AI Project Report CS354

Rahul Pandya (1200122), Darshan Tejani (1200135)

Offline Signature Verification using Support Vector Machine

1. Objective

Signature has been a distinguishing feature for person identification. Even today an increasing number of transactions, especially related to financial and business are being authorized via signatures. Hence the need to have methods of automatic signature verification must be developed if authenticity is to be verified and guaranteed successfully on a regular basis.

The motivation behind the project is growing need for a full proof signature verification scheme which can guarantee maximum possible security from fake signatures. The idea behind the project is also ensure that the proposed scheme can provide comparable and if possible better performance than already established offline signature verification schemes.

2. Literature Survey

The fact that the signature is widely used as a means of personal verification emphasizes the need for an automatic verification system. During last few years, researchers have made great efforts on off-line signature verification. Initially offline verification scheme required perfect alignment of the signature to specified axis. Nowadays there are many approaches available for signature verification. e.g. neural network expert system to identify signature, elastic image matching, methods based on texture analysis of the signature. Some researchers tried to apply wavelet on the system while some tried to compute euclidean distance between signature using global and grid features. Signature learning can also be done using support vector machine (SVM) which is what we are using in this project.

The term **"Signature forgery”** refers to the act of falsely replicating the signature of another person. There are three types of forgery :

1. Random Forgery: Random forgery is done by a person who doesn’t know the shape and structure of the original signature.
2. Simple Forgery: In this type of forgery the person concerned has a vague idea of the actual signature, but is signing without much practice.
3. Skilled Forgery: This type of forgery considers appropriate knowledge about the original signature along with ample time for proper practice.

Our proposed scheme eliminate random and simple forgeries and also reduces skilled forgery to a great extent.

A. **Offline verification :**

A sample signature (photo) is scanned.

Height, width, slope etc (STATIC INFO).

Cost-effective (thats why still in use).

Lower degree of accuracy.

B. **Online Verification :**

A special hardware (a tablet and a stylus with sensors) is needed (So easy to implement)

Pressure points, breakpoints and time taken to generate the signature (DYNAMIC)

Along with static info (like in offline)

High level of accuracy.

Expensive due to sensors cost.

Support Vector Machine :

SVM (also known as support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns used for classification and regression analysis. Given a set of training examples each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Few More Terms :

FAR – False Acceptance Ratio .

The false acceptance ratio is given by the number of fake signatures accepted by the system with respect to the total number of comparisons made.

FRR – False Rejection Ratio.

The false rejection ratio is the total number of genuine signatures rejected by the system with respect to the total number of comparisons made.

Both FAR and FRR depend on the threshold variance parameter taken to decide the genuineness of an image. If we choose a high threshold variance then the FRR is reduced, but at the same time the FAR also increases.

If we choose a low threshold variance then the FAR is reduced, but at the same time the FRR also increases.

ERR – Error rejection Rate.

If the FAR of a system is same as the FRR then the system is said to be in an optimal state. In this condition, the FAR and FRR are also known as ERR.

3. Problem Specifications

We are given an image file of a handwritten signature which needs to be tested for it's authenticity. In modern signature verification, the task of identifying the genuineness of the signature is based on the hardness of solving some hard number of theoretical problems.

The input signature should be subjected to various image processing techniques for refinement (eg Noise reduction).

The shape of a person's signature remains similar in all translational, scaled and rotational alignments of the sign. That is the number of crests, troughs and curves remains the same irrespective of the size and orientation of the image. So the system should be independent of translational , scaled and rotational alignments of the signature. The success of the proposed system can be determined from the appreciable FARs and FAAs. System should minimize the error rejection rate (ERR) as much as possible.

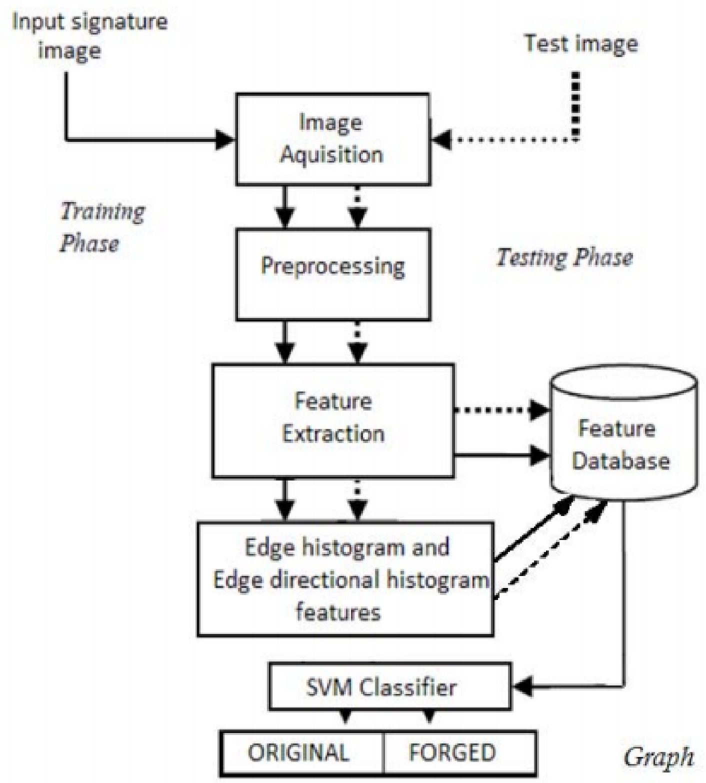
Different parameters from the signature (eg height, width, curve etc.) from the original dataset should be compared with that from the test signature and the trained Support Vector Machine will classify it as “GENUINE” or “NOT GENUINE”.

Our system must ensure robustness, user-friendliness, responsiveness and security. System should be designed to allow possible future extension.

4. Analysis and Design (Architectural Design)

We are going to use novel approach for signature verification based on support vector machine. This system will use static as well as dynamic features for verification. The static features include moment features and 4-direction distribution, While the dynamic features include gray distribution and stroke width distribution. At last support vector machine is used to classify the signature.

The flow of the system can be seen briefly in the following chart:



Preprocessing stage can be divided further in five parts:

1. Noise reduction
2. image binarization
3. data area cropping
4. width normalization
5. image thinning

Let's have an insight in SVM. Signatures are typically represented by sparse vectors under the vector space model. When training classifiers on large collections of signature, both the time and memory requirements connected with these vectors may be prohibitive. This calls for the use of a feature selection method not only to reduce the number of features but also to increase the sparsity of vectors. We propose a feature selection method based on linear Support Vector Machines (SVMs). Linear SVM is used on a subset of training data to train a linear classifier which is characterized by the normal to the hyper-plane dividing positive and negative instances. Components of the normal with higher absolute values have a larger impact on data classification. Instead of predefining the number of highest scoring features to be included in a classifier we apply feature selection that aims at a pre-defined average sparsity level across documents and classifiers for a given training set. After the feature set is determined, the model is trained on the full training data set represented within the selected feature set.

The test signature is then, based on its value for the parameters from the feature set, is mapped and classified as “GENUINE” or “FORGED”.

4. References:

[1] Banshidhar Majhi, Y Santhosh Reddy, D Prasanna Babu, “Novel Features for Off-line Signature Verification” International Journal of Computers, Communications & Control, Vol. I, No. 1, pp. 17-24, 2006.

[2] “Offline Signature Verification Using Support Vector Machine” by Kruthi, C and Shet, D.C., Dept. of Inf. Sci. & Eng., PES Inst. of Technol., Bangalore, India (Jan 2014)

[3] Statistical Learning Theory Lecture: 6 , Non-separable (soft) SVMs, Lecturer: Peter Bartlett ,Scribe: Joseph Austerweil , (Spring 2008)

[4] T.Y.Zhang, C.Y. Ceun “A fast algorithm for Thinning Digital Patterns”.Communications of ACM, March 198 Concepts of Image Processing.