This is a preprint of a paper that will be published in the IEEE ISBI, 2013.

Classification of Colorectal Polyp Regions in Optical Projection Tomography

Wenqi Li*, Jianguo Zhang*, Stephen J. McKenna*,
Maria Coats†, Frank A. Carey‡

* CVIP, School of Computing, University of Dundee, Dundee, DD1 4HN
†School of Medicine, Ninewells Hospital and Medical School, Dundee, DD1 9SY
‡Department of Pathology, Ninewells Hospital and Medical School, Dundee, DD1 9SY

Abstract. The potential of optical projection tomography (OPT) to enhance colorectal polyp diagnosis is beginning to be explored. This paper presents, to the best of our knowledge, the *first study* on automatic image analysis of OPT images of colorectal polyps. 3D regions are classified using the bag of visual words framework and support vector machines. Independent subspace analysis is used to learn a domain-specific feature dictionary. This is compared to the use of raw patches (after random projection) and local binary patterns. Classification experiments (*across patients*) at the patch level and at the region level are presented using a set of 30 expert-annotated OPT images. Results show that accurate classification of 3D OPT image regions is feasible using this approach; regions of low-grade dysplasia and invasive cancer were discriminated with approximately 90% accuracy.

Keywords: Unsupervised Learning, Optical Projection Tomography, Colorectal Polyps.

1 Introduction

Optical projection tomography is a relatively new 3D imaging modality [1]. It is well suited to specimens between 1mm and 10mm across and has an effective resolution of 5-10μm. The potential of OPT to enhance colorectal polyp diagnosis is beginning to be explored [2]. Subtle changes have been noted when unstained human tissue specimens have been scanned, reflecting their different cellular composition.

Colorectal cancer (CRC) is the third most common cancer in the world and the fourth most common cause of cancer death worldwide [3]. Screening has lowered CRC mortality and detected large numbers of adenomas and polypoid cancers. However, diagnosis using conventional 2D histopathology can exhibit marked inter-observer variation, in the case of categorization of adenomas for example [4].

This paper is concerned with automated classification of regions of 3D OPT images of colorectal polyps. There is, to the best of our knowledge, no prior literature on automated image analysis of polyps using OPT. Specifically, we focus

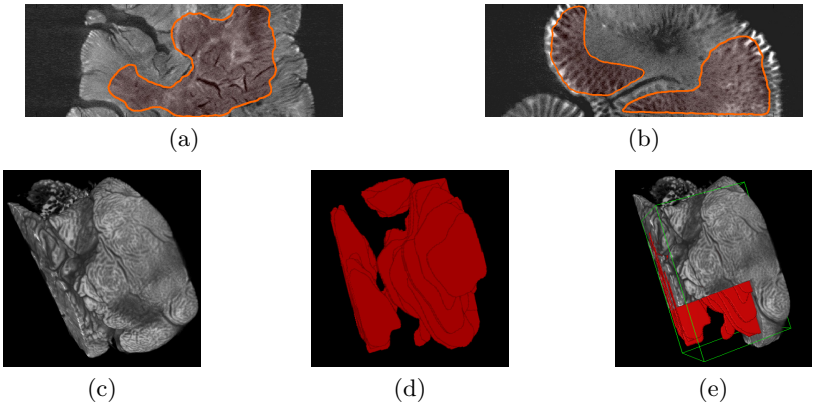

Fig. 1. OPT slices showing regions of (a) invasive cancer and (b) low-grade dysplasia. (c) Volume rendering of a polyp. (d) Annotated region of polyp in (c). (e) Combined volume rendering of the polyp and its annotation. (This figure is best viewed in color.)

here on the task of discriminating between regions of low-grade dysplasia and invasive cancer. Fig. 1 shows example regions annotated by a histopathologist. Regions of invasive cancer tend to have a more dense and homogeneous texture than low-grade dysplasia (see (a) and (b) in Fig. 1).

We adopt the *bag of visual words* framework [5]. Each 3D image patch is represented by extracting local features from it, quantising these features using a learnt visual word dictionary, and histogramming them. We investigate the use of unsupervised learning to obtain domain-specific features. Specifically, we use forms of independent subspace learning (ISA), motivated by recent promising results using ISA for video classification [6] and classification of H&E stained histology images of Glioblastoma Multiforme [7]. We compare this approach with using raw image patches (after dimensionality reduction using random projection [8]) and with local binary pattern descriptors in 3D [9–11]. These methods represent three different categories of low-level feature extraction method, i.e., learning domain-specific features, use of raw image patches and use of hand-crafted features.

2 Feature Extraction

2.1 Independent Subspace Analysis

The Independent Subspace Analysis (ISA) model is described as follows [12]. An input image x^t can be modelled as a linear combination of features:

$$x^t = \sum_k \sum_{i \in S(k)} A_i^t s_i \tag{1}$$

where $S(k)$ is the set of indices i of A_i^t that belongs to the k-th subspace. A^t can be viewed as a non-linear image filter bank. This is a generalized version

of Independent Component Analysis (ICA). It is different from ICA in that components are divided into subspaces; subspaces are assumed independent, whereas components in the same subspace need not be independent of each other. By maximizing independence with a grouping strategy, ISA models of image patches resemble complex cells in visual cortex. Features learnt by ISA show phase-and shift-invariant properties [12]. We investigate whether analysis of 3D images with ISA would benefit from these properties as well.

We deploy a model developed for a visual action recognition problem [6]. In this model, the non-linear filter A_i^t is represented by a two-level network with weights W and V respectively. The first level weights, W, represent filters within subspaces. The second level weights, V, are fixed to represent the structure of subspaces. Features extracted from this non-linear network can be expressed as:

$$s_i(x^t; W, V) = \sqrt{\sum_{k=1}^{d} V_{ik} (\sum_{j=1}^{n} W_{kj} x_j^t)^2} \tag{2}$$

in which W can be learnt from a training set $\{x^1, x^2, \ldots, x^T\}$ by optimising the following:

$$\begin{aligned} \min_{W} \quad & \sum_{t=1}^{T} \sum_{i=1}^{m} s_i(x^t; W, V), \\ \text{subject to} \quad & WW^T = I. \end{aligned} \tag{3}$$

where n, d and m are the input dimensionality, number of linear components in each subspace and number of subspaces respectively.

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

Fig. 2. (a)-(d) Visualisations of 3D filters, W, learnt from 40,000 3D OPT image patches using ISA. (Size: $19 \times 19 \times 19$ voxels)

To apply ISA as a local feature extractor, we first learn weights of the two-layer network, W, from a training set with fixed V (by solving Eq.(3)). Then final local descriptors are extracted by applying Eq.(2) to every 3D image patch. Some of the learnt filters are visualised in Fig. 2. ISA network can also be stacked in a convolutional manner, i.e., the output of first ISA model is the direct input of second ISA. Here we propose to use convolutional ISA as a feature extraction method for OPT images. We follow the implementation described in [6] as a comparison to Bag of Words framework.

2.2 Raw Patches and Random Projection

In our raw 3D image patches method, intensity values are used as local features. These can be of very high dimensionality. For example, in the case of 3D patch

features, the dimensionality is $10 \times 10 \times 10 = 1000$ when sampling window size is 10. Therefore we use randomly projected patches (RP) as a local feature descriptor in our study. Random projection is a simple but powerful method for dimensionality reduction [13]. It has been shown to be useful for texture analysis in [8]. The original n-dimensional data $X = \{x^t\}_{t=1}^{T}$ can be projected to a k-dimensional subspace, using a random $k \times n$ matrix R whose columns have unit lengths:

$$\hat{X}_{k \times T} = R_{k \times n} X_{n \times T} \tag{4}$$

Distances between data points in n-dimensional space ($X_{n \times T}$) are approximately preserved in k-dimensional space ($\hat{X}_{k \times T}$). To construct the random projection matrix R, we generate the elements r_{ij} of R simply by drawing samples from a Gaussian distribution with zero mean and unit variance. The complexity of this process in Eq.(4) is only $O(dkn)$.

2.3 3D Local Binary Patterns

Gray-scale and rotation-invariant LBP operators were introduced for 2D texture analysis [10]. They have shown high discriminative power in certain medical image analysis tasks [11, 14]. For a 2D image patch, the operator encodes a 3×3 neighbourhood of each pixel as a binary number. These binary numbers (called LBP codes) correspond to primitive features of the image such as edges, corners and spots. The histogram of LBP codes computed over the image patch can be used as a texture descriptor.

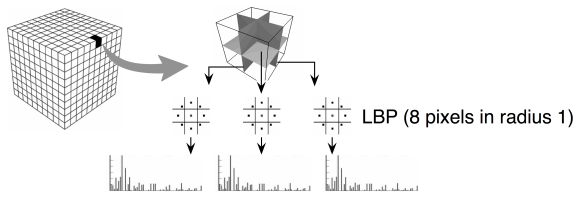

Fig. 3. Process of encoding local patches with volumetric LBP

We applied the LBP operator on three orthogonal planes for volumetric texture characterisation (VLBP) similarly to [9]. LBP features from the three orthogonal planes were concatenated to form the local feature vector as shown in Fig 3.

3 Classification

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

to the learnt visual word dictionary. The window size used for feature extraction is an important parameter. Given a central pixel located at (x, y, z) and window size 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 tested for inter-polyp generalisation. Two classification tasks are considered:

Patch-level classification. Each 3D patch in the test region is classified based on its histogram without taking into account other 3D patches in the region.

Region-level classification. Each region annotated by the expert is classified as a whole based on all the patches it contains. Treating patch labels, y, as conditionally independent given the region class, the task is to decide whether the class, C, of a 3D annotated region is invasive cancer ($C = I$) or low-grade dysplasia ($C = L$) based on the bag of labels, $\mathcal{Y}$, assigned to its patches by the SVM. We assume equal class prior probabilities since our dataset contains equal numbers of regions in each class. Therefore, the likelihood ratio in Eq.(5) is an appropriate quantity on which to base the classification decision,

$$R = \frac{P(\mathcal{Y}|C = I)}{P(\mathcal{Y}|C = L)} \tag{5}$$

Since we assume patch labels are conditionally independent, the likelihoods are modelled as binomial distributions and the likelihood ratio can be written as:

$$R = \frac{(P(y = I|C = I))^M (P(y = L|C = I))^N}{(P(y = I|C = L))^M (P(y = L|C = L))^N} \tag{6}$$

where M and N are the number of patches classified as invasive cancer and low-grade dysplasia in the 3D region. Denoting invasive cancer as positive and low-grade dysplasia as negative, the likelihood function $P(y|C)$ is estimated by counting the number of true positives, true negatives, false positives and false negatives.The classification decision can be made by thresholding at $\log R = 0$.

4 Experiments

Classification experiments were performed on a set of 30 volumetric OPT images from 30 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 15 images, 3D regions judged to consist entirely of low-grade dysplasia were annotated by a trained pathologist. In the other 15 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 [15]. A polyp typically extended across $700 \sim 800$

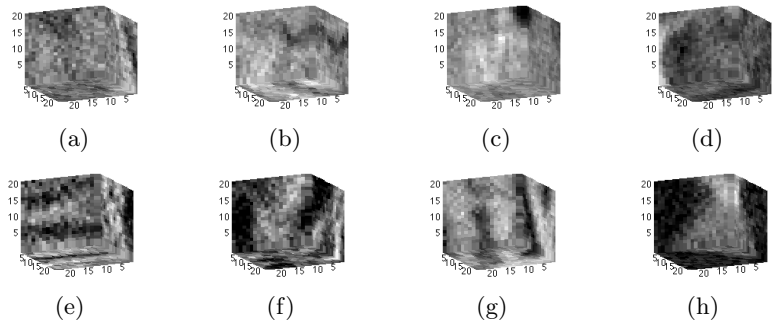

Fig. 4. (a)-(d) Patches sampled from invasive cancer regions. (e)-(h) Patches sampled from low-grade dysplasia regions.

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 2000 non-overlapping 3D image patches with size $21 \times 21 \times 21$ strictly within the annotated regions for each image. Fig 4 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[16] 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 ConvISA method, output of one ISA model serves as input basis of another ISA model. These two ISA in ConvISA could be viewed as a local feature extractor and a feature encoder (in analogy to BoW) respectively. Therefore we do not specifically embed ConvISA in Bag of Words framework. Instead the output feature code of ConvISA is directly input to the classifier.

For the RP method we reduced the dimensionality to 150 using random projection if the dimensionality was greater than 150. In convolutional ISA and ISA with Bag of Words methods, the dimensionanlity was reduced according the same rule but using PCA (following [6]). For all Bag of Words encoding processes, visual word dictionaries were learnt using k-means clustering with k fixed to 200. To form a fair comparison, the second level ISA features in convolutional ISA were also set to 200.

4.1 Results

Patch-level classification. Classification accuracy is calculated as $\frac{TP+TN}{P+N}$ where TP, TN, P, N denote number of true positives, true negatives, and total number of positive and negative samples respectively. Table 1 reports classification rates when using the different feature descriptors with various sizes of

feature window. The accuracies are averaged over 10-fold validation as recommended by [17]. We randomly divided the dataset into 10 folds, with 3 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 40,000 3D patches and tested on 4000 3D patches.

	RP+BoW	ISA+BoW	ConvISA	VLBP+BoW
3	63.0 ± 6.9	52.9 ± 7.8	57.0 ± 6.3	50.8 ± 9.0
5	69.8 ± 5.8	53.0 ± 7.9	68.5 ± 4.2	$\mathbf{73.1 \pm 3.5}$
7	72.4 ± 5.1	61.1 ± 6.0	63.0 ± 3.6	70.1 ± 4.0
9	$\mathbf{73.4 \pm 3.3}$	65.9 ± 5.2	$\mathbf{71.0 \pm 3.0}$	66.8 ± 4.7
11	72.2 ± 3.4	66.9 ± 5.1	65.3 ± 3.9	69.7 ± 4.3
13	70.1 ± 3.5	69.4 ± 5.2	68.8 ± 3.6	68.2 ± 4.7
15	68.1 ± 3.2	68.9 ± 4.9	67.0 ± 3.8	68.4 ± 2.9
17	62.9 ± 3.5	$\mathbf{72.0 \pm 4.3}$	66.9 ± 3.9	67.5 ± 3.5
19	59.0 ± 3.4	67.4 ± 5.3	56.8 ± 6.6	65.9 ± 3.3

Table 1. Classification accuracies (%) $\pm$ standard errors at different feature window sizes ranging from 3 to 19.

There were no significant differences between the best accuracies of these methods. For the window size settings that gave the highest accuracies for each method, we compare sensitivity and specificity in Table 2, where sensitivity $= \frac{TP}{P}$; specificity $= \frac{TN}{N}$.

	Sensitivity	Specificity	Accuracy
RP+BoW	67.6 ± 6.2	78.9 ± 3.4	73.4 ± 3.3
ISA+BoW	86.3 ± 2.0	57.0 ± 6.9	72.0 ± 4.3
ConvISA	62.1 ± 4.5	77.2 ± 4.0	71.0 ± 3.0
VLBP+BoW	84.6 ± 2.6	59.8 ± 5.0	73.1 ± 3.5

Table 2. Sensitivities (%) and specificities (%) $\pm$ standard errors for the window sizes that gave the highest accuracies.

Region-level classification. We evaluated region classification using the likelihood ratio test of Eq.(6) with a threshold $\log R = 0$. The patch labels from patch-level classification (described above) were used in this evaluation. Fig. 5 reports these results.

4.2 Discussion

At patch level, for RP and ConvISA descriptors, the best accuracies were generated at window size $9 \times 9 \times 9$. VLBP tended to have better performance when

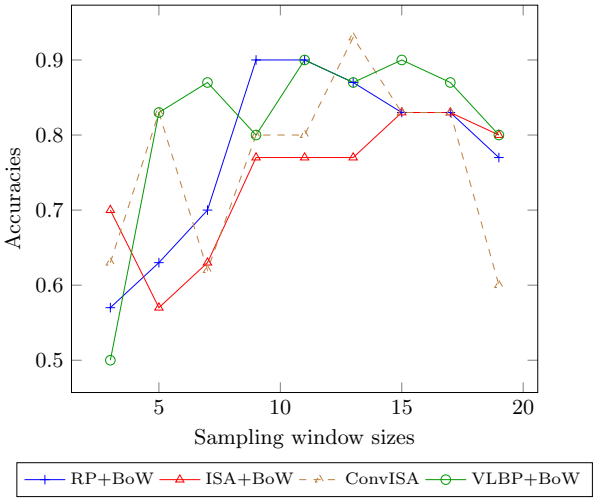

Fig. 5. Classification accuracies based on annotated regions depending on sampling window size.

window size was smaller while ISA had the opposite trend. This indicates that VLBP is more powerful for capturing local textural details while ISA is able to generate discriminative features when the feature window is relatively large. The RP method shows notable results as it takes image intensity values directly as input. The highest accuracies for these methods are very close, but sensitivities and specificities of these methods are quite different. Overall, the results indicate that using raw 3D patches directly as features has equal discriminative power compared with hand-designed descriptors (VLBP) and automatically learnt descriptors (ISA). This finding on our OPT dataset is consistent with studies on other applications of texture analysis [5]. There was no advantage to sophisticated descriptors such as ISA and VLBP for patch classifications in these OPT images. The RP method can serve as an effective local feature extractor.

At region level, the trends of classification accuracies are very similar to the trends of accuracies at patch level (discussed above). Good performance in patch-level classification usually leads to good performance in region-level classification. RP gave classification accuracy of 90% with window sizes of $9 \times 9 \times 9$ or $11 \times 11 \times 11$. ConvISA gave a slightly better accuracy of 93% with $13 \times 13 \times 13$ windows albeit at greater computational expense. Almost at any sampling window size, classification accuracies at region level are higher than at patch level. This is because region-level classification combines information over a larger spatial extent than patch-level classification.

Our classification problem has similarities to the scenario in [7]. However, as well as differences in data dimensionality and modality, there is a clear distinction in that we evaluate *cross patient* classification at both patch level and region level. In [7], training and testing patches might come from the same patients.

5 Conclusion and Future Work

We compared four methods for discriminating between invasive cancer and low-grade dysplasia in OPT images of colorectal polyps. Results showed that automatic classification of annotated regions in OPT images is feasible. In future we would extend this work in three directions. 1) combining different local feature sets for feature learning; 2) experimenting with more tissue classes such as different grades of dysplasia, and 3) incorporating positional and contextual information, for example using autocontext or conditional random field models, to improve classification and segmentation of colorectal polyps in OPT.

Acknowledgements The authors would like to thank members of the CVIP group, especially Shazia Akbar, Emanuele Trucco and Ruixuan Wang, for their valuable comments. This work is partially supported by Dundee Cancer Centre (DCC) Development Fund and RSE-NSFC Joint Project (RSE Reference: 443570/NNS/INT).

References

1. J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, "Optical projection tomography as a tool for 3D microscopy and gene expression studies.," *Science*, vol. 296, no. 5567, pp. 541–545, 2002.
2. M. V. Coats, S. E. Wedden, J. Farrell, G. Cranston, L. Mitchell, J. Wilson, R. J. Steele, and F. A. Carey, "Optical projection tomography: Can it help diagnose the colorectal polypoid cancer?," *Gastroenterology*, vol. 142, no. 5, pp. S178–S179, 2012.
3. J. Ferlay, H. Shin, F. Bray, D. Forman, C. Mathers, and D. Parkin, "Globocan 2008 v2.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 10 [internet]," `http://globocan.iarc.fr`, 2010, Accessed on 26/10/2012.
4. 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*, vol. 58, no. 6, pp. 974–9819, 2011.
5. 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*, vol. 73, no. 2, pp. 213–238, 2007.
6. Q. Le, W. Zou, S. Yeung, and A. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," *in CVPR*, pp. 3361–3368, 2011.
7. Q. Le, J. Han, J. Gray, P. Spellman, A. Borowsky, and B. Parvin, "Learning invariant features of tumor signatures," *in ISBI*, pp. 302 –305, 2012.
8. L. Liu and P. Fieguth, "Texture classification from random features," *in TPAMI*, vol. 34, no. 3, pp. 574 –586, 2012.
9. G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *in TPAMI*, vol. 29, no. 6, pp. 915–928, 2007.
10. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *in TPAMI*, vol. 24, no. 7, pp. 971–987, 2002.

11. L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.
12. A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, 2000.
13. E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," *in ACM SIGKDD*, pp. 245–250, 2001.
14. D. Unay, A. Ekin, M. Cetin, R. Jasinschi, and A. Ercil, "Robustness of local binary patterns in brain MR image analysis," *in EMBS*, pp. 2098–2101, 2007.
15. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, pp. 1116–1128, 2006.
16. R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
17. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *in IJCAI*, pp. 1137–1143, 1995.