Extracting Features for OCR

Feature Extraction

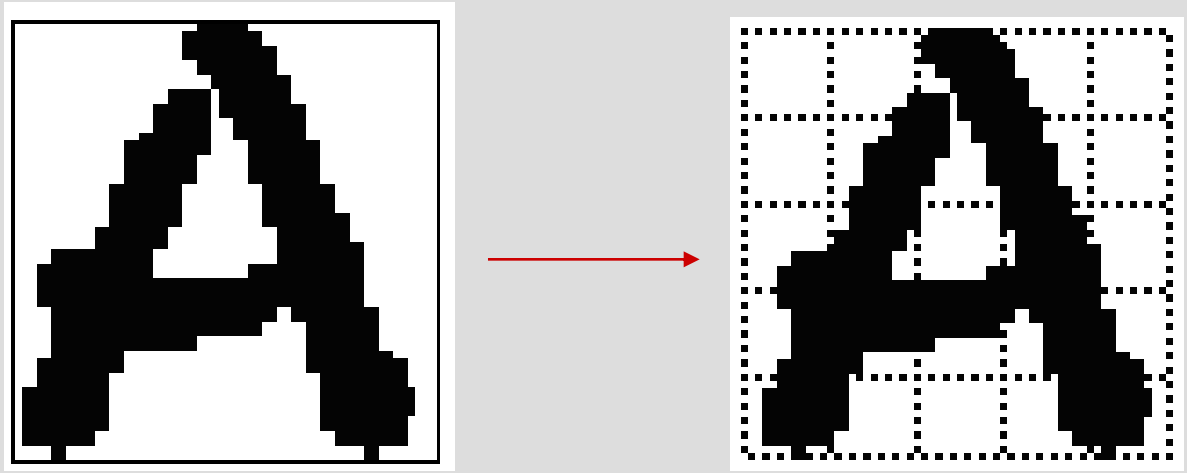
- The major statistical features used for character representation are:
  - Centre Of Mass
  - Zoning
  - Projections and profiles
  - Crossings and distances

Centre Of Mass

- The centre of mass of the character

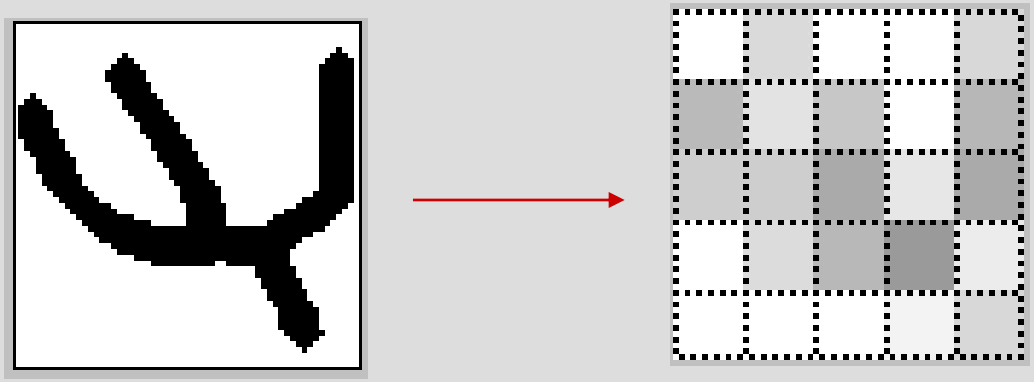
Zoning

- The character image is divided into 4x4 zones. From each zone features are extracted to form the feature vector. The goal of zoning is to obtain the local characteristics instead of global characteristics



Zoning – Density Features

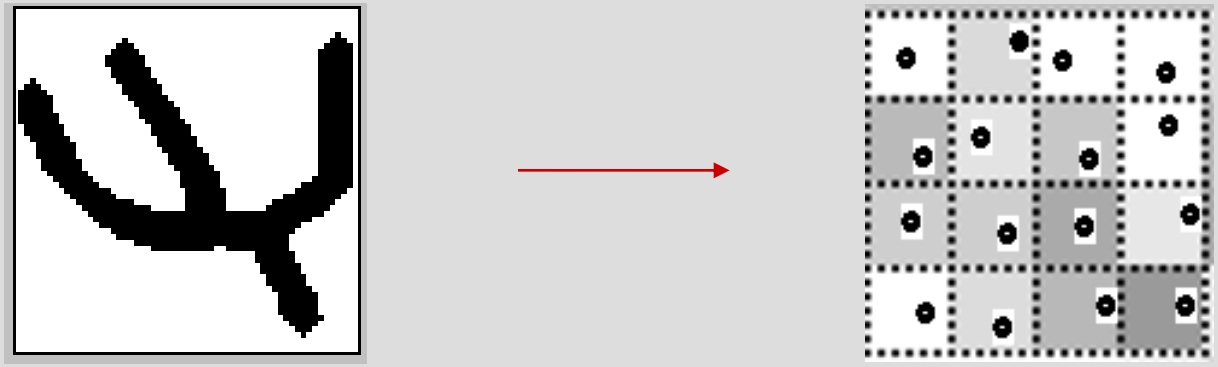
- The number of foreground pixels, or the normalized number of foreground pixels, in each cell is considered a feature.



Darker squares indicate higher density of zone pixels.

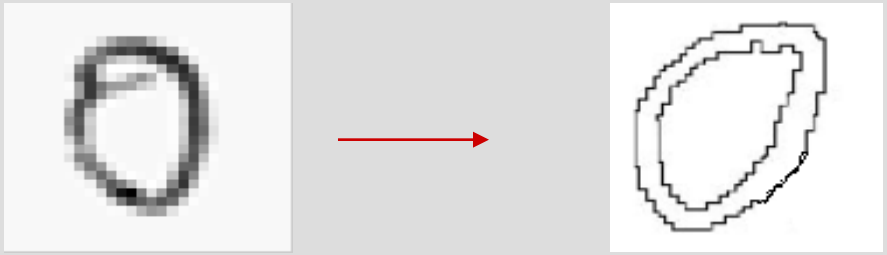
Zoning – Centre Of Mass

- The mean of all positive points in a zone.

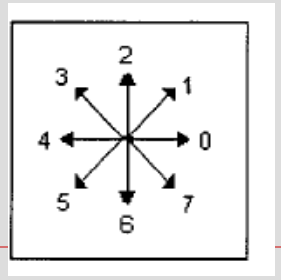


Zoning – Direction Features

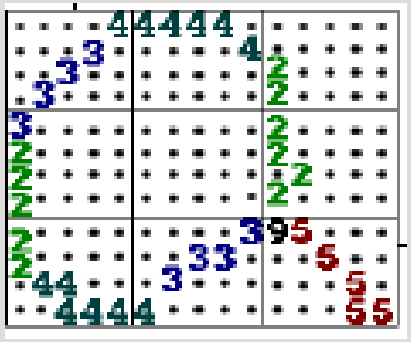
- Based on the contour of the character image



- For each zone the contour is followed and a directional histogram is obtained by analyzing the adjacent pixels in a 3x3 neighborhood



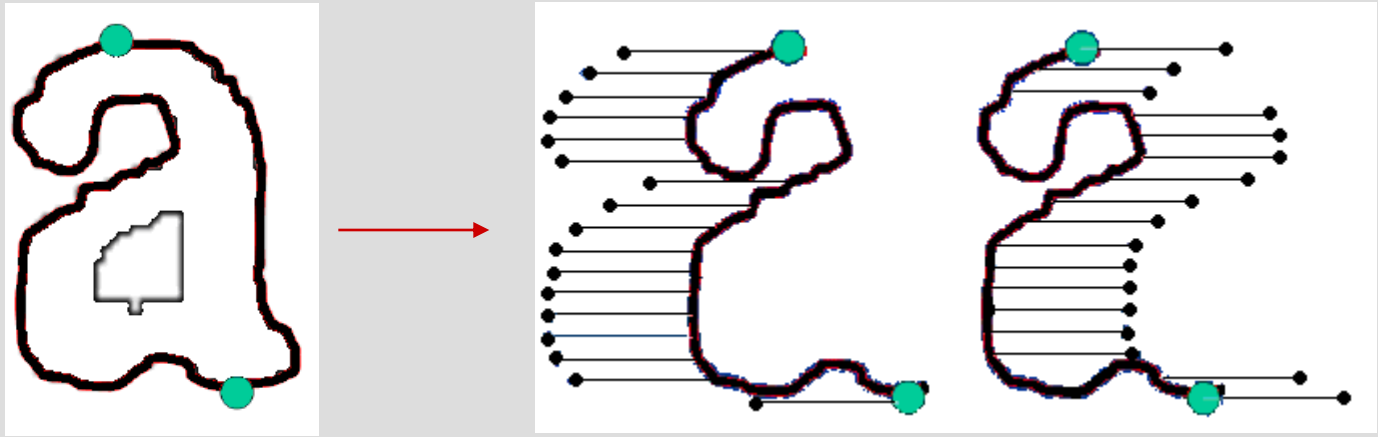
Zoning – Direction Features



- Line segments are coded with a direction number
  - 2 = vertical line segment
  - 3 = right diagonal line segment
  - 4 = horizontal line segment
  - 5 = left diagonal line segment

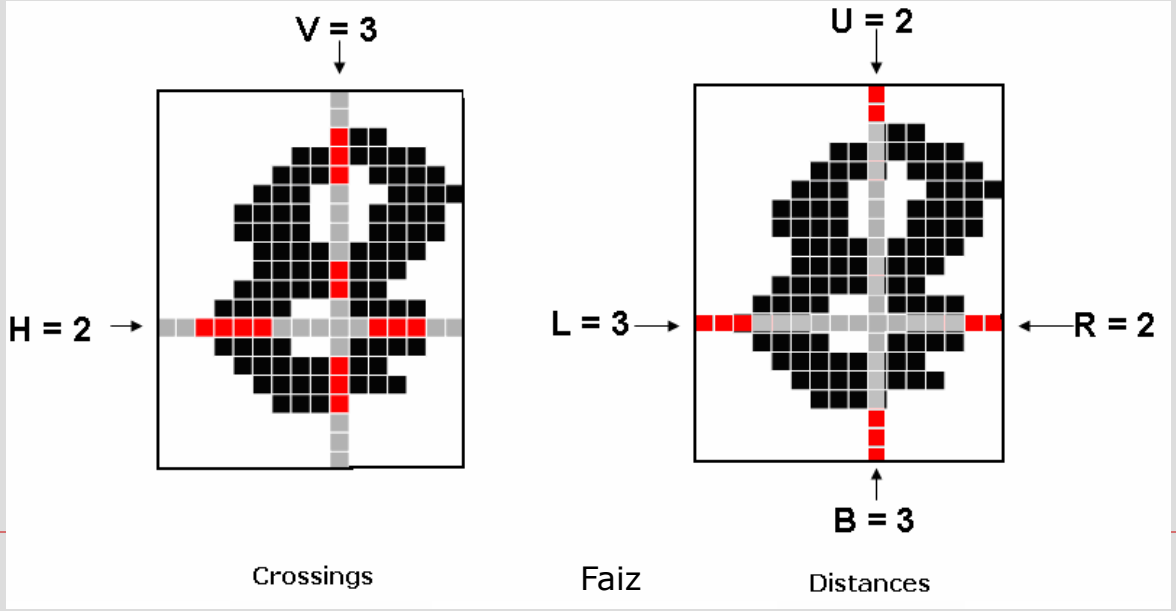
Profiles(Horizontal/Vertical)

- Profiles can also be used to the contour of the character image
  - for each row, find the character starting pixel and then the ending pixel.



Crossings and Distances

- **Crossings** count the number of transitions from background to foreground pixels along vertical and horizontal lines through the character image and **Distances** calculate the distances of the first image pixel detected from the upper and lower boundaries, of the image, along vertical lines and from the left and right boundaries along horizontal lines



Experiments and Results

Number Of Hidden Neurons	All Features	Features excluding Direction Zoning
10	66.6%	28.7%
15	86.3%	88.1%
20	87.2%	90.3%
25	87.0%	90.8%
30	86.5%	91.0%
35	85.6%	90.2%