Local Context in Nonlinear Deformation Models for Handwritten Character Recognition

Daniel Keysers • **Christian Gollan** • **Hermann Ney**

Lehrstuhl für Informatik VI
Human Language Technology and Pattern Recognition
Computer Science Department
RWTH Aachen University
D-52056 Aachen, Germany

Overview

Introduction

State-of-the-Art

Nonlinear Matching

Local Context

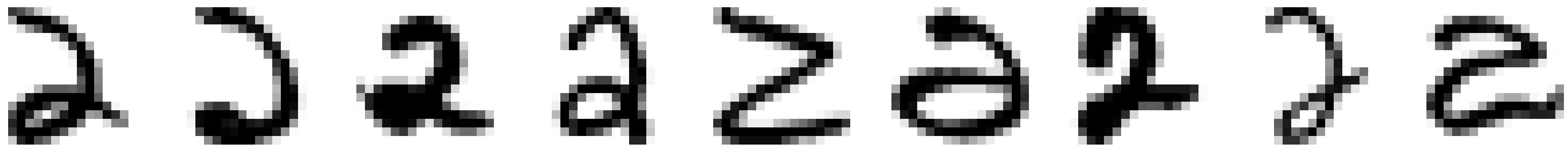
Experiments and Results

Conclusion

Introduction

object recognition $\Rightarrow$ suitable models of image transformation important

recognition of handwritten characters is prototypical:
– different writing styles $\Rightarrow$ varying appearances
– reference model for each class is well-defined



evaluate two-dimensional nonlinear deformation models
(true two-dimensional, pseudo-two-dimensional, zero-order)
using local image context of each pixel

hope: fewer 2D constraints can be compensated by using local context
$\Rightarrow$ confirmed by experimental results

very competitive results across five different tasks (e.g. 0.54% on MNIST)

State-of-the-Art in Character Recognition

Belonging & Malik^{+} 2002 (Berkeley) — shape context matching:
shape contexts = log-polar histograms of contour points
iterative matching with 2D-splines and the Hungarian algorithm

DeCoste & Schölkopf^{+} 2002 (Caltech / MPI Tübingen) — invariant SVM:
use virtual data and kernel jittering in support vector machine

Simard 2003 (Microsoft Research) — convolutional neural net:
generate large amount of virtual data on the fly (e.g. factor 1000)
during training of a well-designed neural network

Nonlinear Matching

test image $A = \{a_{ij}\}$ **reference image** $B = \{b_{xy}\}$ $a_{ij}, b_{xy} \in \mathbb{R}^U$

image deformation mapping $(x_{11}^{IJ}, y_{11}^{IJ}) : (i,j) \mapsto (x_{ij}, y_{ij})$

mappings must fulfill constraints: $(x_{11}^{IJ}, y_{11}^{IJ}) \in \mathcal{M}$

decision rule:

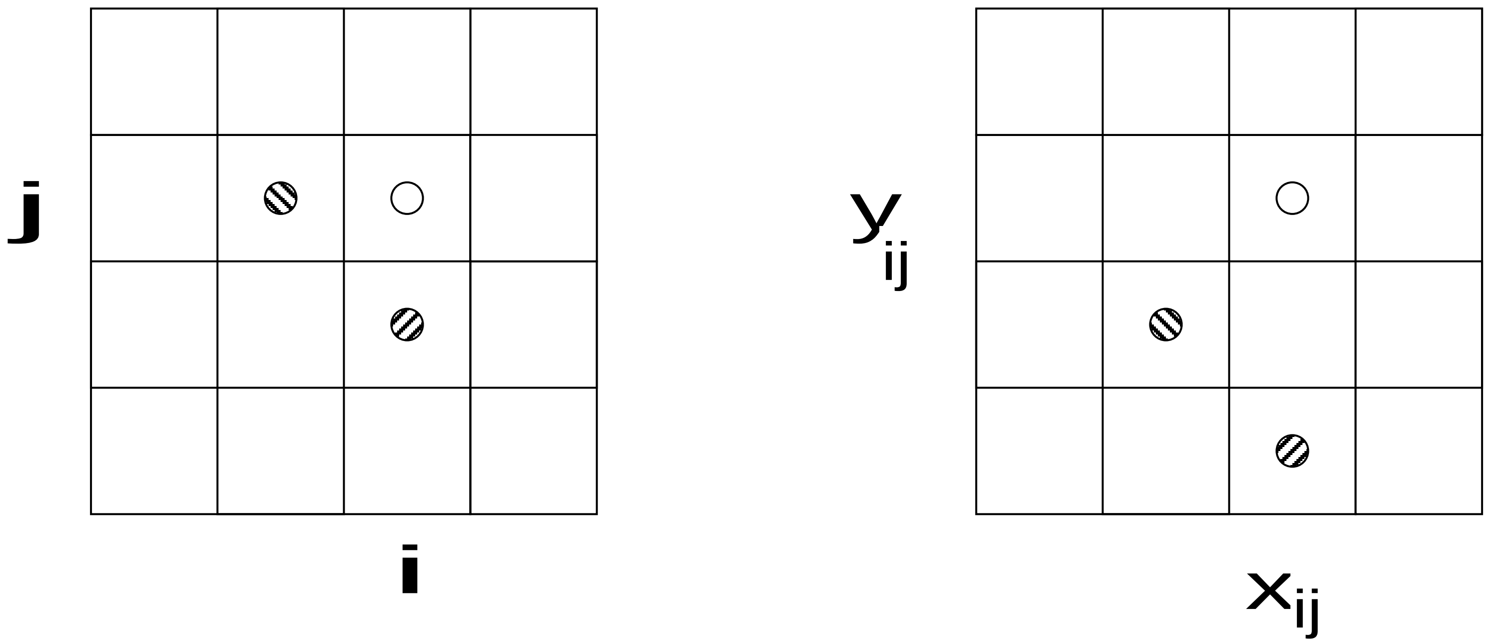
$$r(A) = \arg\min_k \left\{ \min_{n=1,\dots,N_k} d(A, B_{nk}) \right\}$$

$$d(A,B) = \min_{(x_{11}^{IJ}, y_{11}^{IJ}) \in \mathcal{M}} \left\{ d'\left(A, B_{(x_{11}^{IJ}, y_{11}^{IJ})}\right) \right\}$$

$$d'\left(A, B_{(x_{11}^{IJ}, y_{11}^{IJ})}\right) = \sum_{i,j} \sum_u ||a_{ij}^u - b_{x_{ij}y_{ij}}^u||^2$$

2D Dependencies

image deformation mapping $(x_{11}^{IJ}, y_{11}^{IJ}) \in \mathcal{M} : (i,j) \mapsto (x_{ij}, y_{ij})$



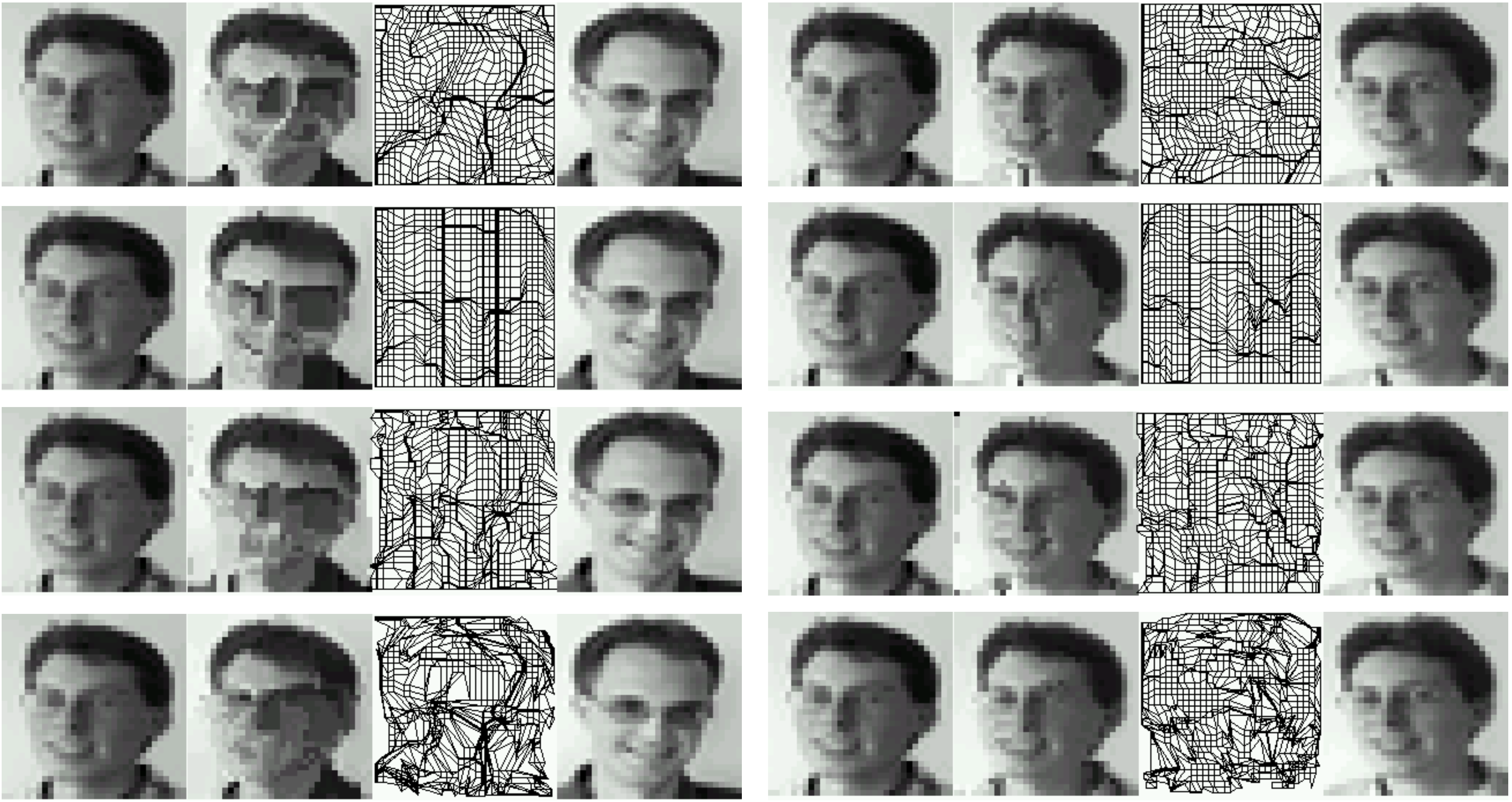
Models

informal descriptions of the used models:

2DW	**2-Dimensional Warping (order 2)** **complete 2D constraints, minimization NP-complete**
P2DHMM	**Pseudo 2-Dimensional Hidden Markov Model (order 1)** **match columns on columns, regard columns as independent**
P2DHMDM	**Pseudo 2-Dimensional Hidden Markov Distortion Model (order 1)** **allow additional horizontal displacements in P2DHMM**
IDM	**Image Distortion Model (order 0)** **disregard relative displacements of neighboring pixels** **restrict absolute displacement**

Examples

'different classes' 'same class'



2DW

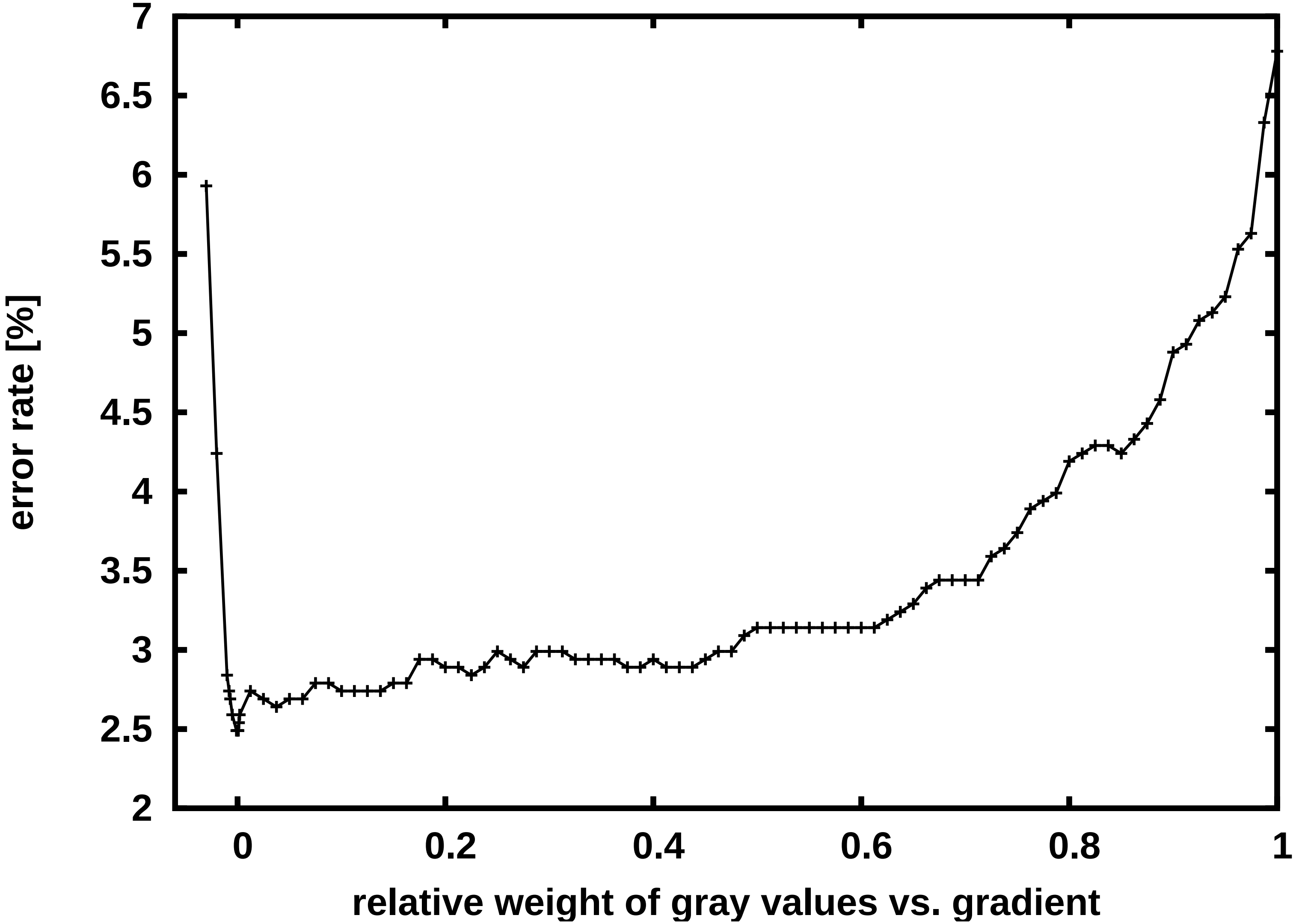
P2DHMM

P2DHMDM

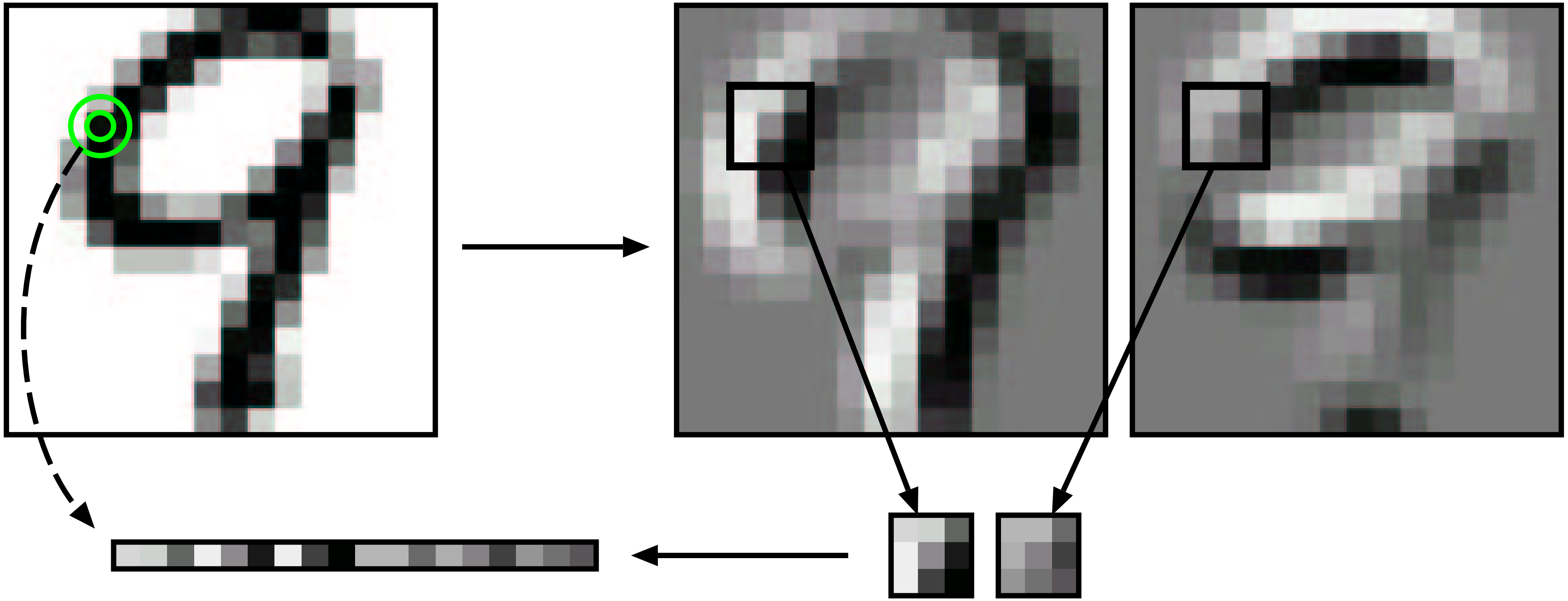
IDM

Local Context 1: Image Gradient

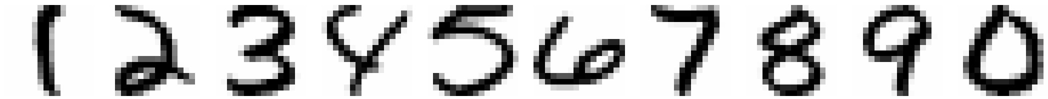
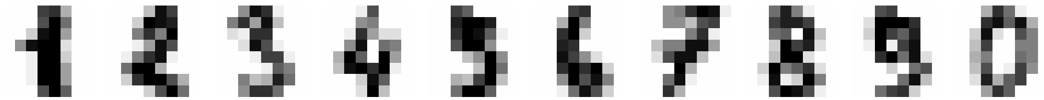
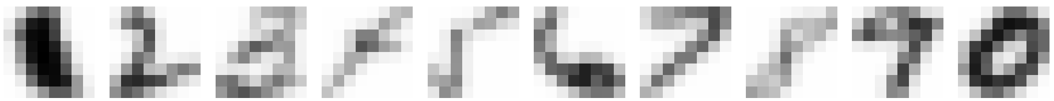
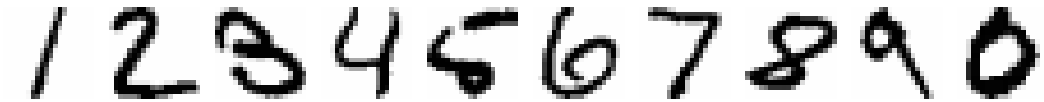
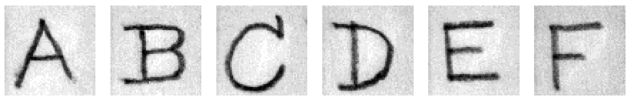
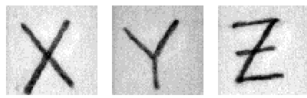
error rate on USPS for P2DHMM using gray values and image gradients (Sobel)



Local Context 2: Windows

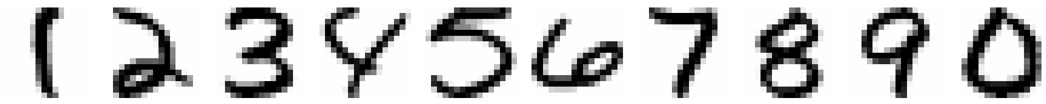


Experiments: Character Corpora

name	example images	size	# train	# test
USPS		16×16	7 291	2 007
UCI		8×8	3 823	1 797
MCEDAR		8×8	11 000	2 711
MNIST		28×28	60 000	10 000
ETL6A	 · · · 	64×63	15 600	13 000

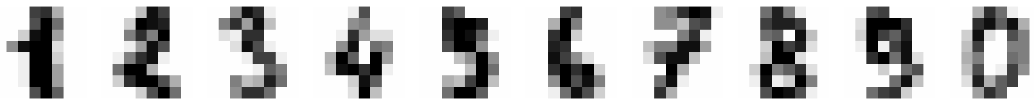
all matching experiments use 3×3 context of gradients

USPS

name	example images	size	# train	# test
USPS		16×16	7 291	2 007

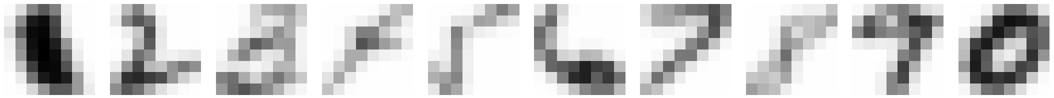
method		ER[%]
no matching, 1-NN		5.6
...		...
invariant SVM	external	3.0
2DW, 1-NN	RWTH	2.7
tangent distance, KD, virtual data	RWTH	2.4
IDM, 1-NN	RWTH	2.4
extended SVM training	external	2.2
local features + tangent distance	RWTH	2.0
P2DHMDM, 3-NN	RWTH	1.9

UCI

name	example images	size	# train	# test
UCI		8×8	3 823	1 797

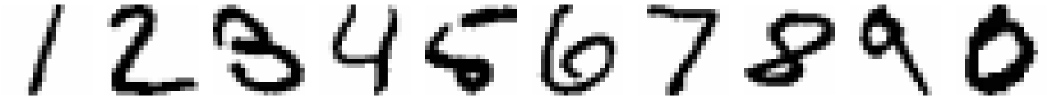
method		ER[%]
no matching, 1-NN		2.0
PCA mixture	external	1.5
P2DHMDM, 1-NN	RWTH	0.8
IDM, 1-NN	RWTH	0.8

MCEDAR

name	example images	size	# train	# test
MCEDAR		8×8	11 000	2 711

method		ER[%]
no matching, 1-NN		5.7
PCA	external	4.9
factor analysis	external	4.7
probabilistic PCA	external	4.6
IDM, 3-NN	RWTH	3.5
P2DHMDM, 3-NN	RWTH	3.3

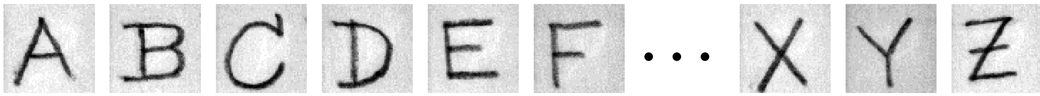
MNIST

name	example images	size	# train	# test
MNIST		28×28	60 000	10 000

method		ER[%]
no matching, 1-NN		3.1
deslant, Euclidean distance, k-NN	external	2.4
tangent distance, KD, virtual data	RWTH	1.0
distortions, neural net, boosting	external	0.7
shape context matching, 3-NN	external	0.63
invariant SVM	external	0.56
IDM, 3-NN$*$	RWTH	0.54
distortions+, neural net$*$	external	0.42
combination of best four systems		0.35

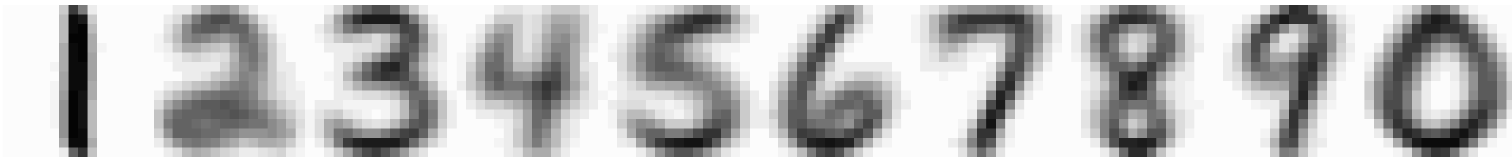
$*$: differences not statistically significant according to bootstrap analysis

ETL6A

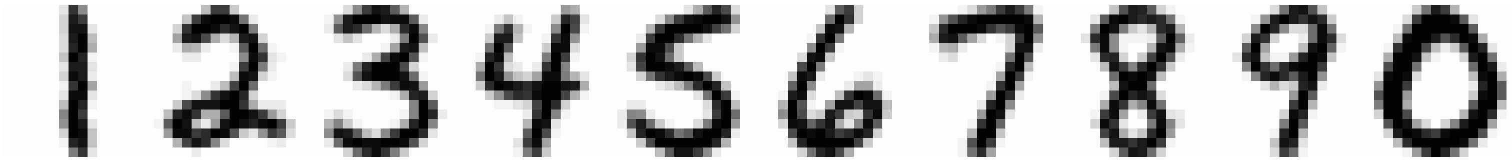
name	example images	size	# train	# test
ETL6A		64×63	15 600	13 000

method		ER[%]
no matching, 1-NN		4.5
piece-wise linear 2D-HMM	external	0.9
Eigen-deformations	external	0.5
IDM, 3-NN	RWTH	0.5

Learning Prototypes



means without deformation



means with deformation

USPS error rates [%] in key experiment

distance	deformation in mean calculation	
	no	yes
no deformation	18.6	26.1
P2DHMDM	25.3	4.9

best other single-prototype result: 5.7%
(Gaussian single densities, discriminative training, no deformation)

Conclusions

Conclusions

– performance increase with local context for all models
– context information $\Rightarrow$ very good results even with simple distortion models
– excellent results for handwritten character recognition, generalizes across tasks
– [combination of four best systems: 0.35% on MNIST]

software + parameter examples can be downloaded from:
`http://www-i6.informatik.rwth-aachen.de/~keysers`

Thank you for your attention.

`http://www-i6.informatik.rwth-aachen.de/~keysers`

Matching Constraints

model (order)	restrictions on $(x_{11}^{IJ}, y_{11}^{IJ})$
2DW (2)	$x_{1j}=1, x_{Ij}=X, y_{i1}=1, y_{iJ}=Y,$ $x_{i+1,j}-x_{ij}\in\{0,1,2\}, x_{i,j+1}-x_{ij}\in\{-1,0,1\},$ $y_{i,j+1}-y_{ij}\in\{0,1,2\}, y_{i+1,j}-y_{ij}\in\{-1,0,1\}$
P2DHMM (1)	$x_{1j}=1, x_{Ij}=X, y_{i1}=1, y_{iJ}=Y,$ $\exists\{\hat{x}_1,\ldots,\hat{x}_I\}: \hat{x}_{i+1}-\hat{x}_i\in\{0,1,2\},$ $x_{ij}-\hat{x}_i=0,\ y_{i,j+1}-y_{ij}\in\{0,1,2\}$
P2DHMDM (1-)	$x_{1j}=1, x_{Ij}=X, y_{i1}=1, y_{iJ}=Y,$ $\exists\{\hat{x}_1,\ldots,\hat{x}_I\}: \hat{x}_{i+1}-\hat{x}_i\in\{0,1,2\},$ $x_{ij}-\hat{x}_i\in\{-1,0,1\},\ y_{i,j+1}-y_{ij}\in\{0,1,2\}$
IDM (0)	$x_{ij}\in\{1,\ldots,X\}\cap\{i'-w,\ldots,i'+w\}, i'=[i\frac{X}{I}],$ $y_{ij}\in\{1,\ldots,Y\}\cap\{j'-w,\ldots,j'+w\}, j'=[j\frac{Y}{J}],$ with warp range w, e.g. $w=3$

IDM: image distortion model; P2DHM(D)M: pseudo 2D hidden Markov (distortion) model; 2DW: 2D warping (NP-complete)

hierarchical search in nearest neighbor:

first use Euclidean distance, then more costly distance on set of closest references