Real Time Scene Text Localization and Recognition [Neumann, Matas (CVPR 2012)]

This paper addresses the problem of locating and recognizing text from natural images in real time.

- Compute “extremal regions” (ER) : a region whose boundary pixels have strivtly higher than the region itself.
- A two stage sequential “Sequential Classifier” is employed which selects regions by estimating class-conditional probability p(r | character)
- In the first stage, Real Adaboost classifier is used, with the following features:
  - aspect ratio
  - compactness
  - number of holes
  - horizontal crossings
- The feature descriptors chosen for the first stage can be computed “incrementally”
- Inclusion relation of ER s at all thresholds is used to track the value of p(r | character)
- While selecting ER s local maxima of probability is considered
- In the second stage, SVM with RBF kernel is used classifier, with the following features:
  - Hole area ratio
  - Convex hull ratio
  - The number of outer boundary inflexion points
- The second stage features are computationally expensive
- The second stage features are scale invariant but not rotationally invariant. Hence, character rotations are used in the training set.
- The detector is incorporated into a system with exhaustive search with pruning.

- Experiments

  Obtained results comparable to state of the art on:

  - ICDAR 2011 Dataset
  - Street View Text Dataset