Classification of Handwritten Mathematical Symbols Subjected to Shear Distortions

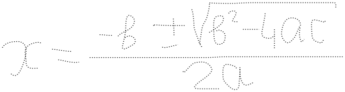
Oleg Golubitsky, Vadim Mazalov and Stephen M. Watt

Ontario Research Center for Computer Algebra
Department of Computer Science
University of Western Ontario

Joint Lab Meeting, University of Waterloo
January 22, 2010

Introduction

- We are working towards online recognition of handwritten mathematics in pen-based environment.



$$\longrightarrow \qquad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

- We were able to achieve 96.3% recognition rate of handwritten samples, subjected to rotation.

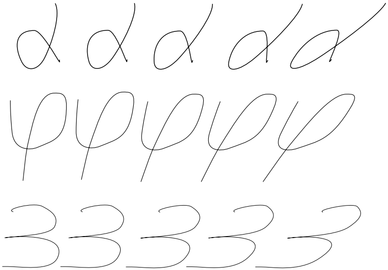
- Now we ask, whether the algorithm can include shear-invariance without drop in the classification rate.

Organization of the Presentation

- Shear.
- Integral invariants.
- Approximation of invariants.
- Algorithm.
- Size normalization of samples.
- Parameterization of coordinate functions.
- Mixed parameterization.
- Towards affine recognition.
- Conclusion.

Shear

- Commonly occurs in practice, possibly even more often than rotation.



- Easily recognizable by a human, even for a large degree of transformation (more than 1 radian).

Integral Invariants

- Provide an elegant approach to planar and spatial curve classification under affine transformations.
- Relatively insensitive to local random noise, as opposed to differential invariants.
- Computed from coordinate functions, which are approximated online, i.e. when the curve is being written, with a minor overhead after pen-up.

As the name suggests, integral invariants are constructed from quantities obtained by integration.

Functional Representation

- I_1 and I_2 are invariant under special linear group $SL(2,R)$ and defined as:

$$I_1(\lambda) = \int_0^\lambda X(\tau)dY(\tau) - \frac{1}{2}X(\lambda)Y(\lambda)$$

$$I_2(\lambda) = X(\lambda)\int_0^\lambda X(\tau)Y(\tau)dY(\tau) - \frac{1}{2}Y(\lambda)\int_0^\lambda X^2(\tau)dY(\tau) - \frac{1}{6}X^2(\lambda)Y^2(\lambda)$$

where $X(\lambda)$, $Y(\lambda)$ are coordinate functions.

Numerical Representation

- After approximation of coordinate functions with truncated Legendre-Sobolev polynomial series, invariants take form

$$I_1(\lambda) \approx \sum_{i,j=1}^{d} x_i y_j \left[\int_0^\lambda P_i(\tau) P_j'(\tau) d\tau - \frac{1}{2} P_i(\lambda) P_j(\lambda) \right]$$

$$I_2(\lambda) \approx \sum_{i,j,k,l=1}^{d} x_i x_j y_k y_l \mu_{ijkl}$$

$$\begin{aligned}\mu_{ijkl} = P_i(\lambda) \int_0^\lambda P_j(\tau) P_k(\tau) P_l'(\tau) d\tau \\ - \frac{1}{2} P_l(\lambda) \int_0^\lambda P_i(\tau) P_j(\tau) P_k'(\tau) d\tau \\ - \frac{1}{6} P_i(\lambda) P_j(\lambda) P_k(\lambda) P_l(\lambda).\end{aligned}$$

Approximation of invariants

Maximum absolute and average relative errors in coefficients of invariants of initial samples and sheared by 1 radian

	I_1		I_2	
Degree	Abs. Err.	Rel. Err.	Abs. Err.	Rel. Err.
2	9×10^{-12}	3×10^{-19}	3×10^{-11}	9×10^{-20}
3	1×10^{-11}	4×10^{-19}	8×10^{-10}	2×10^{-19}
4	5×10^{-11}	9×10^{-19}	1×10^{-9}	4×10^{-19}
5	6×10^{-11}	3×10^{-18}	3×10^{-9}	1×10^{-18}
6	3×10^{-10}	1×10^{-17}	9×10^{-9}	5×10^{-18}
7	2×10^{-9}	5×10^{-17}	7×10^{-8}	2×10^{-17}
8	3×10^{-8}	2×10^{-16}	1×10^{-7}	1×10^{-16}
9	2×10^{-7}	1×10^{-15}	6×10^{-7}	5×10^{-16}
10	2×10^{-6}	6×10^{-15}	4×10^{-6}	2×10^{-15}
11	5×10^{-6}	3×10^{-14}	2×10^{-5}	1×10^{-14}
12	1×10^{-5}	1×10^{-13}	7×10^{-5}	6×10^{-14}
13	4×10^{-5}	7×10^{-13}	5×10^{-4}	3×10^{-13}
14	3×10^{-4}	4×10^{-12}	3×10^{-3}	1×10^{-12}
15	1×10^{-3}	2×10^{-11}	7×10^{-3}	8×10^{-12}
16	1×10^{-2}	1×10^{-10}	2×10^{-2}	5×10^{-11}
17	5×10^{-2}	5×10^{-10}	3×10^{-2}	6×10^{-10}
18	4×10^{-1}	3×10^{-9}	3×10^{-1}	4×10^{-9}

We take degree=12.

Coefficients of Invariants

- Taking the discussed representation of $I_1(\lambda)$, coefficients of the invariant are computed as

$$I_{1,i} = \frac{\langle I_1, P_i \rangle}{\langle P_i, P_i \rangle} \quad i = 1..d$$

  where $\langle \cdot, \cdot \rangle$ is the Legendre-Sobolev inner product.

- Similarly, we calculate coefficients for $I_2(\lambda)$ and obtain a $2d$-dimensional vector for each sample

$$(I_{1,1}, \ldots, I_{1,d}, I_{2,1}, \ldots, I_{2,d}).$$

Classification

- We first find top N candidate classes with the distance from the subject sample to training samples in the space of coefficients of approximation of invariants.

- We consider each of N classes to find the minimal distance to the subject sample with respect to different levels of shear.

- To do so, we solve minimization problem

$$\min_{\varphi}\left(\sum_k (X_k - (x_k + y_k \tan\varphi))^2\right)$$

Size Normalization: Existing Methods

Traditionally implemented by rescaling a sample to achieve standard values of certain parameters, such as

- Euclidean norm of the vector of Legendre-Sobolev coefficients of coordinate functions.
- Height of a sample (works for horizontal shear, but not for rotation).
- Aspect-ratio size normalization (may work for rotation, but becomes inaccurate for horizontal shear).

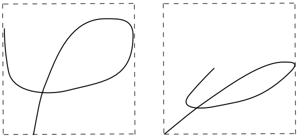


Figure: Aspect ratio size normalization.

Size Normalization: Our Approach

- For the case of shear (and affine) recognition we look at the norm $\|I_1\|$ of the coefficient vector of I_1.

- Coefficients of coordinate functions are normalized by multiplication by $1/\sqrt{\|I_1\|}$.

- Computing the norm of I_1 allows to extend the invariance of I_1 and I_2 from the special linear group, $SL(2,R)$, to the general linear group, $GL(2,R)$. Invariance under the general affine group, $Aff(2,R)$, is obtained by dropping the first (order-0) coefficients from the coefficient vectors of the coordinate functions.

Linear Symbols

Taking the norm of I_1 may perform poorly for samples that have a linear shape, such as "-", "\", "/", "l", ".". Being an area between the curve and its secant, I_1 is close to zero for such characters.

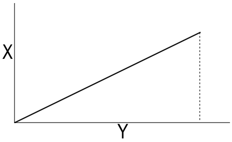


Figure: I_1 for a linear symbol.

Therefore, some of the methods, designed for 2-dimensional curves, are indeed not suitable for 1-dimensional symbols, and such linear characters require special treatment.

Parameterization of Coordinate Functions

We look at the following parameterization choices

- By time.
- By Euclidean arc length

$$F(\lambda) = \int_0^\lambda \sqrt{(X'(\tau))^2 + (Y'(\tau))^2} d\tau.$$

- By affine arc length

$$\hat{F}(\lambda) = \int_0^\lambda \sqrt[3]{|X'(\tau)Y''(\tau) - X''(\tau)Y'(\tau)|} d\tau.$$

Evaluation of Results

The results we obtained can be viewed on the table for different types of parameterization of coordinate functions: by time, by Euclidean arc length (AL) and by affine arc length (AAL); and for discussed size normalization approaches.

(a) Size normalization by height

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAL	82.2	82.2	82.2	82.1	82.1	82.1	82.1	82.1	82.1	82
AL	96.4	96.4	96.1	95.6	95	94.1	93	91.9	90.2	88
Time	94.8	94.9	94.9	94.7	94.5	94.4	94.4	94.4	94.4	94.3

(b) Aspect ratio size normalization

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAL	81.9	81.8	81.6	81.4	81.2	81	80.8	80.2	79.4	77.4
AL	96.3	96.4	96.1	95.5	94.7	93.7	92.3	90.1	85.7	77.5
Time	94.7	94.7	94.6	94.3	94.1	93.9	93.7	93.2	91.9	89.

(c) Size normalization by l_1

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
AAL	83	83.1	83	82.9	82.9	82.8	82.8	82.8	82.8	82.7
AL	**96.3**	**96.3**	**96.1**	**95.7**	**95.1**	**94.4**	**93.3**	**91.9**	**90.2**	**87.9**
Time	94.6	94.7	94.6	94.5	94.5	94.5	94.5	94.5	94.5	94.4

Mixed Parameterization

Parameterization by time gives low recognition rate, while remains affine-invariant. The picture is opposite with parameterization by arc length. We therefore propose to unite these two parameterization approaches in the form of mixed parameterization as follows

- Divide the curve in N equal time intervals, and parameterize each interval by arc length.
- Smooth the transition from time to arc length with a mixed metric of the form $kdt^2 + dx^2 + dy^2$ inside the subintervals, where k is a parameter.
- The optimal values of N and k are found by cross-validation.

Recognition rate for different N and k

We increased the number of selected classes to 50 and obtained the following rates

Table: Recognition rate (%) for mixed parameterization for corresponding values of N and k

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$N = 2,\ k = 1$	96.1	96.2	96	95.9	95.8	95.7	95.6	95.3	95.2	94.9
$N = 2,\ k = 2$	95.8	95.9	96	95.8	95.7	95.7	95.7	95.6	95.5	95.5
$N = 3,\ k = 0.5$	96.1	96.2	96	95.9	95.9	95.7	95.5	95.6	95.3	95
$N = 3,\ k = 1$	95.9	96.2	96	95.8	96	95.8	95.7	95.6	95.5	95.2
$N = 4,\ k = 1$	95.9	96.1	95.9	95.7	95.8	95.6	95.6	95.7	95.6	95.4

We take $N = 2,\ k = 2$ as these values allow to obtain high rate for non-sheared samples and stay invariant when significant distortions take place.

Recognition rate for different N and k

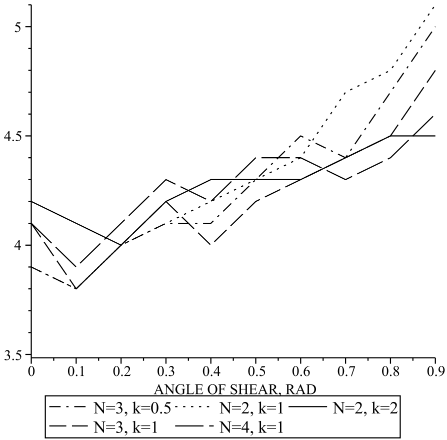


Figure: Error (%) for the mixed parameterization for different values of skew.

Towards Affine Recognition

We propose two approaches to deploy integral invariants for affine recognition

- Consider both rotation and horizontal shear in the same system. The minimization problem would have to include two equations in two variables.

- Define a distance that consists of weighted coefficients of coordinate functions and integral invariants. Size of a sample may also be included in the distance metrics.

Conclusion

We have developed the following

- Size normalization of samples, when shear (and more generally, affine) transformations take place.
- Mixed parameterization of coordinate functions that allows to obtain high recognition rate, while being invariant to affine distortions.
- Shear-invariant algorithm that in conjunction with rotation-invariance allows to model recognition of samples, subjected to the most common transformations.

References

1. S. Feng, I. Kogan and H. Krim, Classification of Curves in 2D and 3D via Affine Integral Signatures, to appear in *Acta Appl. Math.*, 2008, `http://arxiv.org/abs/0806.1984`.