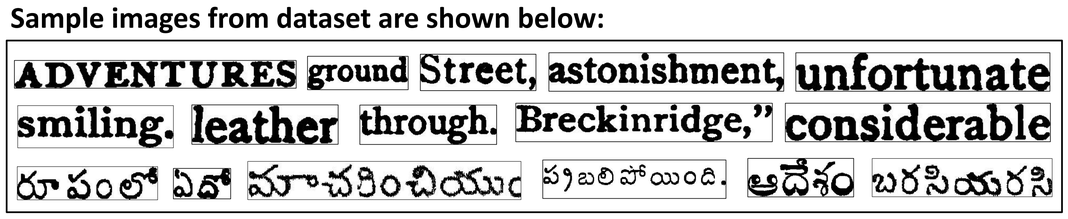
Document Specific Sparse Coding for Word Retrieval

Ravi Shekhar and C.V. Jawahar

CVIT, IIIT-Hyderabad, INDIA

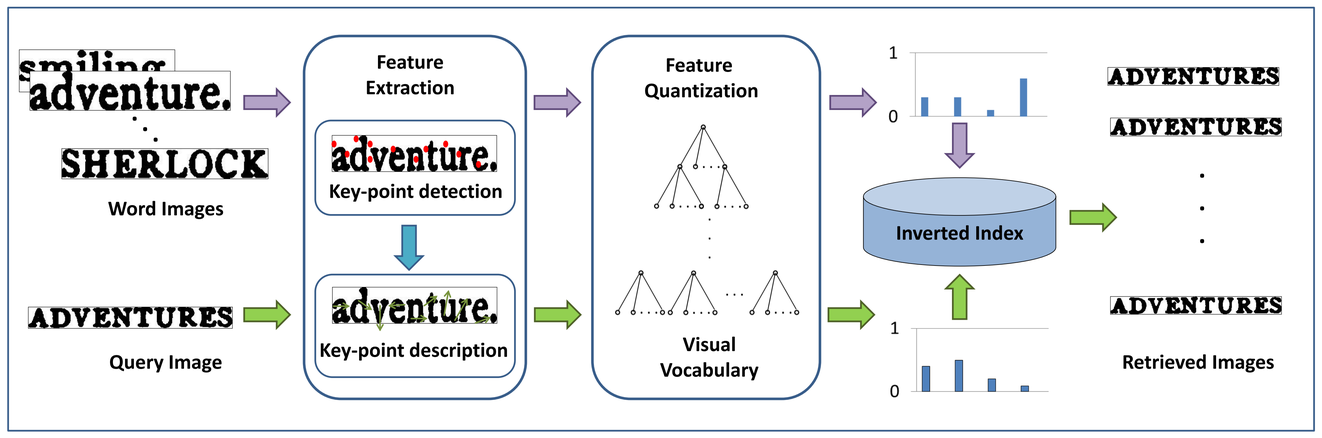
Introduction

- Robust OCRs are unavailable for many non-Latin languages
- Recognition free approach based on Bag of Words outperforms OCR based methods
- Performance is further improved by defining document specific sparse coding
- Coding is provided by using locality constraint at word level
- Next level coding is provided at character level representation
- Text query support is provided using pre-learnt character representation
- Sample images from dataset are shown below:



Bag of Words

- Represents word image as a histogram of visual words
- Histogram is indexed using inverted file index

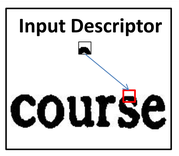


Document Specific Sparse Coding

Coding and Pooling in BoW

- Generates code from raw descriptor
- Single descriptor is assigned to singe visual word
- Has codeword uncertainty and codeword plausibilty problems

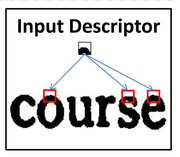
$$\arg\min_{c} \sum_{i=1}^{N} \| x_i - Bc_i \|^2 \; s.t. \| c_0 \|_{l^0} = 1, \| c_i \|_{l^1} = 1 \; \forall i \quad *$$



Sparse Coding

- Single descriptor is assigned to one or more visual words
- Has less quantization error but slow in computation

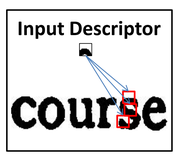
$$\arg\min_{c} \sum_{i=1}^{N} \| x_i - Bc_i \|^2 + \lambda \| c_i \|_{l^1} \quad *$$



Sparsity due to Locality in Coding

- Locality constrained linear coding (LLC) learns representation using nearest codewords (Wang et al. CVPR 2010)
- Locality of character is more important than sparsity
- Similar patches will have similar codewords

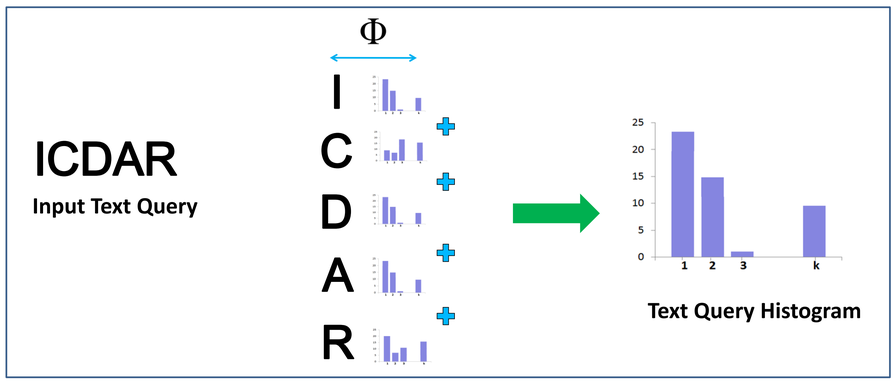
$$\arg\min_{c} \sum_{i=1}^{N} \| x_i - Bc_i \|^2 + \lambda \| d_i \odot c_i \|_{l^1} \; s.t. 1^T i = 1, \forall i \quad *$$



∗ *where* $\boldsymbol{x_i}$*:* i^{th} *descriptor of image X=[x1,...,x2],* $\boldsymbol{c_i}$*:* i^{th} *code for image X from C=[c1,...,c2],* $\boldsymbol{N}$*: #data points,* $\boldsymbol{B}$*: codebook,* $\boldsymbol{l^0}$*:L0 norm,* $\boldsymbol{l^1}$*:L1 norm,* $\|.\|$*:norm,* $\boldsymbol{\lambda}$*: regularization parameter,* $\odot$*:element-wise multiplication,* $\boldsymbol{d_i}$*:* i^{th} *locality adapter .*

Text Query Support

- Originally, retrieval is performed by example
- Character specific representation is learned offline
- Word representation is calculated using learned representation



Retrieved Results

Data set Details

Book	# Pages	# Words
English-1601	363	113008
Telugu-1718	100	21345
Telugu-1716	102	4121

mAP of Text Query Support

Language	mAP
English	0.87
Telugu	0.90

Retrieval time ~ 850msec

mAP of Different Methods

Book	BoW	BoW + SIFT Re-ranking	BoW + LCS Re-ranking	Doc. Coding	Doc. Coding + LCS Re-ranking
English-1601	0.8015	0.8645	0.92	0.8765	**0.9451**
Telugu-1718	0.7834	0.8861	0.918	0.92	**0.96**
Telugu-1716	0.8173	0.8531	0.9036	0.91	**0.95**

Sample Outputs

Query Image	Retrieved Images				
ADVENTURES	ADVENTURES	ADVENTURES	ADVENTURES	ADVENTURE
beggar	beggar	beggar.	beggar,	begging
Holmes	Holmes	Holmes,	Holmes.	Holmes,"
కలిగి	కలిగి	కలిగి	కలిగి	కలిగి
భూస్వామ్య	భూస్వామ్య	భూస్వామ్య	"భూస్వామ్య	భూస్వాములు,

Implementation Details

- Retrieval is performed based on similarity and re-ranking score
- Similarity score is given by

$$Sim\_Score(Q, I) = \frac{\sum_{i \in Q \cap I} w_i}{\sum_{j \in I} w_j} \quad where \;\; w_i = \frac{1}{log(f_i + 1)} \quad ¥$$

- Re-ranking is performed based on Longest common sub-sequence of visual words and given by

$$LCS\_Score(Q, I) = \frac{\sum_{i \in LCS(Q,I)} w_i}{\sum_{j \in Q} w_j} \quad ¥$$

- Final score is given by

$$Score(Q, I) = [\gamma \times Sim\_Score(Q, I)] + [(1 - \gamma) \times LCS\_Score(Q, I)] \quad ¥$$

¥ *where* $\boldsymbol{Q}$*: visual words corresponding to query image,* $\boldsymbol{I}$*: visual words corresponding to candidate image,* $\boldsymbol{w_i}$*: weight of* i^{th} *visual word,* $\boldsymbol{f_i}$*: frequency of* i^{th} *visual word,* $\boldsymbol{\Upsilon}$*: weighting parameter*

Achievements

- An efficient document specific coding technique is proposed
- Text query support is provided using pre-learned representation
- Experimentation is performed on English and Telugu scripts to achieve state of the art performance

Future Work

- Selecting codebook specific to particular documents
- Designing document specific descriptor
- Learning character representation by using automatic mining of codebook

Center for Visual Information Technology
International Institute of Information Technology, Hyderabad, India
jawahar@iiit.ac.in
http://cvit.iiit.ac.in/