signMine algorithm for conditioning and analysis of human handwriting

Abstract

We describe the signMine algorithm, which relates to methods for conditioning, representation, modeling, and analysis of variables, and is specially adapted for analysis of parametric line objects such as human handwritten signatures. signMine has applicability in all areas where signature identification and/or verification is desirable or required. We also present the signMine software package, which includes a searchable signature database and is designed for performing signature identification and/or verification.

In addition, we outline an approach to the analysis and modeling of human image biometrics through analog representation.

Outline

Part I: signMine algorithm

- Conditioning & Representation
- Analysis, Comparison, and Identification
- signMine at *www.contourmine.com*

Part II: Analysis and modeling of biometric information through analog representation

- Two-dimensional modulated linear densities
- Robustness vs. selectivity in comparison of densities
- Conclusion and future work

Conditioning & Representation

Conditioning & Representation

We would like to have:

- Piecewise-continuous (‘analog’) curve (e.g., as a simple scalar function of one variable)
- Independence from any choice of coordinates and/or parameterization
- Invariance with respect to such transformations as translation, rotation, and (isotropic) scaling

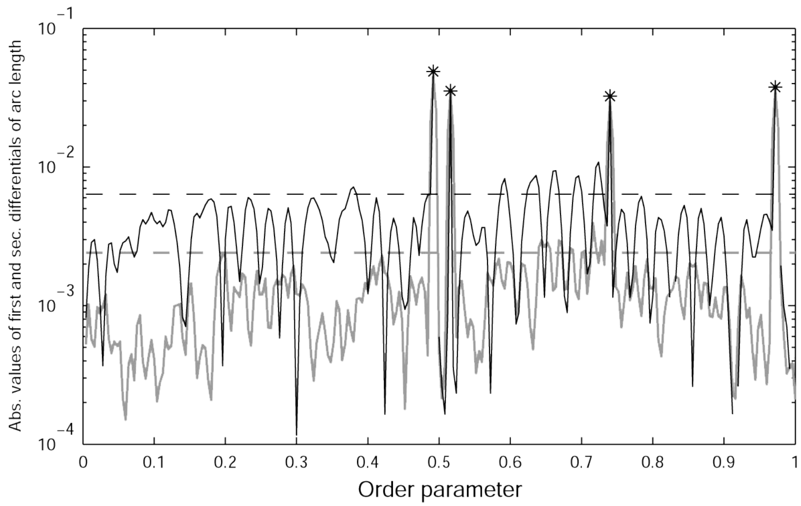
We typically have:

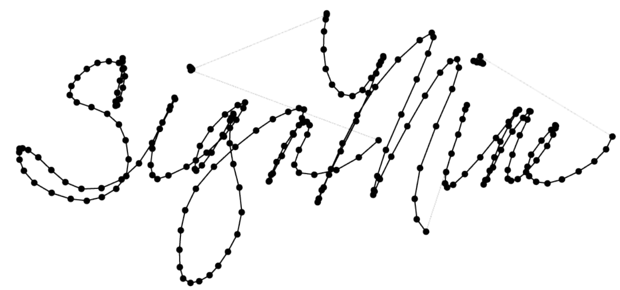
- Coarse and/or noisy digital record
- Discrete anisotropic grid (e.g., in 2D Cartesian coordinates)
- Uncertain segmentation
- Uncertain scale, position, and orientation
- Irregularities (i.e., vanishing speed)

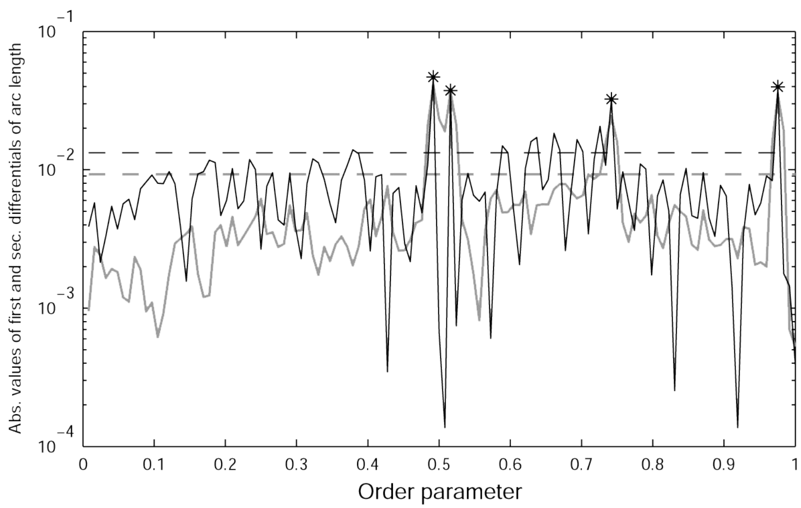
Typical steps of conditioning and representation:

- Robust (coincidence) segmentation
- Smoothing and/or tangential interpolation in order index
- Intrinsic equation and other analog representations of a piecewise-continuous curve

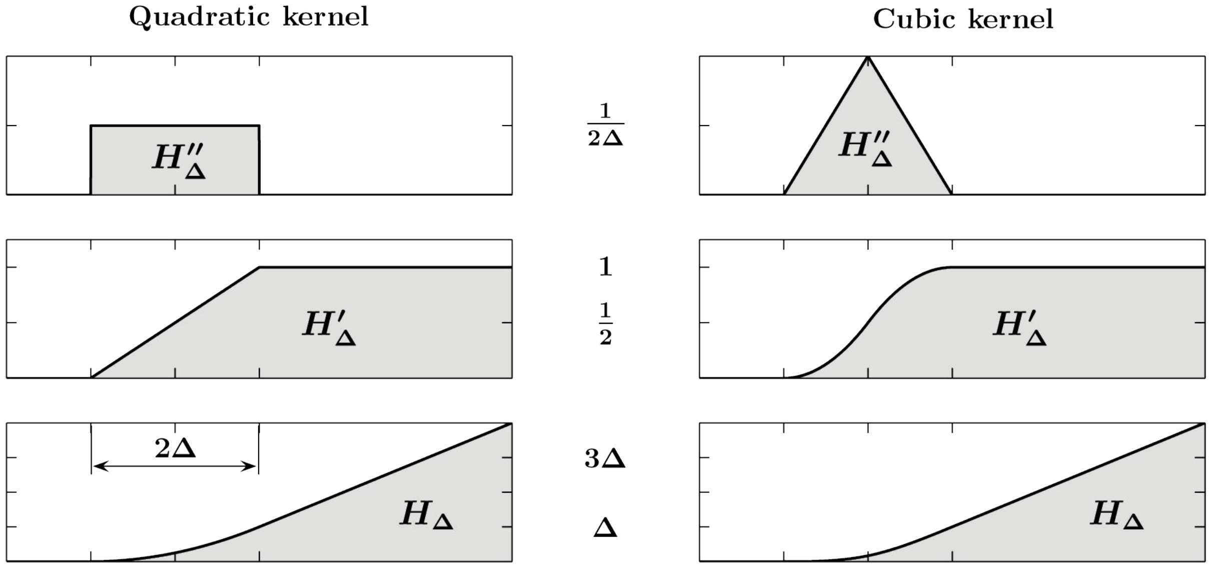
Robust (coincidence) segmentation





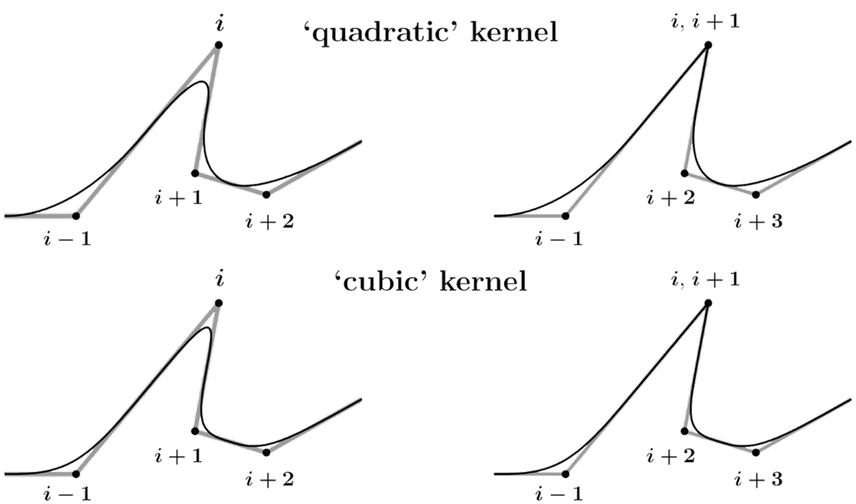


Smoothing and/or tangential interpolation in order index



$$\frac{d^n}{dx^n}\left[y(x) - y_0\right] = \begin{cases} \sum_{i=0}^{N-1} \Delta y_i \frac{d^{n+1}}{dx^{n+1}} H_\Delta(x - x_i) \\ \text{if} \quad \Delta x_i = 0 \\ \sum_{i=0}^{N-1} \frac{\Delta y_i}{\Delta x_i} \frac{d^n}{dx^n} \left[H_\Delta(x - x_i) - H_\Delta(x - x_{i+1})\right] \\ \text{otherwise} \end{cases}$$

Smoothing and tangential interpolations



Tangential interpolating curves constructed using quadratic and cubic kernels

Tangential (upper panel) and smoothing (lower panel) interpolations with a quadratic kernel

Intrinsic equation and other analog representations of a piecewise-continuous curve

Intrinsic (Whewell) equation:

$$z(s) = \int_0^s \mathrm{d}s'\, \mathrm{e}^{\mathrm{i}\phi(s')} + \sum_i \delta l(s_i)\, \mathrm{e}^{\mathrm{i}\phi(s_i)}\, \theta(s - s_i)$$

Same, but kinematic description:

$$z(t) = \int_0^t \mathrm{d}t'\, \dot{s}(t')\, \mathrm{e}^{\mathrm{i}\phi(t')} + \sum_i \delta l(t_i)\, \mathrm{e}^{\mathrm{i}\phi(t_i)}\, \theta(t - t_i)$$

Other representations: …

Analysis, Comparison, and Identification

Typical steps of analysis, comparison, and identification:

- Construction of various (e.g., circular and linear) distributions and their respective densities
- Introduction of various descriptive statistics and distance measures

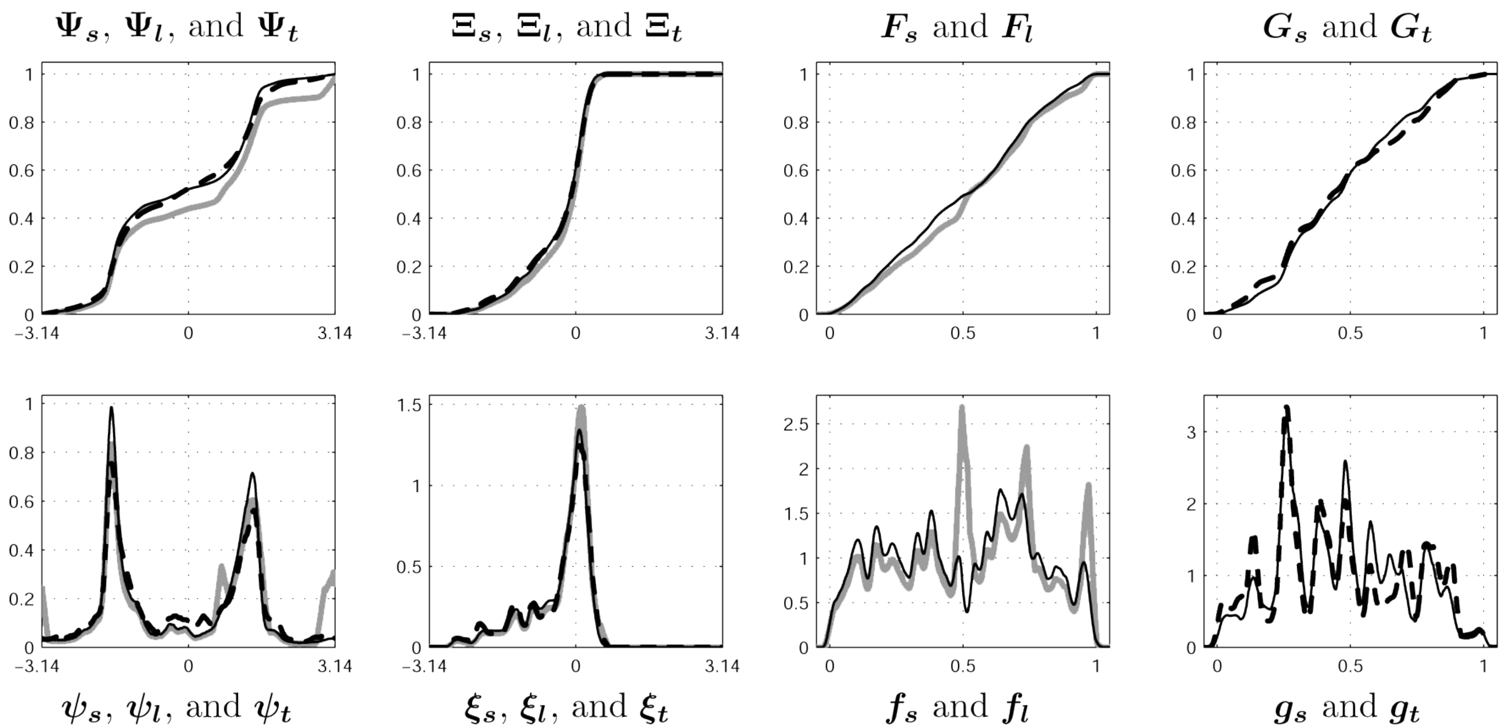
Example of a general modulated distribution:

$$\Phi(D) = \frac{\int_0^S \mathrm{d}s\, K(s)\, \mathcal{F}_{\Delta D}\left[D - x(s)\right]}{\int_0^S \mathrm{d}s\, K(s)},$$

where ***K*(*s*)** is a unipolar modulating signal, and

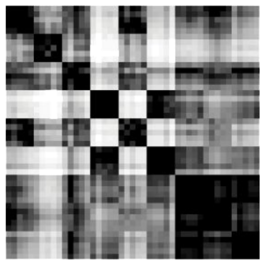
$$\lim_{\Delta D \to 0} \mathcal{F}_{\Delta D}(x) = \theta(x)\,.$$

Examples of angular and linear distributions and their respective densities

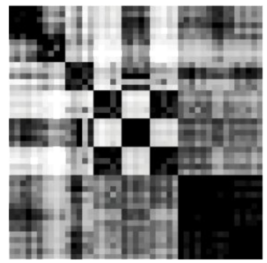


Examples of comparison through two-sample statistics

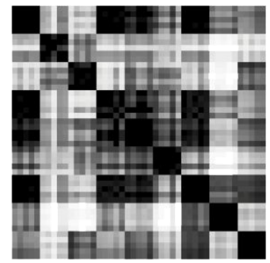
Ψ_s (Watson)



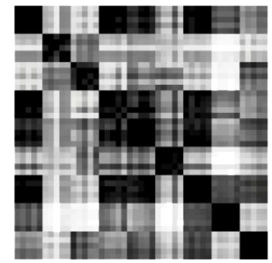
Ψ_l (Watson)



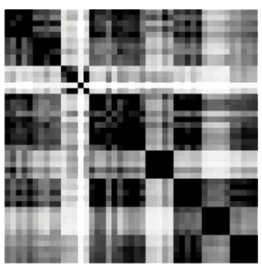
Ψ_t (Watson)



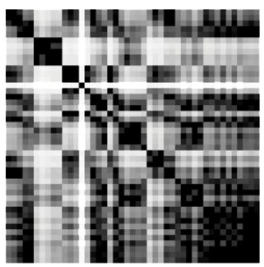
Ξ_s (Watson)



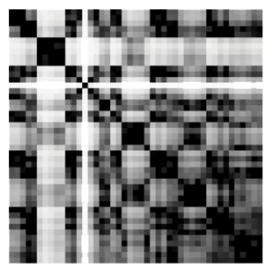
Ξ_l (Watson)



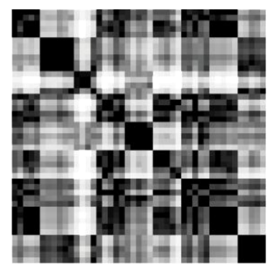
Ξ_t (Watson)



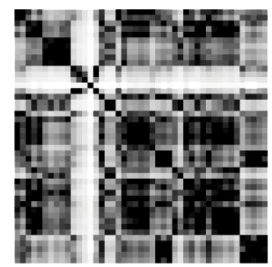
F_s (Cramér-von Mises)



F_l (Cramér-von Mises)

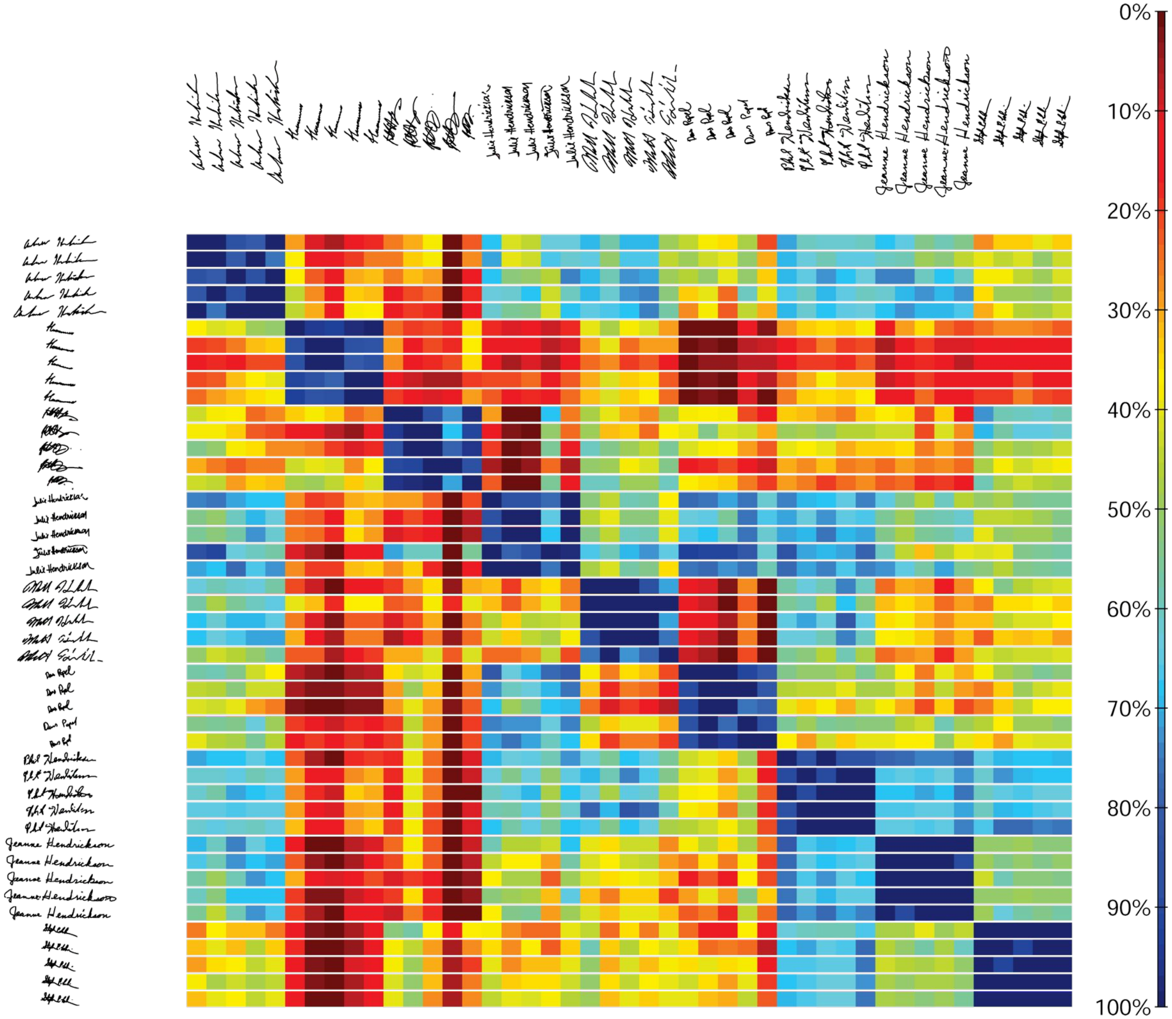


G_s (Cramér-von Mises)



G_t (Cramér-von Mises)

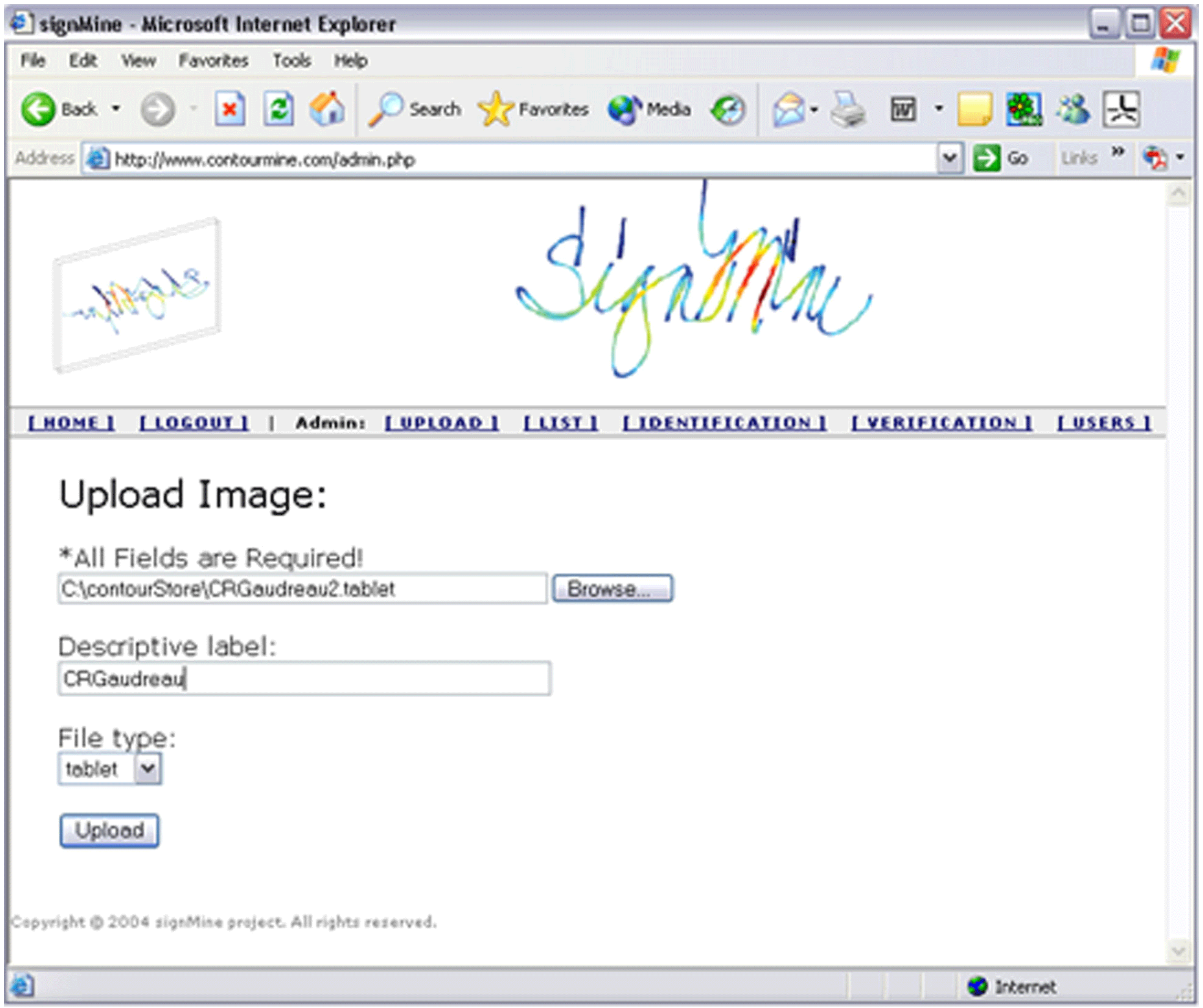
Example of combined percentile comparison



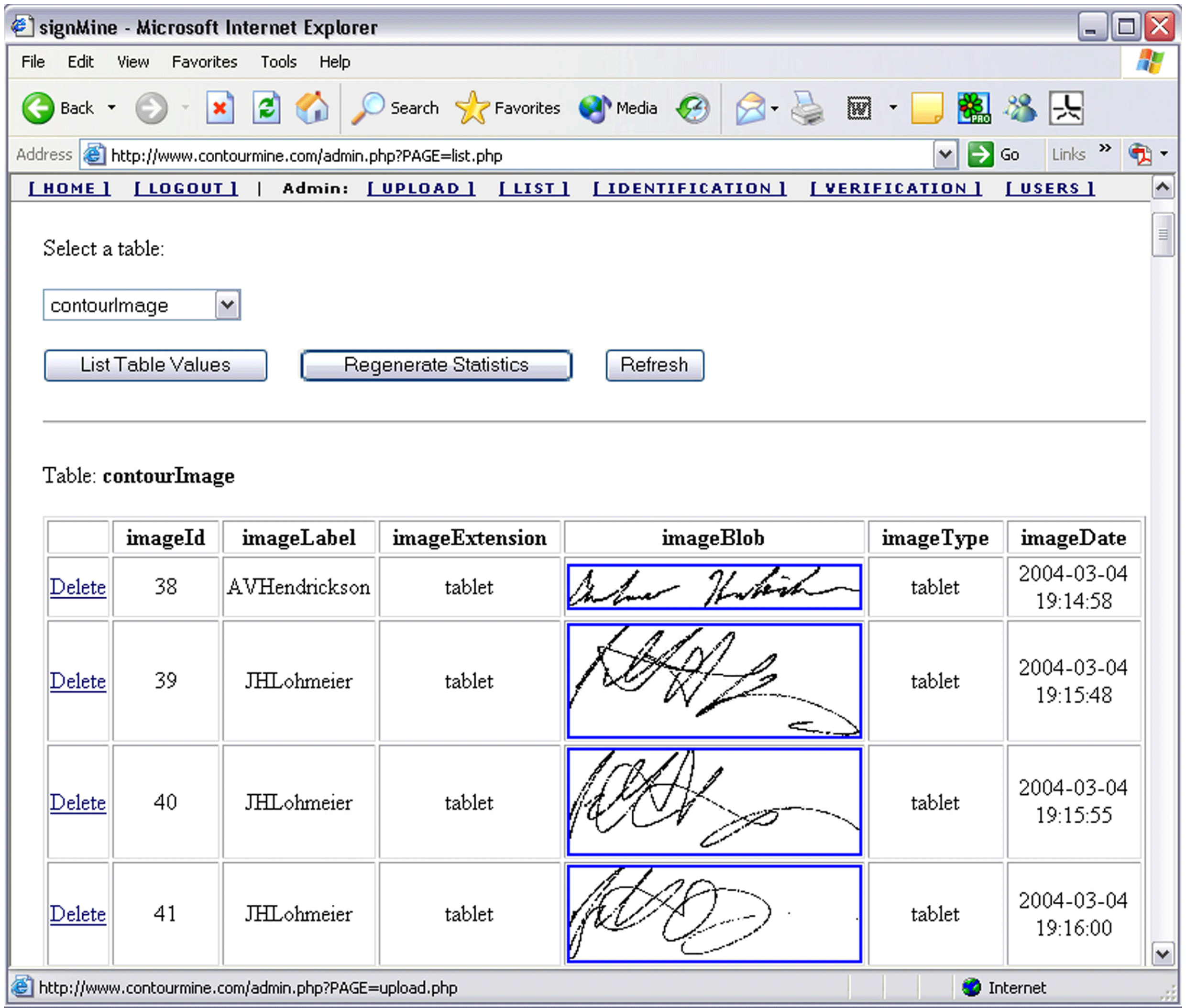
signMine at *http://www.contourmine.com*

signMine includes (i) signature acquisition tools, (ii) a searchable signature database (the signMine engine), and (iii) an online interface

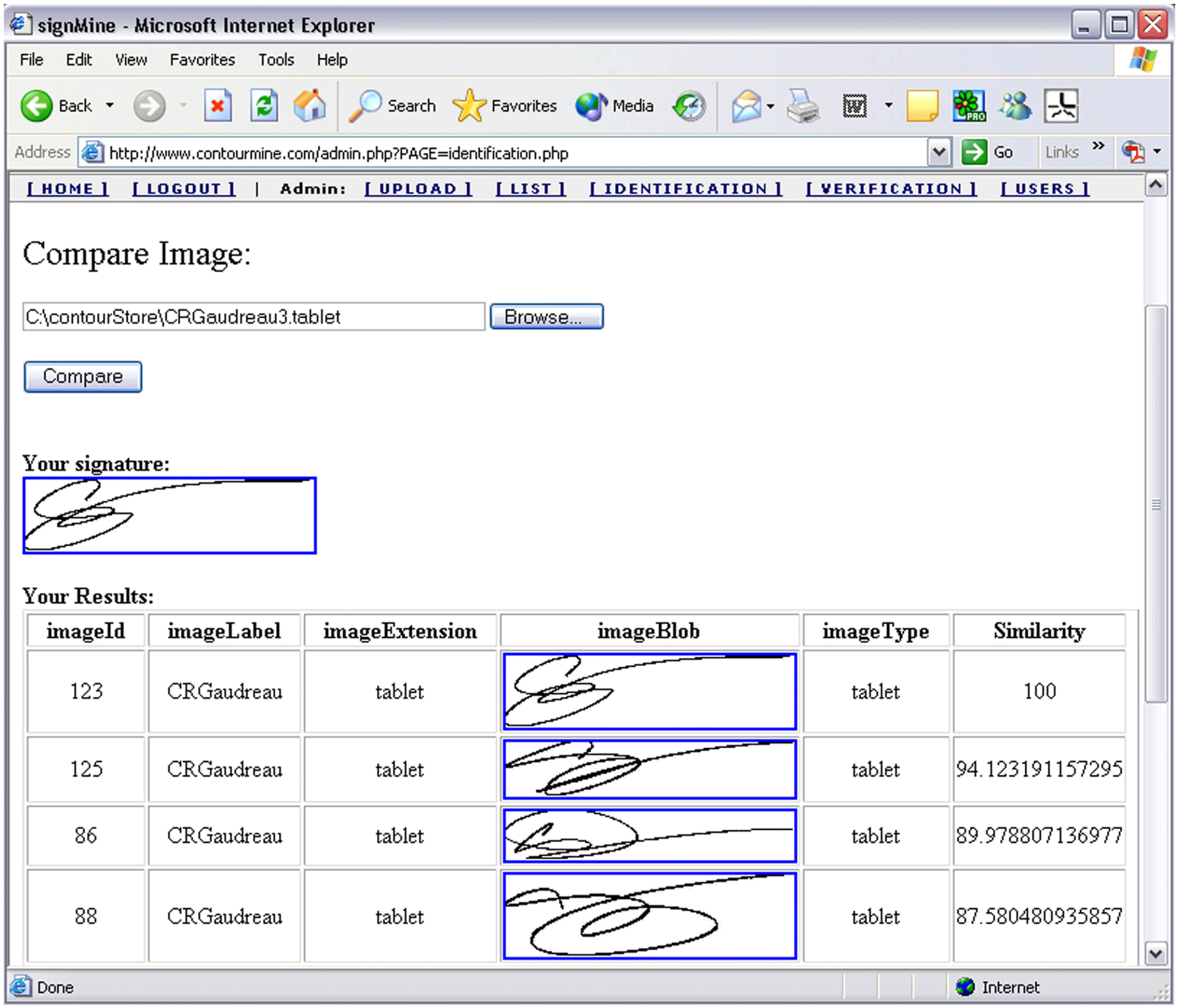
Screenshot of the upload **module**



Screenshot of the list module



Screenshot of the identification **module**



imageId	imageLabel	imageExtension	imageBlob	imageType	Similarity
123	CRGaudreau	tablet		tablet	100
125	CRGaudreau	tablet		tablet	94.123191157295
86	CRGaudreau	tablet		tablet	89.978807136977
88	CRGaudreau	tablet		tablet	87.580480935857

Two-dimensional modulated linear densities

2D density of a line drawn by a realistic instrument

The modulated linear density function $\Phi(\mathbf{R})$ of a line drawn by a writing utensil with the tip profile f_d can be represented as

$$\Phi(\mathbf{R}) = \frac{1}{M} \int_0^T \mathrm{d}t \, \mu(t) \, |\dot{\mathbf{r}}(t)| \, f_d \left(|\mathbf{R} - \mathbf{r}(t)| \right) \ ,$$

where $\mu(t)$ is the modulating parameter along the line of uniform density, $|\dot{r}(t)|$ is the speed of the movement of the tip, T is the duration of writing, and

$$M = \int_0^T \mathrm{d}t \, \mu(t) \, |\dot{\mathbf{r}}(t)|$$

is the total "pseudomass" of the trajectory.

Modulated linear densities

Example: Pen with radial tip profile $f_d(r) = \frac{4}{\pi d^2}\,\theta(d - 2r)$

Density of ink left on paper:

$$\Phi(\mathbf{R}) = \frac{4}{\pi d^2 \Lambda} \int_0^T \mathrm{d}t\, \lambda(t)\, \theta\left(d - 2|\mathbf{R} - \mathbf{r}(t)|\right) ,$$

where $\Lambda = \int_0^T \mathrm{d}t\, \lambda(t)$ is the total amount of used ink

Notice that in this example the modulation is expressed as

$$\mu(t) = \lambda(t) / |\dot{\mathbf{r}}(t)| ,$$

and thus the thickness of the line (the amount of ink per unit length) is inversely proportional to the speed of movement of the pen

Computation in finite differences:

Given a relatively short parametric record of a line (typically of order 10^3 points), we convert this record into a high resolution image which can be numerically treated as a continuous object. This can be done through a convolution with a kernel f_d (representing the writing utensil and/or the reading instrument) such that its characteristic width is large in comparison with the cell of the spatial grid R_{ij}:

$$\Phi(\mathbf{R}_{ij}) = \frac{\sum_{k=1}^{N} \mu_k \left|\mathbf{r}_{k+1} - \mathbf{r}_{k-1}\right| f_d\left(\left|\mathbf{R}_{ij} - \mathbf{r}_k\right|\right)}{\sum_{k=1}^{N} \mu_k \left|\mathbf{r}_{k+1} - \mathbf{r}_{k-1}\right|}$$

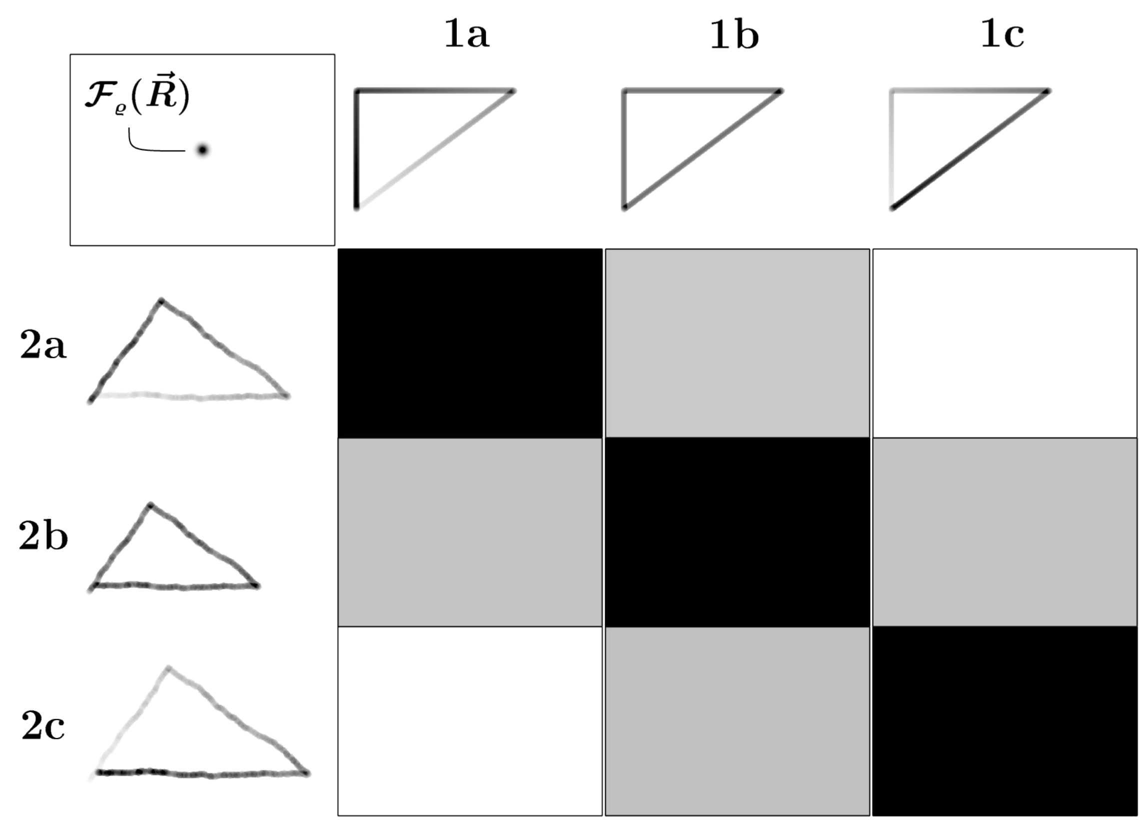
Modulated linear densities

Comparison of densities using statistic

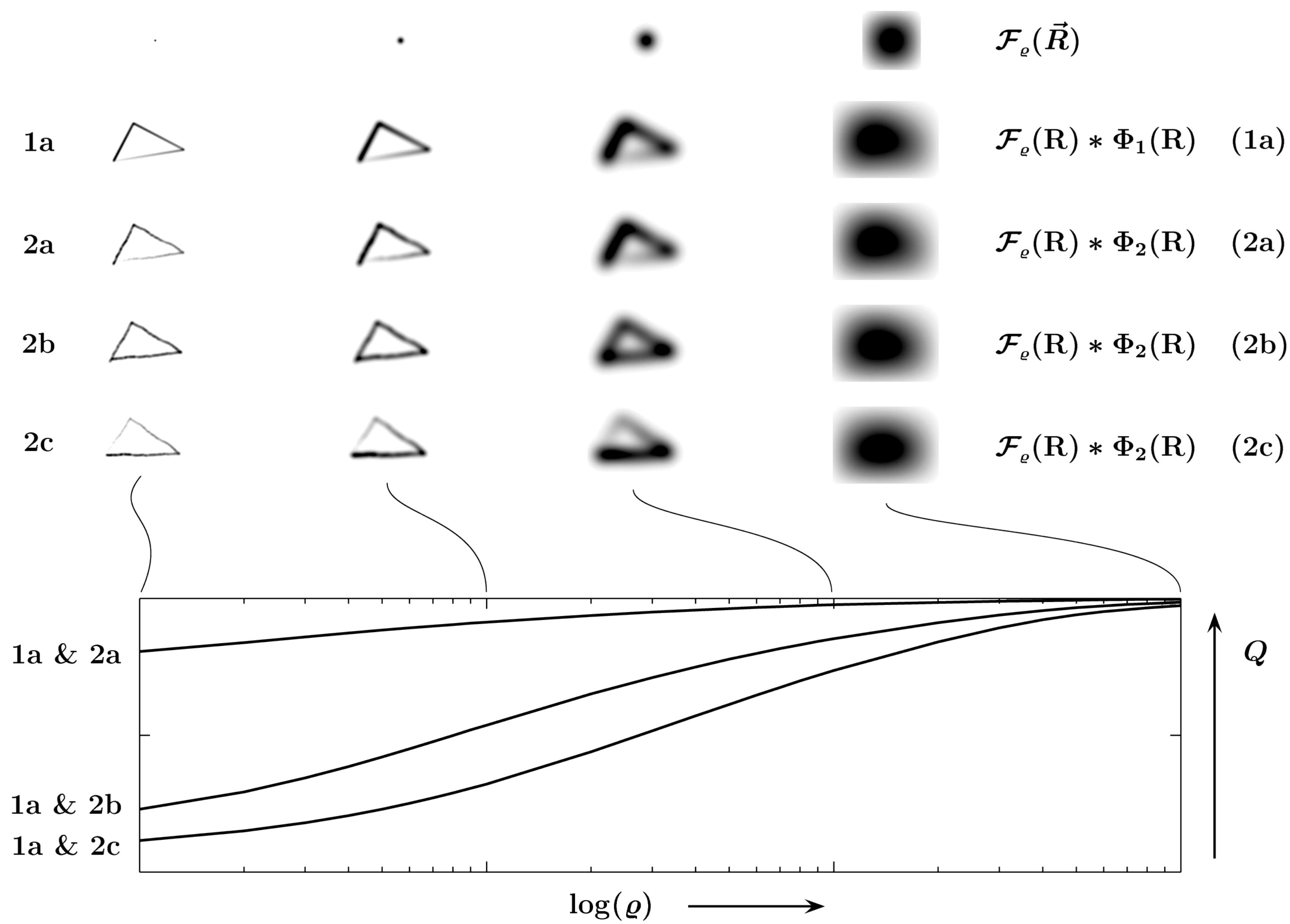
$$1 \geq Q = 1 - \frac{1}{2}\int_{-\infty}^{\infty} \mathbf{d}^2\mathbf{r}\ |\Psi_1(\mathbf{r}) - \Psi_2(\mathbf{r})| \geq 0\,,$$

where

$$\Psi(\mathrm{R}) = \mathcal{F}_\varrho(\mathrm{R}) * \Phi(\mathrm{R})$$



Robustness vs. selectivity in comparison of densities



Conclusion and future work

Conclusion:

- The signMine algorithm uses analog representation for conditioning, modeling, and analysis of human handwritten signatures
- The algorithms and techniques developed in signMine can be used to analyze and model other image biometrics (fingerprints, facial characteristics, etc.)

Future work may include:

- Formalizing the approach for implementation of a self-learning data storage
- Building a larger database of signatures (possible artificially generated) to test the scalability and robustness of the algorithm
- Extending the applicability of the algorithm by developing tools for signatures acquired from flatbed scanners and touch screens