Word Image Retrieval Using Bag of Visual Words

Ravi Shekhar and C.V. Jawahar

CVIT, IIIT-Hyderabad, INDIA

Motivation

Why Recognition free Retrieval?
- Robust OCRs are unavailable for many non-latin languages.
- These languages have rich heritage and there is a need for content level search.
- Word Spotting based methods are too slow for real time system.
- Most of the existing retrieval methods are memory intensive.
- Scalability is an immediate challenge.

Why Bag of Visual Words?
- Bag of Words (BoW) representation is the most popular representation for text retrieval.
- Bag of Visual Words (BoVW) performs excellently for image and video retrieval.
- Ignores the spatial relationships between visual words
- Represent word image as histogram of visual words
- Handles noisy document images: Demonstrated on dataset for which commercial OCRs fail

Bag of Visual Words

- Represents word image as histogram of visual words
- Ignores the spatial relationships between visual words

BoW based system is flexible, powerful and scalable to Billions of images.

Retrieval Results

Why Bag of Visual Words?
- Bag of Words (BoW) representation is the most popular representation for text retrieval.
- Bag of Visual Words (BoVW) performs excellently for image and video retrieval.
- Ignores the spatial relationships between visual words
- Based on number of SIFT matches
- Higher the Total Score, better the match

Total Score(I, j) = \frac{\sum_{i=1}^{3} \text{Score}(I_i, I_j)}{3}

where, \text{Score}(I_i, I_j): Score for `i` part of the image

System Overview

Indexing and Retrieval

Key Point Detection

Key Point Description

Feature Extraction

Indexing

Retrieved Images

Spatial Verification

• To provide spatial order/structure of characters in word

Sample Outputs

Query Image

Retrieved Images

Code Book generation

• Using Hierarchical K-Means (HKM)
• HKM is faster compare to K-Means

Implementation Details

Keypoint Detection:
- Harris Corner Detection – Invariant to rotation, scale and image noise

Keypoint Description:
- Scale Invariant Feature Transform (SIFT) on third scale and zero degree

Indexing:
- Inverted File Index using Lucene – Popular, reliable and open source search engine

Future Work

Learning document-specific local descriptors
Use of noisy OCR outputs along with BoVW
Improve/remove the re-ranking methodology
Mining character specific pattern to support text query