Biometrics laboratory exercise
Exercise 1: Dynamic signature

Jiří Wild, Pavel Vostatek
contact: vostapav@fel.cvut.cz

8. ledna 2013

1 Introduction

The aim of this task is to introduce identity verification using dynamic signature. The basis of verification method is formed by comparison of given signature to verified templates, stored in a database. Various methods can be used for such comparison, out of which two will be presented in this lab exercise: statistical modelling-based algorithm called **Gaussian Mixture Model** (GMM) and an algorithm based on direct comparison of two sequences: **Dynamic Time Warping** (DTW). You will be provided with Signature Toolbox for Matlab, which contains basic functions together with already implemented method for handling GMM and a skeleton function for Dynamic Time Warping. It will be your task to implement the latter, which should provide you with deep understanding of this method, applicable on a wide range of problems. Then, you will compare performance of both algorithms on a SVC2004 database (description in [3]). The last task will be to attempt to falsify selected signature and evaluate the results.

2 Theoretical background

Classical signature, as a biometrical means for document verification is notorically known. On the contrary, dynamic signature extends the classical form by temporal information, as well as other personal properties such as pressure sensitivity or pen tilt. To capture this information, electronical tablet is commonly used.

Considering two basic tasks in biometrics: identification and verification, dynamic signature applies exclusively to verification - i.e. verification of person's identity using his/her signature. Thus, input to the verification algorithm is signature itself and corresponding claimed identity. To carry out the verification process, model of the signature is necessary, which is typically formed by a database of sampled signatures (commonly about 10 samples). The higher the number of signatures in the database, the less the intra-class variability in signature of the same person matters. In the verification procedure itself, the model for claimed identity is compared to the given signature and a similarity measure is

calculated and used to determine whether the provided matches the database record or not. In this particular exercise you will calculate similarity using the already mentioned GMM and DTW methods. The decision threshold, necessary to decide what value of the similarity measure is appropriate to claim match/non-match is set to conform to desired False accept rate (FAR) and False reject rate (FRR) requirements.

2.1 Input signatures, preprocessing

As stated above, a digitally sampled dynamic signature provides information about contour and its formation in time. Moreover, information about pressure level and pen tilt and other properties can be obtained. These properties are commonly referred to as *features* and are captured at discrete time steps $t = 1..T$. Let us denote the individual features: x_t, y_t for position and p_t, ϕ_t, θ_t for pressure tilt and skewness. It is also possible generate more features as a combination or transformation of basic ones. Examples of commonly used derived features are velocity $v_t = \|[\frac{dx}{dt}, \frac{dy}{dt}]\|$ and acceleration $a_t = \|[\frac{d^2x}{dt^2}, \frac{d^2y}{dt^2}]\|$.

All the features mentioned in the previous paragraph form a feature set. For the purpose of verification, it turns out very useful to use only a subset of those features. Only the features with maximum discrimination power in determination of true or falsified signature are kept, the rest is discarded. Commonly, the basic features are position, velocity and acceleration or pressure. Experiments with different feature sets will be one of your tasks in this assignment.

Another factor, influencing the model, is preprocessing: It is necessary to align the signatures to comparable center (e.g. the center of gravity) and normalize their size to a defined range. The last significant issue is the number of signatures, used for model creation. As a part of this exercise, you will experiment with all these parameters and test their influence on performance of the system.

2.2 Signature Toolbox

You are provided with Signature Toolbox for Matlab, which maintains basic functionality for signature import and processing. The toolbox includes functions for data import and feature extraction, as well as calculation of derived features. Implemeted functions for GMM modeling provide model creation and signature scoring functionality. There is also empty skeleton function for DTW which you will implement yourself. More detailed toolbox description can be found in Appendix A.

2.3 Similarity measure using Gaussian mixture model

The principle of a GMM based system can be described as follows:

- n features are derived from the signature $x_{1,t} \dots x_{n,t}$, each captured at discrete times $t = 1 \dots T$.
- During the learning phase, probability distribution for each feature $x_1 \dots x_n$ for each time step is estimated from the input data. If there are more signatures in the training

set, they are all taken into the calculation for each feature. This way, we obtaion n probability distributions $P_1 \dots P_n$

- Then, when evaluating match of the model to given signature, we use log-likelihood function: $score = \sum_{i=1}^{n} \sum_{j=1}^{T} ln[P_i(X = x_{i,j})]$

2.4 Similarity measure using Dynamic time warping

In the case of similarity measure using DTW, there is no model learning, just storing chosen number of signatures to database. In the recognition phase, given signature is compared to records in database using DTW measure(see lectures or [2]). The algorithm for calculation of similarity between two (in general n-dimensional) vectors is as follows:

- Let us assume two signatures s_1, s_2 formed by n features, with lengths t and s. Notation is the same as in previous case.
$$s_1 = \begin{bmatrix} x_{1,1}, x_{2,1}, \dots, x_{n,1} \\ x_{1,2}, x_{2,2}, \dots, x_{n,2} \\ \dots \\ x_{1,t}, x_{2,t}, \dots, x_{n,t} \end{bmatrix}, s_2 = \begin{bmatrix} x_{1,1}, x_{2,1}, \dots, x_{n,1} \\ x_{1,2}, x_{2,2}, \dots, x_{n,2} \\ \dots \\ x_{1,s}, x_{2,s}, \dots, x_{n,s} \end{bmatrix}.$$

- Comparison using DTW can be calculated using the following algorithm: *i)* A matrix D of size $(t+1, s+1)$ is created, initiated with $D(1,1) = 0, D(i,1) = \inf, D(1,j) = \inf, i = 2 \dots n, j = 2 \dots m$. *ii)* The remaining fields in the matrix D are calculated as follows:
$$D(i,j) = ||s_1(i-1,:) - s_2(j-1,:)|| + min \begin{cases} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{cases},$$
$||s_1(i-1,:) - s_2(j-1,:)||$ denotes euclidean distance between (i-1)-th row (sample) of signature s_1 and (j-1)-th row of signature s_2, i.e. $\sqrt{\sum_{d=1}^{n} [s_1(i-1,d) - s_2(j-1,d)]^2}$.

- In the end, mutual distance of both vectors is accumulated in $dist = D(t+1, s+1)$.

To obtain the similarity measure, we calculate distances between the tested signature and all signatures stored in the database, consecutively. Then, we calculate similarity as a negative mean value of the calculated distances. The negative serves to maintain higher (less negative) similarity value for closer samples.

3 Assignment

1. Each student obtains 15 signatures of different persons, including corresponding identification. In the first step, create a model for each person using provided GMM method and choose similarity threshold (minimum *score*, at which we claim signature matches the database record). Evaluate, what is the False accept rate (FAR, or ratio of incorrectly matched signatures) and False reject rate (FRR, or a ratio of incorrectly non-matched signatures) of the system at threshold value you chose. [3 b]

2. Next step will be to implement DTW method to Signature Toolboxu. Implement the method, as described in section 2.4 to the respective skeleton function in the toolboxu. Then, perform signature comparison and statistical processing, same as in the previous step. [7 b]

3. The last step will be system optimization. Experiment with how does prior normalization of input signatures (zero mean value and variance equalling one for each feature) influence the obtained statistics. Further, try adding additional features such as pressure, pen tilt or skewness. [4b]

4. When you have fine-tuned feature set to gain best performance, experiment with higher number of training signatures for model creation. How does it change system performance? [1 b]

5. Try to intrude your system by falsifying selected signature. You will use a graphical tablet and our dedicated application for signature recording [5 b].

Write a brief report, documenting your solution to each step of the assignment.
You are expected to solve the task individually!

3.1 Bonus task:

1. Implement another method of signature verification, from state-of-the-art literature (comprehensive overview of current methods is given in [1]). [5 b]

A Appendix: Structure of the Sigature Toolboxu.

You are provided with a Signature Toolbox in Matlab, implementing GMM verification and containing also a signature database. The whole system can be used through following methods:

- *load_data* — Function for loading raw signatures. The output is a cell-array with imported signatures, e.g. $S = \{s_1, s_2, \ldots\}$, where s_i corresponds to signature as defined in the previous chapter on DTW. Individual signatures can be accessed in Matlab as follows: $s_i = S\{i\}$.

- *preprocess* — function for preprocessing of raw data.

- *extract_features* — function for feature extraction from the raw data.

- *make_model* — function for model learning.

- *score* — function for calculation of score (similarity) for given signature.

- *ukazka* — function with demo code.

All functions and their use is documented directly in the help of respective m-files

Reference

[1] Donato Impedovo. Automatic signature verification: The state of the art. Download: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4603099 (accessible through dialog.cvut.cz).

[2] J M Pascual-Gaspar. Practical on-line signature verification. Download: http://www.springerlink.com/content/n0x73333061702u4/ (accessible through dialog.cvut.cz).

[3] SVC2004. Svc 2004: First international signature verification competition, detailed instructions for participants. Download: http://www.cse.ust.hk/svc2004/instructions.pdf.