Comparative Analysis of Feature Extraction Methods for Colorectal Polyp Images in Optical Projection Tomography

Wenqi Li[1]
wyli@dundee.ac.uk
Maria Coats[2]
mcoats@dundee.ac.uk
Jianguo Zhang[1]
jgzhang@computing.dundee.ac.uk
Stephen J. McKenna[1]
stephen@computing.dundee.ac.uk

[1] CVIP, School of Computing
University of Dundee
Dundee, DD1 4HN

[2] School of Medicine
Ninewells Hospital & Medical School
University of Dundee
Dundee, DD1 9SY

Abstract

Optical Projection Tomography imaging has potential to enhance diagnostic analysis of colorectal polyps. In this paper, the problem of feature extraction for automated classification of optical projection tomography images of colorectal polyp is addressed. 3D patches are classified using the *bag of visual words* framework and support vector machines. We compare the utility of dimensionality reduction by random projections with two prominent techniques for 3D texture analysis: independent subspace analysis and volumetric local binary patterns. By analysing classification performance on a dataset of 59 colorectal polyp images containing annotated regions of low-grade dysplasia and invasive cancer we show that features based on random projection produce the best result (area under ROC curve: 0.87) with lower computational cost than the other methods.

1 Introduction

Colorectal cancer screening has reduced mortality and detected large numbers of adenomas and polypoid cancers. However, diagnosis using conventional 2D histopathology exhibits marked inter-observer variation [12]. Recently, optical projection tomography (OPT) has been used to image colorectal polyps in 3D [8]. This paper investigates automated classification of 3D patches in such images. Specifically, we focus on discriminating between low-grade dysplasia and invasive cancer. Figure 1 shows example OPT polyp images with regions annotated by a histopathologist. Regions of invasive cancer tend to have a more dense and homogeneous texture than low-grade dysplasia.

We investigate the use of random projections in order to obtain feature vectors of reduced dimensionality [4]. This is compared with three alternative methods for feature extraction: 3D local binary pattern descriptors [10, 11, 15] and two forms of independent subspace analysis (ISA). ISA has previously been used for classification of H&E stained histology images [7]. These methods represent contrasting approaches to low-level feature

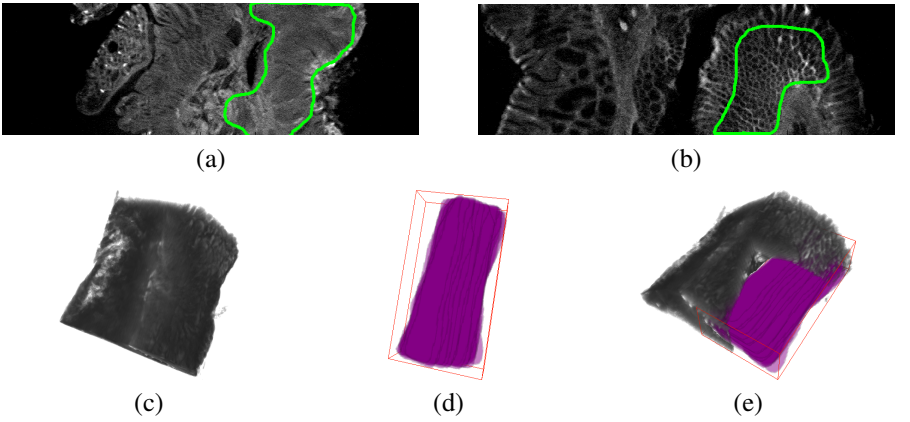

Figure 1: OPT virtual sections showing regions of (a) invasive cancer and (b) low-grade dysplasia. (c) Volume rendering of a polyp. (d) Annotated region of polyp in (c). (e) Combined volume rendering of the polyp and its annotation.

extraction, i.e., hand-crafted features (LBP) and learned domain-specific features (ISA). The popular *bag of visual words* framework [14] was used; each 3D image patch was represented by extracting local feature vectors from multiple 3D windows within the patch, quantising these feature vectors using a learned visual word dictionary, and histogramming these visual words. In the case of feature vectors obtained by random projection, we explore the effect of varying their dimensionality as well as of varying the size of the dictionary.

2 Random Projection

The intensity values of the voxels in an $n \times n \times n$-voxel subvolume (3D window) can be considered directly to be elements in a vector of dimensionality N ($N = n^3$). Random projection (RP) can provide a simple yet surprisingly effective method to reduce the dimensionality of such 'raw' features [7]. Liu and Fieguth [9] performed texture classification using such a method. Let $\mathbf{X}$ be an $N \times T$ data matrix in which the T columns are the N-dimensional window vectors. The RP method simply maps these vectors onto a D-dimensional subspace using a suitably generated $D \times N$ random projection matrix $\mathbf{R}$:

$$\hat{\mathbf{X}}_{D\times T} = \mathbf{R}_{D\times N}\mathbf{X}_{N\times T} \tag{1}$$

Each element in the projection matrix is a sample from a Gaussian distribution with zero mean and unit variance. According to the Johnson-Lindenstrauss lemma and the proofs in [1], the T data points in $\mathbb{R}^N$ are embedded into a lower dimensional Euclidean space $\mathbb{R}^D$ such that the relative distances between any two of these points are approximately preserved. The complexity of this process is only $O(DNT)$. In Liu and Fieguth's 2D image patch classification experiments they compared random features ($\hat{\mathbf{X}}_{D\times T}$) with the use of raw patches ($\mathbf{X}_{N\times T}$), and some hand-crafted feature extractors (LBP and filter banks with 38 filters). Surprisingly, the result based on random features with a simple nearest neighbor classifier matched or even surpassed the state-of-the-art methods on three commonly used texture datasets. Moreover, they observed that approximately one-third of the dimensionality of the original patch space was needed to preserve the salient information contained in the original local patch; any further increase in the number of features yielded only marginal improvements in classification performance. Random projection of texture features has been used previously in the context of tumor segmentation in 2D histopathology images [5].

3 3D Local Binary Patterns

Local binary patterns (LBPs) are popular, computationally simple texture decriptors that exhibit robustness to monotonic changes in intensity. They are computed by thresholding each 3×3-pixel neighbourhood at the value of its central pixel, considering the result as an 8-bit binary code, and histogramming these codes over a 2D image window. Ojala *et al.* [11] found that the vast majority of the binary codes in a local neighborhood are so called "uniform patterns". To achieve rotation-invariance (around the central pixel) using uniform patterns, all non-uniform LBP patterns are stored in a single bin in the histogram computation. In a 3D volumetric image, LBP descriptors can be computed based on 3×3 neighbourhoods in each of three orthogonal planes (taken to be aligned with the image axes for convenience) [15].

4 Independent Subspace Analysis

Independent Subspace Analysis (ISA) is a generalized version of Independent Component Analysis (ICA) in which components are divided into subspaces; subspaces are assumed independent, whereas components in the same subspace need not be independent of each other [4]. Features learned by ISA show phase- and translation-invariant properties. An input image $\mathbf{x}^t$ can be modelled as a linear combination of features:

$$\mathbf{x}^t = \sum_l \sum_{m \in S(l)} \mathbf{A}^t_m \mathbf{s}_m \tag{2}$$

where $S(l)$ is the set of indices m of $\mathbf{A}^t_m$ that belong to the l-th subspace. We used a model in which the non-linear filter $\mathbf{A}^t_m$ is represented by a two-level network with weights $\mathbf{W}$ and $\mathbf{V}$ respectively [6]. The first level weights, $\mathbf{W}$, represent filters within subspaces whereas the second level weights, $\mathbf{V}$, are fixed to represent the structure of subspaces. Features extracted from this non-linear network can be expressed as:

$$\mathbf{s}_m(\mathbf{x}^t; \mathbf{W}, \mathbf{V}) = \sqrt{\sum_{l=1}^{L} \mathbf{V}_{ml} (\sum_{j=1}^{N} \mathbf{W}_{lj} \mathbf{x}^t_j)^2} \tag{3}$$

in which $\mathbf{W}$ is learned from a training set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^T\}$ by minimising $\sum_{t=1}^{T} \sum_{m=1}^{M}$ $\mathbf{s}_m(\mathbf{x}^t; \mathbf{W}, \mathbf{V})$ subject to $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, where N, L and M are the input dimensionality, number of linear components in each subspace and number of subspaces respectively. Some learned filters are visualised in Figure 2. After training, local descriptors are extracted by applying Eq.(3) to each 3D image window. For comparison, we also used convolutional ISA in which ISA networks are stacked in a convolutional manner, following the implementation described in [6].

(a) (b) (c) (d)

Figure 2: 17^3-voxel 3D filters, $\mathbf{W}$, learned from 9,000 3D OPT image windows using ISA.

5 Experiments

Each 3D patch was represented by its bag-of-visual-words histogram. This histogram was formed by binning the feature vectors extracted from each sub-window of a fixed size contained within the patch. Histogram bins corresponded to the learned visual word dictionary.

(a) (b) (c) (d)

Figure 3: 3D patches sampled from invasive cancer regions (a,b), and low-grade dysplasia regions (c,d).

The window size used for feature extraction is an important parameter. Given a central pixel located at (x,y,z) and window size $n \times n \times n$, the neighbouring pixels are defined within the cubic region from $(x-(n-1)/2, y-(n-1)/2, z-(n-1)/2)$ to $(x+(n-1)/2, y+(n-1)/2, z+(n-1)/2)$. Linear support vector machine (SVM) classifiers were trained on sets of 3D patches sampled from the annotated OPT image regions. Testing was always performed on OPT images not used for training. Thus the experiments reported in this paper are tested for inter-polyp generalisation.

Classification experiments were performed on a set of 59 volumetric OPT images from 59 patients. These images were acquired using ultraviolet light and Cy3 dye. Each image was of one colorectal polyp specimen and had $1024 \times 1024 \times 1024$ voxels with aspect ratio of $1:1:1$. In 30 images, 3D regions judged to consist entirely of low-grade dysplasia were annotated by a trained pathologist. In the other 29 images, 3D regions judged to consist entirely of invasive cancer (IC) were similarly annotated. The 3D regions were annotated as 2D regions in sequences of 2D slices using ITK-SNAP [13]. A polyp typically extended across $700 \sim 800$ slices of a volumetric image. The 2D regions were delineated every 4 or 5 slices and the region volume was interpolated in the intervening slices.

We randomly sampled 9,000 non-overlapping 3D image patches with size $21 \times 21 \times 21$ strictly within the annotated regions for each image. Figure 3 shows some example patches. In order to test the generalization capability of our approach *across patients*, we separated the 3D patches sampled from different polyps during experiments. Samples from one polyp were only presented in either training set or testing set. We used the Matlab interface of L2-SVM[3] for the SVM classifiers.

In our experiments, four different feature extraction methods are tested: 1) ISA with Bag of Words, denoted by 'ISA+BoW'; 2) RP with Bag of Words, denoted by 'RP+BoW'; 3) VLBP with Bag of Words, denoted by 'VLBP+BoW' and 4) Convolutional ISA, denoted by 'ConvISA'. In the ConvISA method, the output of one ISA model serves as the input basis of another ISA model. These two ISA models in ConvISA could be viewed as a local feature extractor and a feature encoder respectively (in analogy to BoW). Therefore, we do not specifically embed ConvISA in the Bag of Words framework. Instead, the output features of ConvISA are directly input to the classifier. For the RP method we reduced the dimensionality of raw patch features to 150 using random projection. In convolutional ISA and ISA with Bag of Words methods, the dimensionality was reduced according the same rule but using PCA (exactly following [6]). For all Bag of Words encoding processes, visual word dictionaries were learned using K-means clustering with K fixed to 200 unless otherwise stated. To form a fair comparison, the number of second level ISA features in convolutional ISA was also set to 200, the same value as K.

6 Results

Patch-level classification. We obtained Receiver Operating Characteristic (ROC) curves for each method by varying thresholds of the linear SVM output. Figure 4 shows ROC

curves for each method at the respective optimal window sizes, n, chosen according to the best parameters reported in [8]. The classification experiment was conducted in a 10-fold validation setting similarly to [8]. That is, we randomly divided the dataset into 10 folds, with about 6 images per fold. For each fold evaluation, we trained models with 9 folds and tested on the remaining one. With non-overlapping random sampling, for each evaluation routine the classifier was trained with about 8,100 3D patches and tested on 900 3D patches. For the SVM classifier outputs we also report the value of Area Under ROC Curve (AUC)

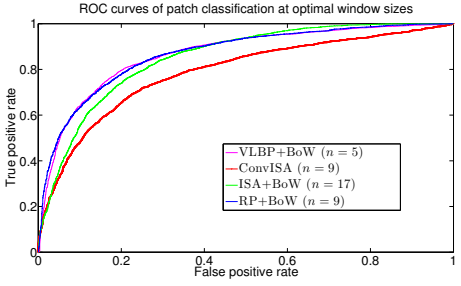

Figure 4: ROC curves given optimal window size.

n	Method	AUC	EER	Time (s)
5	VLBP+BoW	**0.87**	**0.20**	407
9	ConvISA	0.78	0.27	788
17	ISA+BoW	0.85	0.23	472
9	RP+BoW	**0.87**	0.21	**242**

Table 1: AUC and EER given optimal window size.

and Equal Error Rate (EER) in Table 1. All feature extraction methods were implemented in Matlab. The averaged computational time (in seconds) of feature extraction processes on a 2GHz Intel i7 CPU are listed in the last column of Table 1.

Random projection feature evaluation. The performance of classifier with RP+BoW method was affected by two critical parameters: the number of visual words in the dictionary (K) and the dimensionality of the random projection matrix (D). Given the optimal window size of $n = 9$ for RP+BoW, the extent to which D and K affect OPT image patch classification remained unclear. To test this effect, further experiments were conducted by varying D and K. All the other parameters were set the same as in the previous RP+BoW evaluation. The AUC values against D and K are reported in Figure 5. Note that D is shown in a log scale in order to have a better illustration of the trend.

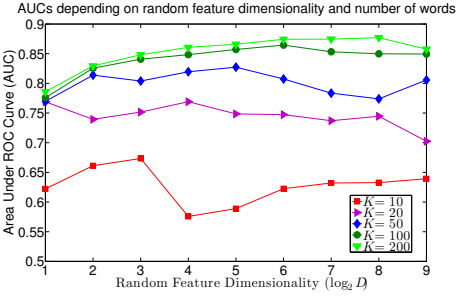

Figure 5: AUCs depending on random feature dimensionality (D) and number of words (K).

7 Discussion and Conclusion

We compared four methods for discriminating between invasive cancer and low-grade dysplasia in OPT images of colorectal polyps. Figure 4 suggests that the simple RP+BoW approach outperforms domain-specific feature learning method (ConvISA and ISA+BoW).

In comparison with hand-crafted features (VLBP), random features showed similar performance but with significant reduction in computational cost (Table 1). In Figure 5, classification performance (AUC) is dominated by number of visual words (K) when K is relatively small ($K \leq 50$). The highest value of K tried ($K = 200$) gave the best AUC. The effect of dimensionality (D) was apparent when $K \geq 100$. Liu and Fieguth [9] found that approximately one-third the dimensionality of the original 2D patch space was needed to preserve the salient information contained in the original local patch. Our experiments indicated that with 3D OPT image patches, given enough words, there is no harm to further reduce the dimensionality to $\frac{2^6}{729} \approx \frac{1}{10}$ the dimensionality of the original patch space.

Acknowledgement

This work is partially supported by Dundee Cancer Centre (DCC) Development Fund and RSE-NSFC Joint Project (RSE Reference: 443570/NNS/INT).

References

[1] Dimitris Achlioptas. Database-friendly random projections. In *ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 274–281. ACM, 2001.

[2] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. *in ACM SIGKDD*, 2001.

[3] RongEn Fan, KaiWei Chang, ChoJui Hsieh, XiangRui Wang, and ChihJen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[4] Aapo Hyvarinen and Patrik Hoyer. Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.*, 12(7):1705–1720, 2000.

[5] A.M. Khan, H. El-Daly, and N. Rajpoot. Ranpec: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images. In *MIUA*, 2012.

[6] Q.V. Le, W.Y. Zou, S.Y. Yeung, and A.Y. Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. *in CVPR*, 2011.

[7] Q.V. Le, Ju Han, J.W. Gray, P.T. Spellman, A. Borowsky, and B. Parvin. Learning invariant features of tumor signatures. *in ISBI*, 2012.

[8] W. Li, J. Zhang, S. J. McKenna, M. Coats, and F. A. Carey. Classification of colorectal polyp regions in optical projection tomography. In *ISBI*, 2013.

[9] L. Liu and P. Fieguth. Texture classification from random features. *in TPAMI*, 34(3):574 –586, 2012.

[10] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.*, 49(2):117–125, 2010.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *in TPAMI*, 24(7):971–987, 2002.

[12] P. G. van Putten, L. Hol, H. van Dekken, J. Han van Krieken, M. van Ballegooijen, E. J. Kuipers, and M. E. van Leerdam. Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening. *Histopathology*, 58(6):974–9819, 2011.

[13] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31:1116–1128, 2006.

[14] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *in IJCV*, 73(2):213–238, 2007.

[15] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *in TPAMI*, 29(6):915–928, 2007.