

Word Spotter

Praveen Balireddy (2018201052)

Aman Joshi (2018201097)

Abhijeet Panda (2018201044)

Objective

Find all instances of a given word in a potentially large dataset of document images.

Queries:

- 1) Query by example(Image)
- 2) Query by string(Text)

Challenges with previous works:

- 1) Out of Vocabulary words(words not there in training data, but exist in test)
- 2) Time taken for the search
- 3) Same word, different handwritings

Current approach:

- 1) Instead of learning models for particular keywords, learning what makes words and letters unique independently of their writers' style.
- 2) Using an attribute based representation for each word.

Fisher Vector

- 1) A Gaussian Mixture Model (GMM) is used to model the distribution of features (e.g. SIFT) extracted all over the image
- 2) The Fisher Vector (FV) encodes the gradients of the log-likelihood of the features under the GMM, with respect to the GMM parameters.

PHOC(pyramidal histogram of characters)

- 1) This binary histogram encodes whether a particular character appears in the represented word or not.
- 2) Spatial pyramid representation ensures that the information of the characters order is preserved

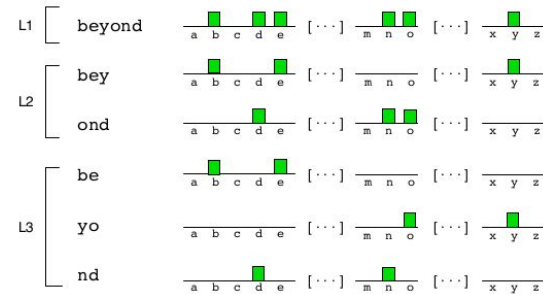


Figure 1. PHOC histogram at levels 1, 2, and 3. The final PHOC histogram is the concatenation of these partial histograms.

Algorithm:

- 1) Get the SIFT features of the image and find their Fisher vectors
- 2) Predict/train the PHOC attributes using a classifier(SVM), given the FV
- 3) Since we have the image and string for a word, actual PHOC attributes can be found by using the string input
- 4) Using CCA (Canonical Correlation Analysis) tool, get the projections of the predicted scores and the ground truth values
- 5) Use distance/cosine similarity to compute the mean average precision

Implementation parameters

- 1) No of Gaussians for training GMM = 16
- 2) Used PCA to reduce dimensions of each SIFT vector from 128 to 64

Work So far:

- 1) Able to get FV representation from SIFT vectors
- 2) Trained a GMM with 16 gaussians using ~35k examples on the FVs
- 3) Did a benchmark test(QBE), by getting the cosine similarity scores between the test data FVs and the training dataset FVs
- 4) MAP score on seen test dataset(~10k examples): 0.20044
- 5) MAP score on unseen test dataset(~10k examples): 0.2418

Thank You!

—