Optical Character Recognition

Who? Gleb Peregud

When? 26 maja 2011

What is OCR?

It is the mechanical or electronic translation of scanned images of handwritten, typewritten or printed text into machine-encoded text.

History

- 1929 Gustav Tauschek, Germany, first patent. Templates and photodetector
- 1950 David H. Shepard, USA, Army Forces Security Agency. Intelligent Machines Research Corporation
- 1955 Readers Digest. First commercial system
- 1965 US Postal Service. Sort mail
- 1965 UK General Post Office. Sort mail
- 1974 Ray Kurzwell. Omni-font optical OCR. For blind

Parts of the problem

- Character recognition
- Word recognition
- Grammar analysis
- Image detection
- Page layout recognition
- Recreation of original page

Character recognition

What? Classifying image samples into defined set of classes, corresponding to characters available in a given script.

Problems

- Scripts
- Fonts
- Distortions
  - skew
  - noise
  - occlusions
  - rotation
- Handwriting style

Preparations

- Scanning
- Binarization or greyscaling
- Character separation
- Deskewing
- Noise removal
- Occlusion removal
- Blurring
- Centering
- Normalization

Approaches

- ANN
  - linear classifier
  - multilayer NN
  - convolution nets
- SVM
- Decision trees
  - stumps
  - Haar features
- Other
  - Contour-based (contour matching, direction codes)
  - Feature extraction
  - Contour matching
  - Histograms

Invariants

- augment training set with modified samples transformed with invariants
- build in invariances by extracting invariant features
- add regularization term to error function
- change structure of classifier to ensure invariance

MNIST database

Name THE MNIST DATABASE of handwritten digits

Who Yann LeCun, Courant Institute, NYU
Corinna Cortes, Google Labs, New York

- 28x28 images
- centered by center of mass
- gray-scale
- 60000 training set
- 10000 test set

Neural networks 0

Who? Gleb Peregud, 2011

What?

- 2-layer NN
- 784-10
- quickprop

Result 15%

Neural networks 1

Who?	LeCun et al. 1998
What?	
■	3-layer NN
■	784-1000-10
■	no preparations
Result	4.5%

Neural networks 2

Who? Ciresan et al., arXiv 1003.0358, 2010

What?

- 7-layer NN
- 784-2500-2000-1500-1000-500-10
- trained on GPU
- additional training with elastic distortions

Result 0.35%

Convolution networks

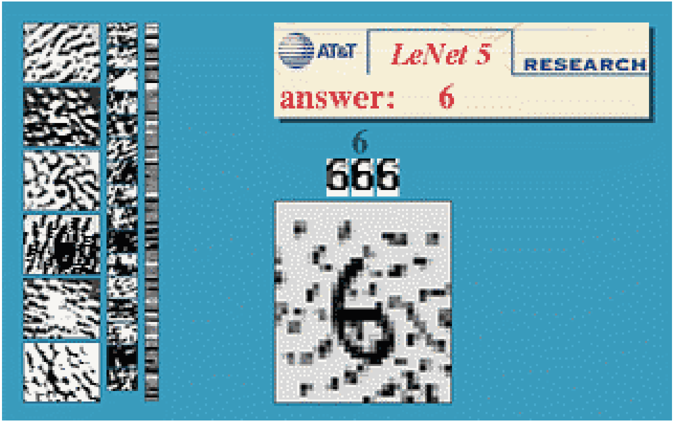
- special neural network where only pixel values around a central pixels go into the network's hidden units (sparse network) and they share their weights
- network performs a nonlinear filtering operation
- do subsampling - the outputs of the hidden units go into a smaller set of hidden units
- optional - repeat until a normal output layer
- contain less parameters than full networks, and are therefore easier to train
- their structure is modeled on the human visual system

Convolution networks - example

Who? LeCun et al. 1998

What?

- convolutional network LeNet-5
- trained with distortions
- works with little-to-no image preprocessing
- works directly on RGB images
- can recognize patterns with extreme variability


AT&T
LeNet 5
RESEARCH
answer: 6
6
666

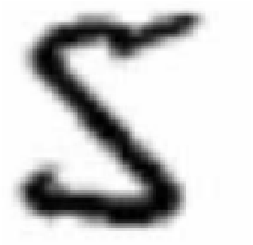
K-NN, contour matching using shape contexts

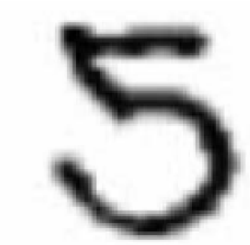
Who? Belongie et al. IEEE PAMI 2002

What?

- contour detection
- shape context calculation
- contour point matching
- iterative transformations (TPS, affine)
- selecting of prototypes
- K-NN for classification

Result 0.63%



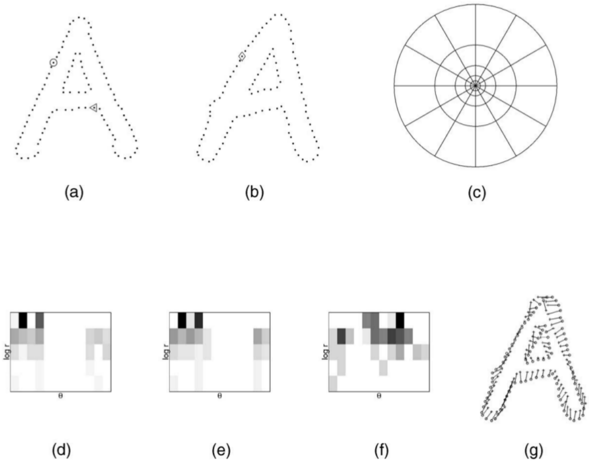


Shape context

What? Histogram in log-polar space

- counts number of points in each bin relative to a selected point
- compact and rich local descriptor
- similar points on a contour have similar shape contexts
- similarity in terms of χ^2

Shape context



(a) (b) (c)

(d) (e) (f) (g)

Shape context matching cost

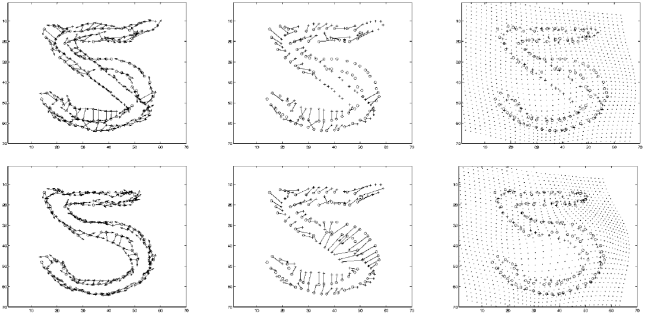
$$C_{ij} = (1 - \beta) C_{ij}^{sc} + \beta C_{ij}^{tan}$$

$$C_{ij}^{tan} = 0.5(1 - cos(\theta_i - \theta_j))$$

Matching points on contours

- Hungarian algorithm
- minimizes
$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})$$
- $O(n^3)$

Shape context, transformations



Shapes similarity distance

- $$D = 1.6D_{ac} + D_{sc} + 0.3D_{be}$$
- D_{ac} is defined as the sum of squared brightness differences in Gaussian windows around corresponding image points
- D_{sc} is defined as as the symmetric sum of shape context matching costs over best matching points
- D_{be} is a bending energy of TPS transformation

Classifier

- select k prototypes for each digit
- use PAM / k-medoids algorithm
- run k-nearest neighbours algorithm for input on prototypes

Thank you!