface, is used for bullet matching used in the ABI systems. In this study, we developed a bullet data synthesis approach to generate LEA signatures by capturing the characteristics of each brand and the details of each barrel. The approach consists of a deterministic component and a random component. The deterministic component includes periodicities in a base curve profile, and the random component is best represented by the fractal model of an irregular curve. Such bullet data synthesis tool had never been developed before.

The parameters used in the generation of the base and fractal curves are obtained from signal analysis based on wavelet transformation. After applying wavelet decomposition to an existing bullet LEA signature at a certain decomposition level, the wavelet approximation (WA) part plus all the wavelet detail (WD) parts can be used to fully reconstruct the original signal. The WA part provides us many useful features regarding the deterministic part of the signature, such as peak/valley points, positive/negative slopes and zero-crossing length, while the fractal dimension of the WD part discloses unique non-stationary effects caused by the randomness of the barrel manufacturing process. By utilizing the wavelet and fractal techniques as tools to analyze all the existing bullet LEA signatures and extract useful features, we synthesize a new signature in a random fashion within predefined parameter distributions.

An optimization procedure was also applied to the synthesis process to ensure the matched and non-matched correlation distribution of the new synthesized data set resembles that of the existing data of the same brand. We also extended the data synthesis process from a 1D LEA signature signal to a 2D LEA image, further to, even, a 3D bullet surface.

According to Our preliminary data synthesis results, newly generated LEAs for a brand show strong similarity to the LEAs in the same brand, but dissimilarity to those in different brands. Also, the correlation distributions for matched and non-matched LEAs in the same brand suggest the synthesized data approaches the same distribution as the experimental ones.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Accomplishments

Most recently, automated ballistics identification systems have been developed [1]~[5]. Although promising results can be found in recently developed automated ballistics identification systems, only limited data have been used to develop and evaluate these systems. Moreover, to present firearm matching evidence in court, we have to answer a fundamental question associated with positive identifications: what is the probability that two bullets with a given level of similarity could originate from two different barrels? To experimentally quantify and evaluate that probability, it would be necessary to compare a very large numbers of bullets fired by barrels of a population of interest until such an occurrence takes place a few times.

Since it is impractical to fire such a large number of bullets for this purpose, developing a synthetic LEA signature generation methodology for a class of guns based on a set of fired sample bullets is significantly important. The method can be used to generalize the results achieved based on the limited data to a much larger population of firearms. In this report, we describe new approaches and tools to address this issue in creating synthetic Land Engraved Areas (LEAs). These synthetic LEAs would be constructed so as to resemble a large population of guns of a common background. Hence, the synthetically generated LEAs would be representative of all those LEAs having the same background as the LEAs present in the test set.

Another motivation for bullet data synthesis is to overcome the challenges in developing a good estimate of the similarity measure towards the "tails" of the matching/non-matching distributions. With the synthetic LEA signature generation engine, we can easily generate large amount of LEAs that can be used to repeatedly test and evaluate the developed algorithms.

Though there is no published study on bullet data synthesis, based on the philosophy and framework of the existing data synthesis applications such as 3D face modeling [13], fingerprint identification [12] and voice morphing [14], we identified two important strategies for data synthesis: key feature identification, and signal decomposition and reconstruction. Using these strategies, we developed a two-step data synthesis approach to generate LEA signature by capturing the characteristics of each brand and also the details for each barrel.

The synthetic LEA signature generation is based on an assumption that the bullets fired by the same barrel or the barrels of the same brand have similar patterns in LEA signatures, and the bullets from different barrels or brands have significantly different LEAs. Therefore, we first characterize the patterns for the LEAs from the same sample set of barrel or brand. Once characterized, we are able to synthesize new LEAs that share the same characteristics as those in the sample set.

To implement the two-step strategy of data synthesis, we developed a superposition approach. The approach consists of a deterministic component and a random component. The deterministic component includes periodicities in a signature base curve profile, and the random component is best represented by the fractal model of an irregular curve.

The parameters used in the generation of the base and fractal curves are obtained from signal analysis. The wavelet transformation is a multi-resolution tool for signal analysis. The use of wavelets to analyze bullet striations had already been undertaken in [7]. Wavelets preserve time (or position) and frequency information and decompose a 2D LEA signature into the wavelet domain so that the pattern can be better analyzed in the new space. After applying wavelet decomposition to an existing bullet LEA signature at certain decomposition level, the wavelet approximation part plus all the wavelet detail parts can be used to fully reconstruct the original signal. The wavelet approximation part provides us many useful features regarding the deterministic part of the signature, while the fractal dimension of the wave detail part discloses unique non-stationary effects caused by the randomness of the barrel manufacturing process. If we can utilize the wavelet and fractal techniques as tools to analyze all the existing bullet LEA signatures and extract useful features, then we would be able to synthesize new signature in a random fashion within predefined parameter distributions. Through observation of matched LEAs in groups, we successfully identified the zero-crossing point length, slopes, and peak-valley features as the deterministic component and fractal features as the random component. In particular, the approximation (a smoothed curve) in the wavelet decomposition is be analyzed to extract the statistical features for base curve generation. The details in the wavelet decomposition are analyzed to extract the fractal parameters such as fractal dimension and magnitude factor. With these parameters, new LEA signatures are constructed in a random fashion.

Based on our intensive study, we found the two-step data synthesis approach can be well justified in a practical bullet firing and firearm examination. No signatures of a bullet LEA have the same pattern of imperfections; however different barrels in the same brand would hold some kind of common (i.e. deterministic) characteristics such as total signature length, and periodicities. Generally speaking, any signature from a bullet LEA have deterministic, as well as random characteristics.

In the framework of LEA data synthesis, we developed different methods to improve the quality of data synthesis. An optimization procedure was also applied to the synthesis process to ensure the matched and non-matched correlation distribution among the new synthesized data set resembles that of the existing data in the same brand. We also extended the data synthesis process from a 1D LEA signature signal to a 2D LEA image, and to even a 3D bullet surface. Our preliminary data synthesis results suggests that newly generated LEAs for a brand shows strong similarity to the LEAs in the same brand, but difference

to those in different brands. Also, the correlation distributions for matched and non-matched LEAs in the same brand suggest the synthesized data approaches the same distribution as the experimental ones.

The ability to generate synthetic land impressions would be of tremendous benefit to the community of tool mark researchers, because it would allow the study of large numbers of samples using a rather reduced number of actual test fires. Another important benefit is the creation of a parametric model of land impressions belonging to bullets of the same brand/model as it is required during the process of generating synthetic land impressions. Such a model could be extremely valuable to address a variety of related problems, such as the potential presence of sub-class characteristics (another challenging question in tool mark identification). For example, if a particular fundamental component is present in a large percentage of land impressions associated with a particular set of bullets fired by guns of the same model/make, then there is a possibility that said fundamental component could be associated with a sub-class characteristic.

Based on the above discussion, the main goals of the current study are as follows:

a) Develop a methodology to generate synthetic land impression (LEA) signatures based on a small set of test fired bullets/LEAs.

b) Develop a methodology to quantify the reproducibility of the features transferred between an individual gun and the bullets it fired.

I.   Introduction

In this section, we will discuss the background information about data synthesis research, which includes background on bullet matching, the motivation of data synthesis, and the strategy that we will use to synthesize data.

1. Statement of the Problem

The automated comparison of bullets requires the generation of LEA signatures. The generation of LEA signatures requires a sequence of steps: data pre-processing, normalization and signature generation, as shown in **Figure 1**. The pre-processing step is responsible for the identification and preliminary handling of unreliable data points (dropouts and outliers). The purpose of normalization is to "flatten out" the curved surface of the LEA to compensate for the curvature of all bullets of the same caliber (which is, in fact, a class characteristic). The normalization process is also designed to compensate for systematic errors during the acquisition process (such as variations in tilt during acquisition). After the normalization process, the signature generation is responsible for isolating the features that are unique to the LEA under consideration (individual features) while rejecting those features which are common to all LEAs of bullets of the same class (class characteristics). After band pass filtering, these normalized LEA cross section become the signatures of the bullet. Details of the signature generation can be found in [1]. As an example, **Figure 2** shows the LEA image on the left and the signature on the right.

Figure 1: LEA signature generation

Figure 2: From LEA image to LEA signature

Before comparing any two signatures, we must align them in the common topology. Signature alignment is an important concept for us to understanding the organization of LEA signatures and provides the basic

methodology for the validation of synthesized data. Signature alignment is only meaningful for the matched LEAs.

This alignment procedure includes two steps [1]. First, we find the common length of the signature signals of a group of matched LEAs with the aim to achieve the maximum correlation between signatures. To elaborate, we shift each LEA signature signal to achieve the best correlation with a reference LEA signature, and then calculate the overlay region for all the signals. Next, there are several LEAs in each bullet and bullets can be compared in a number of relative orientations. **Figure 3** shows the cross section of two bullets to be compared. In both of these cross sections, the LEAs have been labeled. For example, one such orientation is consistent with comparing LEA 1 of bullet 1 with LEA 1 of bullet 2, LEA 2 of bullet 1 with LEA 2 of bullet 2, up to LEA 5 of bullet 1 with LEA 5 of bullet 2. However, if we "rotate" bullet 2 counterclockwise by one LEA, the resulting relative orientation would be consistent with comparing LEA 1 of bullet 1 with LEA 2 of bullet 2, LEA 2 of bullet 1 with LEA 3 of bullet 2, up to LEA 5 of bullet 1 with LEA 1 of bullet 2. In other words, because the pair of bullets under consideration has five rifling grooves, they can be compared in five possible relative orientations. For each of these orientations, a LEA-to-LEA similarity measure is computed according to the number of LEAs present.
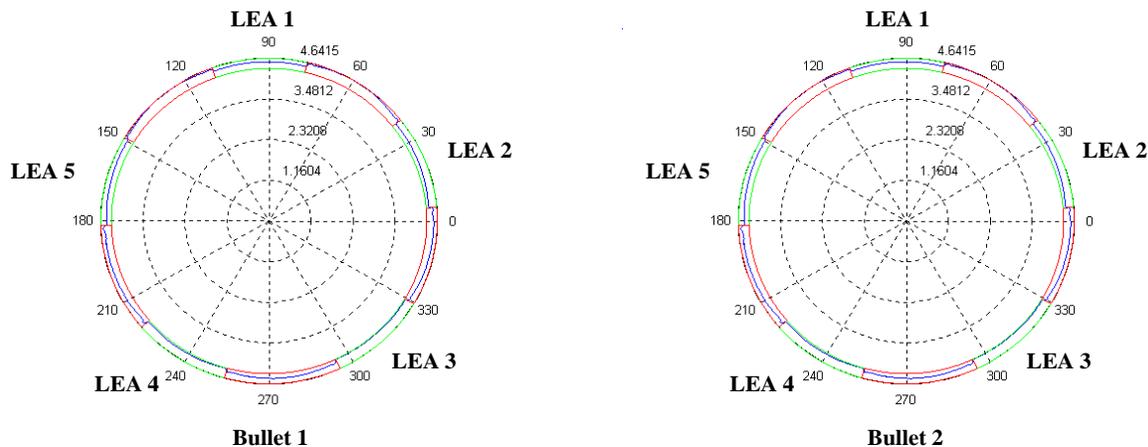
Figure 3: Relative orientation between a pair of bullets

These LEA-to-LEA similarity measures are weight-averaged to compute an orientation similarity measure for each possible orientation. In this manner, if a pair of bullets to be compared has *n* rifling impressions (and therefore *n* possible relative orientations,) *n* similarity measure values corresponding to each of the possible relative orientations are obtained. These orientation similarity measures are ranked to identify the best (highest) and second-best (second highest) orientation similarity measure for the pair of bullets under comparison. The relative orientation rendering the best similarity measure is conjectured to be the relative orientation at which the two bullets under comparison are aligned. The overall similarity measure for the bullet pair under comparison is given by this value.

To present firearm matching evidence in court, we need to answer a fundamental question associated with positive identifications: *what is the probability that two bullets with this level of similarity could originate from two different barrels?* To experimentally quantify and evaluate this probability (FPR: False Positive Rate), it would be necessary to compare a very large numbers of bullets fired by barrels of a population of interest until such an occurrence takes place a few times. This would require a very large database of test fired bullets.

However, it is impractical to fire such a large number of bullets for this purpose. Therefore, it is significantly important to develop a synthetic LEA signature generation methodology for a class of guns based on a set of fired sample bullets, which can be used to generalize the probability of false positive identification estimates based on a small set of data.

Another motivation for bullet data synthesis is to overcome the challenges of developing a good estimate of the similarity measure towards the "tails" of the matching/non-matching distributions (shown in red and blue in Figure 4). Figure 4 illustrates the concept of matching/non-matching distribution of similarity measures resulting from the comparison of pairs of matching/non-matching LEAs fired by guns of different brands. As we can see, for the matching LEAs, the similarity measures are usually higher than those of non-matching LEAs [1][2]. The red and blue region in Figure 4 is precisely the region of greatest interest when we attempt to estimate the probability of false positives and false negatives.

Figure 4: Probability distribution of matching/non-matching bullets

With the synthetic LEA signature generation engine, we can easily generate the LEAs that have similarity measures falling into this overlapping region for further study: 1) randomly generate a synthetic LEA using the synthetic engine; 2) compare the generated LEA with a randomly picked LEA from the fired bullet; and 3) if the similarity measure is within the overlapping region, use this synthetic LEA; otherwise, repeat the whole process.

2. Literature Citations and Review

Firearm and toolmark identification in criminal investigation has been challenged over the years for its subjectivity and difficulty to articulate how identification is made. In 1997, in an attempt to provide a quantitative methodology to describe an observed pattern match based on statistical foundation, Biasotti

---

and Murdock jointly published their conservative quantitative criteria for identification as expressed in terms of Consecutively Matching Striae (CMS) [19], which are expressed as follows:

1. In three-dimensional toolmarks when at least two different groups of at least three consecutive matching striae appear in the same relative position, or one group of six consecutive matching striae are in agreement in an evidence toolmark compared to a test toolmark; and

2. In two-dimensional toolmarks when at least two groups of at least five consecutive matching striae appear in the same relative position, or one group of eight consecutive matching striae are in agreement in an evidence toolmark compared to a test toolmark.

While an innovative and conceptually successful approach, the introduction of the CMS criterion has not completely settled the controversy. Its subjectivity and lack of statistical proof is still being criticized. From the early 1990's, automated identification systems for the comparison of firearms evidence have been developed and used with considerable success. Automated systems are tools that assist the firearms examiner process large numbers of evidence. They rank the bullets stored in the database in light of a certain similarity metric with a subject bullet.

For example, Xie et al. developed an automated bullet-identification system based on surface topography techniques [4]. In this system, bullet surface topography data were acquired for analysis and identification and advanced surface topography analysis, such as surface data segmentation techniques, was applied to extract class characteristics and individual characteristics to improve firearm identification.

Leo´n presented a method by which the comparison of bullets based on gray level images can be automated, thus facilitating the identification of firearms [5]. The acquired gray level images contain special patterns, "fingerprints", which are groove-shaped marks left by the firearm on the cylindrical surface of the bullet. The fine grooves on the bullet surface can be used for bullet comparison. In this approach, a spatial homogenization of the local average gray level is performed followed by a model-based abstraction to project the image intensities of relevant grooves in striations direction. The resulting one-dimensional signals are then filtered by a powerful morphological filter to separate the class characteristics from the individual marks of interest.

Song et al. has also developed a 2D and 3D Topography measurement and correlation system for ballistics signature measurements [20]. This research was based on a joint project from NIST and ATF to establish the National ballistics measurement traceability and quality system using the NIST SRM Bulletsand Casings. This project calls for about 24 ballistics signature acquisitions of both an SRM Bullet and SRM Casing in 13 participating laboratories in the United States and the topography measurement and correlation system was evaluated using this large database.

While automated systems serve an important role, they have not progressed to the point of providing substantive objective evidence that can be presented in court since a universally derived and statistically valid objective criterion has not yet been developed to enable automated systems to partition matching and non-matching bullets. Most of existing studies included a number of barrels of a variety of manufacturing qualities as well as pristine and damaged bullets, and they comprise a minute portion of those found in the street. The limited size of data for development and evaluation makes it challenging to develop high performance and reliable approaches.

The goal of data synthesis is to generalize the existing bullet analysis results to a much larger population so as to overcome the limitation of a false positive identification for firearm examination based on a minute portion of collected bullets. Though there is no published study on bullet data synthesis, the techniques of data synthesis have been applied in several areas such as 3D face modeling [13], fingerprint identification [12], and voice morphing [14].

3. Rationale for the Research

A usual strategy for these data synthesis applications is to identify key features of the underlying signal (image) data that capture common characteristics for the same categories and individual characteristics for different subcategories. The clustering behaviors of those key features will help to generate new features using interpolation and extrapolation approaches. Another strategy of data synthesis is to decompose the underlying signal (image) into low-frequency and high-frequency components and identify features respectively. The low-frequency component usually represents for the contour or frame; while the high-frequency component usually for the details or texture of the original signals (images). The decomposition strategy provides the flexibility of modeling low-frequency and high-frequency parts of signals using different methods.

II. Methods

To determine the probability that two bullets with a given level of similarity could originate from two different barrels, using limited test fired bullets for analysis is far from enough. Based on our intensive study, we developed a two-step approach to generate LEA signature. In this section, we first present an overall architecture of data synthesis, and then discuss the two-step superposition approach of data synthesis based on our observation over data and understanding over physical processes of LEA generation, then discuss the wavelet analysis and fractals, then distribution estimation of key parameters, then the matched and non-matched distribution matching algorithm, and finally LEA image reconstruction.

1. Overall Architecture of Data Synthesis
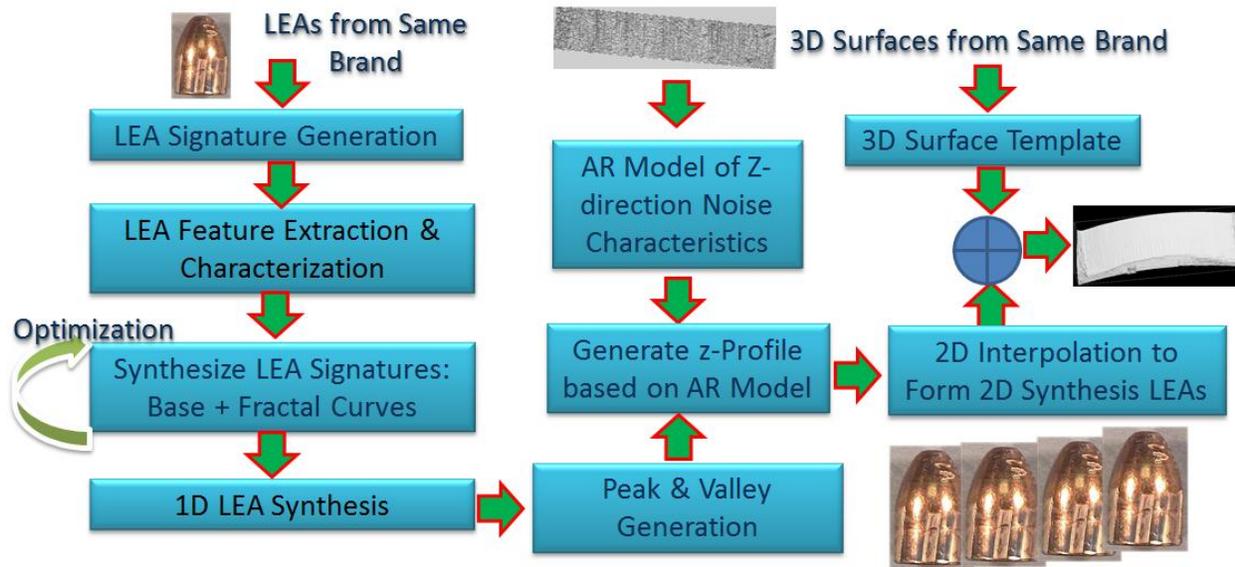
Figure 5: Overall architecture of data synthesis for LEA signatures

The overall architecture of the LEA signature data synthesis is illustrated in **Figure 5**. First, all the LEA images from the same brand are converted to LEA signatures. These signatures are decomposed into approximations and details using wavelet decomposition. The approximation in the wavelet decomposition is analyzed to extract the statistical features; the details in the wavelet decomposition are analyzed to extract the fractal parameters. The peak/valley points, positive/negative slopes, and zero-crossing length and fractal dimension are used to form the parameters of the synthesis model. With these parameters, new LEA signatures are constructed. The LEA signatures converted from the experimental data and those generated from data synthesis engine have been fed into the same validation tool to see whether their distributions match or not. The synthetic 1D signature is then extended to a 2D LEA image using an autoregressive model in which parameters are obtained from existing LEA images. Finally, the LEA image is superposed onto a 3D surface template to generate the 3D LEA surface.

In particular, we adopt a two-step approach with two independent random processes corresponding to the decomposition strategy. The final new signature is the superposition of the signals from these two processes. These two steps are in the following:

STEP 1: *Generate base curve for a new signature. We generate a list of control points such as peak and valley points to define the outline of a signature curve. Based on our observation, the similarity of different signatures is mainly determined by the locations and values of those control points. Every new set of control points generated corresponds to the LEA for a new barrel. We also make perturbations on those control points to emulate the*

signature of the matched LEAs for the same barrel. The final base curve connects these points in a continuous manner.

STEP 2: *Generate the detailed curve for a new signature. We generate the variations on the base curve. These variations are modeled using some random distribution such as Gaussian distribution or some naturally similar process such as fractals.*

STEP 3: *Extend the 1D signature a 2D LEA image using an autoregressive model, and superpose the image onto a 3D surface template to generate the final synthetic 3D LEA surface.*

The three-step data synthesis approach can be well justified in the practical bullet firing and firearm examination. No signatures of bullet LEA have the same pattern of imperfections; however different barrels in the same brand would hold some kind of common (i.e. deterministic) characteristics such as total signature length, and periodicities. Generally speaking, any signature from a bullet LEA will have deterministic, as well as random characteristics.

2. Superposition Approach for LEA Signature Synthesis

To implement the two-step strategy of data synthesis, we developed a superposition approach for LEA signature synthesis. The superposition approach consists of a deterministic component and a random component. The deterministic component includes periodicities in a signature profile, while the random component can be best represented by the fractal model of an irregular curve. In the following sections, we apply multi-resolution techniques to the signature synthesis. By combining the wavelet transformation techniques and fractal estimation and generation approach, a new signature in a known brand can be synthesized based on experiment distributions from the existing bullet signature data.

The wavelet transformation is a multi-resolution tool for signal analysis. After applying wavelet decomposition to a bullet LEA signature, at certain decomposition level, the wavelet approximation part plus all the wavelet detail parts can be used to fully reconstruct the original signal. The wavelet approximation part provides us many useful features regarding the deterministic part of the signature, while the fractal dimension of the wavelet detail part discloses unique non-stationary effects caused by the randomness of the barrel manufacturing process. If we can utilize the wavelet and fractal techniques as tools to analyze all the existing bullet LEA signatures and extract useful features, then we would be able to synthesize new signature in a randomness fashion within predefined parameter distributions.

As illustrated in **Figure 6**, a superposition model is used to synthesize the entire signature. Basically, the deterministic component is implemented in a base curve, and the random component is implemented in a fractal curve. Assuming independence of the two components identified above, a superposition model is defined as follows:

$$y(x) = y_b(x) + y_f(x)$$

Where $y_b(x)$ and $y_f(x)$ are the base and fractal components respectively.
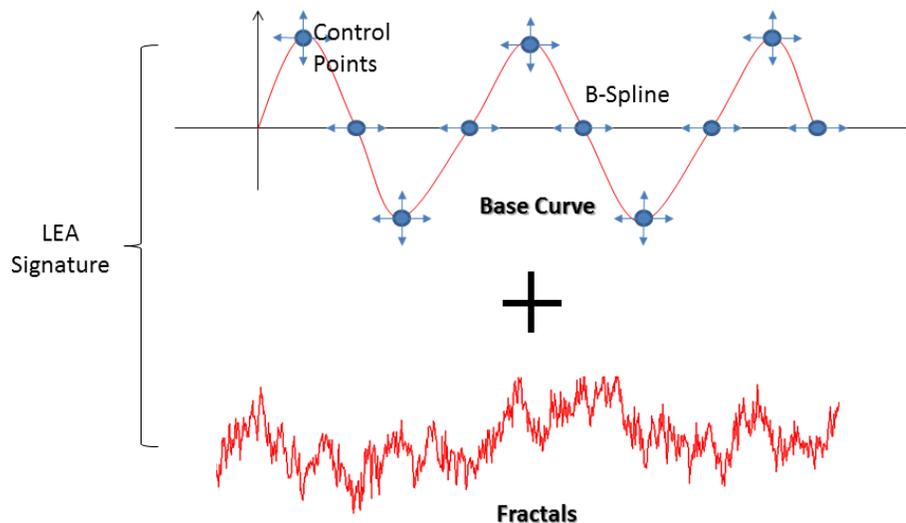
Figure 6: The superposition model for signature synthesis

2.1 Base Curve

The base curve is actually a B-Spline curve, with all the control points generated by a random process, in which is controlled by predefined distributions of deterministic feature parameters. The feature parameters used to generate control points are curve total length, zero-crossing length, positive and negative slopes, and peak and valley points, as shown in **Figure 7**. Note that zero-crossing lengths correspond to peak-valley spacing and the slopes can be derived from the values of peak-valley spacing and peak valley values.
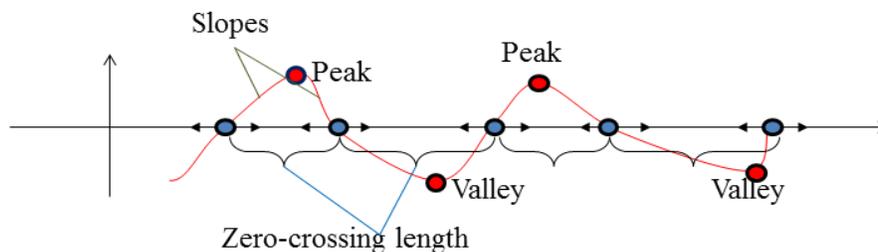
Figure 7: Deterministic features in LEA signature

The base curve generation procedure includes four basic steps. It is illustrated in **Figure 8**.

Algorithm 1:

1) *Randomly generate zero-crossing points using the predetermined distribution until the total curve length is reached.*

2) *Randomly and alternatively generate positive and negative slopes using the predetermined distributions, and use these slopes and zero-crossing points to produce lines, and then the intersection points between two consecutive lines are calculated as control points.*

3) *Randomly perturb zero-crossing points along x axis within a predetermined range.*

4) *Generate B-Spline base curve using the generated points as control points.*
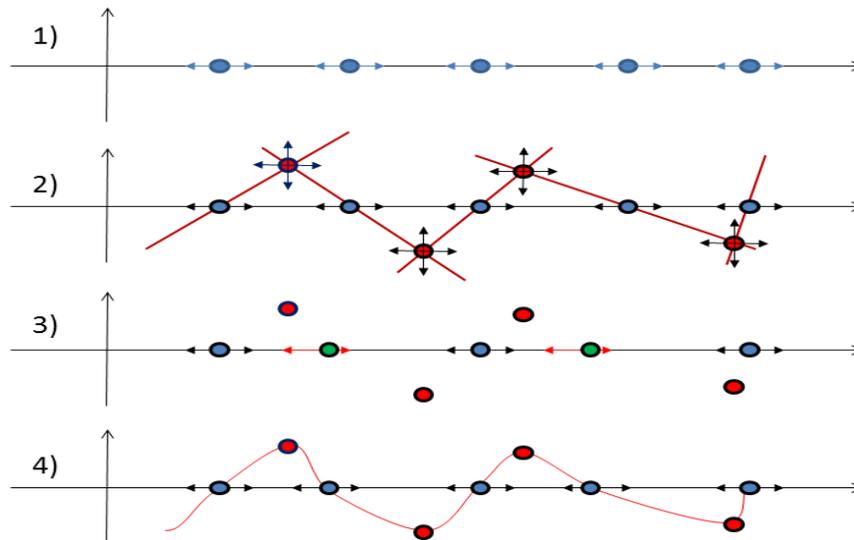
Figure 8: Base curve generation procedures

The generated base curve would certainly follow the predetermined distribution of all feature parameters. These predetermined distributions are calculated from all the existing bullet LEA signature data. The approaches used in calculation will be discussed later. In order to improve the realistic characteristics, more control points can be inserted into the control point list between peak (valley) points and zero-crossing points.

2.2 Fractal Component

After the generation of the deterministic component, the fractal-based method is used to generate the randomness component. The key components in the fractal generation are fractal dimension (or fractal Hurst index) and magnitude factor. The fractal dimension effectively describes the structure resulting from complex processes. As a result, the fractal component takes into account effects that cannot be represented with the deterministic component.

Figure 9 shows several Fractal Gaussian Noise (**fGn**) signals generated using different fractal Hurst indices. For each different barrel brand, we can estimate its fractal dimension distribution, which will be used in signature synthesis. The approach to estimate fractal dimension is based on the idea that the

details from the wavelet analysis are related to the fractal information. The fractal Gaussian Noise generation approach is based on the simulation of fractional Brownian motion. Both approaches will be discussed in the following section.
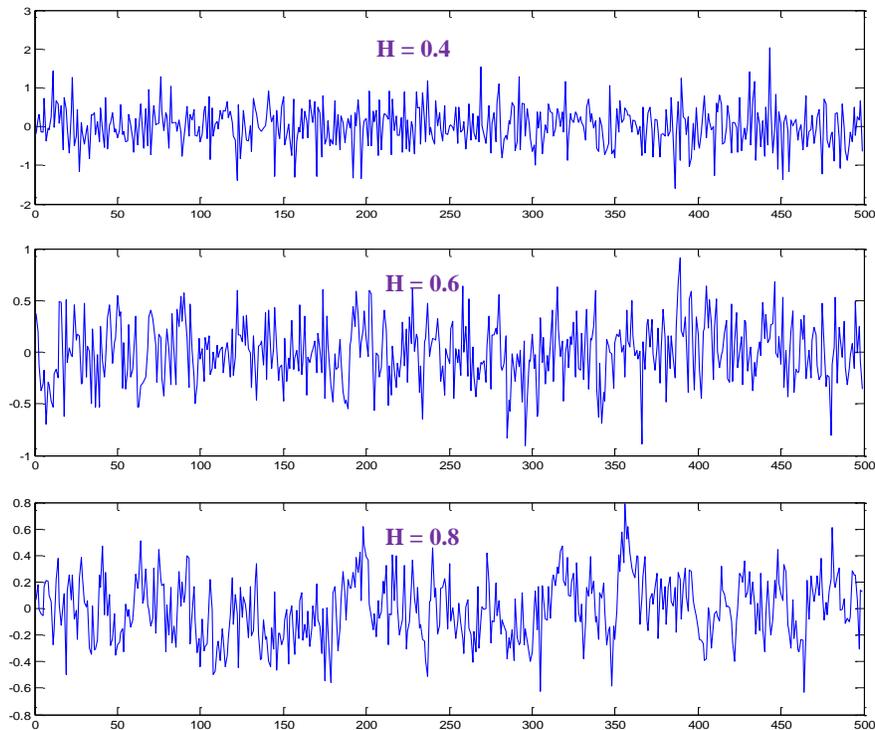
Figure 9: Fractal Gaussian Noise with different Fractal Hurst Index

In summary, using the synthesis method discussed above, a minimum set of measures for any signature synthesis includes the following parameters: *fractal dimension* and *magnitude factor* from the fractal generation method; *signature curve length, zero-crossing length, positive and negative slopes* for the base curve generation. All these parameters will be drawn from their corresponding probability distribution.

2.3 Two Categories of Data Synthesis Problem

By observing all the existing bullet signature data, we knew that there are several barrels in one brand, and each barrel has several fired bullets, and each bullet have six or more LEAs. There exist common base characteristics (by common base characteristics we refer to the characteristics of the base curve as described in **Figure 6**) only for matched LEAs in each barrel. There are no common base characteristics for non-matched LEAs in each barrel and for matched LEAs among different barrels. There are even no common base characteristics for different LEAs in the same bullet. Thus, the data synthesis is scaled down to two categories as shown in **Figure 10** below: 1) same brand/different barrels and non-matched LEAs; 2) same brand/same barrel/matched LEAs.
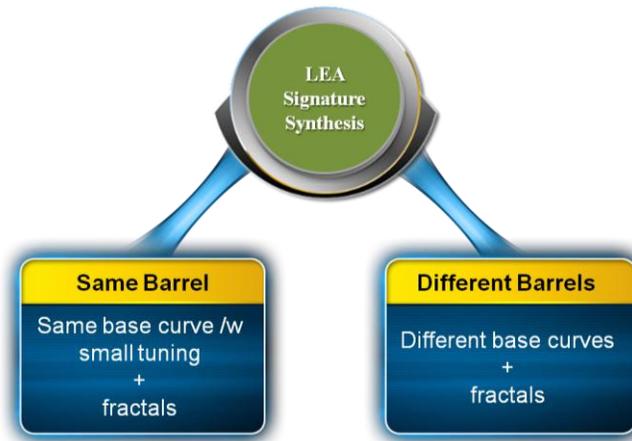
Figure 10: Two categories of LEA signature synthesis

For both categories, the superposition approach using the base curve plus fractal curve remains correct. However, there will be slight differences for different situations. For the first category, the data synthesis can be carried out using the different base curves plus fractals approach; for the second category, the approach is changed to use same base curve with small tuning plus fractals approach.

The second category data synthesis approach is inspired by the experiments results listed in **Figure 11**. We select four matched LEAs of same barrel in the same Beretta brand, as shown in **Figure 11**(a), then extract their wavelet approximation at level 3, and then align them together. Apparently, all base curves share the same basic structure, with only slight differences in magnitude and zero-crossing points.

(a) Matched LEAs   (b) Wavelet Approximation at Level 3
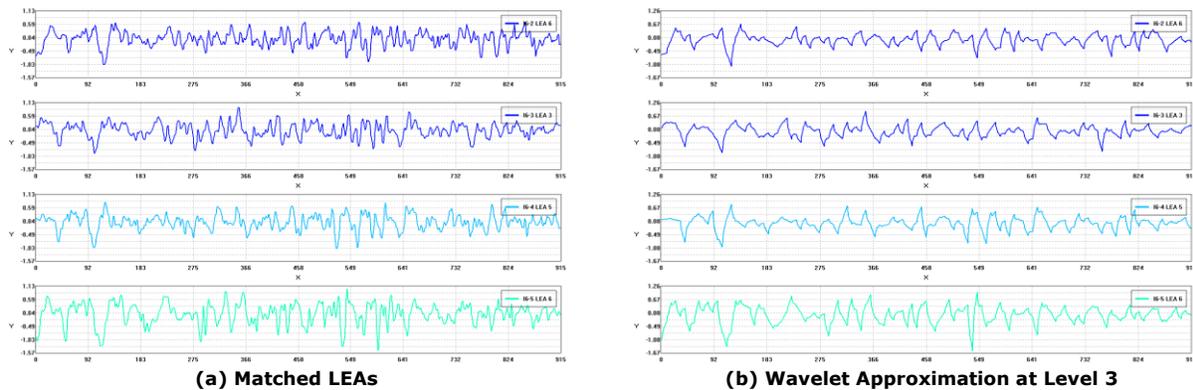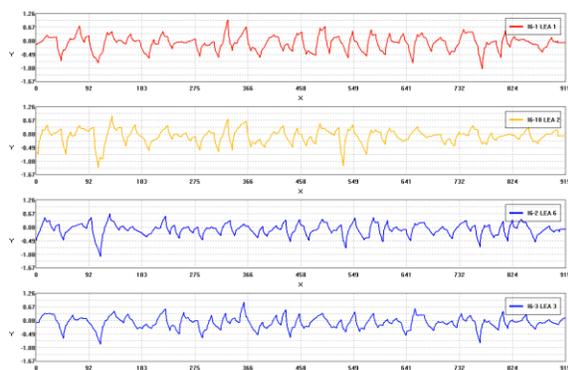
(c) Aligned wavelet approximation   **(d) Overlay each curve**
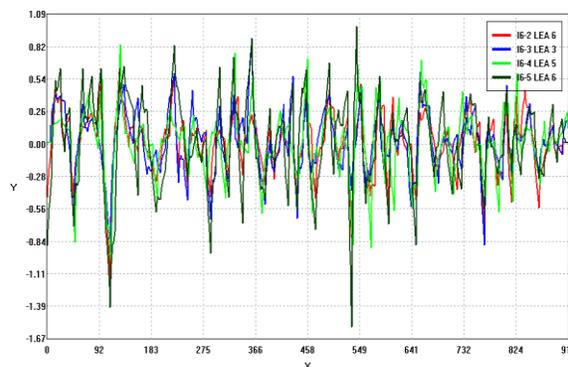
Figure 11: Same barrel LEA signature Experiments

3. Wavelet Analysis and Fractals

The wavelet-based analysis procedure can be used to decompose complex signals into multi-scale approximations and details. The corresponding reconstruction procedure preserves and produces the original data exactly. Following the above decomposition of the structural information in the bullet LEA signature, the next step is to identify and isolate each component, using mathematical tools, though the mathematical model for the fractal component utilizes the concepts of fractal dimensions and wavelet.

3.1 Wavelet Analysis

The fundamental idea behind wavelet analysis is to analyze data at different scales or resolutions using mother wavelets. The mother wavelets are also called wavelet prototype functions, which are functions that satisfy certain mathematical requirements and are used in representing data or other functions. In wavelet analysis, temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients.

Mathematically, it is helpful to think of these wavelet coefficients as a filter. In implementing the discrete wavelet transformation (DWT), these coefficients are placed in a transformation matrix, which is applied to a raw data vector in a hierarchical algorithm. The wavelet coefficients are arranged using two dominant patterns, one that works as a smoothing filter to preserve concrete "approximation" information, and another pattern that works to bring out the data's "detail" information. The approximation is the minimum information needed to represent a signal, while the detail is the additional information needed to exactly reconstruct a signal. The matrix is first applied to the original, full-length vector. Then the vector is

smoothed and decimated by half and the matrix is applied again. Then the smoothed, halved vector is smoothed, and halved again, and the matrix applied once more. This process continues until a trivial number of "smooth-smooth- smooth..." data remain. That is, each matrix application brings out a higher resolution of the data while at the same time smoothing the remaining data. The output of the DWT consists of the remaining "approximation" components, and all of the accumulated "detail" components, as represented in the equation below:

$$s = a_n + \sum_{i=1}^{n} d_i$$

where s is the original signal, a_n is the approximation component at level n, and d_i is the detail component at level i.
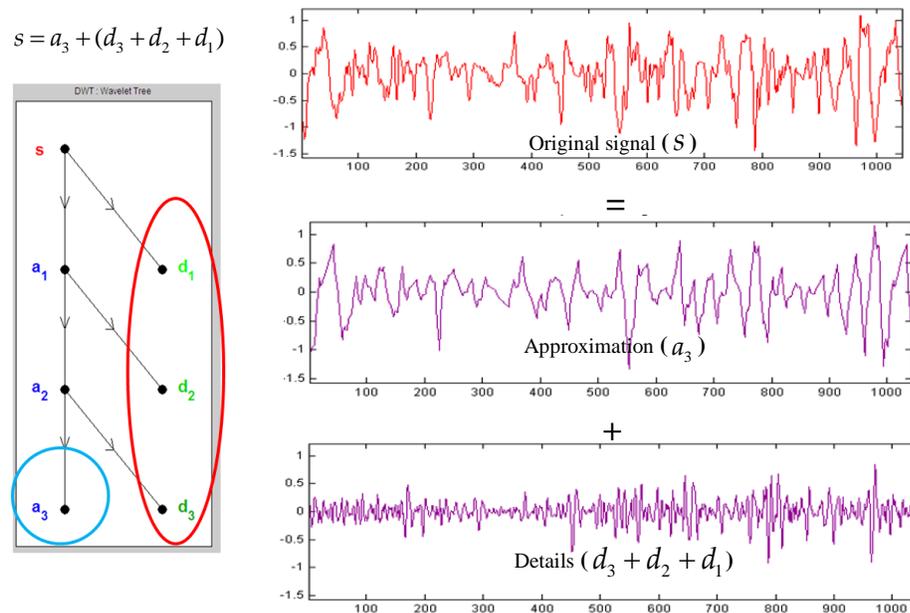
Figure 12: Wavelet decomposition

Figure 12 illustrates the wavelet transformation procedure using '"db2" as mother wavelet and 3 levels. For all data analysis using DWT herein, we will use a_3 as the "bone" data for the distribution estimation of feature parameters for base curve generation, and use $d_3 + d_2 + d_1$ for the distribution estimation of fractal dimension and magnitude scale for fractal curve.

3.2 Fractals

Mandelbrot used the concept of the Hausdorff dimension to define a fractal as an object whose Hausdorff dimension is greater than its topological dimension [10]. Physically, a fractal is "a rough or fragmented geometric shape that can be split into parts, each of which is (at least approximately) a reduced-size copy of the whole", a property called self-similarity. Several natural objects exhibit fractal characteristics, e.g.,

coastlines, clouds, and geological formations. There has been an unprecedented surge of applications for fractals in widely diverse fields, ranging from economic data analysis through analysis of lightning patterns, to structural studies of human lungs. One of the reasons for the wide-ranging applications of fractals is that they abstract the essence of the interplay between the geometry and the dynamics underlying the problem. This central idea is the motivation to use fractals in this research, i.e., to simulate the randomness characteristics of bullet LEA signature caused by imperfections of barrel manufacturing processes, and to relate it to the structure of the differences between LEA signature and deterministic base curve, abstracting it in terms of fractal-based parameters.

3.2.1 Fractal Dimension Estimation

A number of techniques exist to estimate the fractal dimension of signals, such as 1) The Yardstick method, 2) Box Counting method and 3) Spectral method. All these three are found in Mandelbrot's premier reference on the subject.

Both fractals and wavelets share the properties of non-stationary and self-similarity. A number of researchers have examined this link. It indicates that wavelets can be used to estimate fractal or spectral parameters [11]. The variance of wavelet coefficients within each scale of the DWT is related to the scale,

$$\sigma^2(d_{*,j}) \approx V_0(2^j)^\beta$$

Where $d_{i,j}$ is the *i*-th detail wavelet coefficient in the *j*-th level, thus $d_{*,j}$ are all the detail wavelet coefficients in the *j*-th level, and

$$\sigma^2(d_{*,j}) = \frac{1}{N-1}\sum_k (d_{k,j} - \bar{d}_{*,j})^2$$

β is the spectral exponent, a function of the fractal dimension D_f,

$$D_f = \frac{3 - \beta}{2}$$

The parameter V_0 is a constant, which plays a crucial role in determining the magnitude of variance. With this equation, the fractal dimension is calculated from the slope of the log-log plot of the variance versus the scale 2^j, i.e.

$$\log_2(\sigma^2(d_{*,j})) = c + j\beta + \varepsilon_j$$

Where $c = \log_2^{V_0}$ and ε_j is the error term which is negligible.

3.2.2 Fractional Brownian motion and Fractional Gaussian Noise

The above method describes the analysis of signature, wherein the fractal dimension is estimated from a given data set. We now need a method to synthesize signal with the computed fractal dimension. Mathematical fractals can usually be synthesized by an initiator and a generator, along with a set of recursive rules. However, physical fractals cannot be generated by such deterministic methods, and must be based on some underlying model. The trajectory of a particle undergoing Brownian motion is a canonical model for one-dimensional objects, and the most general model for Brownian motion is the fractional Brownian motion (fBm).

The fBm usually considers a dependent-variable trace as a function of time t. For one-dimensional signal, we introduce a spatial variable x. The fBm trace $y_H(x)$ is defined by:

$$y_H(x) = \frac{1}{\Gamma(H+0.5)}[\int_{-\infty}^{0}(|x-s|^{H-0.5} - |s|^{H-0.5})dy(s) + \int_{0}^{x}(|x-s|^{H-0.5})dy(s)]$$

Where $\Gamma(\cdot)$ is the gamma function, $y(x)$ is ordinary Brownian motion, and $H \in [0,1]$ is the **Hurst parameter,** which is related to the **fractal dimension** as:

$$H = 2 - D_f$$

An increment in $y_H(x)$ is related to the change in x by:

$$\Delta y_H(x) \propto \Delta x^H$$

Where H = 0.5 corresponds to the classical Brownian motion trace. A special property of fBm is that the increments $\Delta y_H(x)$ are stationary, while $y_H(x)$ is non-stationary. This equation represents a spatial relationship between different points in the signal, and provides one basis for synthesizing a signal with a given fractal dimension. The sequence of the increments $\Delta y_H(x)$ can be directly used to generate fGn from fBm.

Based on the above insight, we propose to use fractional Gaussian noise (fGn) to simulate our fractal signal. The point of interest is its power spectrum, which presents a different relationship between the spectral exponent and the Hurst parameter:

$$\beta = 2H - 1$$

The corresponding fractal dimension is calculated as:

$$D_f = \frac{3 - \beta}{2}$$

Using different H value, different fractal signal can be generated as shown in Figure 9.

4. Distribution Estimation of Key Parameters

In order to synthesize signature data, our superposition algorithm combining base curves and fractals discussed in Section 2 needs various parameter distributions. For base curve generation, the distributions of zero-crossing length and slopes are needed. For fractal curve generation, the Fractal Hurst index distributions for each specific barrel brand are required. As mentioned in Section 3.1, wavelet decomposition, as a crucial tool, will help us split any existing bullet LEA signature into two components: one is smoothed base data, and another is the detail data. The base data will be used in estimating parameter distribution for base curve generation, and the detail data is used to estimate fractal dimensions. In this section, we will talk about how to estimate these distributions of key parameters.

4.1 Distribution Estimation for Base Curve

In order to use the features of base curves for data synthesis, we need to calculate these features and model their distribution. Based on the synthesis approach for base curves, we need to calculate the features of zero-crossing lengths and slopes.

The calculation of zero-crossing lengths is based on the results of zero-crossing point detection. By looping through each point in the signature vector, we can easily find all the cases where a line segment connected by two consecutive points, i.e. zero-crossing points, is intersected with x axis. Let's say the coordinates of these two points are (x_0, y_0) and (x_1, y_1), then the coordinate of the zero-crossing point would be easily calculated by,

$$\begin{cases} x = (x_1 y_0 - x_0 y_1)/(x_1 - x_0) \\ y = 0 \end{cases}$$

The calculation of slopes is relatively challenging compared to the zero-crossing point detection. In order to calculate slopes, we need to firstly detect all the peak and valley values along the whole signal. The peak point is defined as local maxima and the valley point is defined as local minima. It is so obvious for human to detect peaks and valleys, but it is a challenging task for a computer. Due to the noise in real-life signals, the well-known zero-derivate method might not be always right. Accidental zero-crossings of the first derivate often occur, yielding false detections. The typical solution is to smooth the curve with some low-pass filter, usually killing the original signal at the same time.

The trick here is to realize, that a peak is the highest point between "valleys". What makes a peak is the fact that there are lower points around it. This strategy is adopted in our peak detection algorithm.

(a) Find all the peaks and valleys

(b) Remove peaks < 0 and valleys > 0

(c)Keep peak and valley point alternatively
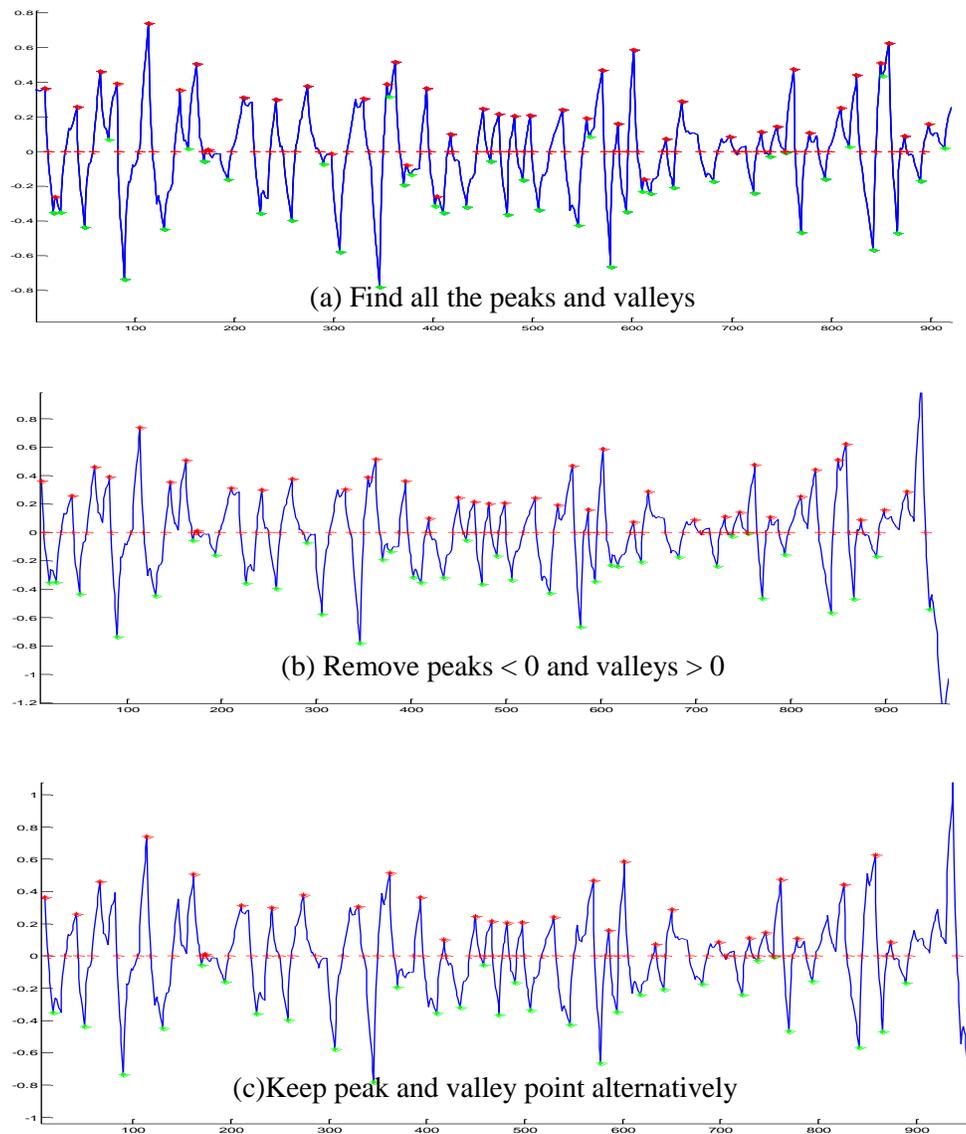
Figure 13: Zero-crossing point and peak point detection

The slope detection procedure for a wavelet approximation of a bullet LEA signature is shown in **Figure 13**. In step (a), we find the entire peak and valley points using the approach mentioned above, and then in step (b), we remove all the invalid peak and valley points by requiring that all the $y_{peak} > 0$ and all $y_{valley} < 0$. In the final step (c), we remove excessive peak and valley points along the zero x-axis to make sure only one peak point between two consecutive valley points, vice versa.

The approaches discussed above will be applied to all the LEA signature data under each brand. Table 1 and Table 2 show the statistical information of the base curve features for all six brands. They clearly show apparent differences among all the brands.

Table 1: Zero-crossing length distribution

Brand	Zero –crossing length distribution	
	Mean	Standard Deviation
Beretta	15.916255	9.146383
Browning	16.742649	10.195792
Bryco	19.494606	12.429748
HiPoint	18.812483	11.823937
Sig	15.518603	9.440852
Taurus	15.310292	8.194318

Table 2: Slope distribution

Brand	Slope distribution			
	Mean	Standard Deviation	Max	Min
Beretta	0.060840	0.055054	1.196841	0.000114
	-0.060144	0.051165	-0.000156	-0.786151
Browning	0.020968	0.019076	0.430644	0.000282
	-0.020741	0.017866	-0.000401	-0.228075
Bryco	0.030061	0.039811	0.557924	0.000641
	-0.032473	0.047814	-0.001198	-0.559317
HiPoint	0.027204	0.046012	1.183929	0.000056
	-0.026589	0.037446	-0.000204	-0.798353
Sig	0.018618	0.019378	0.523168	0.000139
	-0.019315	0.025546	-0.001032	-0.959769
Taurus	0.034296	0.031336	0.286915	0.000141
	-0.034151	0.031230	-0.000988	-0.451545

We have also calculated the base feature distributions for different barrels of one brand named Beretta. The results are listed in Table 3 and Table 4. From the results, we cannot clearly distinguish from barrel to barrel. Such observation provided valid evidence that the base features are suitable representative features for each brand.

Table 3: Zero-crossing length distribution for different barrels

Beretta	Zero-crossing length distribution	
	Mean	Standard Deviation
Barrel 1	15.534972	8.947869
Barrel 2	16.076855	9.000587
Barrel 3	15.848721	9.255147
Barrel 4	15.336478	8.507813
Barrel 5	15.726011	9.133241
Barrel 6	15.984325	9.238554
Barrel 7	15.985072	9.214276
Barrel 8	16.588253	9.786658
Barrel 9	16.672823	9.518230
Barrel 10	15.745819	9.037264
Barrel 11	15.678578	8.893054

Table 4: Slope distribution for different barrels

Beretta	Slope distribution			
	Mean	Standard Deviation	Max	Min
Barrel 1	0.055483	0.051735	0.548080	0.000496
	-0.055147	0.053597	-0.000508	-0.786151
Barrel 2	0.054941	0.043017	0.329780	0.000114
	-0.055081	0.043300	-0.002243	-0.378107
Barrel 3	0.056433	0.049249	0.430410	0.000845
	-0.055403	0.046699	-0.001404	-0.574601
Barrel 4	0.070504	0.053330	0.515149	0.002064
	-0.071660	0.054221	-0.004677	-0.461111
Barrel 5	0.067888	0.058658	0.556391	0.001589
	-0.65499	0.056377	-0.000156	-0.587961
Barrel 6	0.060948	0.043325	0.327225	0.002876
	-0.061767	0.050772	-0.001386	-0.472183
Barrel 7	0.059711	0.050809	0.646553	0.001860
	-0.058853	0.046611	-0.000878	-0.433223

Accordingly, we accumulate all computed feature data to easily plot its distribution as shown in Figure 14 and Figure 15.


(a) Beretta   (b)Browning   (c)Bryco

(d)HiPoint   (e)Sig   (f)Taurus
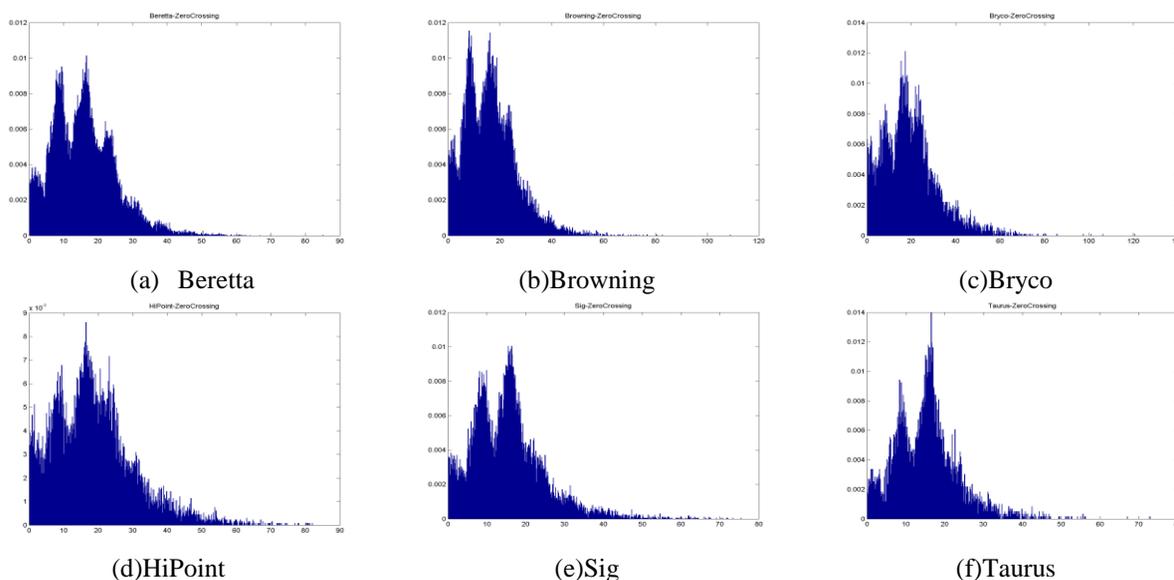
Figure 14: Zero-crossing length distribution

(a) Beretta   **(b) Browning**

(c) Bryco

(d) HiPoint

(e) Sig

(f) Taurus
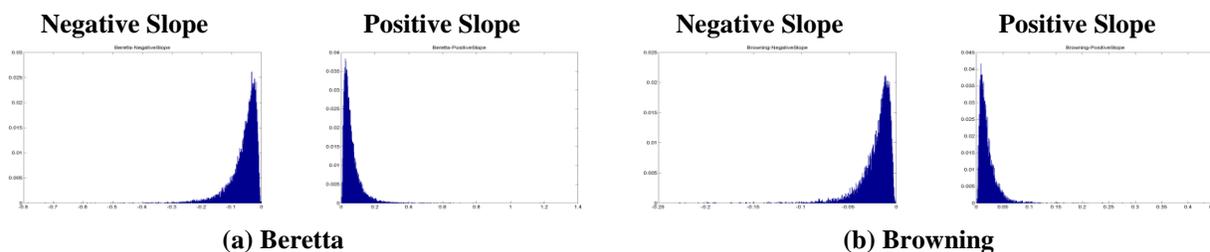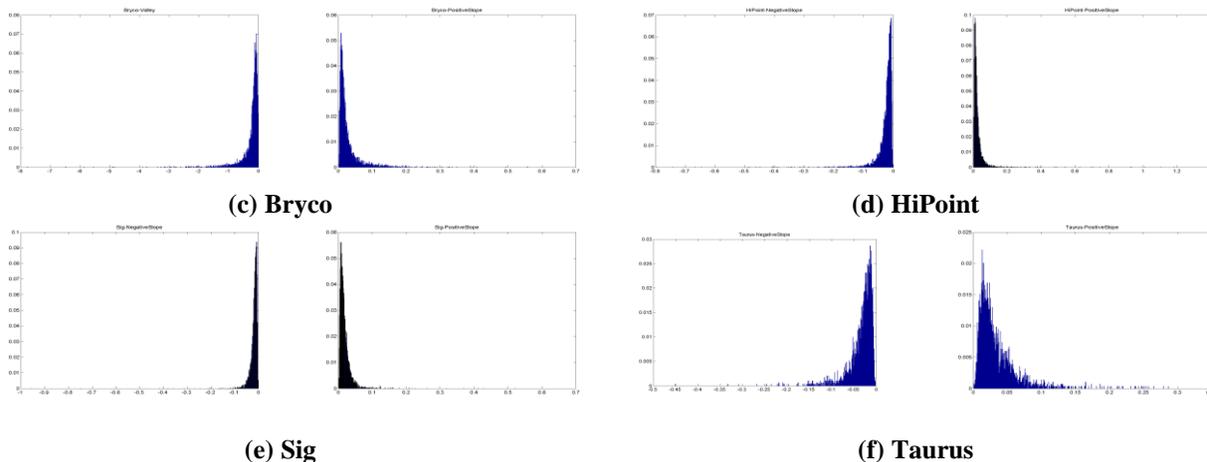
Figure 15: Slope (peak-valley) distribution

4.2 Distribution Estimation of Fractal Dimension

Using the approach described in Section 3.2, we estimated the distribution of fractal Hurst Index for all barrel brands. The results are listed in Table 5. Apparently, the fractal dimension results differentiate each brand very well.

Table 5: Fractal Hurst Index distribution

Brand	Fractal Parameter Distribution			
	Fractal Hurst Index		Magnitude Factor	
	Mean	Standard Deviation	Mean	Standard Deviation
Beretta	0.655617	0.081033	0.009911	0.004480
Browning	0.507806	0.075185	0.003578	0.002007
Bryco	0.624133	0.089258	0.001462	0.001836
HiPoint	0.604183	0.094664	0.001575	0.001105
Sig	0.553442	0.084182	0.002289	0.001656
Taurus	0.696786	0.082796	0.002122	0.000962

Accordingly, we accumulate all computed data to easily plot its distribution as shown in Figure 16.

(b) Beretta

(b)Browning

(c)Bryco

(d)HiPoint   (e)Sig   (f)Taurus
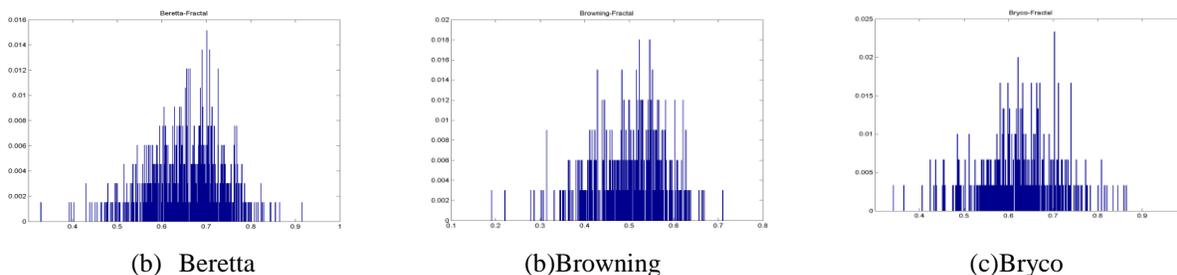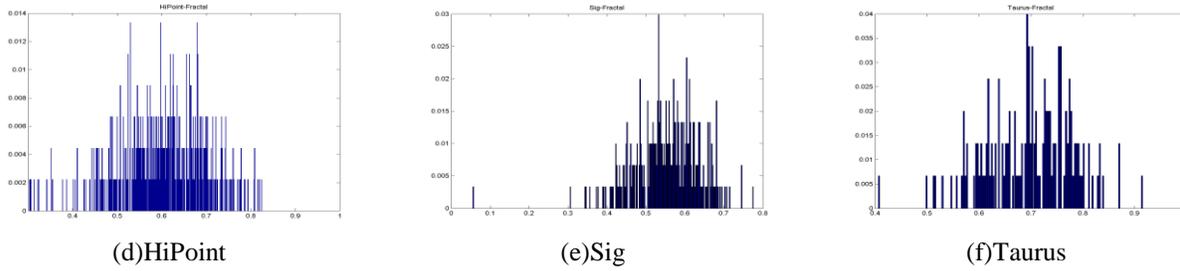
Figure 16: Fractal Hurst Index distribution

4.3 Model Feature Distribution

According to our observation over various distributions of these features, normal distributions oversimplified statistic information of the LEA signatures. Therefore, we developed a new approach based on the fitness proportionate selection to model the distribution. The fitness proportionate selection, also known as Roulette-wheel selection [15], is a genetic operator used in genetic algorithms for selecting potentially useful solutions for recombination. Similar to a Roulette wheel in a casino, a proportion of the wheel is assigned to each of the possible selection based on their fitness values. This could be achieved by dividing the fitness of a selection by the total fitness of all the selections, thereby normalizing them to 1. Then a random selection is made similar to how the roulette wheel is rotated. The concept is illustrated in Figure 17.
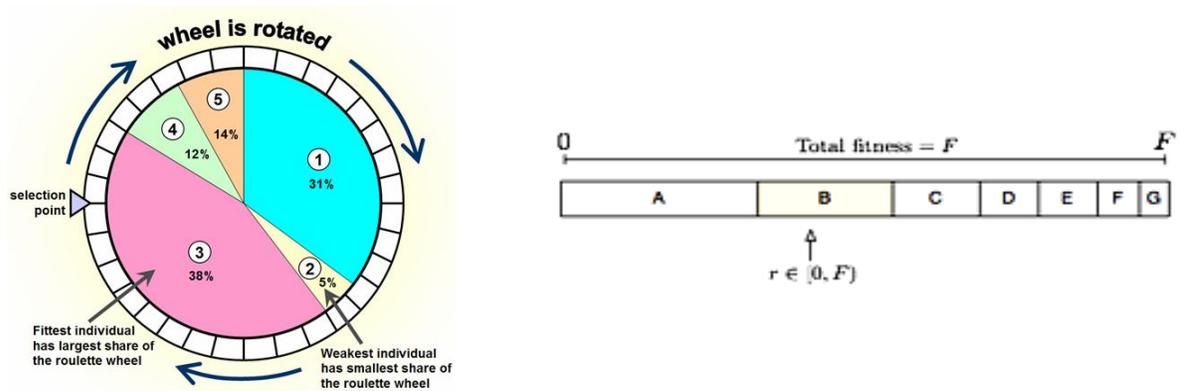
Figure 17: Roulette-wheel selection

Specifically, the fitness level is used to associate a probability of selection with each individual element. If f_i is the fitness of individual i in the population, its probability of being selected is

$$p_i = f_i / (\sum_{j=1}^{N} f_j)$$

where N is the number of individuals in the population.

The following steps were developed based on the concept described above to randomly generate a number that complies with the underline distribution shown in Figure 14 ~ Figure 16.

Algorithm 2:

1) *Use 256 bin to approximate the whole distribution (the bin number is determined by experiments)*
2) *Each bin contains a fitness score which is the percentage of possibility*
3) *Align all the bin into one vector*
4) *Generate a random number in [0,1] with uniform distribution*
5) *Translate this random number to the corresponding member in the vector such that the vector member is the first in the row that makes the total score larger than the number*
6) *Use the center value of the selected bin as the output*

We applied the procedure to simulate all the statistical features used in the LEA signature synthesis process. An example is shown in Figure 18. The left one is the fractal index distribution of Beretta, and the right one is the randomly generated distribution (1000 times). We can see that the generated signal is very similar to the original signal from the observation.

(a)Original Signal   (b)Randomly generated (1000 times)
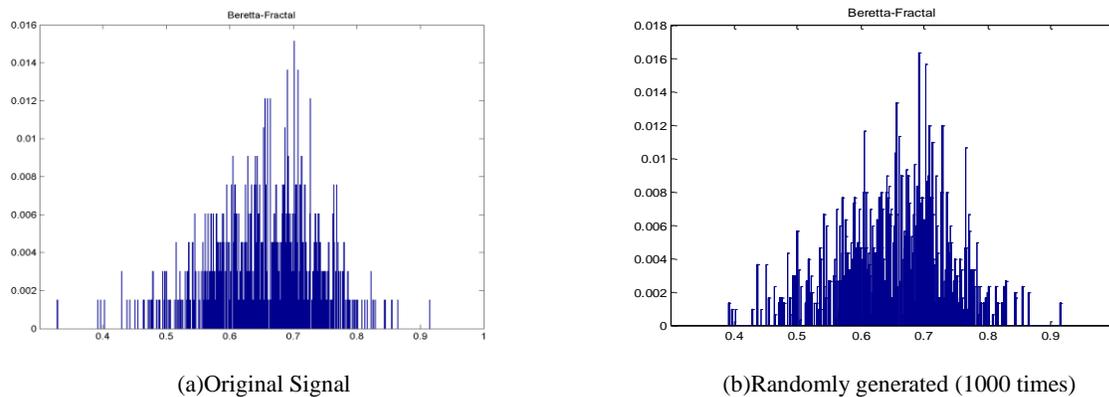
Figure 18: Feature simulation

5. Correlation Distribution Matching

To compare the correlation distribution for the synthesized data and experimental data, we generate the same amount of data as the amount for the brand Beretta. The synthesized data include *11* barrels, *10* bullets for each barrel, and *6* LEAs for each bullet. We fed all these data to the correlation analysis tool we developed in previous study to observe the distribution. **Figure 19** shows the side-by-side comparison of the probability distribution for the matched and non-matched LEAs respectively. The left chart is the correlation distribution for synthesized data, while the right chart is the correlation distribution for experimental data. We observed that the non-matched part is similar to each other, but the matched part deviate to some extent.
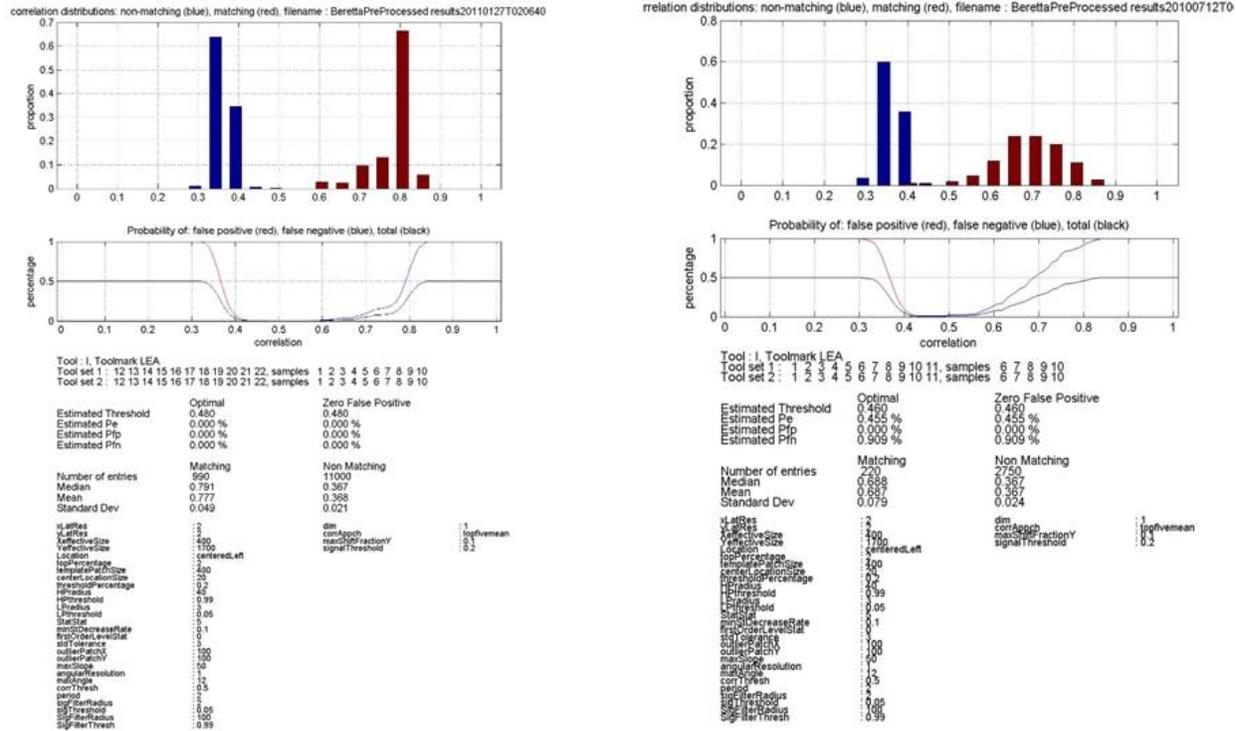
Figure 19: Comparison of correlation distribution for brand Berretta

In the process of LEA signature data synthesis, in order to match the correlation distribution (matched or non-matched) of the synthesized data with that of the original data, we developed an optimization procedure to optimize the data synthesis process. The cost function of the optimization is defined as a similarity measurement between the synthesized distribution and the original distribution. The bigger the similarity is, the better the matching is.

5.1 Preparation of the Target Distribution

The bullet-to-bullet similarity measure is defined as the maximum value of all LEA-to-LEA similarity measure or the mean of the LEA-to-LEA similarity measures for each possible pair of LEAs for the pair of bullets under comparison.

$$Sim(bullet_a, bullet_b) = \max(Sim(LEA_a^i, LEA_b^i)) \quad or \quad \frac{\sum_{i=1}^{N} Sim(LEA_a^i, LEA_b^i)}{N}$$

The LEA-to-LEA similarity measure is computed according to the following equation:

$$Sim(a,b) = \max_{|\Delta x| < \Delta x_{max}} \left[1 - \frac{\left\| l_1(x + \Delta x) - l_s(x) \right\|^2}{\left\| l_1(x + \Delta x) + l_s(x) \right\|^2} \right]$$

where l_1 and l_s correspond to the zero-mean one-dimensional signatures associated with the LEAs under comparison, and the norm $\|\bullet\|$ corresponds to the Euclidean norm:

$$\|l\| = \sqrt{\sum l_i^2}$$

and Δx_{mx} is a maximum amount of lateral displacement allowed for comparison. The maximum correlation is found by displacing (shifting) one dataset with respect to the other by Δx. We refer to this similarity metric as a "relative distance metric". The relative distance metric is a time-domain similarity metric, and it offers advantages in terms of being suited to deal with signatures of different lengths, as well as signatures with missing data points (dropped points, outliers, etc).

The matching distribution is defined as the similarity measure distribution among all matched bullets, and the non-matching distribution is defined as the similarity measure distribution among all non-matched bullets. For instance, the brand X has 10 barrels, and each barrel has about 6 bullets. For one-to-one comparison, we have $(60 \times 59)/2$ possible similarity measures. Among them, the matching measures are $[(6 \times 5)/2] \times 10$, and the rest is the non-matching measures.

We pre-calculated the matching and non-matching correlation distribution among all existing bullet LEA signature data as shown in Figure 20. The data is listed in **Table 6**. The distribution is represented as a 20-bin histogram. In the plot, the x-axis is the similarity score, and the y-axis is the percentage probability.

Brand	Non-matching Distribution	Matching Distribution
Beretta		
Browning		
Bryco		
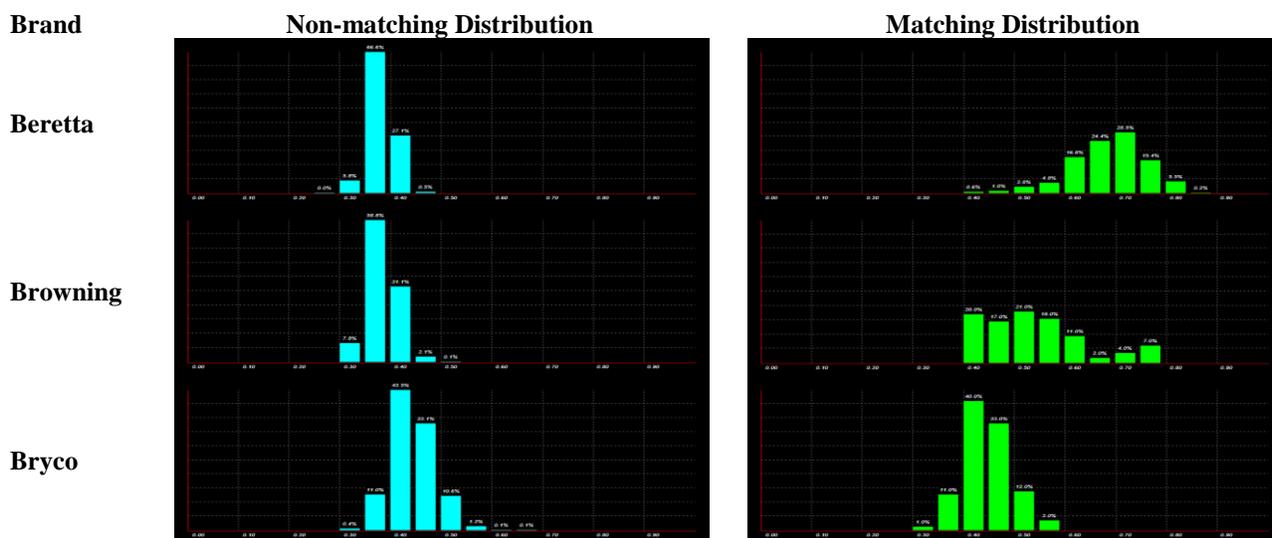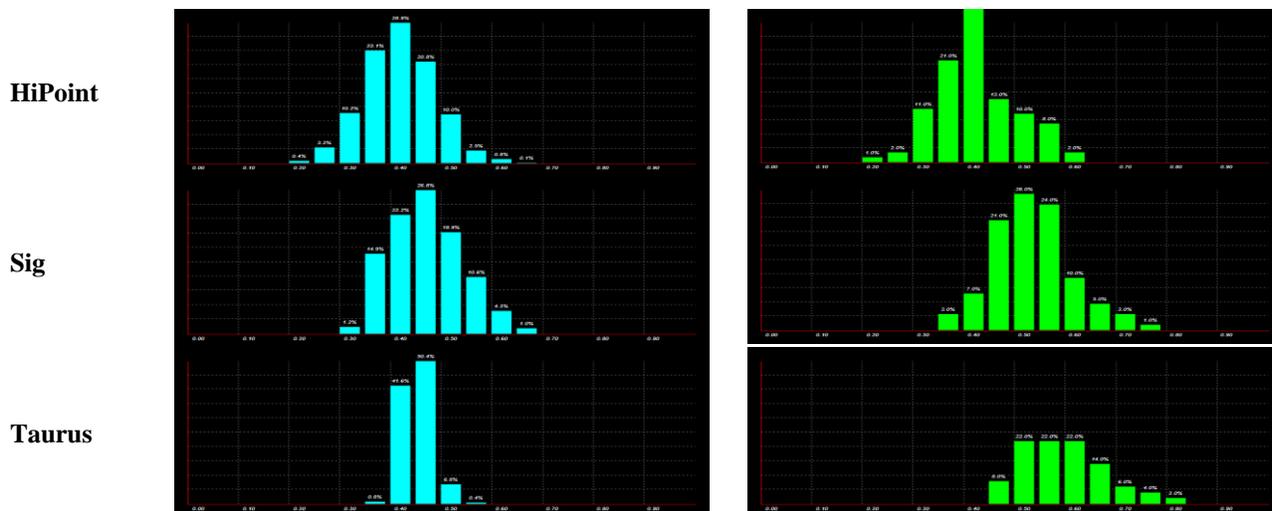
Figure 20: Bullet-based correlation histogram for all brands

Table 6: Bullet-based histogram for all brands

Label	Beretta		Browning		Bryco		HiPoint		Sig		Taurus	
	M	Non	M	Non	M	Non	M	Non	M	Non	M	Non
0.00	0	0	0	0	0	0	0	0	0	0	0	0
0.05	0	0	0	0	0	0	0	0	0	0	0	0
0.10	0	0	0	0	0	0	0	0	0	0	0	0
0.15	0	0	0	0	0	0	0	0	0	0	0	0
0.20	0	0	0	0	0	0	0.01	0.004	0	0	0	0
0.25	0	0.001	0	0	0	0	0.02	0.032	0	0	0	0
0.30	0	0.064	0	0.078	0.01	0.004	0.11	0.102	0	0.012	0	0
0.35	0	0.645	0	0.588	0.11	0.11	0.21	0.231	0.03	0.149	0	0.008
0.40	0.009	0.284	0.2	0.311	0.4	0.435	0.32	0.289	0.07	0.222	0	0.416
0.45	0	0.006	0.17	0.021	0.33	0.331	0.13	0.208	0.21	0.268	0.08	0.504
0.50	0.027	0	0.21	0.001	0.12	0.106	0.1	0.1	0.26	0.189	0.22	0.068
0.55	0.109	0	0.18	0	0.03	0.012	0.08	0.025	0.24	0.106	0.22	0.004
0.60	0.164	0	0.11	0	0	0.001	0.02	0.008	0.1	0.043	0.22	0
0.65	0.236	0	0.02	0	0	0.001	0	0.001	0.05	0.01	0.14	0
0.70	0.255	0	0.04	0	0	0	0	0	0.03	0	0.06	0
0.75	0.118	0	0.07	0	0	0	0	0	0.01	0	0.04	0
0.80	0.082	0	0	0	0	0	0	0	0	0	0.02	0
0.85	0	0	0	0	0	0	0	0	0	0	0	0
0.90	0	0	0	0	0	0	0	0	0	0	0	0
0.95	0	0	0	0	0	0	0	0	0	0	0	0

5.2 Definition of Cost Function

As described before, a cost function to measure the similarity between the distribution of the synthesized data and the target distribution is needed for the optimization process. The cost function is defined in the following equation:

$$f(H_0, H_1) = 1.0 - \frac{\sum_{i=1}^{n}[(H_0^i - \overline{H}_0) - (H_1^i - \overline{H}_1)]^2}{\sum_{i=1}^{n}[(H_0^i - \overline{H}_0) + (H_1^i - \overline{H}_1)]^2}$$

where H_0 is the target distribution, H_1 is the synthesized distribution, and $\overline{H}_*$ is the estimation of the corresponding distributions. Since we use a 20-bin histogram to represent each distribution, here H_i is a 20-element vector.

Once a fixed amount of LEA signature data is generated, the correlation distribution histogram can be easily calculated based on the similarity score equation listed above.

Another distance measure could also be used in the cost function. The measure is also called Tanimoto coefficient, which was proposed by Jaccard [16] as in the equation below:

$$f_T(H_0, H_1) = \frac{H_0 \cdot H_1}{\|H_0\|^2 + \|H_1\|^2 - (H_0 \cdot H_1)}$$

The well-known Mahalanobis distance is also a good distance measure among two vectors. In our case, suppose we have two distribution H_0 and H_1 with means μ_0 and μ_1, the Mahalanobis distance is given by the following formula

$$D^2 = (\mu_2 - \mu_1)^T \Sigma^{-1} (\mu_2 - \mu_1)$$

where Σ denotes the common (nonsingular) covariance matrix in each group H_0 and H_1. It can be estimated by the pooled estimate as,

$$\Sigma = (\Sigma_1 + \Sigma_2)/2$$

And the sample covariance matrix is calculated by,

$$\Sigma_i = \frac{1}{n-1} \sum_{j=1}^{n} (H_i^j - \mu_i)(H_i^j - \mu_i)^T$$

5.3 Control of the surface roughness of the synthesized LEA data

In the optimization process, we need to find a way to control the surface roughness of the synthesized LEA signature so that its distribution would move towards the direction that could potentially increase the matching score. As described before, for each matched LEA signature, it is generated by using the same base curve (B-spline curve with control points) with small tuning. The base curve is tuned with a certain level of random noise. The noise level is increased to reduce the similarity, and the noise level is decreased to increase the similarity among matched LEAs. Therefore, the noise level becomes the control input to the signature generation process, which is determined by the histogram mean difference between the synthesized distribution and the target distribution. For non-matched LEA signature, it is synthesized by using different base curve, as discussed in section 2.3. We can still control the base curve roughness

through the adjustment of key statistical features using the histogram mean difference between the synthesized non-matching distribution and the target non-matching distribution. The concept is illustrated in Figure 21.
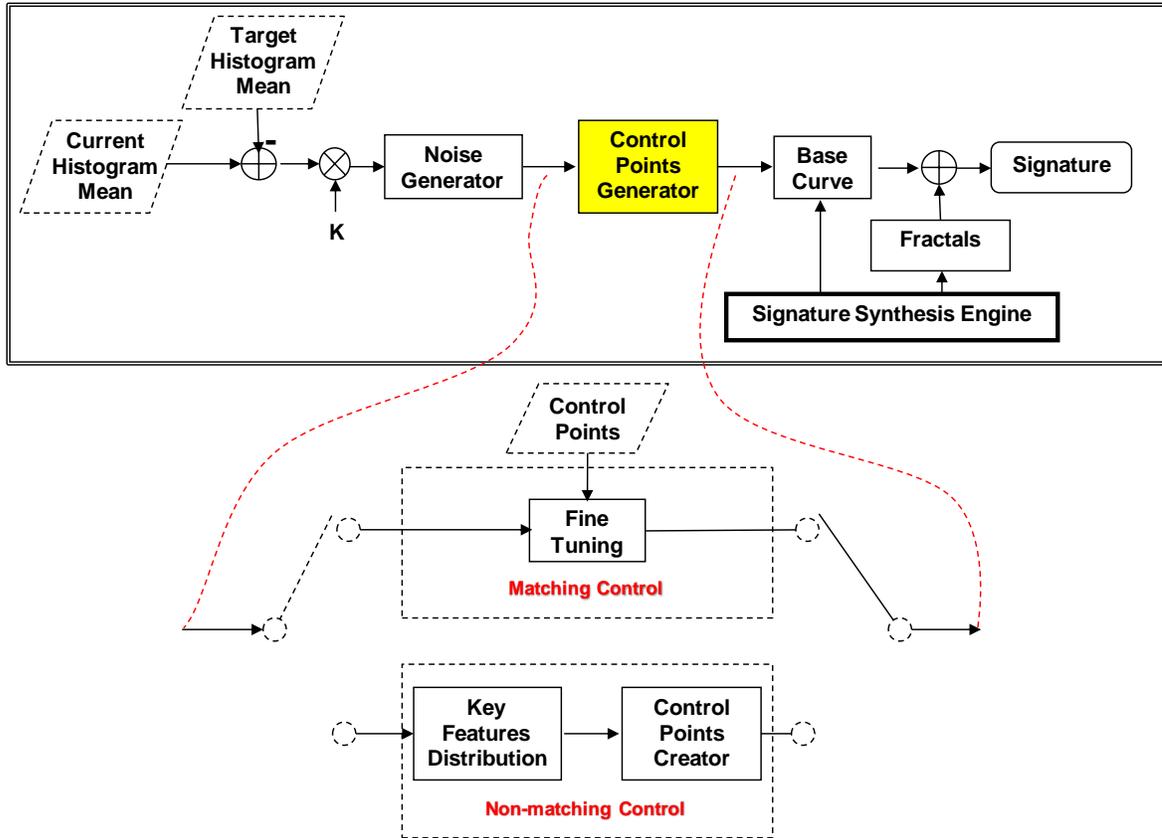

Figure 21: Control the signature generation

For the matched LEA signature generation, the control input is used to fine-tune the same set of control points which was pre-generated for the same barrel. The process is illustrated in Figure 22. The red curve is the original base curve before adjustment, and the blue curve is the adjusted base curve after applying the following rules.

Algorithm 3:

1) *Randomly pick several peak-valley line such as C1-C2 (percentage parameter: A)*

2) *Randomly select the number of adding points such as 0, 1, 2, 3(number parameter: B)*

3) *Randomly select positions along peak-valley line such as B1,B2*

4) *Randomly choose height value ±|A1-B1| to add a new control point A1*

5) *Randomly select number of zero-crossing points and shift it (number parameter: C, shift length: D)*

6) *Connect each points in right order*
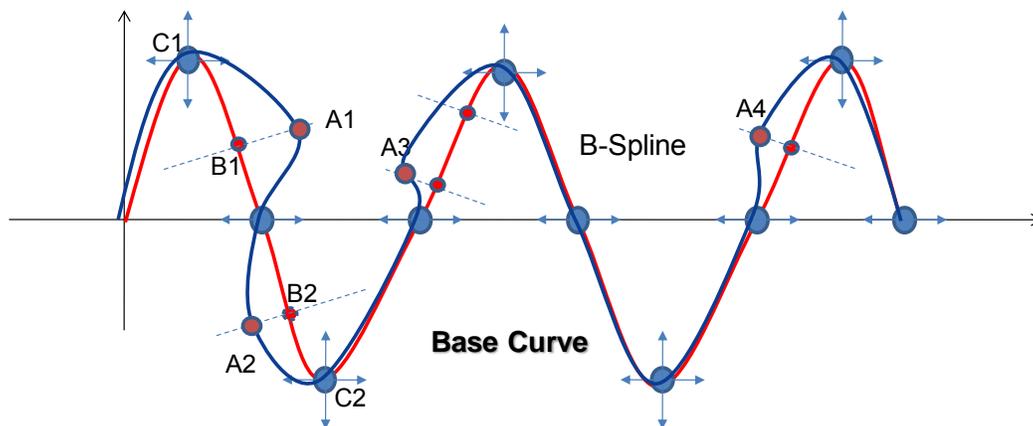
Figure 22: Base curve fine-tuning

For the non-matched LEA signature synthesis, the control input is used to adjust the distributions of several statistical features such as zero-crossing length and slopes used in the base curve generation. As described before, if the distribution of these statistical features is modified, the final base curve would be different. A distribution can be modified to a new distribution by shrinking or expanding operation. We can specify a condensation ratio r as the control parameter: $r > 1$ (expand) and $r < 1$ (shrink). For either case, the distribution curve shall cover the same area as that of the original distribution, as shown in Figure 23. Since the distributions for the zero-crossing (Figure 14) and slopes (Figure 15) are asymmetric, only one side distribution needs to be considered.
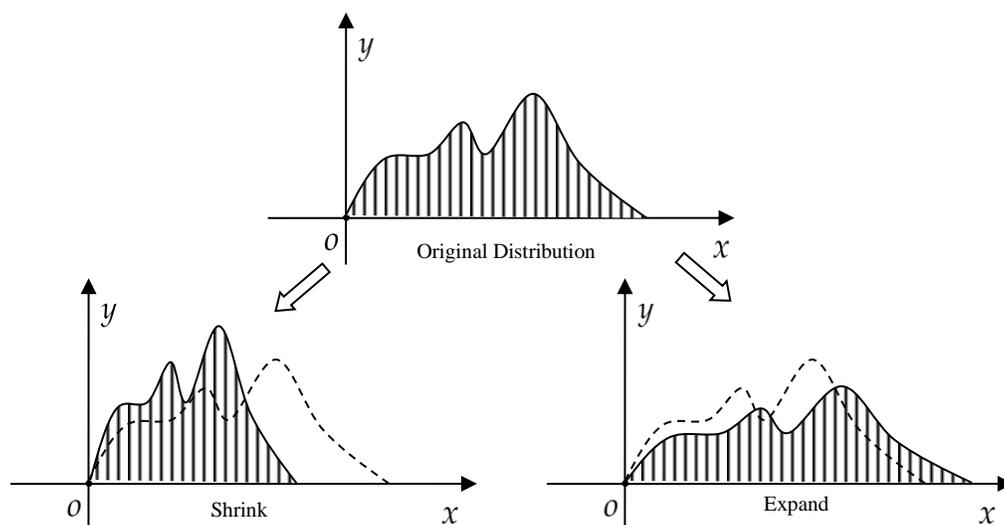

Figure 23: Distribution condensation

A shrunk distribution makes the synthesized signal to be more random thus moves the non-matching distribution against the matching distribution. On the contrary, an expanded distribution makes the

synthesized signal less random thus moves the non-matching distribution towards the matching distribution. We have intentionally specified the condensation factor as 0.5, 1.0 and 2.0 respectively, and generated 240 LEA signature data in total for 4 barrels, and calculated its matching and non-matching distribution. The results are shown in Figure 24. The real experiments clearly verified our perception.
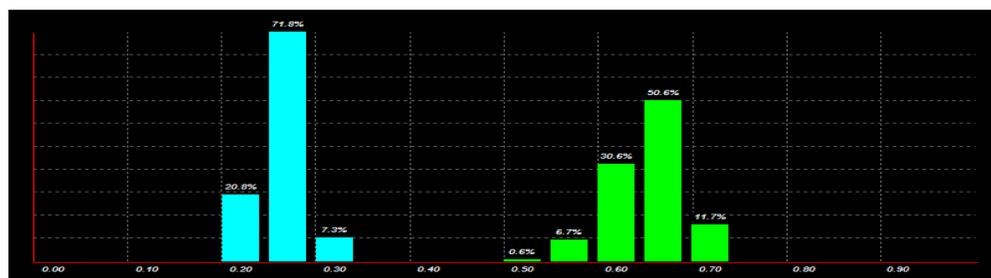
(a) r = 0.5
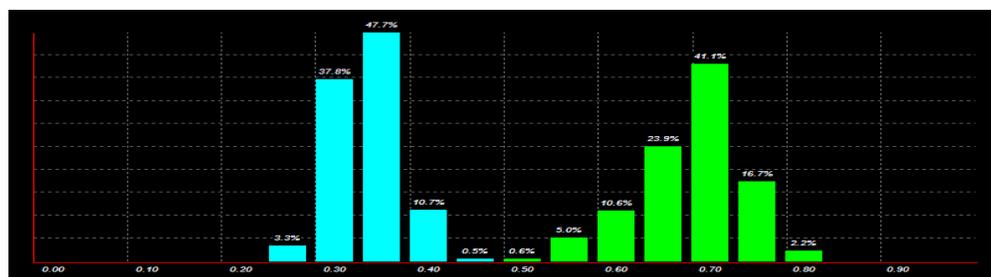
(b) r = 1.0
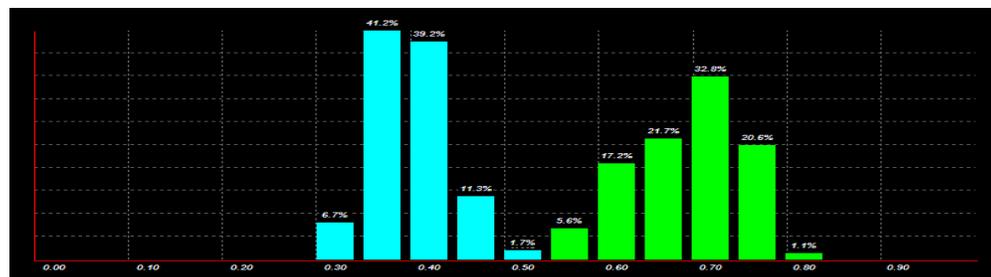
(c) r = 2.0

Figure 24: Matching and Non-matching Distribution for different condensation factor

5.4 Optimization Procedure

The uniqueness of our approach lies in the strategy to generate LEA signature data. The inputs to the algorithm include the number of LEA signature data n and the brand name B. The brand will be used to select the target distribution H_0 and a set of statistical information used in the signature generation. We use this procedure to optimize both matching and non-matching distribution of the synthesized data to match that of the existing data.

As shown in Figure 25, the optimization procedure first randomly generates n LEA signatures and places them into a data pool. Next, the matching or non-matching distribution will be calculated, as well as the matching score. If, based on the score, the generated data satisfies with the pre-determined precision requirement, the optimization is terminated. Otherwise, we go to the adding-and-pruning iterative process. A new noise level is determined by the distribution mean difference between the computed distribution and the target distribution, and a new LEA signature is generated accordingly.
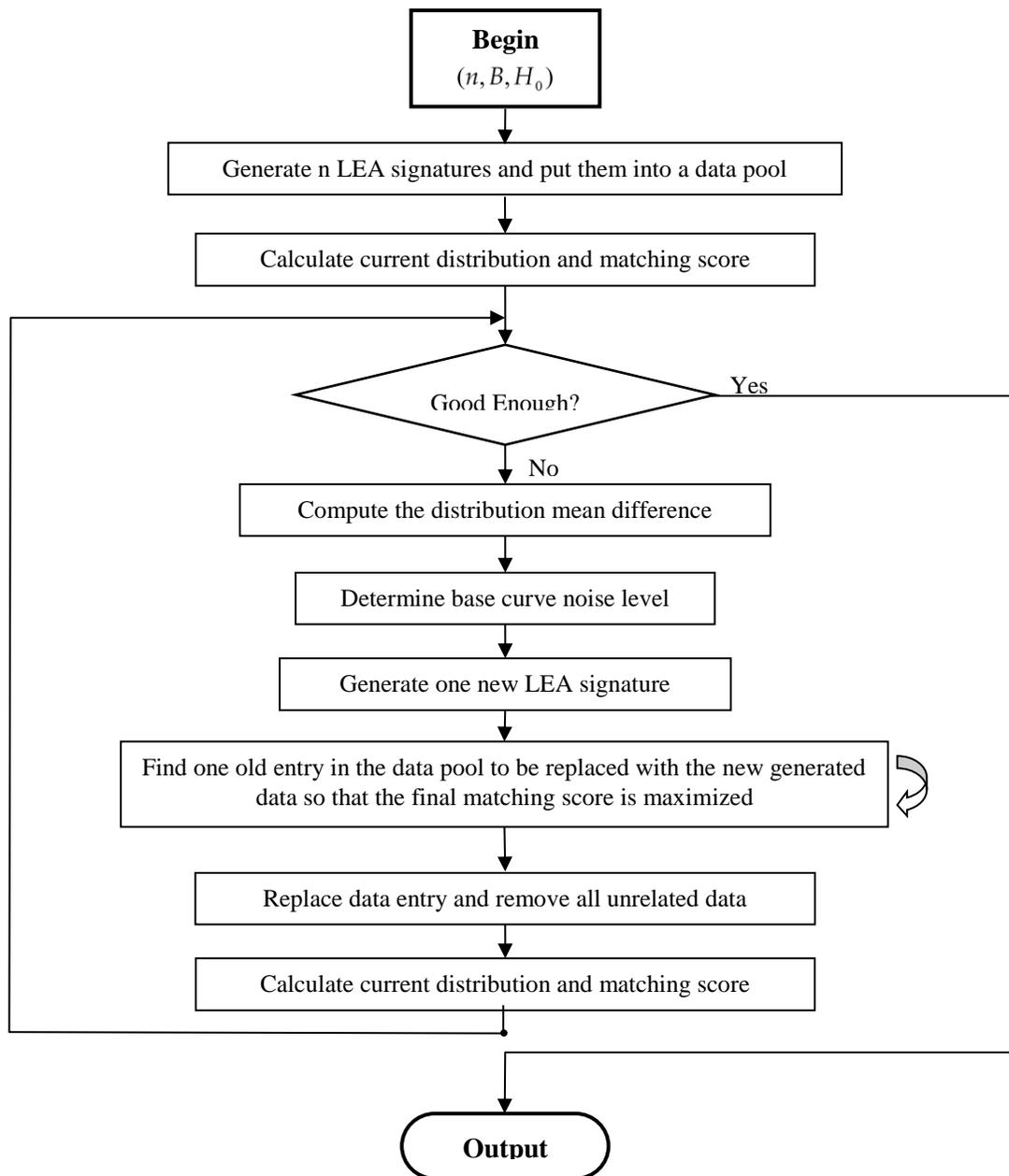

Figure 25: Optimization Process

The adding-and-pruning process is used to update the data pool. We will loop the data pool to find an entry to be replaced with the newly generated data so that the final matching score is maximized. A new score will then be compared with the threshold and the whole process will continue.

During the optimization process, a matching score is calculated in each iteration. If the current matching score is smaller than the previous matching score, a counter is increased by one. Otherwise, the counter is reset to zero. We called this counter as the stable counter because it is an indicator to the stable status of the optimization process. If the count in the stable counter keeps increasing, we think the optimization process has not made any progress during the past consecutive times of trying. In summary, the adding-and-pruning process is terminated when one of the following conditions is satisfied: 1) the total iterative number reaches a defined maximum iteration; 2) the total stable count is larger than the defined number; 3) the pre-defined accuracy is reached.

5.5 Global Optimization Process

The procedure described above only considers the optimization process separately for the matching and non-matching cases. In the following, we developed a global optimization algorithm to synthesize LEA signature data by combining the matching distribution optimization and non-matching distribution optimization into a two-level optimization process. The basic idea is to treat the matching distribution optimization and non-matching distribution optimization as a decoupled problem.

In the inner loop, we perform the following procedures: First, we use non-matching distribution optimization to generate base curves for all barrels to make sure the non-matching correlation distribution of the synthesized data is matched with the target non-matching distribution of the existing data. Here, the target non matching distribution can be adjusted by an external loop. Second, we generate all required LEA signatures using the matching distribution optimization approach described above. Third, the final matching and non-matching correlation distributions of the synthesized LEA signature data are calculated, and compare with the target distribution. In the external loop, the difference between the computed distribution and the target distribution is used to adjust the target non-matching distribution. The whole procedure is illustrated in Figure 26. Good results are obtained and will be reported in the section III.

```
                    ┌─────────────────────────────────────┐
                    │              Begin                  │
                    │  * Brand Name                       │
                    │  * Number of barrel: A              │
                    │  * Number of bullet in each barrel: B│
                    │  * Number of LEA in each bullet: C   │
                    └─────────────────────────────────────┘
```

Begin
* Brand Name
* Number of barrel: **A**
* Number of bullet in each barrel: **B**
* Number of LEA in each bullet: **C**

Yes ◇ First **No**

◇ Good Enough? **Yes**

No

Compute the non-match distribution mean difference

Update the target non-match distribution by shifting

Generate A * C base curves to fit for the target non-matching distribution

For each barrel {
    Update the target matching distribution
    Utilize the generated base curve control points
    Generate all LEA signatures to fit for the target matching distribution
}

Calculate the matching and non-matching distribution

Calculate the combined score

Output

Figure 26: Global optimization process

6. Two-dimensional LEA Image Synthesis

To reconstruct a 2D LEA image from a synthesized LEA signature curve in the X-Y plane, the simplest approach is to repeat each signature curve in z-direction as shown in **Figure 27**.
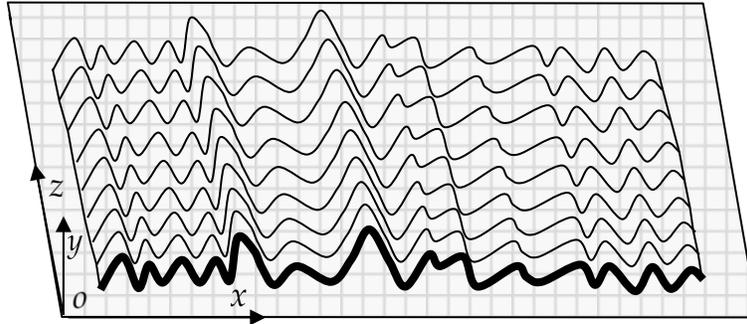

**Figure 27: Expanded LEA signatures to form a LEA image**

To make a generated 2D LEA image realistic, we have to investigate the noise characteristics along the z-direction.

6.1 Data synthesis along z-direction

A sample LEA data along the z-direction is shown in **Figure 28**(a). The data average is the value of the LEA signature at the given point on the x-axis. **Figure 28**(b) shows the distribution of the noise, calculated by subtracting the signature value from the data. The noise distribution can be approximated as a Gaussian distribution. It is important to note that the value of each point on the z-axis will not change abruptly from its closest neighbors, which suggests an independent identical distribution (i.i.d.) noise model cannot be used to generate synthetic data along the z-direction. In this work, we use an Autoregressive (AR) model to characterize data along the z-direction.
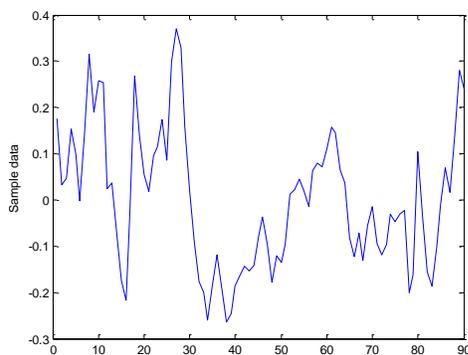
(a) Sample LEA data along z-direction
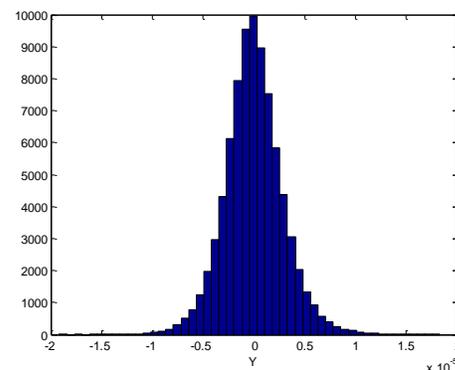
(b) Noise distribution along z-direction

Figure 28 : Noise distribution along z-direction

Autoregressive Model

In statistics and signal processing, an autoregressive (AR) model is a type of random process which is often used to model and predict various types of natural phenomena. The autoregressive model is a group of linear prediction formulas that attempts to predict an output of a system based on the previous outputs [17].

The notation *AR(p)* indicates an autoregressive model of order *p*. The *AR(p)* model is defined as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

where $\varphi_1, \cdots, \varphi_p$ are the parameters of the model, c is a constant (often omitted for simplicity) and ε_t is noise term.

There are many ways to estimate the coefficients. One of the most commonly used methods is *Yule-Walker* equations as given below:

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

where m = 0, ... , *p*, yielding p + 1 equations. γ_m is the autocorrelation function of X, σ_ε is the standard deviation of the input noise process, and $\delta_{m,0}$ is the Kronecker delta function. The Yule-Walker algorithm is implemented in the MATLAB signal processing toolbox with function *aryule()*, which is used in our work.

An example of the AR coefficients calculated by Yule-Walker algorithm is shown in Figure 29. Figure 29(a) shows a number of examples of AR coefficients and noise variance obtained for Berreta barrel 1. We can see that although some high order AR coefficients are fluctuant in small range, the first three orders are basically consistent, which is much more important than high order coefficients. It is worth mentioning that the AR coefficients are not independent, as seen the bivariate distribution of the 2nd and 3rd order of AR coefficients in Figure 29(b).

For each brand, we create a look-up table for trivariate distribution of the first three AR coefficients, as well as a look-up for the distribution of noise variance. When generating synthetic data along z-axis, at each point along x-axis, the first three AR coefficients and the noise variance are obtained randomly from the distribution look-up tables. **Figure 30** shows an example synthetic data along z-axis generated using the AR model.
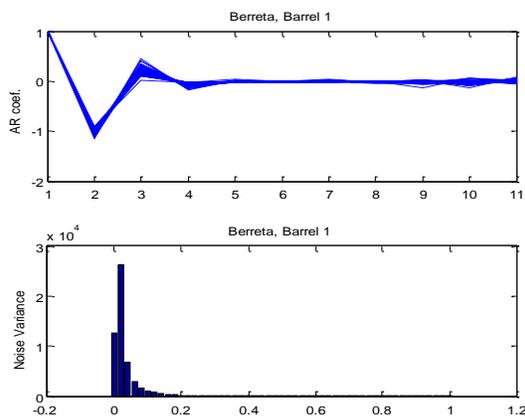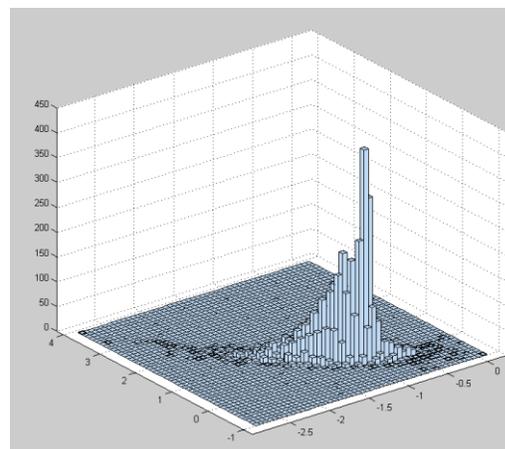
(a) AR coefficients and noise variance   (b) 2nd vs. 3rd AR coefficient

Figure 29: Example of AR coefficients and noise distributions obtained for Berreta Barrel 1.
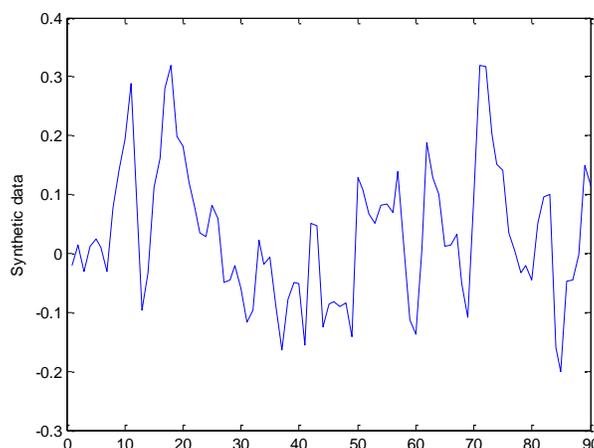

Figure 30: Synthetic noise along z-axis generated by AR model

6.2 Two-dimensional Image Synthesis

A straightforward method for 2D image synthesis is to generate synthetic data at each signature point. A drawback of this method, however, is that the 2D image could be discontinuous along the x-axis. So, we first perform peak and valley detection to find all local maxima and minima as shown in **Figure 31**. Then, we generate synthetic data at each peak/valley points and a 2D grillwork will be obtained as shown in **Figure 32**, where synthetic data is generated as some sample points on LEA signature. Finally, a 2D interpolation will be conducted to obtain the synthetic 2D LEA image.
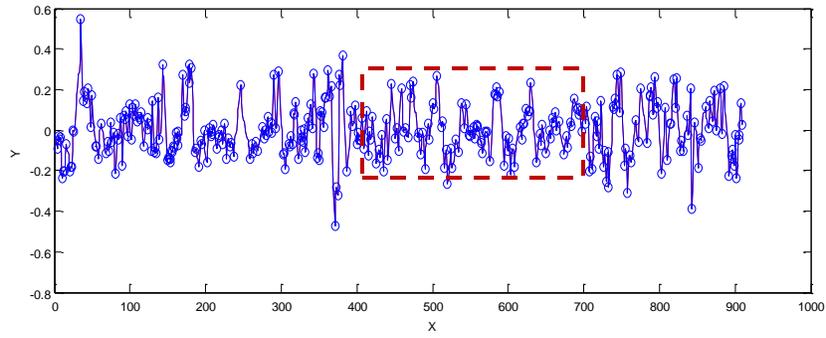
(a) Peak and valley detection
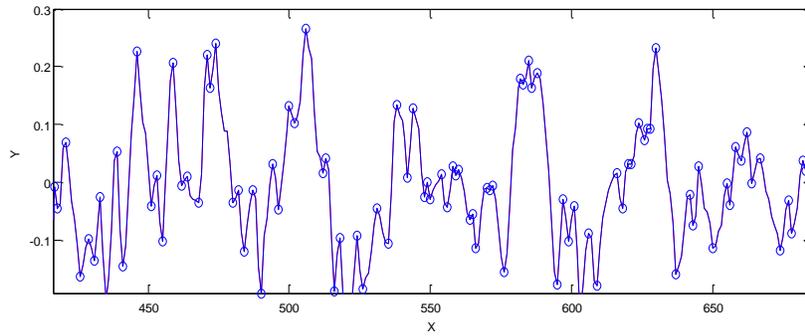
(b) Zoom-in on peak and valley detection

**Figure 31: Peak and valley detection on 1D LEA signature**
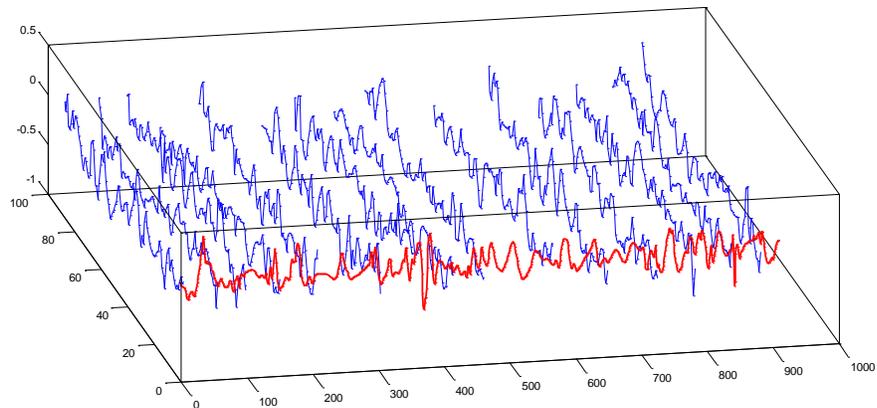
Figure 32: 2D grillwork for interpolation

The 2D LEA image data generation procedures are summarized as follows:

Algorithm 4:

1) *Create a 3D matrix in a grid format, storing a 3D point (x, y, z) in each grid node. The width of the matrix is determined by the length of the signature, and the height of the matrix is determined by a width to height ratio, computed by the user.*

2) *Identify peaks and valleys of the signature*

3) *At each peak and valley location, extend a z-axis profile using an AR model while keeping its mean equal to signature value. The AR coefficients and variance are generated based on known distributions.*

4) *Perform 2D interpolation on the grillwork to form surface*

In **Figure 33**, both the true LEA and synthetic images are shown. Note that the image in **Figure 33** is only a part of whole LEA data; the outliers and the coast line area are removed so that the parameters used in synthesis are estimated based on valid data.
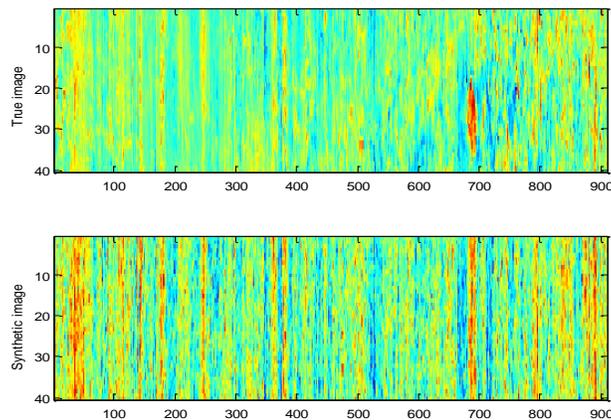

Figure 33: Synthetic 2D LEA image

Figure 34 shows examples of synthetic 2D LEAs for three brands, Browning, Berreta and Bryco. For comparison purpose, we used true 1D signature to generate 2D synthetic LEAs, and showed both the synthetic and true 2D LEAs in **Figure 34**. We can see that the synthetic images match the true image well with different noise characteristics.
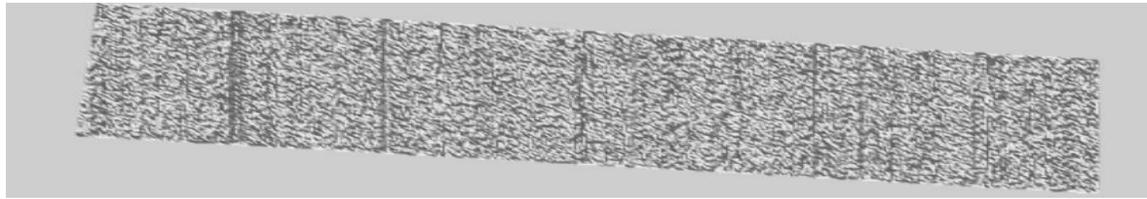
(a) Slide view of a 3D synthetic LEA (Beretta, I1-1 LEA1)
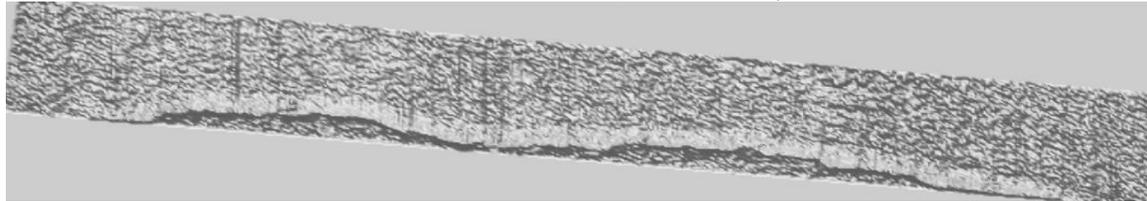
(b) Slide view of a 3D true LEA (Beretta, I1-1 LEA1)
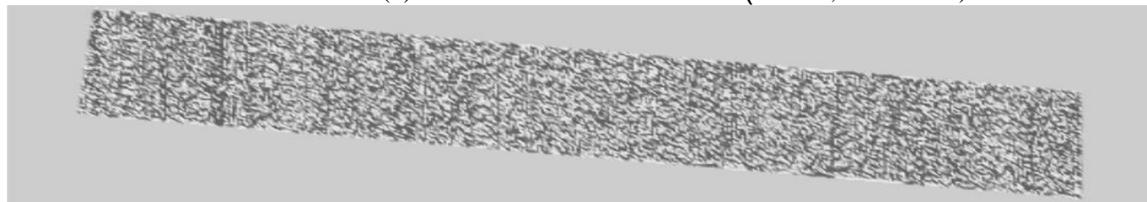
(c) Slide view of a synthetic LEA (Browning, G2-6 LEA2)
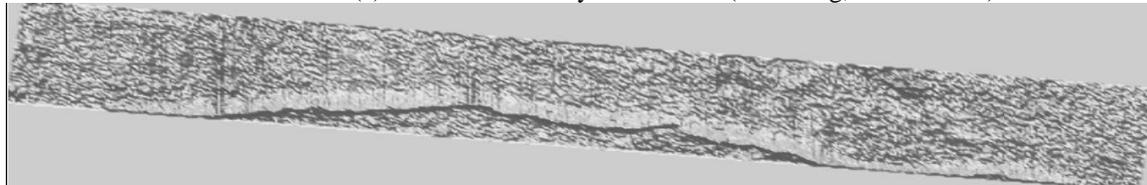
(d) Slide view of a true LEA (Browning, G2-6 LEA2)

(e) Slide view of a synthetic LEA (Bryco, D3-8, LEA3)
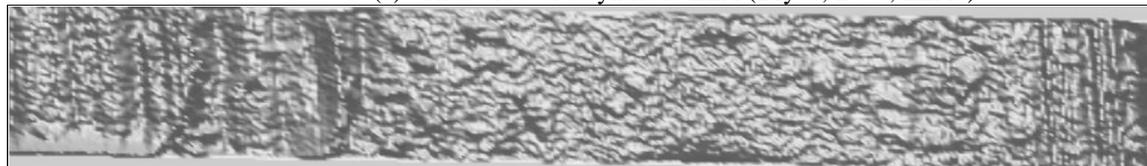
(f) Slide view of a true LEA (Bryco, D3-8, LEA3)

Figure 34: 3D view of synthetic and true LEA

To evaluate the 2D synthesis performance, for each synthetic LEA image, we calculate the correlation coefficient between the signature and each row in the 2D image, as well as the noise variances of synthetic LEA image. In **Figure 35** shows the histograms of the correlation coefficients and noise variances obtained from three barrels of Berreta, four of Browning and four of Bryco, in which we can see very similar distributions. Note that, when generating random AR parameters in the process of 2D

extension, we discard outliers to ensure the synthetic images not too far from normal LEA images. This is the reason why we can see slightly less variance and higher correlations in **Figure 35**.
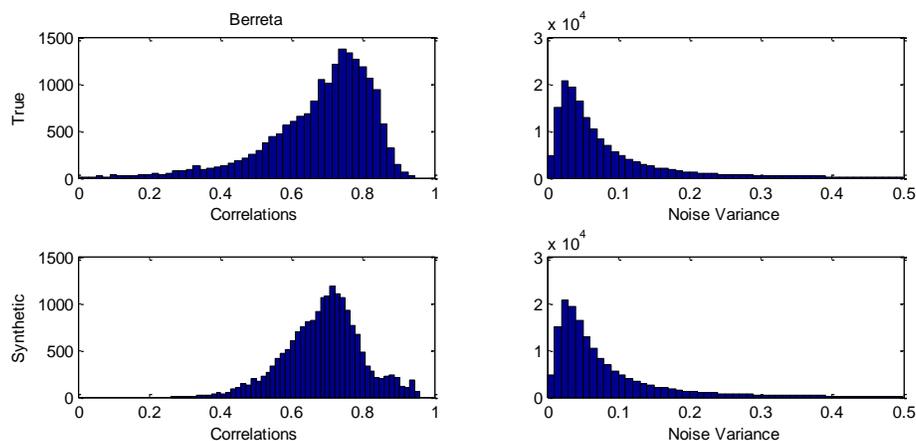
(a) Statistic comparison for Berreta (I1, I2 and I3)
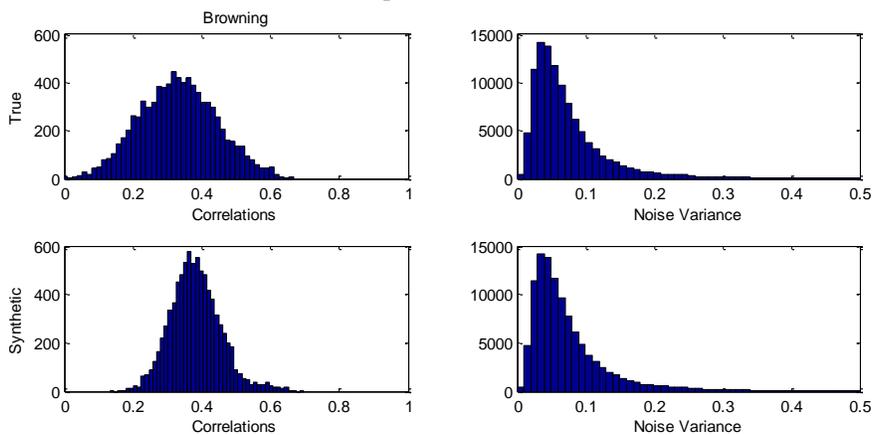
(b) Statistic comparison for Browning (G2, G3, G4 and G5)
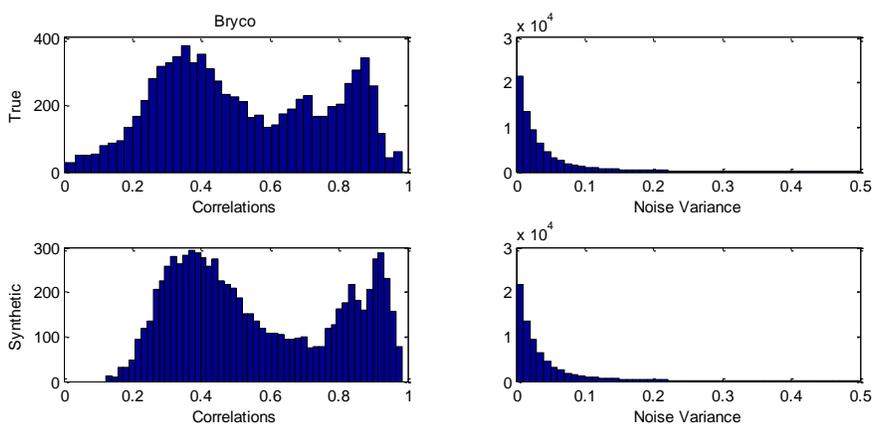
(c) Statistic comparison for Bryco (D2, D3, D4 and D5)

Figure 35: Histograms of the correlations with the LEA signature and noise variances of true and synthetic 2D LEA images. (a) Berreta, (b) Browning, and (c) Bryco

7. Three-dimensional LEA Surface Composition

While Leas generated by different barrels of the same type (brand) differ in fine details, the overall characteristics of the LEAs remain unchanged for the same brand. At the same time, we also know that a LEA carry distinctive information about the specific barrel it has been shot from. This means that a LEA can be seen as a composition of the common 3D surface peculiar to the brand type of the barrel and a 2D LEA containing distinctive feature of each barrel of that brand type. And to synthesize a 3D LEA out of a 1D signature, we can first generate the 2D LEA surface from 1D signature and then overlay it onto the 3D surface template generated already for that particular brand of the barrel. The procedure to carry out the first step is described in section 6.2 already. Here we discuss how we generate the 3D template for each barrel type and how we overlay the 2D surface onto 3D template.

Generating 3D surface Template

For this purpose, we combine the 3D surfaces of the corresponding LEAs from different bullets fired from different barrels of the same brand. The 3D surfaces of two corresponding LEAs, however, are never aligned such that we could simply average them to generate the template. So we will have to, first, align the 3D surfaces before combining them to create the template.

Two LEA surfaces in the database are related to each other by a rigid 3D transformation, caused by the change in the placement of the bullets in the 3D scanning device. To align them, we have to estimate that 3D transformation and then use it to project one of the surfaces on to the other. In our case, we will consider the first LEA in a sequence of LEAs as the reference surface and compute the transformation of the rest of the LEAs to the reference and project them all onto this reference surface.

Before explaining how we estimate the transformation, it is important to note that, due to illumination issues during scanning, 3D surfaces contain some in-valid entries as well. So the data is filtered to remove those known invalid entries. Another issue to note is about the computation complexity of 3D transformation estimation: more points means slower estimation. So we sample the 3D surfaces at a regular spacing of 20 units along the x and y axes, making the point cloud for each surface smaller and computationally more tractable to estimate rigid transformation between them.

Estimation of 3D Rigid Transformation using ICP

We estimate the 3D transformation between two surfaces or point clouds using Iterative Closest Point method (ICP) (introduced by Besl and Mckay [18]), which is one of the popular approaches used for 3D point registration. Assume A and B are two point clouds, and we would like to find the 3D transformation

between A and B. Without loss of generality, we choose to transform A onto B. Then, ICP, which is an iterative process, follows these steps:

1) Using nearest neighbor criteria, the corresponding pairs of points in A and B, represented by a set $C = (i,j)| a_i \in A, b_j \in B$, are found.

2) 3D transformation between corresponding pairs of points is estimated by solving the following optimization problem:

$$argmin_{R,T} \sum_{(i,j) \in C} \|b_j - Ra_i - T\|$$

3) A is transformed using the estimated transformation: $A \leftarrow RA + T$

4) Repeat 1-3 if $dist(A,B)$ is more than a threshold.
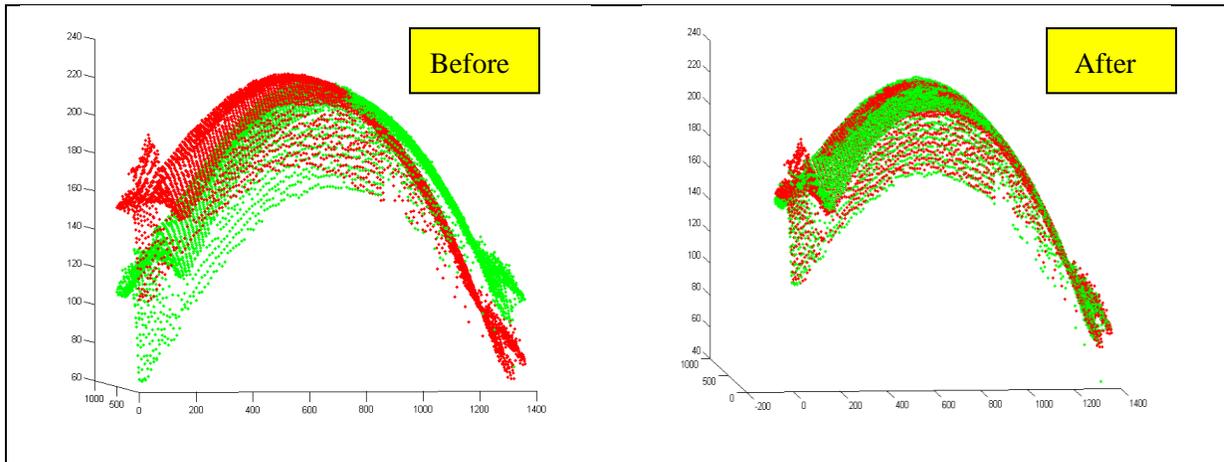
Figure 36: an example of 3D estimation using ICP

Median Surface

For each brand, individual templates are generated for each LEA. To generate a 3D template LEA surface, a set of matching LEAs from different bullets fired using different barrels (see Table 7 for details) is selected. The rigid transformation between the first LEA and the rest of the LEAs is computed as described above. Using the estimated rigid transformation between a LEA and the reference LEA X_0, the LEAs are projected back on the reference LEA and $X'_1, X'_2, \dots X'_9$ represent the transformed LEAs. Now to combine the transformed surfaces with the reference surface, we use the regular x-y grid on which X_0 is defined. The (x,y) co-ordinates of the transformed LEA surfaces are going to lie between the grid locations. Using nearest neighbor, we associate each point on the transformed surface with a particular grid location of X_0. For every grid location of the reference LEA, we have a vector of z-values from

$X'_1, X'_2, \ldots X'_9$ and X_0. Averaging the z-values will result in the final 3D surface template. Since there is going to be at least one z-value for each location on the z-y grid due to X_0, the averaging process will give rise to a smooth 3D template. Next, we are going to overlay the 2D LEA surface generated from signature on the template.

Synthetic LEA surface

So far, we have generated templates for two different brands using a number of LEAs, as described in Table 7. **Figure 45** shows 3D surface of a typical LEA of two barrel brands.

Table 7: Details of LEAs selected for each brand

| Brand name	Number of Bullets	Number of Barrels
Berretta	4	4
Browning	4	4
Bryco	4	4
Taurus	5	4
SIG	4	4
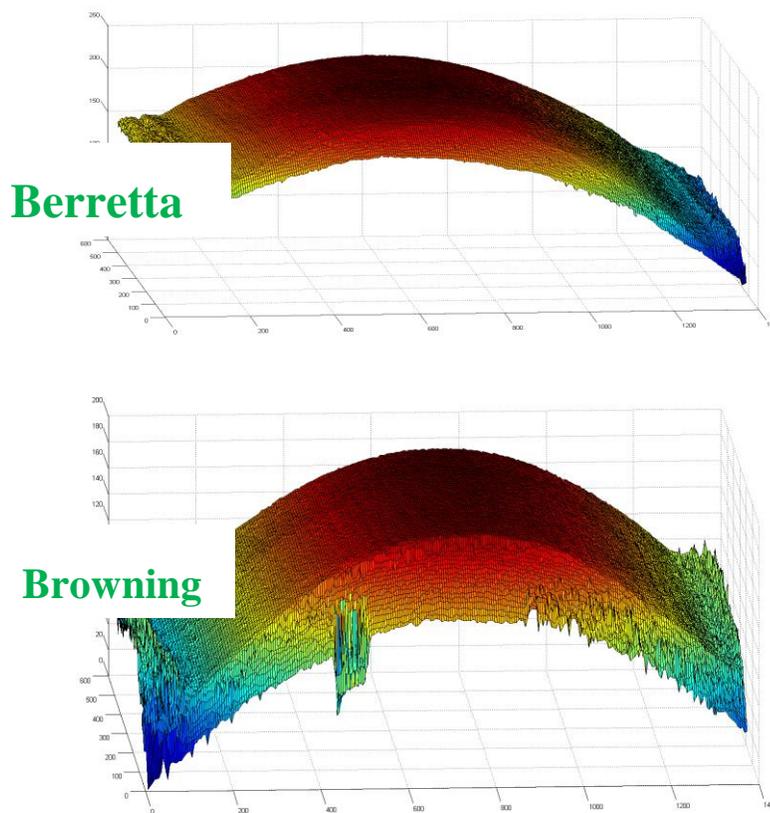
Figure 37 3D LEA surface templates of two specific barrel types

Overlaying 2D LEA on 3D Surface Template

Since the synthesized corresponds to particular brand of barrel, we use that information along with LEA index to select corresponding 3D template which represents the average surface of a bullet fired from each brand of rifles. Then, we superimpose the extended 2D surface onto the 3D template as illustrated in **Figure 38**.

Let's say $z = f(x, y)$ is the 3D template and L_{2D} is the 2D LEA surface generated from the synthesized signature. Since 3D template $z = f(x, y)$ is defined over a uniform x-y grid, we superimpose the synthesized 2D LEA by simply placing it an offset $(x_o y_o)$ on the XY grid and adding their z-values as

$$z = f'(x, y) = f(x, y) + L_{2D}(x - x_0 + \Delta x, y - y_0)$$

where Δx is the additional shift of each row of L_{2D} given by $(y - y_0) \tan \propto$ and $\propto$ is the rifling angle.
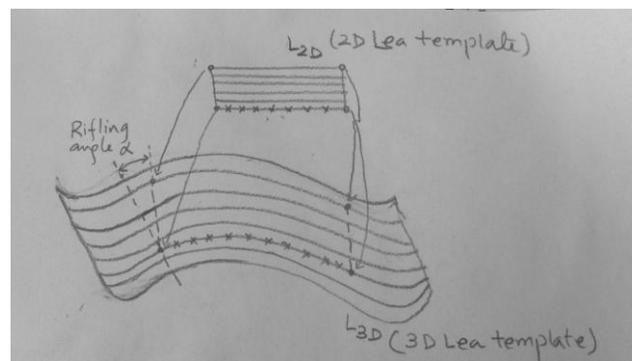

Figure 38: Illustration showing how 2D LEA surface is overlaid on the 3D template

III.   Results

1   Software Development

The software developed as part of this project is written primarily in C++ under the Microsoft Visual Studio 2005 IDE, but some of functionalities are implemented by calling MATLAB functions. A 3D display capability was implemented using OpenGL. Various algorithms, evaluation and visualization tools have been developed. This software provides a powerful tool to investigate various algorithms in data synthesis and tool mark identification.

In developing algorithms, we have discovered that a large amount of data processing procedures can be easily implemented in Matlab. Matlab(c) is well-known scientific software covering a wide range of engineering and mathematical fields. People can easily develop applications on it. However, we still want to utilize our existing visualization tools to enhance the user's experience. Mixing C++ and Matlab programming would be the solution. In this project, we developed an interface between the native

windows program and the MATLAB engine using Microsoft Component Object Model (COM) technique to bridge these two platforms. By utilizing this interface, we can easily call the MATLAB functions in the C++ program. The software will collect data, feed it to Matlab workspace, get back output data from Matlab workspace to the C++ program to process or visualize. All the data processing is run in the background by Matlab engine. In the way, we got fast prototyping and the same kind of user experience as other native C++ software.

Basic functions of the software tool are illustrated in Figure 39, and described as follows:
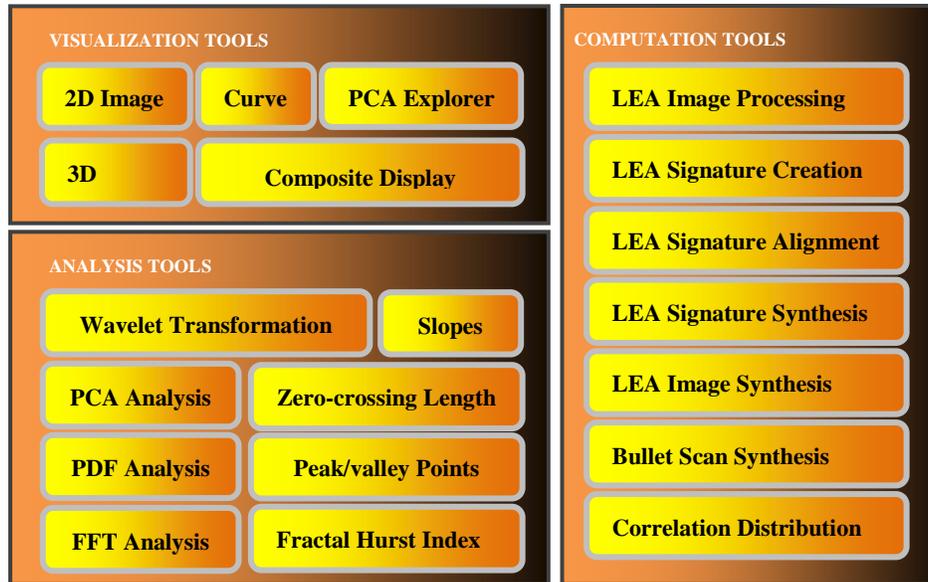

Figure 39: Software functionalities

- *LEA image data processing.*
- *LEA signature creation from LEA image.*
- *LEA signature alignment.*
- *PCA space creation tool for LEA signatures*
- *Visualization tools: LEA data can be displayed in different modes: 2D image, 3D model, 2D curves and combined display.*
- *LEA data management.*
- *Fast correlation distribution calculation.*
- *Synthesis and evaluation of LEA signature/LEA image/Bullet scan.*
- *Wavelet analysis tool for LEA signatures.*
- *Statistical tools for LEA signatures such as FFT, PDF.*
- *Analysis tools to extract key features of peak-valley points, zero crossing length, slopes and fractal Hurst indexes.*

Graphical User Interface (GUI)

The GUI is designed to be a view-oriented system. We use a multiple-tab structure to represent each different view, as shown in **Figure 40**. Users can freely switch to any one view at any time to visually inspect the data. The toolbar button changes accordingly to reflect the change of the view. We also designed a detachable control bar in the software framework for frequently used functionalities.


(a) Image Tab

(b) 3D Model Tab

(c) Curve tab

(d) Combined Display Tab
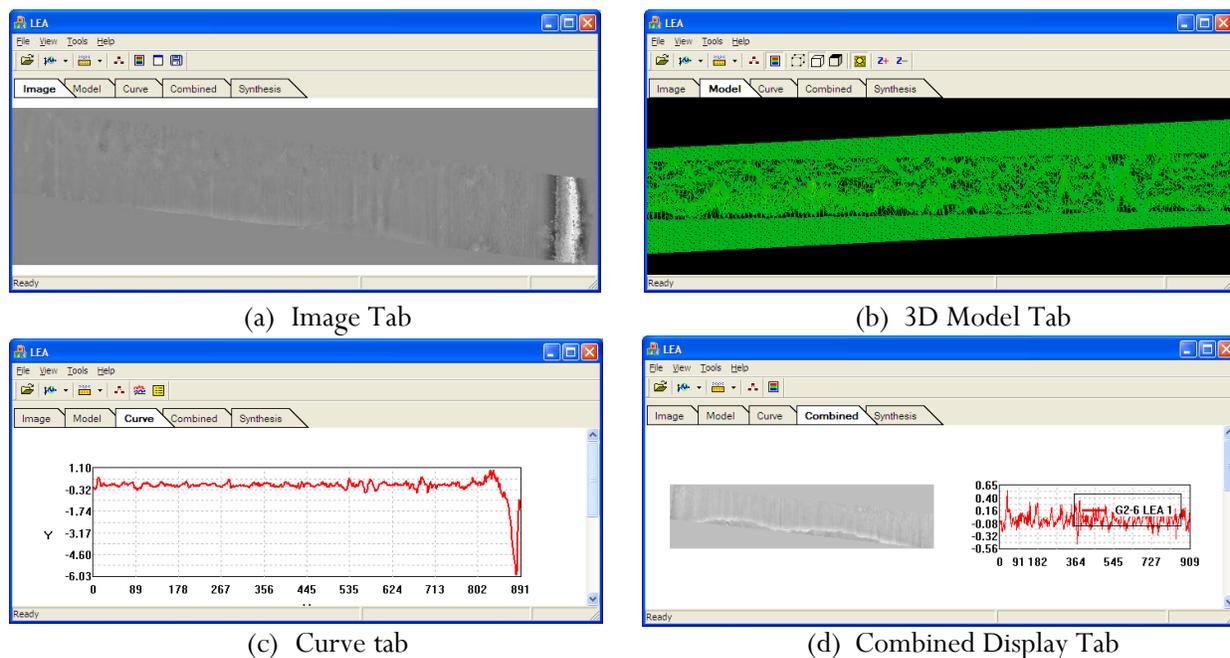
Figure 40: Multi-Tab System

Visualization Tools

(a) 2D Image Display

(b) 3D Mesh Display

(c) 2D Curve Display


(d) Combined overlay display
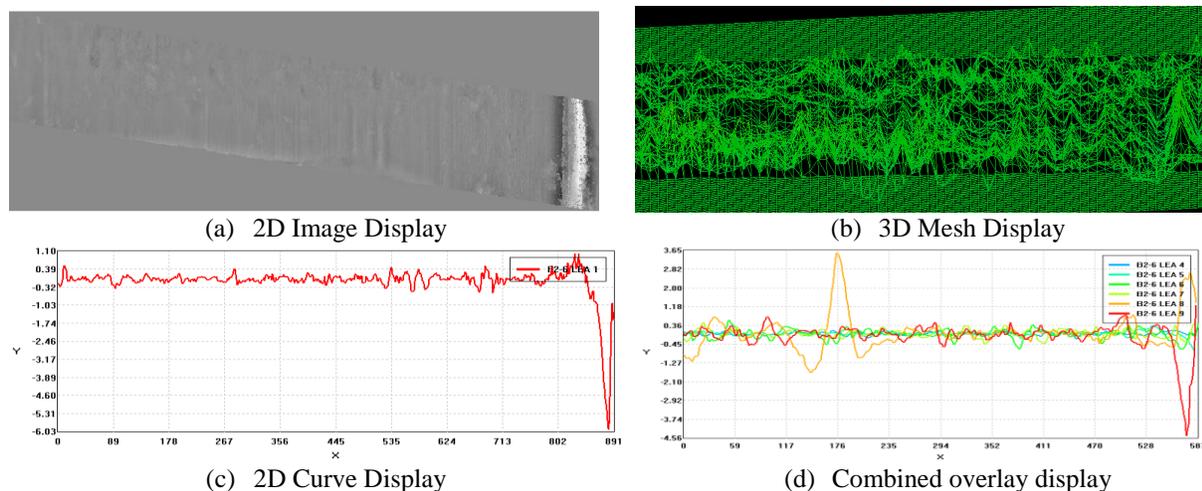
Figure 41: Multiple Visualization Display Modes

Visualization tools in the software are very powerful. The LEA data can be shown in 2D image, 2D curve, and 3D model or in combined display as shown in **Figure 41**.

For the combined display, it allows multiple curves showing row by row or overlaying multiple signatures in one common coordinate system. LEA data image and its signature can also be shown side by side as shown in Figure 42.
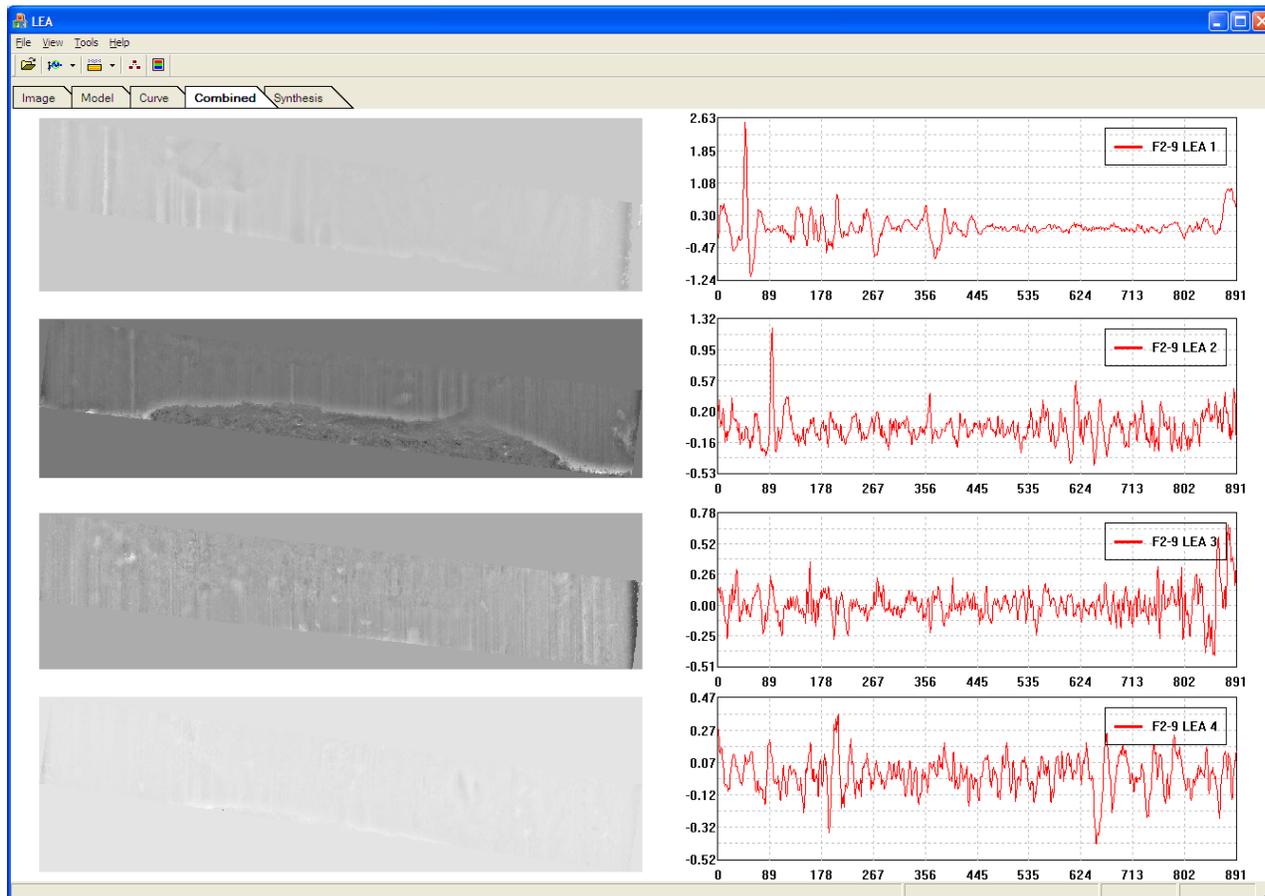

Figure 42: Combined Display for multiple LEA signatures

During the software development, we have found the one-dimensional signal display is very critical for the observing, inspecting and comparing the underline data. Thus, we have dramatically enhanced our visualization tool for one-dimensional signal display and manipulation. We can show multiple curves on the same screen simultaneously either in row by row mode or overlap mode. All the curve properties can be changed on the fly. The statistical information regarding multiple curves can be generated very easily. All the curve data and displayed curve image can be selectively saved to files. The following screenshots shown in Figure 43 demonstrate some of these features.

(a)Multi-curve display in overlap mode	(b)Statistical information
(c) Change Curve properties on the fly	(d)Save Selected curve data
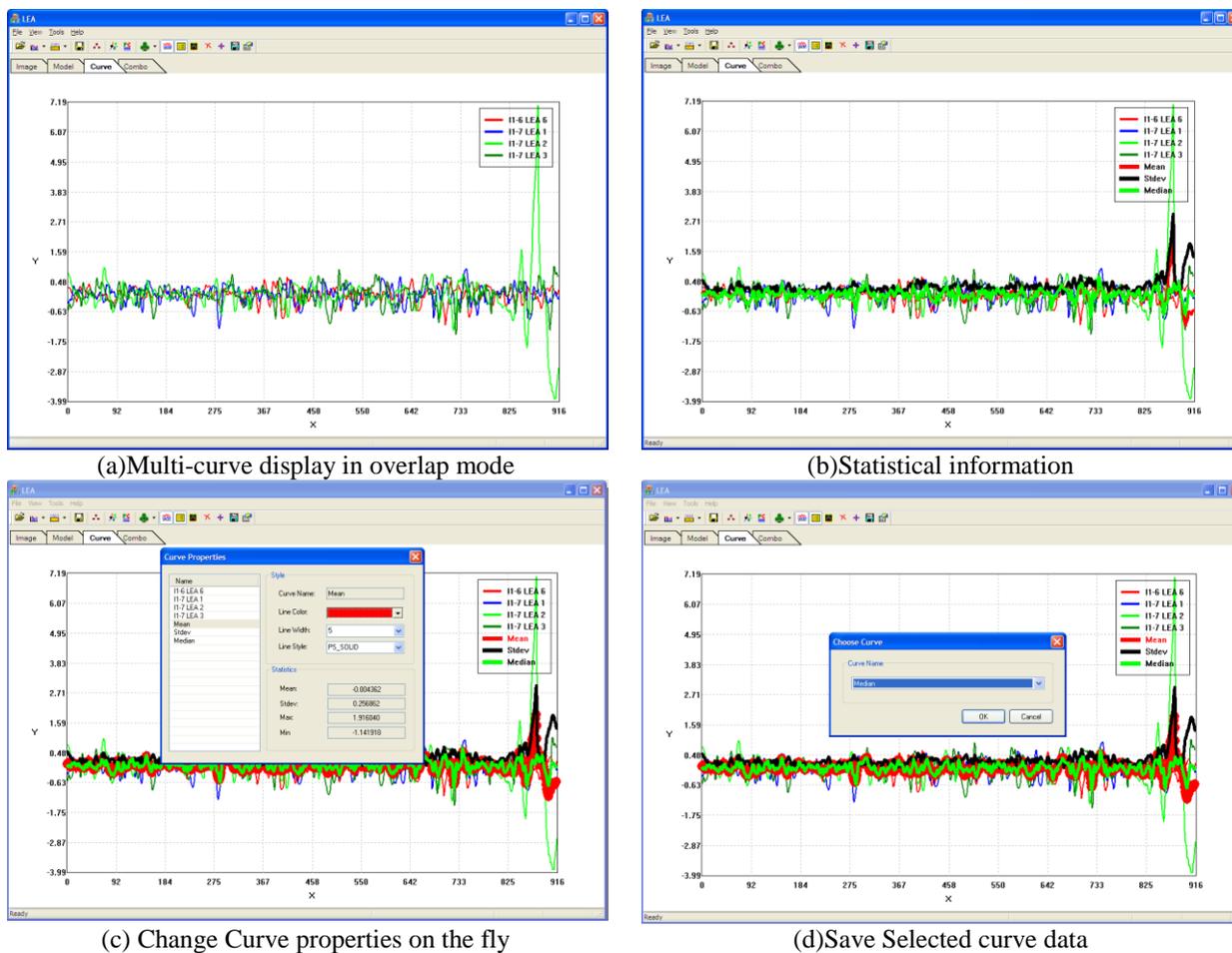
Figure 43: Enhanced one-dimensional curve display

2   Results of LEA Signature Data Synthesis

Using the developed algorithm, we have generated the synthesized data for different brands such as Beretta, Browning, Sig and Taurus. Our synthesis results are presented in two perspectives. First, we align the synthesized LEAs with the real LEAs side by side to compare their common characteristics and individual characteristics. Second, we generate the same amount of data as the amount for the brand Beretta and run the correlation analysis to observe the correlation distribution of the bullets in this brand. The results are shown from **Figure 44** to **Figure 47**, where the top two rows show the real data selected from each brand, and the bottom two rows show the synthesized data. **Figure 48** and **Figure 49** show the synthesized matching and non-matching LEA signatures for Beretta and Browning brand, respectively.
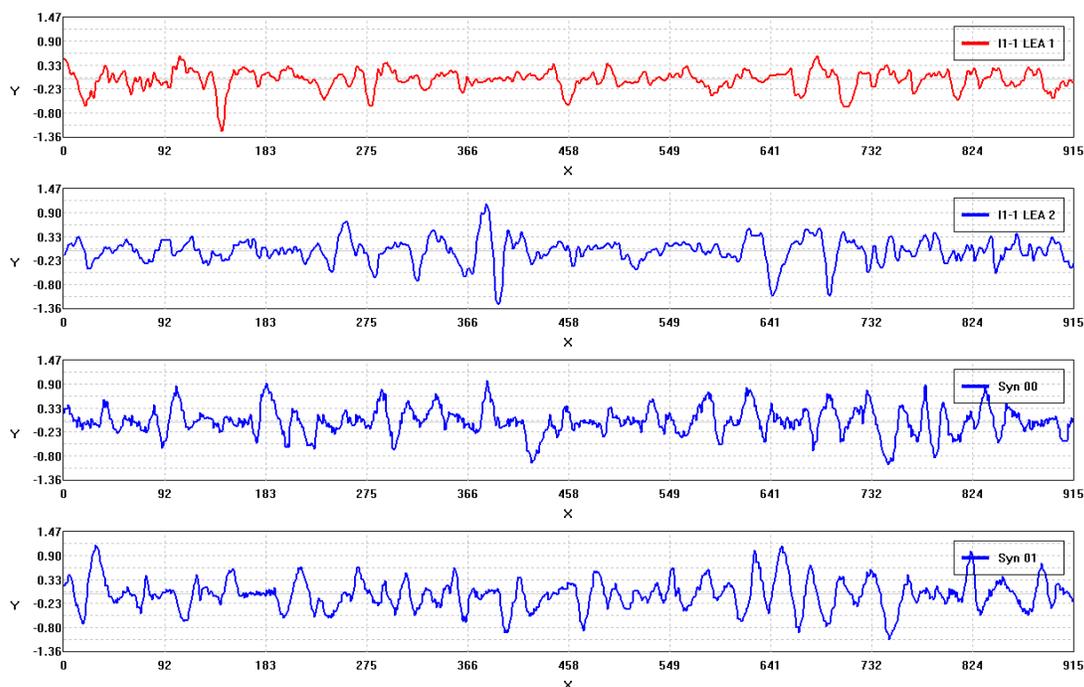
**Figure 44: Beretta existing signatures and synthesized data**
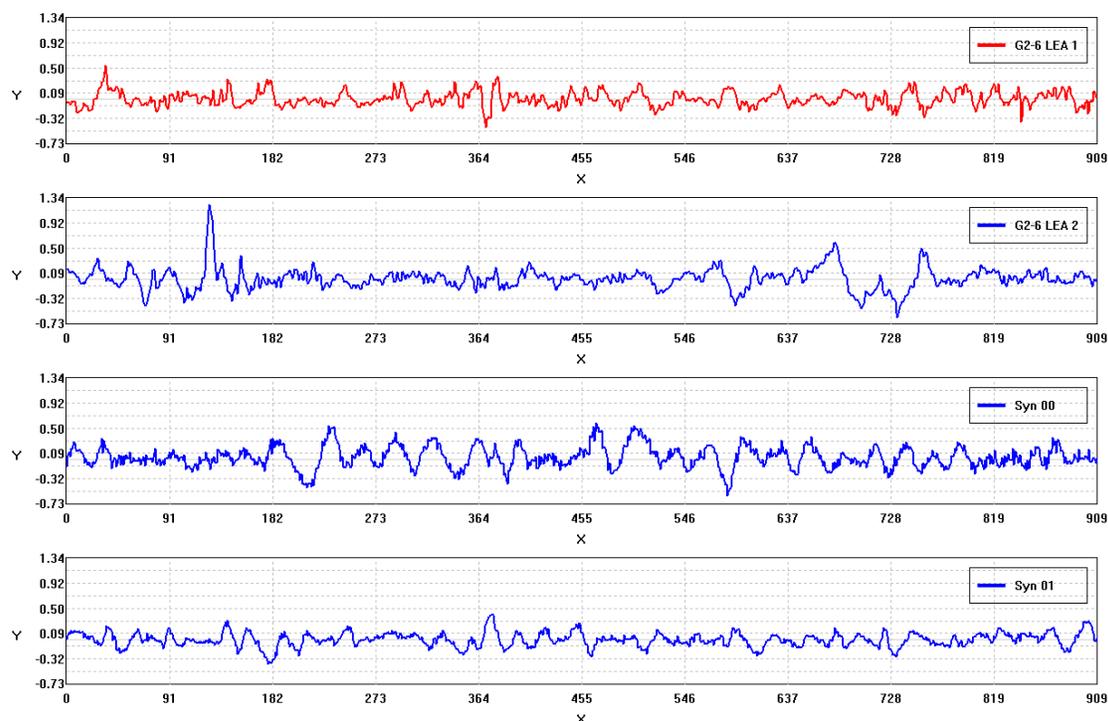
**Figure 45: Browning existing signatures and synthesized data**
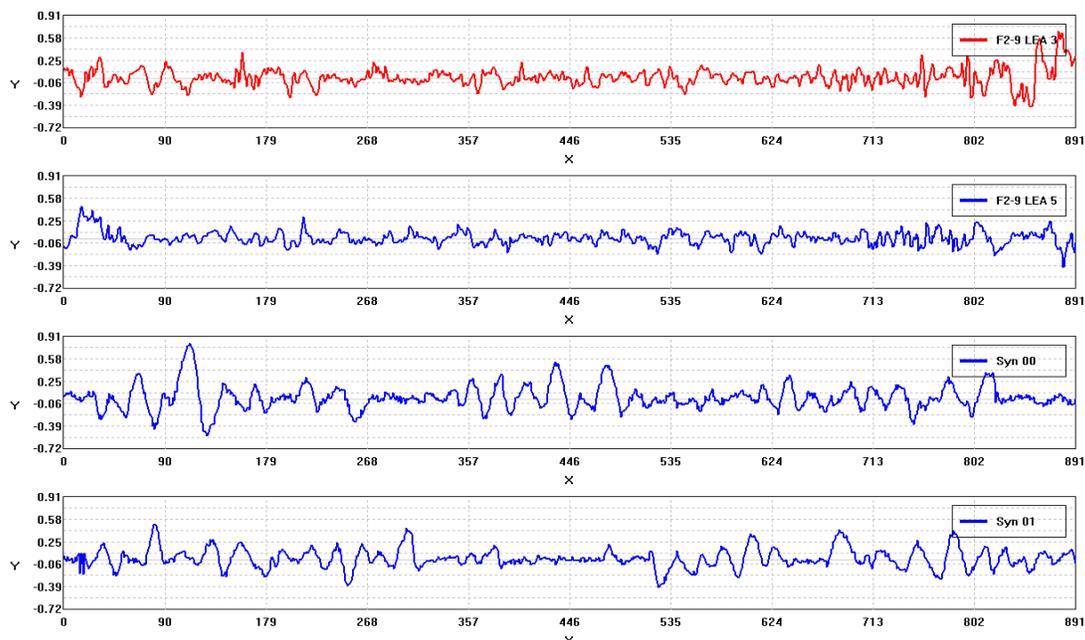
**Figure 46: Sig existing signatures and synthesized data**
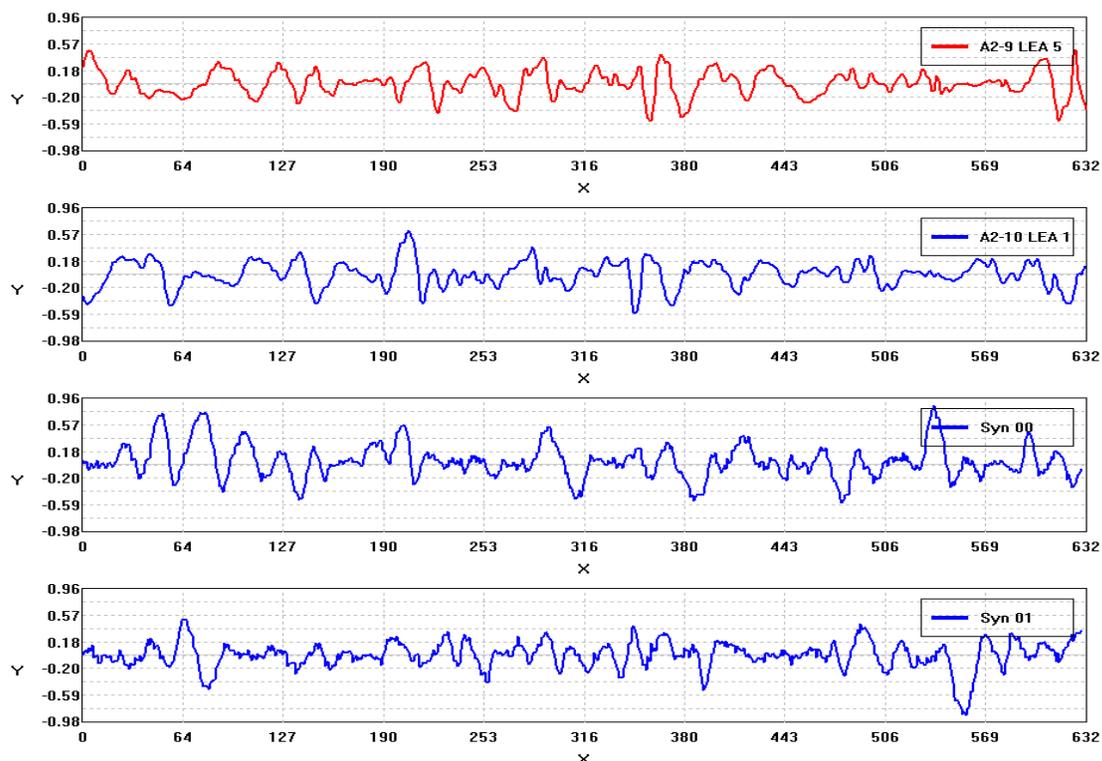
Figure 47: Taurus existing signatures and synthesized data
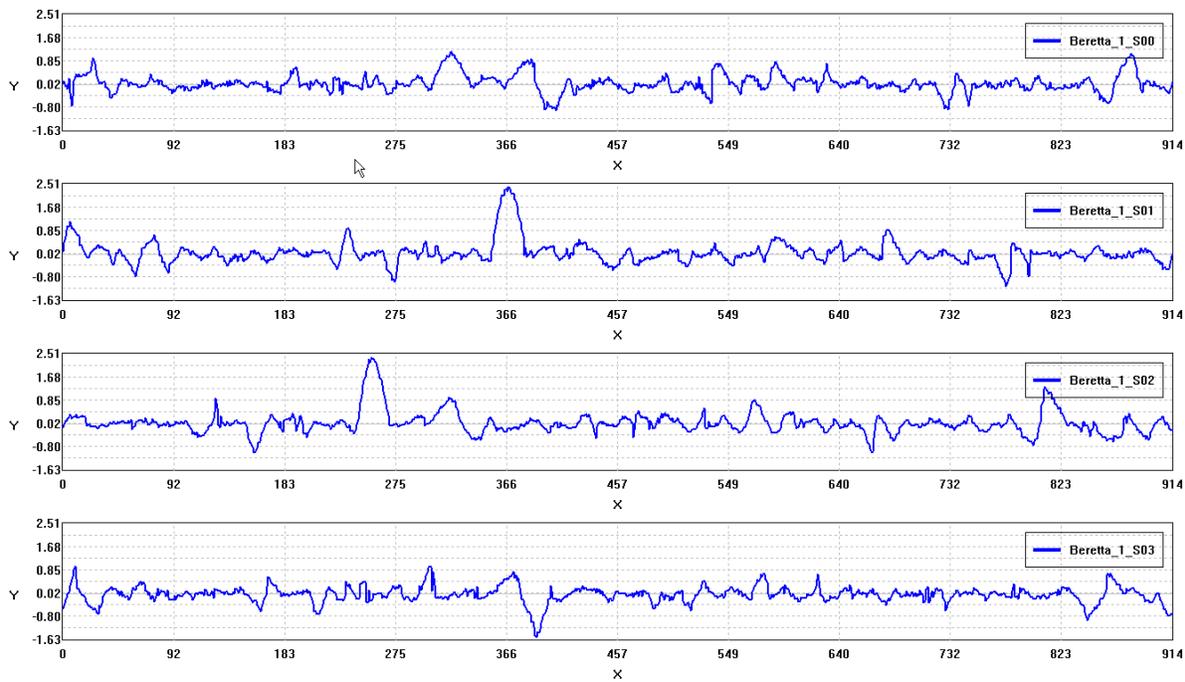
(a) Four matched LEA signature
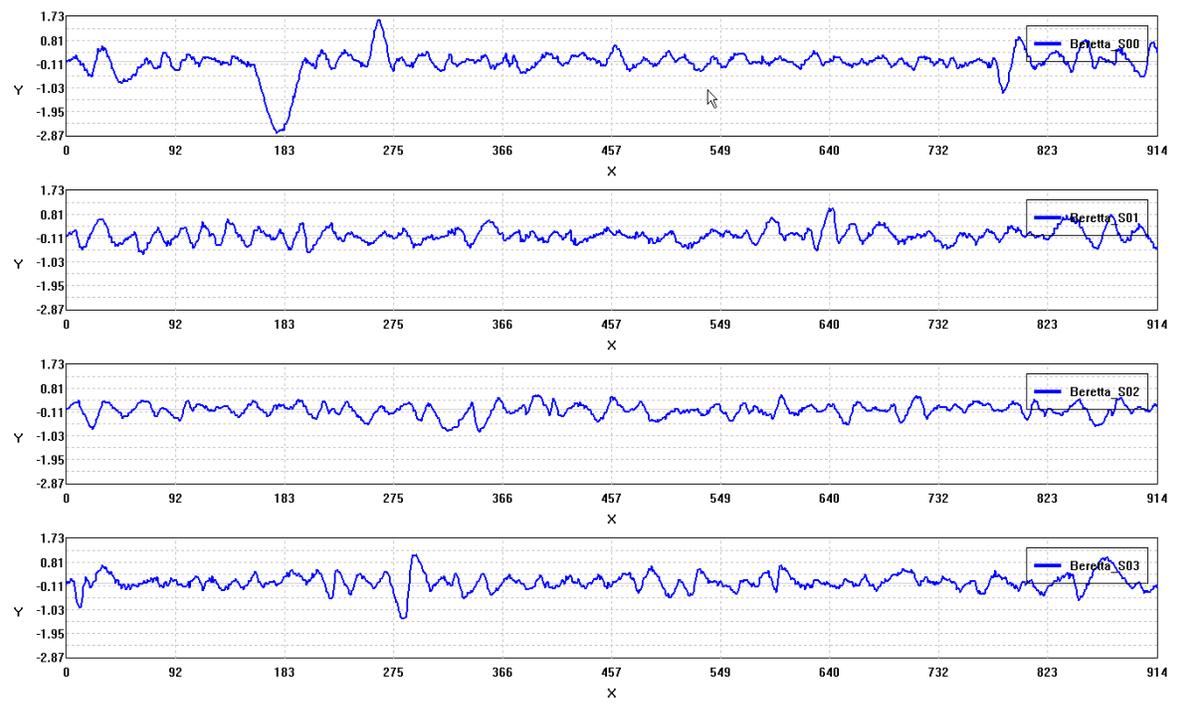
(b) Four non-matched LEA signature

Figure 48: Synthesized LEA signature of Beretta
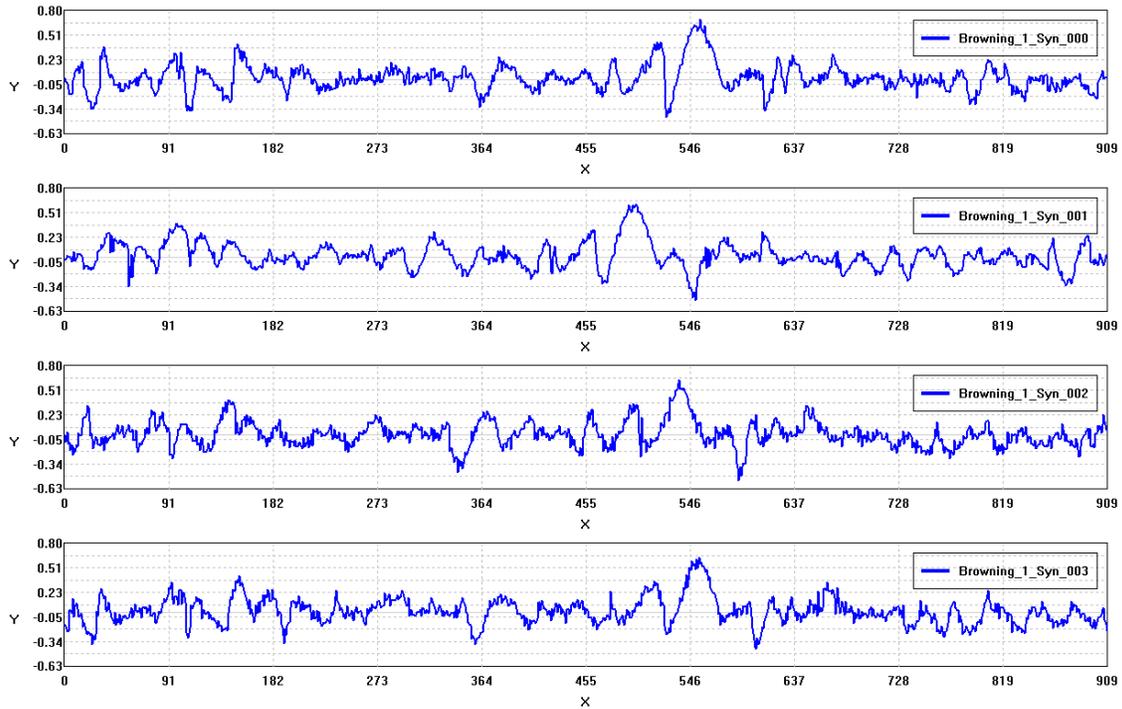
(a) Four matched LEA signature
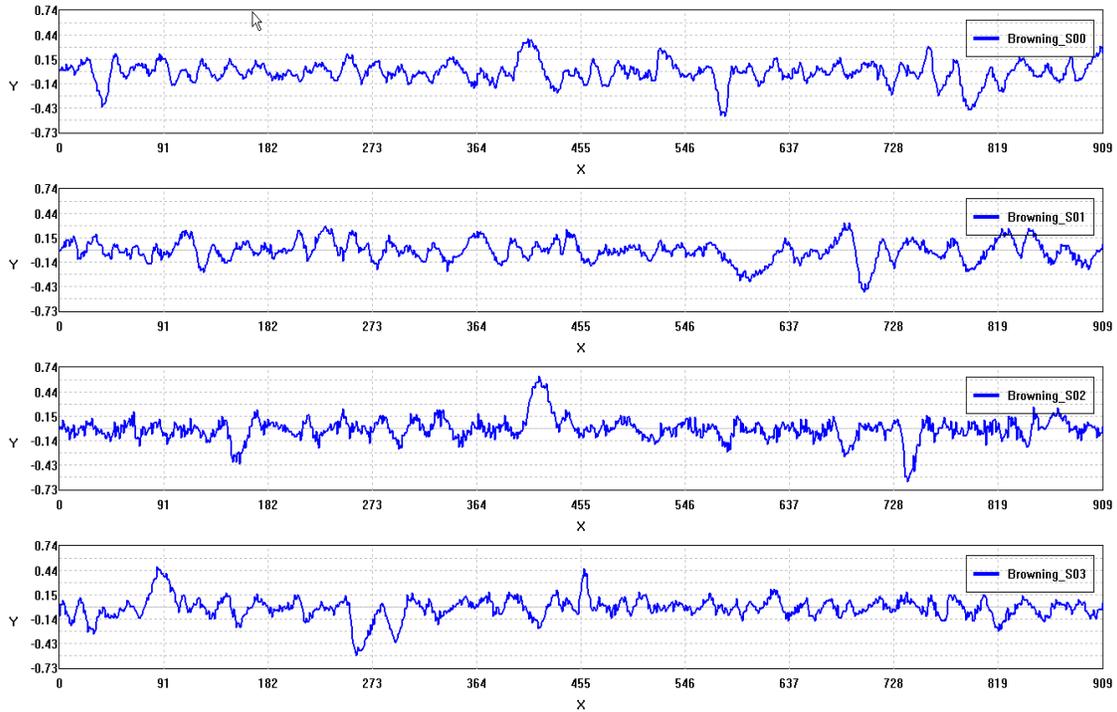
(a) Four non-matched LEA signature

Figure 49: Synthesized matched LEA signature of Browning

3   Results of Distribution Matching

We have applied the above optimization based distribution matching algorithm to the LEA signature synthesis. The optimization process gradually achieved the optimal results after about 200 iterations as illustrated in the plot below.
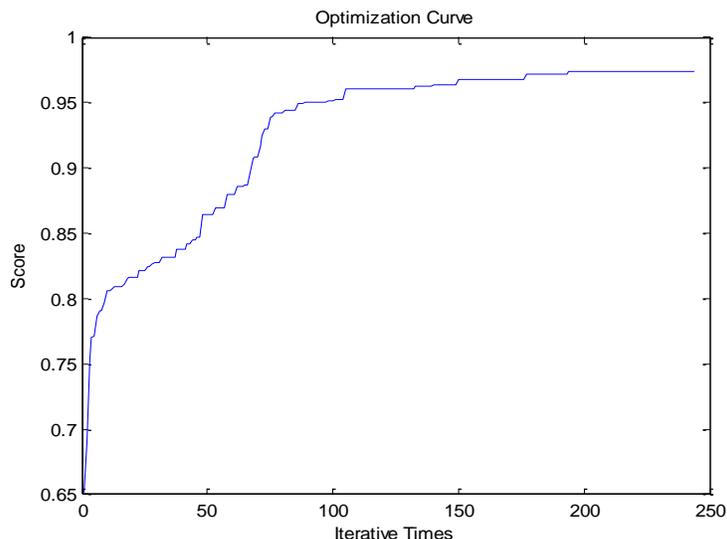
Figure 50: Optimization curve

For evaluation, we specified the target matching and non-matching correlation distribution, and run the optimization approach to synthesize 200 matched LEA signatures and the same amount of non-matched LEA signatures respectively. The results are shown below. Clearly we achieved very good matching for both the matching and non-matching distribution. After about 200 iterations, the matching score reaches as high as 0.9995.

(a) Non-matching distribution

(b) Matching distribution
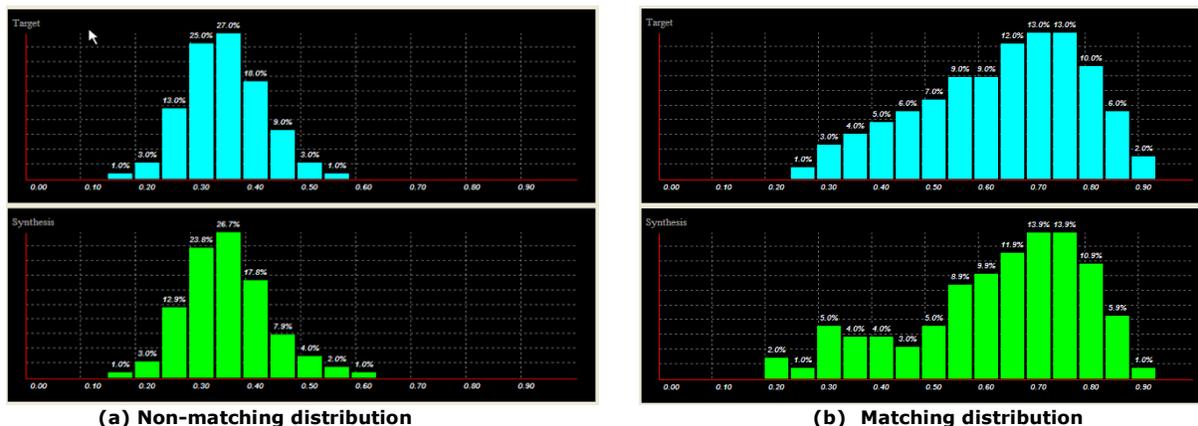
Figure 51: Distribution matching results for Beretta data synthesis

Please note that it will be difficult to achieve 100% matching due to several reasons. First, the distribution bin size is limited (only 20 bins are used); second, the amount of the synthesized data is limited (only 200 LEA data are generated); third, the matching cost function is not very sensitive to minor differences among bin values; fourth, several optimization exiting threshold values are not set to be high enough in order to achieve the balance between matching accuracy and computation time.

To test the global optimization of distribution matching, we generated 660 LEA signatures (11 barrels, each barrel contains 10 bullets, and each bullet has 6 LEAs) for the Beretta brand using the proposed approach. The results are shown in Figure 52. In the figure, (a) shows the target (top) and the synthesized (bottom) correlation distribution of both matching and non-matching, and (b) shows the target (top) and the synthesized (bottom) distribution curve. The statistical information about the correlation distribution curve is listed in the **Table 8**. Apparently, we have achieved very good matching performance between the target distribution and the synthesized distribution in both matching and non-matching correlation distribution.

Table 8: Statistical information of distribution curve

Item	Target		Synthesized	
	Matching	**Non-matching**	**Matching**	**Non-matching**
Entries	495	5500	495	5500
Max	0.851	0.471	0.859	0.559
Min	0.419	0.296	0.519	0.260
Median	0.699	0.386	0.724	0.383
Mean	0.693	0.387	0.717	0.384
Stdev	0.074	0.023	0.060	0.037
Threshold	0.460	0.470	0.500	0.550
Estimated Pe	0.31%	0.51%	0.05%	0.71%
Estimated Pfp	0.02%	0.00%	0.11%	0.00%
Estimated Pfn	0.61%	1.01%	0.00%	1.41%
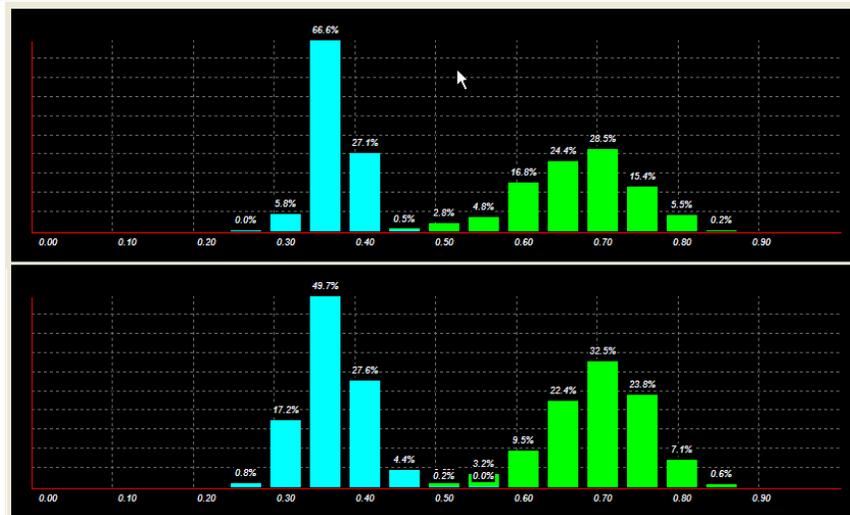
(a) Target (top) and synthesized (bottom) total distribution
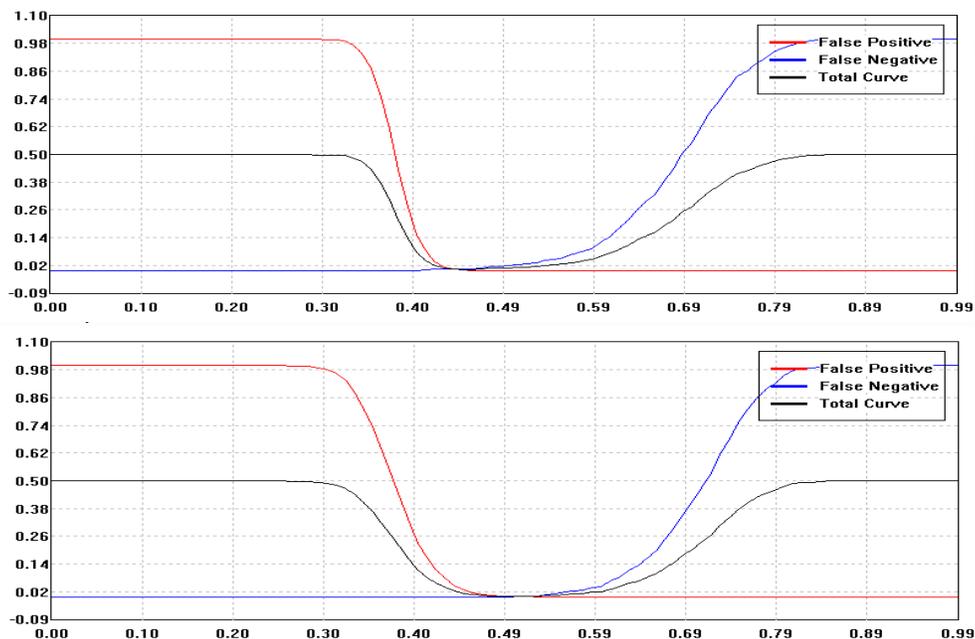
(b) Target (top) and synthesized (bottom) distribution curve

Figure 52: Distribution matching results for Beretta data synthesis

4 3D Composition

Figure 53 shows the results of synthesized 3D LEA surface from different perspectives. From the observation, we can see the relationship between the original signature and the synthesized 3D LEA surfaces.
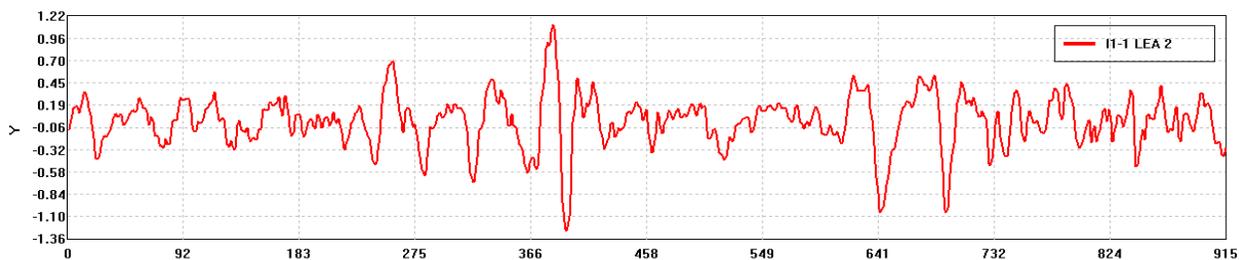
(a) Original LEA signature
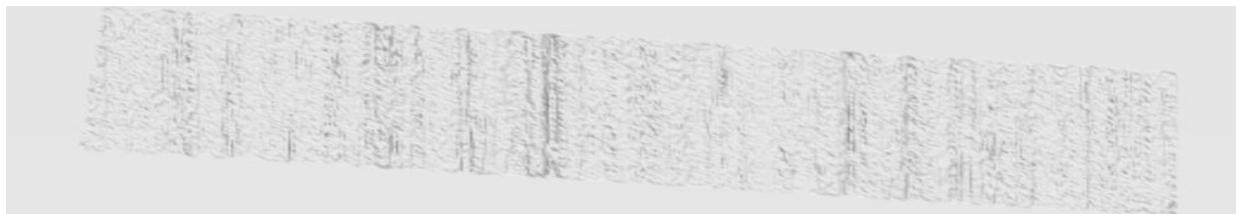
(b) Top view of the 2D Synthesized LEA

(c) 2D LEA overlaid on the 3D template of Beretta

Figure 53: Various views of original and synthesized LEA

IV.   Conclusions

It is a challenging problem to synthesize more data with limited bullet data while modeling the high-level common and low-level individual characteristics of bullets. In this project, we have successfully combined the strategies of feature identification and signal decomposition to form a two-step data synthesis approach. Our two-step data synthesis approach has been integrated with the wavelet decomposition framework to include deterministic and random components for data synthesis. Through observation of matched LEAs in groups, we successfully identified the zero-crossing point length, slopes, and peak-valley features as the deterministic component and fractal features as the random component. Our preliminary data synthesis results suggests that newly generated LEAs for a brand shows strong similarity to the LEAs in the same brand, but difference to those in different brands. Also, the correlation distributions for matched and non-matched LEAs in the same brand suggest the synthesized data approaches the same distribution as the experimental ones.

In the framework of LEA data synthesis, we developed different methods to improve the quality of data synthesis. These methods helped the distribution of the synthesized LEA data to approach the distribution of the original LEA data. We also extended the data synthesis process from a 1D LEA signature signal to a 3D LEA surface image. Our results look promising upon generating realistic LEA surfaces.

For the future work, we will continue to refine our data synthesis process so that the correlation distribution of the non-matched LEAs can be optimized to approach the original distribution. At the same time, we will continue our development on synthesizing 3D LEA surfaces so that we can present our synthesized results to firearm examiners for evaluation.

V.   References

[1]   Bachrach B. Development of a 3-D automated firearms evidence comparison system. J Forensic Sci 2002;47(6):1-12 [paper # JFS2002054_476]

[2]   B Bachrach, "A statistical validation of the individuality of guns using 3D images of bullets," research report, 2006, NIJ

[3]   Wei Chu, et. al., "Pilot study of automated bullet signature identification based on topography measurements and correlations," Journal of Forensic Sciences, Vol. 55, Issue 2, Page 341-347, 19 Jan, 2010

[4]   F. Xie, *et. al.,* "Automated bullet-identification system based on surface topography techniques," Wear, Vol. 266, Issue 5-6, 15 March 2009, pg. 518-522

[5]   Fernado Puente, leon, "Automated comparison of firearm bullets," Forensic Science International, Vol. 156, Issue 1, 6 January 2006, pg. 40-50

[6]   DAUBERT v. MERREL DOW PHARMACEUTICALS, 509 U.S. 579 (1993).

[7]   Niola Vincenzo, Buccelli Claudio, Gervasio Luca, Pilleri Michele, Policino Fabio, Quaremba Giuseppe, "The Impact Of The Wavelet Transform On The Problem Of Firearm Identification: First Evidences", WSEAS TRANSACTIONS on SIGNAL PROCESSING, Issue 2, Volume 2, February 2006

[8]   Jolliffe I.T. Principal Component Analysis, Springer Series in Statistics, 2nd ed., Springer, NY, 2002.

[9]   M. A. Turk and A. P. Pentland, Face recognition using eigenfaces. In Proc. of Computer Vision and Pattern Recognition, pages 586-591. IEEE, June 1991.

[10]   Abry, Patrice and Sellan, Fabrice. (1996) The Wavelet-Based Synthesis for Fractional Brownian Motion Proposed by F. Sellan and Y. Meyer: Remarks and Fast Implementation, Applied and Computational Harmonic Analysis **3**, 377-383.

[11]   Jaffard, S., Functions with prescribed Holder exponent," Applied and Computational Harmonic Analysis, Vol. 2, No. 4, 400{401, 1995.

[12]   Kar-Ann Toh; Wei-Yun Yau; Xudong Jiang; Tai-Pang Chen; Juwei Lu; Eyung Lim, "Minutiae data synthesis for fingerprint identification applications", in Proceedings of International Conference on Image Processing, Vol: 2, Page: 262 -265, 2001.

[13]   Yu Zhang and Shuhong Xu, Data-Driven Feature-Based 3D Face Synthesis, 3-D Digital Imaging and Modeling, 2007, Sixth International Conference on Issue Date: 21-23 Aug. 2007 On page(s): 39 – 46, Montreal, QC ISSN: 1550-6185.

[14]   Ye, H. and S. Young (2004). High Quality Voice Morphing. ICASSP 2004, Montreal, Canada.

[15] http://en.wikipedia.org/wiki/Fitness_proportionate_selection, Fitness proportionate selection, Wikipedia.

[16] Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin (2005), Introduction to Data Mining, ISBN 0-321-32136-7.

[17] http://en.wikipedia.org/wiki/Autoregressive_model

[18] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," TPAMI, Vol.14, No. 2, 1992

[19] Biasotti, A.A, and J.E. Murdock. 1997. Firearms and toolmark identification: Legal issues and scientific status. In Modern Scientific Evidence: The Law and Science of Expert Testimony, ed D.L. Faigman, D.H. Kay, M.J. Saks, and J. Sanders, 124 – 151. St Paul: West Publishing Co.

[20] Song, J.; Vorburger, T. V.; Ballou, S. M.; Ma, L.; Renegar, T. B.; Zheng, X. A.; Ols, M., "Traceability for Ballistics Signature Measurements in Forensic Science," Measurement, Vol 42, pp 1433-1438

Addendum: Response to Reviewers' Comments

1. Substantive Quality

Comment	Response
Reviewer 1: Are the findings supported by the research? Was the methodology appropriate and sound? *Yes, the findings are in agreement with supporting experimental data. Because this is a fair new research idea, there are not sufficient relevant literature was mentioned. Of specific mention is the need for good control point generation for base component, which seems critical and worth further investigation. The investigators also realized that they algorithm still could not well synthesize the non-matched LEAs*	Our algorithm generates control points based on zero-crossing and slop distributions that we extract from the experimental data. It is right that these control points are critical for data synthesis and the performance will depend on the bullet samples that are available. If there is a larger data size or some empirical knowledge, we could further improve the control point generation for each brand. Although the distributions for the non-matched LEAs do not match as well as that for the matched LEAs, given the limited data we collected, the overall performance of the LEA synthesis approach is good. However, the performance can be improved in the future by incorporating barrel characteristics from a larger dataset or expert knowledge.
Reviewer 1: Is the report well-written in terms of style, organization, and format? *The report is well-written and organized. Some minor issues such as putting the figure and captions on separate pages seems unusual, but should be fixed fairly easily.*	Thanks for the comment. We addressed the formatting issue in the final version.
Reviewer 1: Classify the overall quality as one of the following: Poor, Fair, Good, Excellent. *The quality is very good to excellent. It would have been nicer to see a somewhat move validations on success using the synthetic LEA to represent the real LEA. It would be the best if they are willing to disseminate the algorithms to researchers in the academic community to speed up the new algorithm developments for firearm and toolmark*	We are willing to share our research findings with other parties if an appropriate Non-Disclosure Agreement is in place. Also, we will present our work in the 2012 Impression & Pattern Evidence Symposium (IPES 2012) on Aug. 5-9, 2012 in Clearwater, FL.

2. Implications of the Research

Comment	Response
What are the implications, if any, for further research, program development, and evaluation efforts? *Reviewer 1: The findings showed that for the proposed approach worked well for matched LEAs, whilst does not for non-matched LEAs due to the dissimilarity among non-matched LEAs. Another effort, as indicated by the investigators, shall be focused on synthesize 2D and 3D LEAs, which will a lot more complex than 1D cases, but will be of significant interest.*	We developed and implemented algorithms for 2D/3D LEA synthesis in the final period. Please see the following sections in the final report: • Accomplishments->II. Methods->6. Two-dimensional LEA Imagine Synthesis. • Accomplishments->III. Results->4. Results of 3D Composition.

3. Relevance

Comment	Response
How well does this report/deliverable address the practitioner need as stated above? *Reviewer 1: To my knowledge, this reported work is still in its early research stage and is not ready to address the practitioner need.*	Agree. We are looking forward to continuing our research.

4. Recommended Revisions

Comment	Response
Reviewer 1: Some minor revisions are recommended concerning the dissemination of this reported work to benefit the society, and the format of the report. The current report does not include any dissemination. *Reviewer 2: typographical revisions suggested.* *Reviewer 3: literature survey of ballistic identification system.*	Revisions and literature survey have been addressed/added in the final version. Also see the publication list at the end of this addendum.

5. General Comments

Comment	Response
Reviewer 2: This is a highly technical and specific approach to advance the utility of non-human comparison methodologies by developing a method to produce synthetic striated tool mark samples, such as those found in the land engraved areas of fired bullets This approach is limited to striated tool marks only and further limited by the eventual utility of any such automated system for comparing fired bullets which is less practically relevant than a similar tool for comparing cartridge cases that are easier to recover at a shooting scene and less likely to be damaged. Another concern pertaining to relevancy is the gap between genuine tool mark samples and synthetic ones meant to model them. It's always going to be a criticism that limits the relevance to actual casework so ultimately, any approach using synthetic samples will have to be redone at its conclusion with a data set of actual striated tool marks for final validation.	This is an open question regarding applications of the data synthesis methods developed under this grant. We look forward to extending our research work to other areas, as well as performing validation by presenting synthetic LEAs to SMEs.
Reviewer 2: Beyond those global limitations, the approach made use of multiple types of source data for modeling purposes (i.e.- multiple types of firearms representing class characteristic variations) and gave correlation distributions that suggest this modeling was done properly. Features that represent class and individual characteristics were handled separately allowing for more specific modeling of individual brands of firearms based on these differences. Ultimately, the ability to produce a diverse and large set of synthetic/simulated tool mark data of this type has general research applications and more specifically can be used to test the functionality of automated ballistic identification (ABI) systems used to aid examiners in their comparisons or function as database comparators for searching an unknown against large bodies of data.	Agree.
Reviewer 2: The varied techniques used in this project to examine, alter, represent, and evaluate the data features were elegant and well assembled. Charts assembled to illustrate the relationships between false positive and false negative comparison outcomes and approaches to probe Z-axis noise that can affect a comparison were well considered. The graphical user interface (GUI) as represented appears user friendly.	Thanks for the comment.
Reviewer 2: Ultimately, this work is a successful step along a lengthier path towards working with ABI systems and optimizing their performance with large synthetic data sets. Presumably, Intelligent Automation Incorporated will continue along that path. Of concern is the lack of dissemination of this work which could inspire other similar approaches or be used as a technique by others to continue along this same research path.	Thanks for the comment.

Publications, conference papers, and presentations

- Roger Xu, et al., "Development of Synthetically Generated LEA Signatures to Generalize Probability of False Positive Identification Estimates," 2012 Impression & Pattern Evidence Symposium, Aug. 5-9, 2012, Clearwater, FL

- Benjamin B, et al., "Development of Synthetically Generated LEA Signatures to Generalize Probability of False Positive Identification Estimates," 2010 Impression & Pattern Evidence Symposium, Aug. 2-5, 2012, Clearwater, FL