An Efficient Approach for Patch-based Visual Object Classification

Veerapathirapillai Vinoharan and Amirthalingam Ramanan

Department of Computer Science, Faculty of Science, University of Jaffna

{vvinoharan,a.ramanan}@univ.jfn.ac.lk

Introduction

In this work, we propose a two stage approach to optimize the process of building visual codebooks with discriminative power and compactness in the classification of visual objects.

Contribution

- A one-pass feature selection which is followed by an entropy-based feature selection approach is proposed to filter out ambiguous descriptors from initially extracted large descriptors set.
- Statistical-based measures and Visual-bit representation of codewords is proposed to select informative codewords from an initially constructed large codebook.

Methodology

- Unambiguous descriptors are selected from initially extracted SIFT descriptors using a one-pass feature selection (OPFS) method which is then followed by an entropy-based feature selection (EBFS) method to increase the discriminative power of the codebook.
- A codebook is then constructed by means of K-means approach.
- Indistinctive codewords are eliminated based on statistical measures (inter, intra, and combined category confidence) [1] or visual-bit representation of codeword to obtain a compact codebook [2].
- A histogram representation is created for each descriptor set of images and linear SVM classification algorithm is applied to those fixed-length feature vectors.

The overall framework of the proposed method is illustrated in Figure 1.

Methodology...

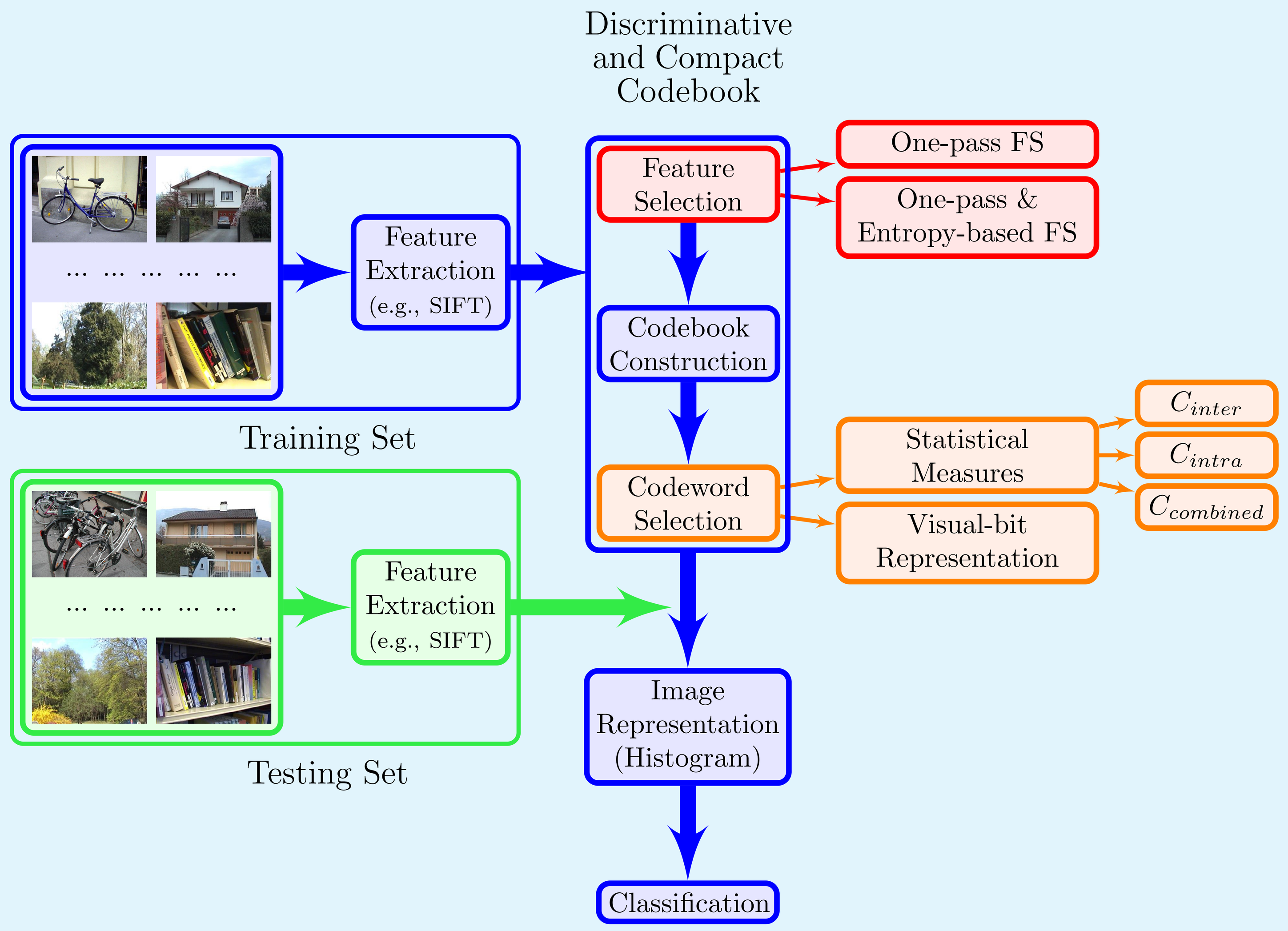

Figure 1: Overview of the proposed framework

One Pass Feature Selection (OPFS)

⋆ **Input:** *trainingFeatures*
⋆ **Output:** *selectedFeatures*
⋆ $r \leftarrow$ radius of the hypersphere,
⋆ $selectedFeatures \leftarrow trainingFeatures\{1\}$,
⋆ **for all** $feature \in trainingFeatures$ **do**
  if $min\| feature - trainingFeatures \|^2 > r^2$ **then**
    Create a new hypersphere of r such that,
    $selectedFeatures \leftarrow \{selectedFeatures \cup feature\}$
  end if
end for

Entropy-based feature selection (EBFS)

$$E(F) = -\sum_{i=0}^{255} p_i(F) log_2 p_i(F) \quad (1)$$

where, $p_i(F)$ - $\frac{|\{k|f_k=i\}|}{128}$, descriptors are treated as 128 samples of discrete random variable k in {0, 1, 2, $\cdots$, 255}.

Statistical Measures

$$C_{inter,i} = \sum_{j=1}^{N} max\left(\frac{f_{ij}}{n_i} - \frac{1}{m_i}, 0\right) \quad (2)$$

$$C_{intra,i} = \frac{1}{\sum_{j=1}^{N} var(h_{ij})} \quad (3)$$

$$C_{combined,i} = \alpha C_{inter,i} + \beta C_{intra,i};\ 0 \leq \alpha, \beta \leq 1 \quad (4)$$

where,
f_{ij} - number of training features in the i^{th} codeword & j^{th} category.
h_{ij} - i^{th} codeword value of each image belonging to the j^{th} category in the BoF histogram domain, i = 1, 2, ..., K, & j = 1, 2, ..., N.
n_i - is the total number of features in the i^{th} codeword.
m_i - is the number of object categories in the i^{th} codeword.
N - is the number of object categories in classification.
K - is the size of the codebook.

Visual-bit representation of codewords

$$h_i = \begin{cases} 1 : \text{if } C_i \geq t_0 \\ 0 : \text{otherwise} \end{cases} \forall i = 1, \ldots, K \quad (5)$$

$$t_1 = \frac{\lambda p_0 + p_1}{\lambda + 1} \quad (6)$$

$$Compact_{CB} = \begin{cases} \text{eleminate } C_i & : \text{if } C_i \geq t_0 \\ \text{retain } C_i & : \text{Otherwise} \end{cases} \quad (7)$$

where,
p_0 is $min_{1\leq i \leq K}(SB_i)$, p_1 is $max_{1\leq i\leq K}(SB_i)$
SB_i - sum of visual bits associated with the i^{th} codeword.
λ - weighting parameter for a rare informative word.
t_i - level of significant activation of a codeword in a codebook.

Test Results

Comparison of average precision (AP) with number of training features and codebook size: Traditional BoF approach and proposed feature selection method *with* and *without* codeword selection (CS)

Approach	Dataset	#Descriptors	Without CS		Statistical Measures with CS						Visual bit with CS	
					inter		intra		combined			
			CB	AP	CB	AP	CB	AP	CB	AP	CB	AP
Traditional	Xerox7	4,046,578	987	84.21	803	83.68	740	87.89	902	82.41	286	83.85
OPFS		212,294	500	94.11	400	93.31	375	94.69	409	93.72	191	93.42
OPFS+EBFS		172,006	500	94.04	400	93.40	**375**	**94.79**	406	93.41	201	94.13
Traditional	UIUCTex	4,543,590	1032	82.73	835	81.94	774	86.40	842	81.53	387	90.25
OPFS		314,724	500	93.73	400	94.56	**375**	**95.51**	401	94.27	264	92.45
OPFS+EBFS		157,094	500	94.17	400	92.95	374	94.08	404	92.88	257	93.48
Traditional	PASCAL VOC 2007	1,760,400	1049	71.78	847	72.41	787	73.71	953	71.99	421	71.69
OPFS		245,327	500	72.93	400	73.16	375	73.47	**405**	**73.91**	262	72.88
OPFS+EBFS		181,248	500	72.58	400	72.90	375	73.64	414	72.71	252	72.20
Traditional	Caltech101	5,659,137	925	84.72	742	82.87	694	84.80	850	82.30	336	84.32
OPFS		393,024	500	86.01	400	85.17	375	85.97	408	85.83	289	85.48
OPFS+EBFS		286,925	500	86.02	400	85.36	**375**	**86.34**	407	85.50	249	85.35

Experimental setup

- For the image sets: Xerox7, UIUCTex, and Caltech101 we used 70% for training and 30% for testing from each class.
- For PASCAL VOC 2007, the training was performed on the provided 'trainval' set and evaluated on the testing set.
- We used SIFT descriptors in extracting the features from those image sets.
- The codebook is constructed by using the K-means algorithm with K = 500 for all datasets.
- The OVA-SVMs with RBF kernel was used for classification and the reported classification rates are of average precision (AP) [3].

Discussion and Conclusion

- The proposed ideas in this paper are to generate a compact and discriminative codebook, that can be obtained by selecting representative keypoints and eliminating indistinctive codewords.
- These processes not only reduces the overall computational complexity but also maintains the BoF model to be efficient with stable performance.
- As a near future work we will incorporate another set of detector-descriptors: SURF and ORB.

References

[1] V. Vinoharan and A. Ramanan, "Keypoints and Codewords Selection for Efficient Bag-of-Features Representation", In IEEE International Conference on Future of Information and Communication Conference (FICC), pp. 203-208, 2018.

[2] T. Kirishanthy and A. Ramanan, "Creating Compact and Discriminative Visual Vocabularies Using Visual Bits", In Proceedings of the IEEE Digital Image Computing: Techniques and Applications (DICTA), pp. 258–263, 2015.

[3] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The Binormal Assumption on Precision-Recall Curves", In Proceedings of the International Conference on Pattern Recognition (ICPR), pp. 4263–4266, 2010.