AN EFFICIENT SCHEME FOR INVARIANT OPTICAL CHARACTER RECOGNITION USING TRIPLE CORRELATIONS

Yannis Avrithis, Anastasios Delopoulos and Stefanos Kollias

Computer Science Division,
Electrical and Computer Engineering Dept.,
National Technical University of Athens,
Athens, GR-15773, Greece
e-mail: `iavr@theseas.ntua.gr`

Optical Character Recognition

- *Analysis* of the document into its constituents (photographs, figures and text).
- *Segmentation* of the text into columns, paragraphs, lines, words and characters.
- *Recognition* of the segmented characters.
- *Ambiguity resolution.*

Also:

- *Preprocessing:* gray scale normalization, noise elimination.
- *Postprocessing:* spelling verification or correction, customized lexicons.
- *Interaction* with human operators.

Recognition of segmented characters

- **Feature extraction**
  Input image $\longrightarrow$ Vector of fixed length, containing all the "essential" characteristics.

- **Classification**
  Class discrimination by generalizing from a training set.

Goals:

1. Derivation of a feature vector which is:
   - Invariant to shift, rotation and scale [SRS] transformations.
   - Insensitive to noise and small distortions.
2. Fast implementation in the case of binary images.

Solution:

- Calculation of the 3^{rd} order correlation of the pattern.
- Appropriate clustering & transformation of the triple-correlation domain.

Definition & properties of 3^{rd} order correlations

$x(\mathbf{t})$: real 2-D signal with support

$$S = \{0, \ldots, N-1\} \times \{0, \ldots, N-1\}$$

$$x_3(\tau_1, \tau_2) \triangleq \frac{1}{N^2} \sum_{S} x(\mathbf{t})\, x(\mathbf{t}+\tau_1)\, x(\mathbf{t}+\tau_2)\ , \quad \tau_1, \tau_2 \in S'$$

$$S' = \{-N+1, \ldots, N-1\} \times \{-N+1, \ldots, N-1\}$$

- *Symmetries*
- *Insensitivity* to noise.
- *One-to-one correspondence* with the original signal.
- *Translation, rotation & scaling:*

  $y(\mathbf{t}) = x(\mathbf{T}_{\alpha,\theta}\, \mathbf{t} + \mathbf{t}_o) \Longrightarrow$

  $y_3(\tau_1, \tau_2) = x_3(\mathbf{T}_{\alpha,\theta}\, \tau_1, \mathbf{T}_{\alpha,\theta}\, \tau_2)$

  where $\mathbf{T}_{\alpha,\theta}$: scaling & rotation matrix

  $\mathbf{t}_o$: shifting vector

The invariant representation

1. *Definition of classes:*

   $C(\tau_1, \tau_2)$: set of all lags which form triangles similar to the one defined by τ_1, τ_2

2. *Arrangement* between members of each class in a log-polar grid:

   $\tilde{x}_3(\rho, \phi; \tau_1, \tau_2) \stackrel{\triangle}{=} x_3(\mathbf{T}_{\beta,\phi}\,\tau_1, \mathbf{T}_{\beta,\phi}\,\tau_2)$

   where $\mathbf{T}_{\beta,\phi} = \beta \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$ and $\rho = \log\beta$

   Then:

   $y(\mathbf{t}) = x(\mathbf{T}_{\alpha,\theta}\,\mathbf{t} + \mathbf{t}_o) \Longleftrightarrow$

   $\tilde{y}_3(\rho, \phi; \tau_1, \tau_2) = \tilde{x}_3(\rho \log\alpha, \phi + \theta; \tau_1, \tau_2)$

3. Calculation of the 2-D Fourier transform $\tilde{X}_3(P, \Phi; \tau_1, \tau_2)$ of $\tilde{x}_3(\rho, \phi; \tau_1, \tau_2)$ w.r.t. ρ and ϕ $\Rightarrow$ shift invariant w.r.t. ρ and ϕ.

4. Appropriate amplitude & phase information of the transform $\rightarrow$ representation F_x.

Properties of F_x:

- Unique correspondence with class of images that are mutually related with SRS transformation.
- Noise insensitivity
- Stand-alone representation $\Rightarrow$ direct input to classifier.

Feature size reduction

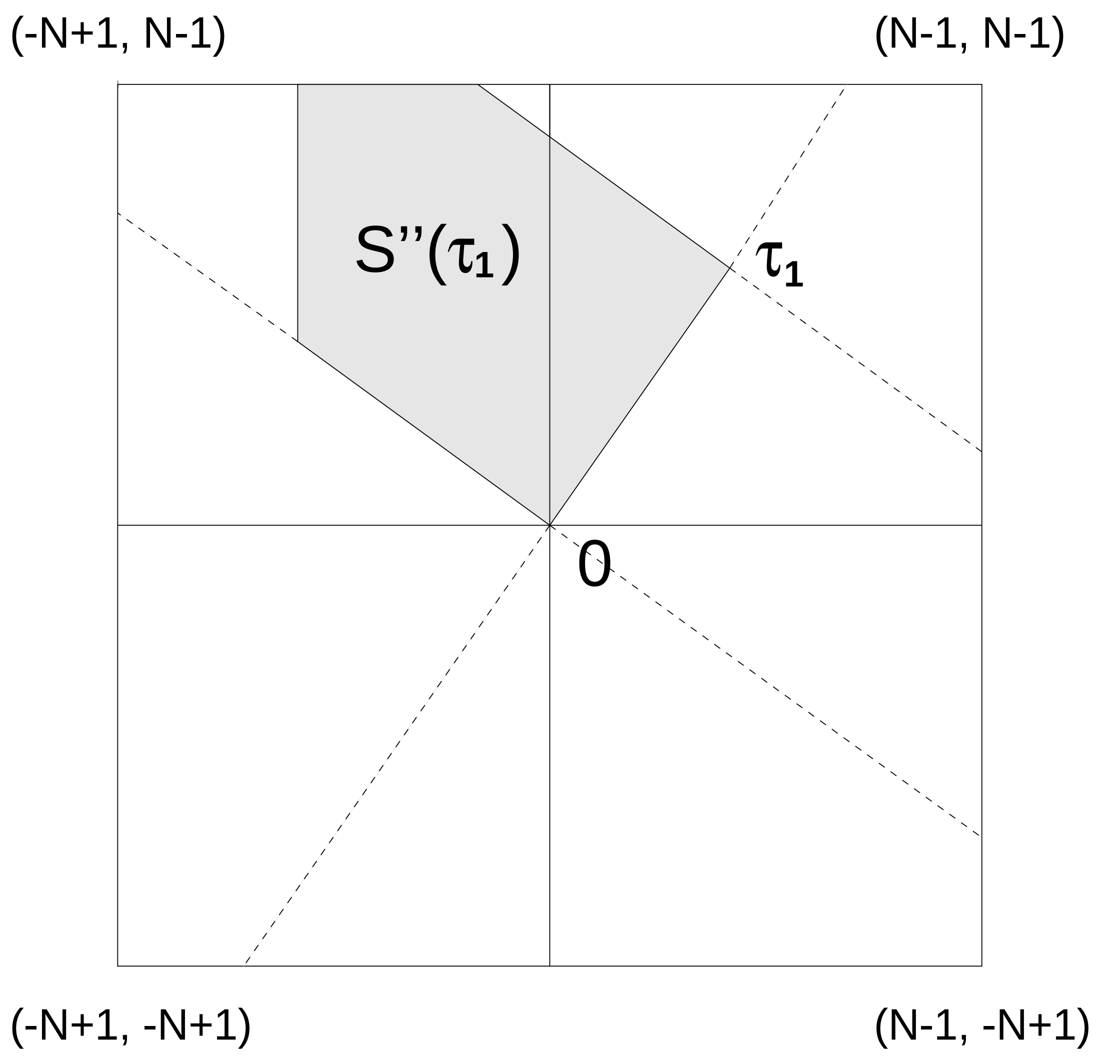
1. *Reduced representation:*
   - amplitude information only, or
   - zero-frequency Fourier coefficient only.

2. *Reduced redundancy:*
   For each $\tau_1 \in S' \Rightarrow \tau_2 \in S''(\tau_1)$

$$S''(\tau_1) = \{\tau_2 \in S' : 0 \leq \tau_1 \cdot \tau_2 \leq |\tau_1|^2, \; 0 \leq \tau_1 \cdot \tau_2{}', \; (\tau_1 - \tau_2) \in S'\}$$

3. *Class parametrization:*
   $(\tau_1, \tau_2) \rightarrow (\theta_1, \theta_2) \qquad \theta_1, \theta_2 \in [0, \frac{\pi}{2}]$
   4-D space $\rightarrow$ 2-D space

The region $S''(\tau_1)$



Discrete implementation

- Binary (black & white) images: $x(\mathbf{t}) \in \{0, 1\}$
  Then:
  - Image stored in binary integers
  - Real multiplication $\Longleftrightarrow$ Boolean "AND".
- Discrete 2-D grid of $(\theta_1,\ \theta_2) \Rightarrow$
  finite number of distinct classes $C(\theta_1, \theta_2)$.
- Discrete 2-D grid of ρ and ϕ $\Rightarrow$
  discretization in the interior of each class.
- $\tilde{x}_3(\rho, \phi; \theta_1, \theta_2) \to \tilde{X}_3(P, \Phi; \theta_1, \theta_2)$
  via FFT algorithm.
  Parrallel implementation: possible.
- $|\tilde{X}_3(P, \Phi; \theta_1, \theta_2)|$: Final 4-D invariant representation.

Algorithm

Step 1: Calculate all the horizontal shifts of the input image.

Step 2: For each $\tau_1 \in S'$:

1. Find & store all possible products $x(\mathbf{t})\,x(\mathbf{t}+\tau_1)$, using the already calculated horizontal shifts and the logical operator AND.
2. Calculate & quantize ρ, ϕ by comparing τ_1 with $\tau_o = (1, 0)$.
3. For each $\tau_2 \in S''(\tau_1)$:
   (a) Find $x_3(\tau_1, \tau_2)$, using the products $x(\mathbf{t})\,x(\mathbf{t}+\tau_1)$.
   (b) If $x_3(\tau_1, \tau_2) = 0$, proceed to the next τ_2.
   (c) Calculate & quantize θ_1, θ_2.
   (d) $\tilde{x}_3(\rho, \phi; \theta_1, \theta_2) := \tilde{x}_3(\rho, \phi; \theta_1, \theta_2) + x_3(\tau_1, \tau_2)$

Step 3: Find $\tilde{X}_3(P, \Phi; \theta_1, \theta_2)$; keep the amplitude of the transform only.

Simulation Results

Sample input images

Invariant representations

Invariant representations of scaled and rotated versions of the letter 'N'.

Euclidean distances between invariant representations