SigComp 2011 Disclaimer

The SigComp11-NFI signature collection was collected by:
Erwin Mattijssen (e.mattijssen@nfi.minjus.nl)
The SigComp11-Chinese signature collection was collected by:
Xiaohong Chen (ccpccxh@hotmail.com)

These data are made publicly available for:
(i) the ICDAR'11 signature verification competition and
(ii) research purposes

Any use of these data should mention reference [1]:

[1] Marcus Liwicki, Michael Blumenstein, Elisa van den Heuvel, Charles E.H. Berger, Reinoud D. Stoel, Bryan Found, Xiaohong Chen, Muhammad Imran Malik. SigComp11: Signature Verification Competition for On- and Offline Skilled Forgeries, Proc. 11th Int. Conference on Document Analysis and Recognition, 2011

It is not allowed to display the images in publications, hand-outs etc.
Copy of the data can be obtained when emailing:
liwicki, vandenhevel, stoel, berger

The executable will not be used for purposes other than the ICDAR2011 competition. Performances will be published in the ICDAR proceedings and in an extended journal publication. After the competition, all executables will be destroyed. If a team chooses to use some expiration mechanism, the expiration should be set to 31 December 2011.

Organizing committee

Marcus Liwicki, German Research Center for Artificial Intelligence
Trippstadter Str. 122, 67663 Kaiserslautern, Germany marcus.liwicki@dfki.de

Michael Blumenstein, Griffith University
Parklands Drive, Southport QLD 4215, Australia m.blumenstein@griffith.edu.au

Elisa van den Heuvel, Charles E.H. Berger, and Reinoud D. Stoel
Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GP The Hague, The Netherlands
e.van.den.heuvel@nfi.minjus.nl c.berger@nfi.minjus.nl reinoud@holmes.nl

Bryan Found, La Trobe University, Melbourne, Australia b.found@latrobe.edu.au

Xiaohong Chen, Forensic Science Institute, Ministry of Justice,
Shanghai 200063, China ccpccxh@hotmail.com

Dataset Description and Code Submission

In the test phase, signatures can either be genuine: written by the reference writer, or simulated: simulation of the signature by another writer than the reference writer.
The collection contains offline and online signature samples. The offline datasets will be constituted of PNG images, scanned at 400 dpi, RGB color. The online dataset will consist of ascii files with the format: X, Y, Z (per line). Sampling rate 200 Hz, resolution 2000 lines/cm, precision of 0.25 mm. Collection device: WACOM Intuos3 A3 Wide USB Pen Tablet. Collection software: MovAlyzer. A preprinted paper was used with 12 numbered boxes (width 59mm, height 23mm). The preprinted paper was placed underneath the blank writing paper. Four extra blank pages were added underneath the first two pages to ascertain a soft writing surface.

Dutch dataset

Total set: 1790 signatures.
Training set: data of 10 reference writers and some skilled forgeries of these signatures.
Additionally, the public data of the 2009 competition may be used (contact Elisa van den Heuvel)
Test set: #.

Chinese dataset

Total set: 960 signatures.
Training set: data of 10 reference writers and some skilled forgeries of these signatures.
Test set: #.

Evaluation

The system will get as an input parameter the mode (online or offline), the questioned signature and up to 12 authentic signatures from the reference writer (Note that 1 till 12 should be possible). In this competition we ask to produce a comparison score (e.g. a degree of similarity or difference), and the evidential value of that score, expressed as the ratio of the probabilities of finding that score when Hypothesis 1 is true and when Hypothesis 2 is true (i.e. the likelihood ratio).

Folder Structure and File Naming
The signatures are arranged according to the following folder structure:

- OfflineSignatures
  - Chinese
    - TrainingSet
      - Offline Genuine (Containing the reference signatures)
      - Offline Forgeries (Containing the simulations)
  - Dutch
    - TrainingSet
      - Offline Genuine
      - Offline Forgeries
- OnlineSignatures
  - … (similar structure)

Note that the online and offline folders do not necessary contain exactly the same signatures, because during acquisition not all samples could be acquired in both modes. Furthermore, note that the online signatures may contain artifacts from the pen-movements (e.g., strokes that do not belong to the actual signature anymore). Systems could recover from those artifacts by applying preprocessing heuristics).
Genuine signatures are named according to the following convention (the same for all data sets):
III_NN.*, where III is the ID of the reference writer and NN is an index of the signature, i.e., it is the NNth authentic signature contributed by writer III.
Simulated signatures (forgeries) are named according to the following conventions:
FFFFIII_NN.*, where FFFF is the ID of the forger, III is the ID of the reference writer and NN is an index, i.e., it is the NNth simulation attempt of writer FFFF to simulate the signature of writer III.

It is advised to optimize your systems by using 12 authentic signatures per writer for training and the other authentic signatures for validation. You could also perform a cross validation. Note that in the first version of the data set there are some online authentic signatures missing (only 12 reference signatures). In a second version of this data set we will provide you with this missing data.

The filename conventions will not be the same for testing the systems, i.e., we will use some random file names.

Code Submission

Participants are required to submit a software tool that is able to compare n (n<=12) specimen signatures against each of the questioned signature, and output the required values.
The interface of the verification tool consist of 4 parameters:

- The path and filename of the output file.
- The mode of the signatures (online/oflline/mixed)
- The path and filenames without extension of the specimen signatures
- The path and filename without extension of the questioned signature

The tool should be named *SigVerify* and should write a result line in the output file that should have the following format:
The filename without extension of the questioned signature; filename without extension of the specimen signatures, separated; A comparison score (e.g. a degree of similarity or difference); the evidential value of that score (see above).

Example of a program call in case a subset of 7 of the specimen signatures are used as reference set:
./SigVerify output.txt oflline questioned/0101020_02 genuine/020_01 genuine/020_02 genuine/020_03 genuine/020_04 genuine/020_05

When the output file *output.txt* already exists, the tool should append the result line to the bottom of the already existing result lines.
Expected output:
0101020_02 020_01 020_02 020_03 020_04 020_05 [score] [evidencial value]

Participants must deliver the verification tool preferably in a single ZIP file. The ZIP file must also contain a readme.txt with all relevant information of installation and running the program. The code will not be used for purposes other than the ICDAR2011 competition and publication. After the competition, all executable will be destroyed. However, if the participant agrees some further investigation and evaluation will be performed, i.e., we might perform a deeper analysis and comparison to the forensic experts or fuse the output scores of several systems – please state during submission when you do not agree to that. If a team chooses to use some expiration mechanism, the expiration should be set to 31 December 2011. Researchers/groups can decide to participate anonymously in the competition and not to be mentioned by name in any publication that follows.

Accessing the data:
The data sets can be downloaded at:
http://www.dfki.de/~liwicki/SigComp2011/trainingSet.zip
The code for opening the zip-file is:
I hereby accept the SigComp 2011 disclaimer.