Adaptive Feature Thresholding for Off-line Signature Verification

Robert Larkins, Michael Mayo

Machine Learning Group, Department of Computer Science, University of Waikato, New Zealand.
Email: {rll6,mmayo}@cs.waikato.ac.nz

Abstract

This paper introduces Adaptive Feature Thresholding (AFT) which is a novel method of person-dependent off-line signature verification. AFT enhances how a simple image feature of a signature is converted to a binary feature vector by significantly improving its representation in relation to the training signatures. The similarity between signatures is then easily computed from their corresponding binary feature vectors. AFT was tested on the CEDAR and GPDS benchmark datasets, with classification using either a manual or an automatic variant. On the CEDAR dataset we achieved a classification accuracy of 92% for manual and 90% for automatic, while on the GPDS dataset we achieved over 87% and 85% respectively. For both datasets AFT is less complex and requires fewer images features than the existing state of the art methods, while achieving competitive results.

Keywords: off-line signature verification, person-dependent, feature thresholding, spatial pyramid

1 Introduction

Hand written signatures have a well-established and accepted place in society as a formal means of personal verification, for both the identification and the intent of the signatory. Because of this, signatures are the most accepted method of verification [1] and are used in government, legal and commercial transactions. A result of this is that signatures are often forged for the purpose of feigning the authenticity of a document. This leads to the problem of being able to correctly verify whether a signature is a genuine or a forgery.

Many different approaches have been employed for accurately classifying whether a signature is a genuine or a forgery. These approaches are split into two categories: on-line and off-line.

Online approaches use a digitising surface to capture dynamic features about how a signature is written. These are features such as pressure, speed and direction, which allow online classification to achieve accuracies of over 95% [2].

Off-line verification deals with signatures that have been written on paper and scanned in to the computer. Because of this, they are unable to use dynamic features. This means that the signatures can only be distinguished from each other by what is visually available.

Person-dependent classification is a commonly used method that is designed to train with genuine signatures from only one person. Two methods that have achieved high person-dependent classification accuracies are graph matching [3] and the Discrete Wavelet Transform (DWT) [4].

The approach that we use is similar to the Gradient, Structural and Concavity (GSC) method [5], except we only use the gradient direction. The improvement that dramatically increased classification accuracy was the way in which we implemented the thresholding of each image feature count. This enhanced the creation of the binary feature vector [6] by adaptively restricting which feature bits are set to 1. In GSC, thresholding is carried out with a fixed value that is manually chosen, with its capability being determined by experimentation, where if the count of an image feature is above this value, then the feature bit is set to 1, otherwise it is set to 0. An adaptation is then made to our thresholding method to achieve automatic classification. As well as these methods, a novel combination of spatial pyramids [7] and equimass sampling grids [8] is also introduced to help boost the classification accuracy.

2 Signature Representation

Each signature, before it is processed, is in the form of a binarised digital image (see Figure 1). This format does not describe the individual aspects of

the signature in a manner that makes it feasible for comparing it to other signatures. This is because digital images are designed to be visually identifiable to humans. As a result, each signature needs to be converted into a format that will allow the similarity of it and another signature to be easily computed.

This section details the implementation of the novel thresholding method, AFT, that this paper presents. Essentially, this method ensures that the comparative similarity of a signature is more accurately represented in contrast to the training signatures when it is converted from a digital image to a binary feature vector.
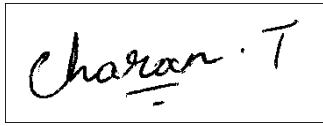

Figure 1: A signature that has been binarised.

2.1 Binary Feature Vector

A binary feature vector is a method of representing a signature by indicating whether a particular feature matches a certain criteria by turning the corresponding feature bit on. The vector structure is shown in equation (1), where V is the feature vector, z is a feature bit and k is the number of elements in the vector.

$$V = (z_1, z_2, \cdots, z_k) \qquad z_i \in \{0, 1\} \qquad (1)$$

2.2 Gradient Direction Extraction

The creation of a binary feature vector is based solely on the gradient direction of each pixel from across a signature. This direction θ of a pixel at coordinates x and y is found by equation (2), where G_x is the Sobel kernel for horizontal change and G_y is the kernel for vertical change.

$$\theta(x, y) = \tan^{-1}\left(\frac{G_y}{G_x}\right) \qquad (2)$$

The resulting direction is a value that ranges from 0.0 to 2π radians. This range can then be split into 18 non-overlapping segments based on $\frac{2\pi}{18}$ radians, allowing a gradient direction histogram to be created from the count of each direction. For the experiments conducted in this paper, 18 segments were chosen because this value proved effective in initial tests and is also used in [9].

2.3 Equimass Spatial Pyramids

Using the gradient direction in its current state only expresses a signature at the global level. This can be improved upon through the use of a spatial pyramid, which defines the signature at increasingly finer levels of granularity, improving the ability to distinguish a signature in relation to other signatures. This is normally achieved through the use of different types of image features which capture particular properties of the signature, but the use of spatial pyramids helps to overcome this need for multiple features.

The levels of granularity are produced by splitting a signature up into increasingly smaller regions, which is usually carried out with a uniform grid. The disadvantage of a uniform grid is that it does not capture the same structural properties of corresponding regions between signatures. A novel and effective approach that improved the capture of these structural properties was achieved by combining spatial pyramids and equimass sampling grids.

Equimass is an adaptive grid based on the number of black pixels or mass M of a signature, where the grid lines are found at the equimass divisions of the horizontal and vertical mass histogram. That is, where the masses between all adjacent points on either the x-axis or the y-axis are equal. This average mass M_A is found by equation (3), where r is the number of horizontal or vertical regions.

$$M_A = \frac{M}{r} \qquad (3)$$

Figure 2 shows how the grid lines will be placed for each of the three levels of the spatial pyramid if the number of regions for the finest level (c) is 8×4. An example of this calculation is if Figure 2(a) has mass $M = 2709$, then the single vertical line will be placed where the number of black pixels in both regions is $\frac{2709}{2} = 1354$.
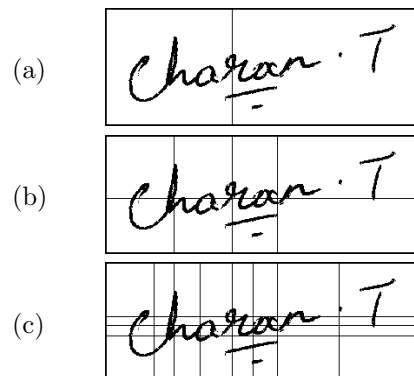
(a)

(b)

(c)

Figure 2: A three level spatial pyramid.

2.4 Novel Feature Thresholding

When thresholding a particular gradient direction of a region, the criteria for determining whether the corresponding feature bit will be 0 or 1 is dependent upon the training signatures, where if the direction count c of a set of pixels abides by $\tau_1 \leq c < \tau_2$,

then the feature bit is set to 1, otherwise it is set to 0.

The lower threshold, τ_1 and the upper threshold, τ_2 are found at one sample standard deviation S either side of μ, where μ is the mean of D, in which D is the count of a particular direction from the same region across each training signature. The problem with using S to calculate both τ_1 and τ_2 is that any possible skew is not taken into account, as the values from D may be spread about μ in a fashion that is not normally distributed. So to adjust for the skew, τ_1 is calculated by the equation (4) and τ_2 is calculated by (5).

$$\tau_1 = \mu - S_L \qquad (4)$$
$$\tau_2 = \mu + S_U \qquad (5)$$

The variable S_L is the average distance that values below μ lie from μ and is found by equation (6). S_U is found in a similar fashion but for values above μ, this is calculated by equation (7).

$$S_L = \sqrt{\frac{1}{D_L - 1}\sum_{i=1}^{n}(D_i - \mu)^2} \quad \forall D < \mu \quad (6)$$

$$S_U = \sqrt{\frac{1}{D_U - 1}\sum_{i=1}^{n}(D_i - \mu)^2} \quad \forall D > \mu \quad (7)$$

D_L is the number of values in D which are less than μ, D_U is the number that have a value greater than μ and n is the size of D.

For example, if $D = \{36, 47, 54, 59, 63, 81\}$ then μ would equal the mean value of D which is 56.67. S_L then equals 16.24 and is calculated by equation (6) using the values 36, 47 and 54, the values from D which are less than μ. S_U equals 17.86 and is found by equation (7), using the values 59, 63 and 81, the values greater than μ. τ_1 then equals 56.67 - 16.24 = 40.43 and τ_2 would equal 56.67 + 17.86 = 74.53. Therefore the corresponding feature bit will be 1 if $40.43 \leq c < 74.53$, otherwise it will be 0.

Using this method, each direction in a gradient direction histogram can be thresholded independently and used to produce one bit of the feature vector. This is then repeated for the histogram of each region across each spatial level for all signatures. Using this thresholding method on the signature in Figure 2 would produce a feature vector that is 756 feature bits in length, consisting of 36 bits for (a), 144 bits for (b) and 576 bits for (c).

3 Signature Classification

The classification of an unknown signature is based heavily on the similarity score of two feature vectors. The comparison of these two vectors produces

$$
\begin{aligned}
V_a &= 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 0\ 1 \\
V_b &= 1\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1 \\
V_c &= 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1 \\
V_u &= 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 1
\end{aligned}
$$

Figure 3: A set of binary feature vectors of length 18

four values based on the sum of the four possible variations at each position in the vectors. These four values are defined by equation (8), where C is the count of times that each of the four outcomes occur, i is the feature bit value for the first vector V_1, and j is the value for the second vector V_2.

$$C_{ij}\ (i, j \in \{0, 1\}) \qquad (8)$$

The similarity between two vectors, that both have the length k, is calculated by the function f and produces a score between 0 and 1, where if the score is 0, the two vectors are completely different, while 1 means they are the same. This scoring method is used in [3] and is calculated by equation (9). The calculation of f only requires C_{00} and C_{11}, as it is designed to take into consideration only the positions where the feature bits are both the same. For example, if the feature vectors V_a and V_b in Figure 3 were compared, the result would be that $C_{00} = 4$ and $C_{11} = 5$. The similarity of these two vectors would be calculated by equation (9) and would equal 0.389.

$$f(V_1, V_2) = \frac{0.5 \times C_{00} + C_{11}}{k} \qquad (9)$$

The similarity score between an unknown vector U and the set of training vectors T is then found as the mean similarity when U is compared to each signature in T. For example, if the vectors V_a, V_b and V_c from Figure 3 are used as the set T, and V_u is an unknown vector, the similarity score between T and V_u would be .491.

The threshold that will be used to determine the classification is produced from the average similarity score μ of the training vectors when they are compared to each other. This method of finding μ is outlined in [3], and [10]. Carrying on from the previous example, if T comprises of the first three vectors from Figure 3, μ would be .398.

Because μ is the mean similarity score of the training vectors, it will tend towards classifying half of the genuine signatures as forgeries, therefore its use as the threshold is not ideal. So to create the threshold, an offset is required to move the mean down the similarity scale, with the intent being to maximise the classification accuracy of both the genuine and forgery signatures. This is shown in Figure 4, where the threshold is the offset of the

mean. Two methods were investigated for finding this offset, these were manual and automatic, both of which are described below.
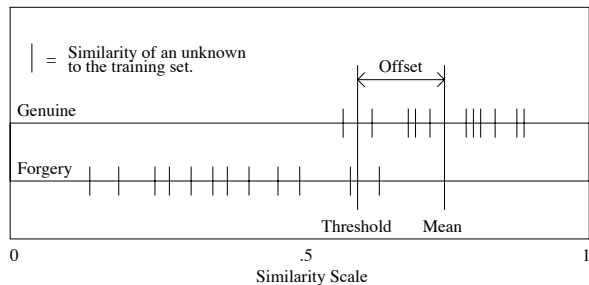

Figure 4: Threshold creation.

3.1 Manual Classification

Manual classification uses a fixed offset o to move μ down the similarity scale. This offset does not compute a threshold directly, but instead is used to generate a false acceptance rate (FAR) versus false rejection rate (FRR) curve, from which the best threshold (found at the equal error rate) is computed. This offset is then used for all signature sets, with the class of an unknown signature being determined by equation (10).

$$class = \begin{cases} genuine & \text{if } score \geq \mu - o \\ forgery & \text{if } score < \mu - o \end{cases} \qquad (10)$$

The disadvantage of the manual method is that it requires a range of experiments to find a value which minimises the FAR and the FRR. The use of this method was to make AFT comparable to the graph matching technique [3], which also uses this manual offset.

3.2 Automatic Classification

Automatic classification is a heuristic method for finding an offset based on the training signatures. The offset is found in the same manner that AFT uses for calculating the lower sample standard deviation S_L, where only the values below μ are used. These values are the similarity scores from equation (9) when all training signatures are compared to each other. The class of an unknown signature is then determined by equation (11).

$$class = \begin{cases} genuine & \text{if } score \geq \mu - S_L \\ forgery & \text{if } score < \mu - S_L \end{cases} \qquad (11)$$

4 Experiments

The evaluation of AFT was carried out on two datasets of signatures, both of which comprise of skilled forgeries. A skilled forgery is one in which the forger has both seen and practised writing a
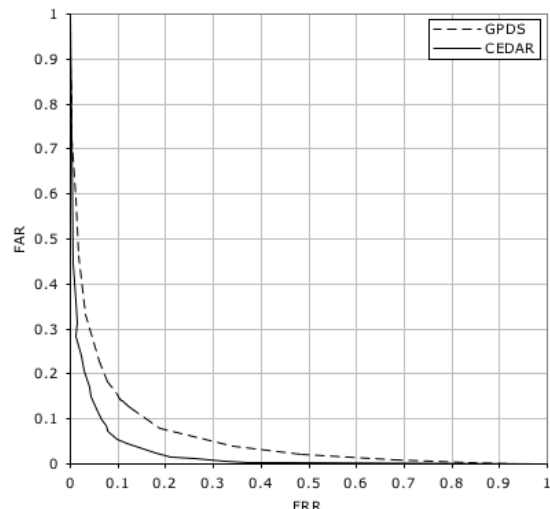

Figure 5: Correlation between FAR and FRR

genuine signature, making it visually similar to the original. Because of this visual similarity, there is a substantial increase in difficulty of being able to classifying an unknown signature, as opposed to signatures of the random or simple variations [11].

The first dataset, CEDAR [12], is made up of 55 signature sets, where each set consists of 24 genuine signatures and 24 forgeries. Initially this dataset was in grey-scale, but was converted to binary using a classic iterative thresholding method [13], which was chosen for its simplicity and robustness. The second dataset was GPDS [14], which was already binarised. This dataset contains 39 signature sets, where each set consists of 24 genuine signatures and 30 forgeries.

The experiments for both of these datasets were carried out using three spatial pyramid levels and an 8×4 grid for the finest level of region sampling. The final accuracy is calculated by equation (12) and is the middle point between the genuine and forgery classification accuracies.

$$accuracy = \frac{(1 - FAR) + (1 - FRR)}{2} \qquad (12)$$

4.1 CEDAR Results

For each signature set in CEDAR, 16 signatures were randomly selected as training samples, while the remaining 8 genuine signatures along with the 24 forgeries were used for testing. The use of 16 signatures for training was to make the results comparable to [3], [5], and [10]. This was then repeated 10 times for each set. By varying the offset, the error trade-off can be plotted (See Figure 5), allowing the offset that minimises both the FAR and FRR to be identified and used for the final classification accuracy.

The results in Table 1 show that AFT produces competitive results to the graph matching method, with the manual method achieving an accuracy approximately 2% greater than the automatic method. AFT, in comparison to GSC, has a significant accuracy increase of approximately 14%.

Table 1: CEDAR Results

Method	1-FAR	1-FRR	Accuracy
GSC [5]	80.5	77.55	78.5
Zernike [10]	83.7	83.4	83.6
Graph Matching [3]	91.8	92.3	92.1
AFT (Auto)	**89.04**	**91.84**	**90.44**
AFT (Manual)	**92.58**	**92.25**	**92.42**

4.2 GPDS Results

In GPDS, training was carried out in the same fashion as CEDAR, except only 12 randomly selected signatures were used, so that the results would be comparable to [4]. The remaining 12 genuine signatures and 30 forgeries were then used for testing. This was once again repeated 10 times for each set. The offset that minimises both the FAR and FRR was then found in the same way as CEDAR, and was used for the final classification accuracy.

Table 2 shows that AFT can match the classification accuracy that the more complex method DWT is able to achieve. Once again, there is approximately a 2% accuracy difference between the manual and automatic methods.

Table 2: GPDS Results

Method	1-FAR	1-FRR	Accuracy
DWT (Random) [4]	82.08	80.69	81.39
DWT [4]	88.11	86.74	87.43
AFT (Auto)	**82.76**	**89.21**	**85.99**
AFT (Manual)	**85.79**	**89.52**	**87.66**

4.3 Reduced Training Size

AFT also remains fairly stable when the number of training signatures is varied, showing that it is remarkably resilient. This is shown in Figures 6 and 7. The manual method in both CEDAR and GPDS required the offset to be adjusted with relation to the change in the number of training signatures. This change allows the classification accuracy to remain stable when 8 or more training signatures were used; when less than 8 signature were used, the accuracy dropped off fairly quickly. The automatic method tended to consistently follow the same pattern, except it dropped off at a much quicker rate. The stability of the classification accuracy when the number of training signatures varied was not tested in [3], [4], [5], and

[10], therefore, the comparable ability of AFT in this regard cannot be determined.
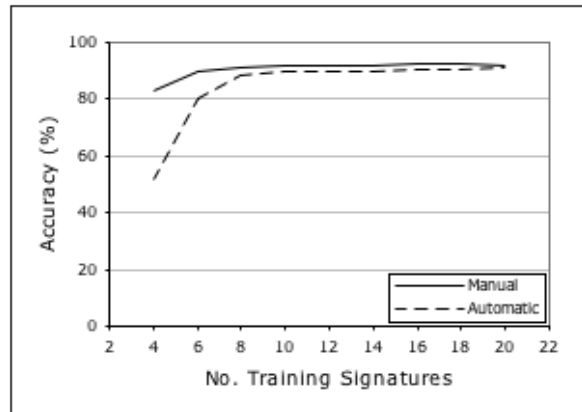

Figure 6: CEDAR stability when the number of training signatures is varied. The standard training amount is 16 signatures.
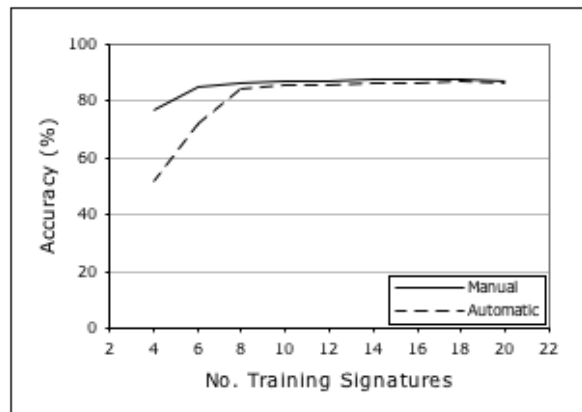

Figure 7: GPDS stability when the number of training signatures is varied. The standard training amount is 12 signatures.

5 Conclusions and Future Work

This paper presented a novel method for off-line signature verification by introducing what has been termed adaptive feature thresholding. AFT is designed to greatly restrict how a binary feature vector is created, improving its representative similarity in relation to the training signatures. Along with AFT, we found that the combination of spatial pyramids and equimass sampling grids helped to improve the extraction and representation of a signature through the use of the gradient direction. Using these methods, AFT achieved a classification accuracy that is competitive to both the graph matching and the DWT methods. As well as this, AFT also remains computationally less complex due to using only one image feature as opposed to the ensemble of image features that other methods tend to use. Experimentation was carried out using

two different approaches, manual and automatic. Manual tended to achieve an accuracy that was 2% greater than automatic. It was also shown that this method remained relatively stable when the number of training signatures was greater than 8.

Possible future work would be to test AFT with other image features, as well as combining it with the graph matching method, as this may further enhance the classification accuracy. The use of AFT in other areas of research is also possible due to its generalist nature, and may prove to be beneficial.

References

[1] A. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 4–20, 2004.

[2] A. Jain, F. Griess, and S. Connell, "On-line signature verification," *Pattern Recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[3] S. Chen and S. Srihari, "A New Off-line Signature Verification Method based on Graph Matching," *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)-Volume 02*, pp. 869–872, 2006.

[4] W. Tian, Y. Qiao, and Z. Ma, "A New Scheme for Off-line Signature Verification Using DWT and Fuzzy Net," *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on*, vol. 3, no. 2, pp. 30–35, 2007.

[5] S. Srihari, A. Xu, and M. Kalera, "Learning strategies and classification methods for offline signature verification," *Proc. of the 7th Int. Workshop on Frontiers in handwriting recognition (IWHR)*, pp. 161–166, 2004.

[6] B. Zhang and S. Srihari, "Properties of Binary Vector Dissimilarity Measures," *Proc. JCIS Int'l Conf. Computer Vision, Pattern Recognition, and Image Processing*, 2003.

[7] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *Proc. CVPR*, vol. 2, no. 2169-2178, p. 1, 2006.

[8] J. Favata and G. Srikantan, "A Multiple Feature/Resolution Approach To Handprinted Digit and Character Recognition," *International journal of imaging systems and technology*, vol. 7, no. 4, pp. 304–311, 1996.

[9] G. Srikantan, S. Lam, and S. Srihari, "Gradient-based contour encoding for character recognition," *Pattern Recognition*, vol. 29, no. 7, pp. 1147–1160, 1996.

[10] S. Chen and S. Srihari, "Use of Exterior Contours and Shape Features in Off-line Signature Verification," *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, pp. 1280–1284, 2005.

[11] B. Zhang, "Off-Line Signature Recognition and Verification by Kernel Principal Component Self-Regression," *Proceedings of the 5th International Conference on Machine Learning and Applications*, pp. 28–33, 2006.

[12] M. Kalera, S. Srihari, and A. Xu, "Offline Signature Verification and Identification Using Distance Statistics," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 7, pp. 1339–1360, 2004.

[13] T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 8, pp. 630–632, 1978.

[14] M. Ferrer, J. Alonso, and C. Travieso, "Offline geometric parameters for automatic signature verification using fixed-point arithmetic." *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 6, pp. 993–997, 2005.