Abstract: Signature verification and recognition can be divided into online and offline, depending on the sensing modality. In an online method, the dynamic information signature features such as direction, slant, baseline, pressure, speed and numbers of pen ups and downs can be captured. Method of extracting features signature depends on the requirement features to be extracted. Thus this paper discusses the construction of an algorithm to extract the baseline from signature. The Signature Extraction Features System (SEFS) uses the proposed algorithm and provides a set of tools that allow the systems to extract baseline features in signature automatically for analysis purposes. Signatures are taken from twenty randomly selected individuals with different background. These individuals will have to sign their signatures on the tablet and the SEFS system will gather and store the raw data. SEFS created the image of each signature and these images were used as samples for a developed questionnaire to be given to human expert for evaluation. These questionnaires are all about identifying the baseline of the signatures. Both results from SEFS and the questionnaire are compared, and it shows that the algorithm is 90% accurate. It can be concluded that the algorithm proposed are acceptable to represent extraction of signature features based on baseline.

Keywords: Signature Verification, Signature Identification, Baseline, Dynamic Signature, Signature Analysis

1 Introduction
Signature is one of a biometric modality that probably the most practiced handwriting in everyday life, the act of signing is almost entirely subconscious and habitual, which means that signature dynamics is extremely individualistic [4]. A signature is a handwriting that people writes on documents as a proof of identity. Signatures are among the most widely accepted personal attributes for identity verification. Signatures reflect quite everything related to the specificity of the biomechanical system used to produce the message, physiological and even the psychological state of the signer [12]. Semantically, it may represent the name of the signatory, but it also reflects the intrinsic and unique features of the signer.

Signature had been used in many applications such in signature authentication [7], signature verification and recognition [11], forensic authentication [1] and many more. Generally there are two categories of signature verification and recognition, there are offline, which signature is usually scanned from paper documents where they were written, and online, which the signatures are written using an electronic instrument device that read the information regarding dynamic characteristic (pen tip location when signing). Usually, online signature systems can also measure pen angle and contact pressure [2]. In order to analyze an online signature, several features need to be extracted from the signature. These features namely are direction, slant, baseline, pressure, speed and numbers of pen ups and downs. Each of these features requires a method or algorithm to extract the raw data of the signature [8, 10, 14, 15].

The aim of this paper is to create an algorithm for detecting baseline in an online signature. Section 2 would discuss signature recognition in general and emphasizing the baseline feature and its assessment. Subsequently, would discuss related work on handwriting and signature algorithm. Meanwhile, section 3 would explain the approach and method that were employed in this research. In section 4, discussion about the online baseline algorithm and followed by the findings and results of the research with respect to the human experts evaluation. Finally in Section 6, some conclusions about the results presented in this research would be discussed besides future directions of research inspired by these results.
2 Related Works

In analyzing signatures as a means of establishing or verifying identity requires the need to formulate robust algorithms for automatic signature verification and for the powers of human perception [3]. Several features analyzed in signature recognition are direction, slant, stroke, pressure, baseline, caliber and shape of individual’s signature. Thus, this paper would explore further the extraction of baseline feature.

Baseline is the imaginary or invisible line, which a signature is assumed to rest on. Baseline is normally straight, ascending or descending baseline. Normally straight baseline is where a baseline is horizontally straight. Ascending baseline is where a baseline is inclining upwards while descending baseline is where the signature is moving downwards. The example of normally straight baseline, ascending and descending baseline are shown in Figure 1.

![Ascending Baseline](a)

![Normally Straight Baseline](b)

![Descending Baseline](c)

Figure 1: (a) Ascending Baseline (b) Normally Straight Baseline (c) Descending Baseline

A baseline is the line on which the letter sits. In our daily life, a baseline must be imagined when signing or writing in an unlined sheet of paper. The straightness and direction of the signature can be changeable features in a signature. For an individual the rise and fall of lines is usually a constant feature. It can be changeable or varies according to the conditions of writing and the focus or attention being given during the process of signing or writing. The emotions and feeling of the individual will also affect the signature.

Many different approaches have being used in order to extract discriminative information from online signature data [13]. The existing methods can be broadly divided into two classes, which are feature-based approaches and function-based approaches. Features-based approaches use a holistic vector representation consisting of a set of global features derived from the signature trajectories [6]. While function-based approaches describe local properties of the signature using time sequences for recognition [3, 5, 9].

In another study by Madabusi et al., [8] presents a relative slope based algorithm for online and offline signature verification. They used a slope based model in which the input signature is divided into many segments using optimized Hidden Markov Models method and then the slope of every segment are calculated with respect to its previous segment after the normalization process of a signature. The work by Tong et al. [15] present a stroke based algorithm for dynamic signature verification. Their algorithm is developed to convert sample signatures to a template by considering their spatial and time domain characteristic and also by extracting features in term of individual strokes. They proposed system call Dynamic Signature Verification (DSV) that consists of four subsystems which is data acquisition, signature preprocessing, feature extraction and signature verification.

3 Approach and Method

There are two major approaches that was employed in this research namely prototype development and conducting experiment.

3.1 Prototype development

In conducting this approach, a few steps have been identified and it is as discussed in the subsection below.

3.1.1 Data Acquisition

The WACOM Model CTE-440 tablet and pen is used in this research. The tablet is capable of sampling data at 100 samples per second. The Wacom’s pen captured samples during the interaction of the pen tip with the tablet. The raw data available from Tablet Pen consists of two dimensional series data which are \(x\) and \(y\) coordinates of signature’s route which are recorded and representing the pen position. In the online signature recognition, stroke sequence strings
are available in the online signature acquisition process. The next step is to group all points into strokes to create image of signature.

3.1.2 Preprocessing
In the preprocessing phase, the position points are counted from when the pen is down until it is lifted up. If the amount of points of data is less then 10 then the set points of data would be discarded from the process of extracting features. It would reduce processing time where only the possible set points of data that have extracted features would be processed.

3.1.3 Features Extraction and Classifications of Baseline
After the preprocessing steps, features are extracted from the pen position respect to the x-axis and y-axis. An algorithm is proposed in this study to extract baseline signature features. Baseline has three attributes, which are normally straight, ascending or descending baseline. Normally straight baseline is where a baseline is horizontally straight. Ascending baseline is where a baseline is inclining upwards while descending baseline is where the signature is moving downwards.

3.2 Conducting Experiment
Meanwhile the next subsection would explain the processes involved in conducting experiment to test the algorithm constructed.

3.2.1 Collection of Signature through Random Sampling
There are a few steps in the process of gathering and collecting data through sampling signature. Firstly, 20 individuals are randomly selected to have their signature taken. These people would have to sign their signatures on the tablet and the SEFS system would eventually gathered the raw data. The data would be stored in binary files that are readable by the SEFS system for future analysis. The image of the signature is created by the SEFS system based on those signatures. The images would be used as samples for the questionnaire to identify the features of baseline, where the questionnaire would be given to human expert for evaluation purposes.

3.2.2 Creation of Questionnaire from Individual Signature Sampling Data for Baseline Identification
The questionnaire was created base on image produce by the SEFS system. It contains 3 parts. The first part contains the introduction and explanation of baseline. The second part comprises of questions regarding the sample signatures and the last part is verification of questionnaire. Questionnaires are distributed to several experts from different organizations as listed in table 1.

Table 1: List of Organizations contributes in Questionnaire

<table>
<thead>
<tr>
<th>No.</th>
<th>Organization</th>
<th>No of Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>BANK ISLAM MALAYSIA BERHAD</td>
<td>6</td>
</tr>
<tr>
<td>2.</td>
<td>AFFIN BANK</td>
<td>8</td>
</tr>
<tr>
<td>3.</td>
<td>MALAYAN BANKING BERHAD (MAYBANK)</td>
<td>6</td>
</tr>
<tr>
<td>4.</td>
<td>BANK RAKYAT</td>
<td>4</td>
</tr>
</tbody>
</table>

3.3 Measurement and analysis
The next research method is the Measure and Analyze procedure. It comprises of the following steps. Firstly the SEFS system would analyze the stored data. It would then store the extracting features of the baseline. The next step is where the result of the questionnaire would be gathered for analysis and comparison. Analyzed result from questionnaire would be used to adjust the classification of baseline in order to suite as close as possible with human expert’s judgment. The testing would be done several times until satisfied classified results are produced.

In the final process of analyzing the data, the results regarding the features of baseline from the SEFS system are compared with the result from the questionnaire. Analysis would be made on the number of similarity and differences between the results from the prototype and from the questionnaires.

4 Baseline Algorithm
The baseline is the invisible line used to guide a signature. Some baselines are relatively straight while others are ascending or descending baseline.

Figure 2: The signature Image with a drawn Baseline
Figure 2 shows how baseline is drawn to examine the properties of baseline in signature. The calculation is based on the angle of start point of baseline to the last point of baseline.

4.1 Raw Data Collection
The raw data available from tablet (WACOM) consists of two dimensional series data:
\[(x_i, y_i) \in \mathbb{R} \quad i = 0,1,2,3,4,5,6,7,8,\ldots, I\]  

Where \((x_i, y_i) \in \mathbb{R}\) is the pen position with respect to the \(x\)- and \(y\)-axis. When \(i\) is equal to zero, it represents the starting point of the data. The raw data gathered from the tablet shows that the smaller value in \(y\)-axis stand as the higher position. For \(x\)-axis, the higher value is towards the right site compared to left.

4.2 Design of Algorithm to Identify Baseline
The baseline is the invisible line used to guide a signature. Some baselines are relatively straight while others are ascending or descending baseline.

4.3 The Angle and Degree of Baseline
For baseline extraction in signature, there are three possible positions that will categorize the type of baseline namely ascending, straight and descending. The technique used to get the degree of baseline is the same as the technique used to calculate the degree of two points in slant identification [16]. However for baseline, the two points in this case are based on the starting and end points of the baseline. Where \(x_i\) and \(x_{i-1}\) refers to the ending baseline point and starting baseline point of \(x\)-axis and \(y_i\) and \(y_{i-1}\) refers to the ending point and starting point of \(y\)-axis. The calculations to get the degree are as shown in Equation 2 up to Equation 4.

Let
\[
\begin{align*}
dy(i) &= y_i - y_{i-1} \\
dx(i) &= x_i - x_{i-1} \\
i &= 1,2,3,4,5,6,7,8,\ldots, I
\end{align*}
\]  

The Angle, of each two point of pen position in radians are given by
\[
\text{Angle} = \arctan\left(\frac{dy(i)}{dx(i)}\right) \quad (3)
\]

Since the Angle is positive when measured counterclockwise and negative when measured clockwise, the degree of each two points are given by
\[
\text{Degree, } \theta_i = \left(\text{Angle} \times 180\right) / \pi \quad (4)
\]
where \(\pi = 3.142\), radians = Degree \((\pi / 180)\).

In this algorithm, the degrees for each baseline have only three possible options since baseline angle is only counted from the left to the right of the signature. The three possible degree, \(\theta_i\) for baseline in this research uses the selected equation below

Right & Up, degree, \(\theta_i = \theta_i\)  
where \(dy(i) < 0\), \(dx(i) > 0\)  
(5)  

Right & Down, degree, \(\theta_i = 360 - \theta_i\)  
where \(dy(i) > 0\), \(dx(i) > 0\)  
(6)  

Right, degree, \(\theta_i = 0^\circ\)  
where \(dy(i) = 0\), \(dx(i) > 0\)  
(7)

4.4 The Algorithm
The algorithm to detect baseline begins by processing raw sequential data from beginning to the end of a signing process. From the raw data, the lowest and highest point in the \(x\)-axis is located to represent the width of the signature. The reference distance is calculated by getting 15% of the difference in the width of the signature. Then, locate the lowest point of the \(y\)-axis from the lowest point of \(x\)-axis up to the length of reference distance in order to find the start baseline point of \(x\) and \(y\). Next will be to find the lowest point of the \(y\)-axis from the highest point of \(x\)-axis up to the length of reference distance to find the end baseline point of \(x\) and \(y\). Calculation are then conducted on the degree of baseline from the start to the end of baseline points. Lastly the baseline will be categorized based on ascending baseline, normally straight baseline and descending baseline. A baseline is categorized as normally straight baseline when the degree of baseline is in the range of less than 4 or more than 356 degrees. If the degree of baseline is more or equal to four, it will be categorized as ascending baseline while descending baseline is when the baseline is less than 356 degree.

5 Findings and Discussions
Table 2 shows the comparison results for baseline between the SEFS system and the answer to the questionnaire by human expert. Out of the twenty questions, two answers are not identical which are
question eleven and thirteen. As for question six, 100% of human expert classified the signature as upright baseline, the same as being classified by the SEFS system. Overall, the below table shows 90 % of the results give identical answers between the SEFS system and the human expert.

Table 2 Result of Comparison Between SEFS vs Questionnaire for Baseline

<table>
<thead>
<tr>
<th>Data</th>
<th>SEFS</th>
<th>Questionnaire Result</th>
<th>SEFS vs Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Ascnd</td>
<td>Strgh</td>
</tr>
<tr>
<td>1</td>
<td>Strgh</td>
<td>8.3%</td>
<td>75%</td>
</tr>
<tr>
<td>2</td>
<td>Ascnd</td>
<td>87.5%</td>
<td>8.3%</td>
</tr>
<tr>
<td>3</td>
<td>Strgh</td>
<td>41.7%</td>
<td>45.8%</td>
</tr>
<tr>
<td>4</td>
<td>Strgh</td>
<td>12.5%</td>
<td>70.8%</td>
</tr>
<tr>
<td>5</td>
<td>Ascnd</td>
<td>79.2%</td>
<td>4.2%</td>
</tr>
<tr>
<td>6</td>
<td>Strgh</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>7</td>
<td>Strgh</td>
<td>4.2%</td>
<td>87.5%</td>
</tr>
<tr>
<td>8</td>
<td>Ascnd</td>
<td>70.8%</td>
<td>0%</td>
</tr>
<tr>
<td>9</td>
<td>Ascnd</td>
<td>58.3%</td>
<td>41.7%</td>
</tr>
<tr>
<td>10</td>
<td>Strgh</td>
<td>25.0%</td>
<td>66.7%</td>
</tr>
<tr>
<td>11</td>
<td>Strgh</td>
<td>75.0%</td>
<td>12.5%</td>
</tr>
<tr>
<td>12</td>
<td>Strgh</td>
<td>37.5%</td>
<td>54.2%</td>
</tr>
<tr>
<td>13</td>
<td>Dscnd</td>
<td>70.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td>14</td>
<td>Ascnd</td>
<td>79.2%</td>
<td>8%</td>
</tr>
<tr>
<td>15</td>
<td>Ascnd</td>
<td>75%</td>
<td>10%</td>
</tr>
<tr>
<td>16</td>
<td>Ascnd</td>
<td>70.8%</td>
<td>25%</td>
</tr>
<tr>
<td>17</td>
<td>Strgh</td>
<td>8.3%</td>
<td>87.5%</td>
</tr>
<tr>
<td>18</td>
<td>Ascnd</td>
<td>70.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td>19</td>
<td>Ascnd</td>
<td>79.2%</td>
<td>12.5%</td>
</tr>
<tr>
<td>20</td>
<td>Ascnd</td>
<td>79.2%</td>
<td>4.2%</td>
</tr>
</tbody>
</table>

Strgh – Straight; Ascnd – Ascending; Dscnd - Descending

The non-identical signature for signature data number 11 in table 2 is as shown in figure 3 below.

Figure 3 The Image of Signature for Signature 11

Meanwhile the second non-identical signature is signature data number 13 in table 2 is as shown in figure 4 below.

Figure 4 The Image of Signature for Signature 13

Table 2 shows the results of the baseline extraction from the SEFS system, which indicate that it is descending baseline. While from the questionnaire, ascending baseline is the most chosen type of baseline for the signature. The human eye seems to see the signature baseline based on every character except for the last character of the signature while in SEFS system the baseline is taken from the whole signature as shown in figure 4. The inclusion of the last character of this signature will affect the overall results of the extracted baseline.

6 Conclusion and Future Works

This research introduces a new algorithm for extracting baseline in signature. The algorithm for extracting baseline has been successfully implemented as demonstrated by the SEFS system. The algorithm of extracting baseline has variable values that can be adjusted to work according to the required application. It shows the flexibility of this algorithm to adapt to requirements regarding baseline extraction. Results from the SEFS system that uses the algorithm are compared with the human expert judgment in questionnaire to show that the algorithm works and produced similar results. Results produced by the algorithm for baseline extraction show 90% identical answers compared with the judgment by human expert. With these results, it can be concluded that the algorithm is a success.

This proposed algorithm will help community in the field of signature verification, signature analysis and signature recognition. It can be used as a tool to extract the required signature features depending on application requirement. In character recognition or handwritten recognition, the algorithm for baseline extraction can be used as baseline correction before the character can be recognized.

The preprocessing steps used in this algorithm might be enhanced to increase its accuracy. Further preprocessing techniques can be explored to reduce noise or jitter during data acquisition before
proceeding to extract the baseline feature. Technique to adjust and realign any small movement of lines towards different direction during signing process should also be considered. Besides that, further study on how human expert look and classified the baseline will help the process of classification. Additional study on classification of baseline in the algorithms may produce a better decision.

Acknowledgement: This research is funded by the Ministry of Higher Education of Malaysia through the Fundamental Research Grant Scheme with the help of Research Management Institute of Universiti Teknologi MARA.

References: