Handwritten Signature Matching using GPUMLib

João Gonçalves
jcgonc@dei.uc.pt
Noel Lopes
noel@ipg.pt
Bernardete Ribeiro
bribeiro@dei.uc.pt

Department of Informatics Engineering
CISUC, University of Coimbra
Coimbra, Portugal

Abstract

Today the programmable Graphics Processing Unit (GPU) has raised a noticeable interest for applications that demand high-computational power. In particular, biometric applications containing thousands of samples and features need efficient tools to process data. GPUMLib is an open source library with machine learning techniques endowed with GPU that is able to handle the significant memory and computational burden needed for signature matching. In this paper, the SVM component imbued with GPUMLib has been used for signature matching yielding good performance results assessed by the F-Score and False Positive Rate (FPR) in the GPDS database.

1 Signature Matching Background

Signature matching is a very important problem in authentication which covers a broad range of areas such as personal identification, security and bank transactions. Many efforts have been put to tackle the verification of signatures which contain biometric information. Often the databases are very large and such big data appears difficult to handle. Additionally, in offline settings, the lack of the dynamic characteristics makes the problem hard to solve.

This problem is very difficult for many reasons. The biometric data is a scanned 2D image. Unlike on-line verification there is lack of dynamic characteristics (e.g. velocity, pen pressure, acceleration, etc.) which reflect each individual motion style and are harder to fake. Additionally the biometric features in original and faked signatures can be extremely similar which makes the problem even harder. Examples are the shapes, sizes and variations of signatures that lead to a confluence of factors extremely tricky to verify. Also, the sheer volume of biometric data in many applications require fast tools for model selection in order to expose better models. Preprocessing of offline handwritten biometric data is complex and motivates the holistic study of many features capable of proper capturing the intra-variational characteristics of the individual signatures and the optimal group of features for building a better model.

Fast machine learning algorithms that are able to extract relevant information from large repositories play an important role. Therefore, we use in this work GPUMLib[1] which is an open-source machine learning library that will be described next.

2 GPUMLib

The GPUMLib framework is endowed with a wide range of machine learning algorithms implemented in CUDA using GPU. At its core, the library contains a set of CUDA kernels that support the execution of Machine Learning (ML) algorithms on the GPU. Usually, in order to implement an ML algorithm on the GPU several kernels are required. However, the same kernel might be used to implement different algorithms.

Each ML algorithm has its own C++ class that is responsible for: transferring the information (inputs) needed by the algorithm to the device (GPU); calling the algorithm kernels in the proper order; and transferring the algorithm outputs and intermediate values back to the host. This model allows non-GPU developers to take advantage of GPUMLib, without requiring them to understand the specific details of CUDA programming.

Moreover, GPUMLib provides a standard memory access framework to support the tasks of memory allocation and data transfer between the host and device (and vice-versa) in an effortless manner.

[1] `http://gpumlib.sourceforge.net/`

Figure 1: Genuine and forged signatures from the GPDS database.

3 Computational Experiments

In this section we describe the dataset, present the experimental setup and indicate the performance metrics for signature matching assessment.

3.1 Dataset

The database contains data from 300 individuals. For each individual there are 24 genuine signatures, plus 30 forgeries of his/her signature making 54 images per individual and a total of 16200 images. The 24 genuine specimens of each signer were collected in single day writing sessions. The forgeries were produced under the following conditions: The forger imitates a genuine signature from the static image of the genuine signature (scanned at 300 DPI) and the forger is allowed to practice writing the signature for as long as s/he wishes. Each forger has to imitate three signatures of five signers in a single day writing session. The genuine signature shown to each forger is chosen randomly from the 24 genuine ones. Therefore, for each genuine signature, there are 30 simple forgeries made by 10 forgers from 10 different genuine specimens. The dataset used consists of 16200 handwritten off-line signature recognition (each signature is a 649×462 pixels image). Additional information on this database can be found in Ferrer et al. [3].

3.2 Experimental Setup

For the test set we used 9 images and the remaining 45 for the training set. Both training and testing sets were randomly generated from the initial data, being the test set composed of 4 genuine signatures and 5 forged. The experiments were run 10 times per configuration.

With regard to feature extraction from the GPDS database previous research can be found in [1], [2] and [3]. From the aforementioned studies the best features so far were extracted from the original dataset. Following these authors and our previous work [5] in Table 1 we present the extracted features and the corresponding number of attributes from the GPDS image dataset.

3.3 Performance Metrics

We defined several measures based on the possible outcomes of the classification, namely, False Positive Rate ($FPR = \frac{FP}{FP+TN}$), and False Discovery Rate ($FDR = \frac{FP}{FP+TP}$), as well as combined measures, such as, the van Rijsbergen F_β measure, which combines recall ($R = \frac{TP}{TP+FN}$) and precision ($P = \frac{TP}{TP+FP}$) in a single score (F-Score $= \mathrm{F1} = \frac{2PR}{P+R}$), yielding an harmonic average between precision and recall.

Table 1: Number of attributes of each feature.

Feature	Attributes
Best Fit	4
Discrete Cosine Transform (DCT)	5
Geometric Parameters (Cartesian)	180
Geometric Parameters (Polar)	192
Gravity Center	1
Histogram Frequencies (hist)	6
K-Means	10
Max Intensity Points (maxint)	1
Modified Direction Feature (MDF)	160
Six-fold-Surface	6
Three-fold-Surface	3
Wavelet Transform Feature	12

Table 2: NVIDIA GeForce 570 GTX characteristics.

Characteristic	Value
Number of SPs	480
IEEE single precision (float) performance	748.8 GFlops
Number of SMs	15
Shading clock speed	1.56 GHz
Memory size	1.25 GB
Memory bandwidth	152.0 GB/s
Shared memory per block	48 KB

4 Results and Discussion

In this section we describe the three experiments to deal with the signature matching problem using the SVM component of GPUMLib [4].

The Experience 1 comprises the identification of original and forged signatures. For that purpose, we used all the 300 individuals and studied each group of features. The best feature combination DCT + MDF was tested for all the supported kernel functions by our GPU-based SVM Classifier [5]. The results obtained by 5-Fold cross-validation are shown in Figure 2.

In the second experience[2], instead of using all the 300 groups of signatures we exploited several groups or combinations of features for each individual. Therefore, this experiment consisted of identifying, for each person, if a signature was either original or forged. Only the Radial Basis Function (RBF) kernel was employed. For validation the test set consisted of genuine 4 original plus 5 forged signature, as specified above.

In the Experience 3, we performed signature matching related to a given individual, using a One-Against-One multi-class classifier, that is, we trained and tested each individual class against one of the others. As the cost involved in the training process is high, we only used the RBF kernel and 5 K-Fold cross-validation procedure. The results are illustrated in Figure 3.

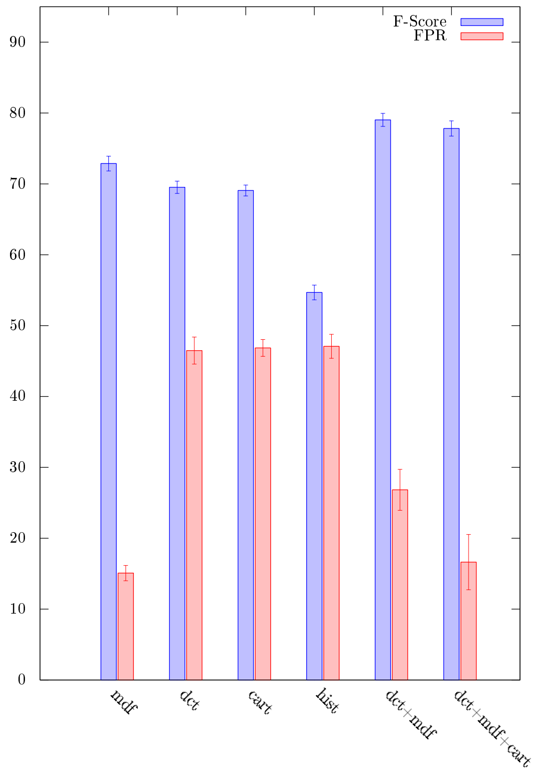

Figure 2: Experience 1: Results with the RBF kernel.

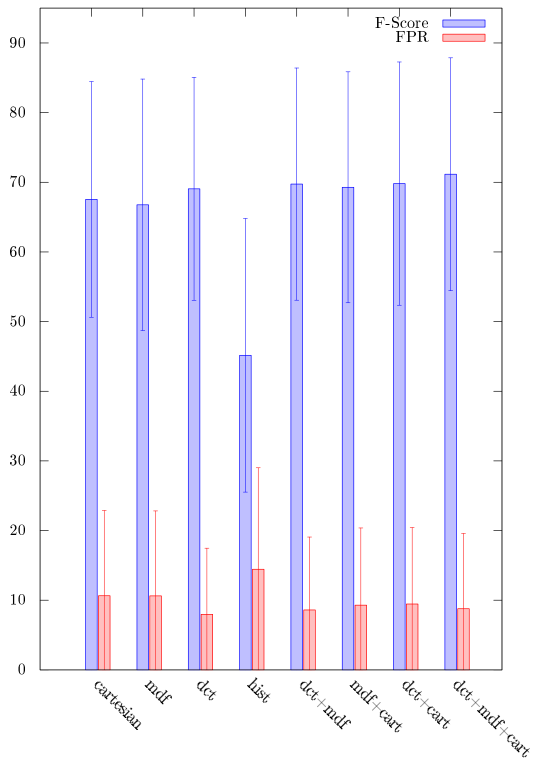

Figure 3: Experience 3: Forged/original signature per individual.

5 Conclusions

Handwritten signature matching plays a crucial role in many important transactions for security and privacy reasons. In this work, we presented the performance analysis of the SVM component of the GPUMLib which yielded good results in the GPDS database. To this end, several experiments were performed using the features extracted in a previous work. One possible direction of future work is to integrate multiple kernel learning in the GPUMLib for this kind of pattern recognition problems.

References

[1] S. Armand, M. Blumenstein, and V. Muthukkumarasamy. Off-line signature verification using an enhanced modified direction feature with single and multi-classifier approaches. *Computational Intelligence Magazine, IEEE*, 2(2):18 –25, 2007.

[2] M. Blumenstein, X.Y. Liu, and B. Verma. A modified direction feature for cursive character recognition. In *Proc of IEEE Int J Conf on Neural Networks*, volume 4, pages 2983–2987, 2004.

[3] M. Ferrer, J. Alonso, and C. Travieso. Offline geometric parameters for automatic signature verification using fixed-point arithmetic. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27 (6):993–997, 2005.

[4] Noel Lopes and Bernardete Ribeiro. GPUMLib: An efficient open-source GPU machine learning library. *Int J of Computer Information Systems and Industrial Management Applications*, 3:355–362, 2011.

[5] Bernardete Ribeiro, Noel Lopes, and João Gonçalves. Signature identification via efficient feature selection and GPU-based SVM classifier. In *Proc. of IEEE Int J Conf on Neural Networks (IJCNN)*, 2014.

[2] Results from Experience 2 are omitted due to paper's space restrictions.