Analysing the Surface Morphology of Colorectal Polyps: Differential Geometry and Pit Pattern Prediction

Jingjing Zhang[1]
ZhangJ68@cardiff.ac.uk
Stephen J. McKenna[1]
stephen@computing.dundee.ac.uk
Jianguo Zhang[1]
jgzhang@computing.dundee.ac.uk
Maria Coats[2]
mariacoats@nhs.net
Frank A. Carey[3]
frank.carey@nhs.net

[1] CVIP, School of Computing, University of Dundee, Dundee DD1 4HN, UK
[2] School of Medicine, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK
[3] Department of Pathology, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK

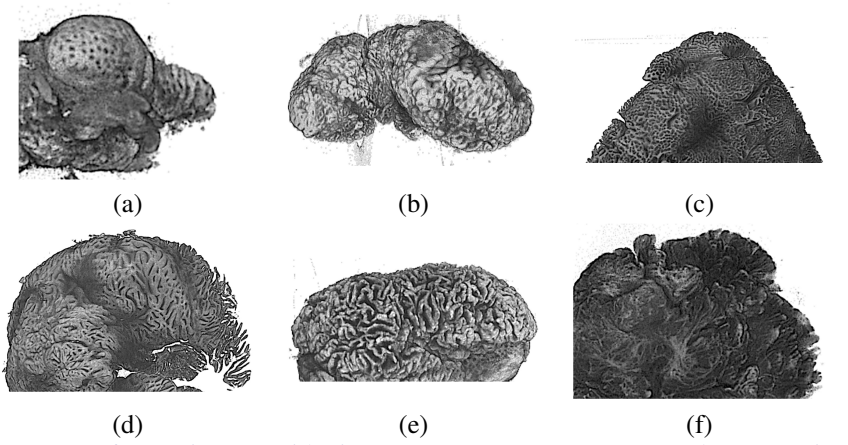
Abstract

We present an initial study analysing the surface morphology of colorectal polyps from optical projection tomography. The differential geometry of polyp surfaces, segmented using a level sets method, is explored in terms of local, multi-scale shape index and curvedness descriptors. A surface region of interest can be represented using histograms of these descriptors. An experiment is described investigating the ability to predict pit pattern categories from these histograms using support vector machines.

1 Introduction

Colorectal cancer is the third most common cancer in men (746k cases, 10.0%) and the second in women (614k cases, 9.2%) worldwide [3]. Screening has reduced mortality by up to 21% and detected large numbers of adenomas and polypoid cancers [7]. However, analysis of histological H&E sections suffers from marked inter-observer variation.

There is a long history of study of polyp surface morphology relating to different stages of cancer development in the medical literature. Polyp surface morphology is often categorised in the pathology lab as being villous (having cerebral-like folds and finger-like protrusions), tubulo-villous or tubular (a smooth surface with pits/tubes in it). Whilst villous polyps are strongly associated with invasive cancer, all three types can and do develop into polypoid cancers[8]. This classification scheme is a coarse categorization of the complex and highly variable surface morphology and normally presents large inter-observer variation. Another categorization, used for in vivo endoscopic assessment based on visual appearance, is Kudo's pit patterns [5]. There are six classes of pit pattern: I, II, III-S, III-L, IV, and V.



(a) (b) (c)

(d) (e) (f)

Figure 1: (a)-(f) OPT images with pit patterns I, II, III-S, III-L, IV and V respectively.

Here we investigate polyp morphology using optical projection tomography (OPT), an in vitro imaging technology capable of producing high-resolution 3D images of small biological specimens [11]. Li et al. [8] reported classification of 3D regions of dysplasia and invasive cancer in colorectal polyps from OPT. Here we analyse polyps' surface morphology. Figure 1 shows surface renderings of some OPT polyp images, one for each pit pattern.

2 Surface Segmentation and Multi-scale Features

We employ a level sets fast marching method for polyp surface segmentation [10] using the implementation in [9]. Segmentation was seeded using 100 voxels randomly selected on a plane within the background. Once the surface is segmented, shape descriptors are constructed based on multi-scale differential geometry features computed at each voxel on the segmented surface. We do not fit a mesh grid which makes assumptions about topology. Instead, we work directly with image data and extract surface curvature estimates based on a 3D local neighbourhood in a similar fashion to [1]. This has the advantage of avoiding meshing errors. The principal curvatures κ_{max} and κ_{min} (maximum and minimum values of curvatures over all of directions) are obtained at each such voxel at several scales obtained by setting the Gaussian smoothing parameter, σ, used in the estimation of image gradients.

Figure 2 shows polyp surface regions rendered according to whether curvature is convex (green) or concave (magenta) at three scales. As σ increases, the connected regions of concave and convex surface become larger and the coarse morphology becomes apparent. The characteristic scales exhibited by the different pit patterns are not all the same. For example, the surface morphology of the polyp region in Fig. 2(a) is perhaps best characterised at a larger scale ($\sigma = 4$) while the morphology of the region in Fig. 2(e) is perhaps best characterised at a smaller scale (e.g. $\sigma = 1$). This observation motivates a multi-scale analysis.

Rather than representing surface curvature directly in terms of principal curvatures κ_{max} and κ_{min}, we compute a local *shape index*, S, and *curvedness*, C, as proposed by Koenderink [4] (Equations (1) and (2)).

$$S = -\frac{2}{\pi} \arctan \frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}} \tag{1}$$

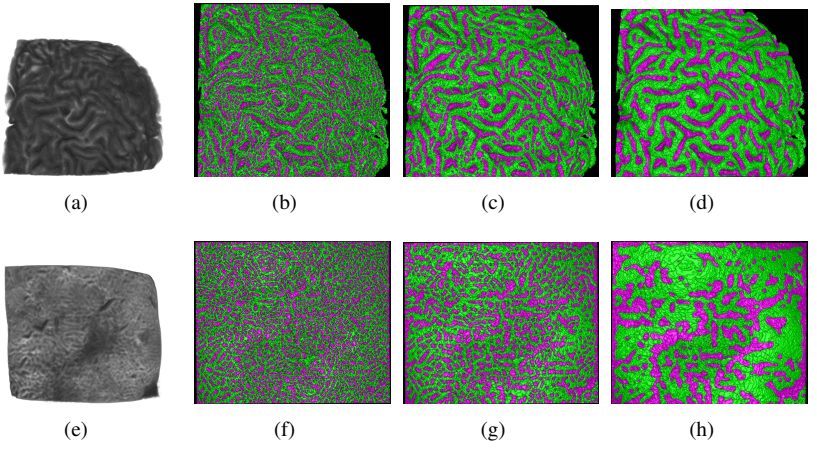


Figure 2: Two colorectal polyp surfaces classified as concave (magenta) and convex (green) at three scales. (a) A pit pattern IV region. (b-d) $\sigma = 1,2,4$ respectively. (e) A pit pattern III-S region. (f)-(h) $\sigma = 1,2,4$ respectively.

$$C = \frac{2}{\pi} \ln \left(\sqrt{\frac{\kappa_{max}^2 + \kappa_{min}^2}{2}} \right) \tag{2}$$

The shape index is a convenient measure of "which" shape, and the curvedness of "how much" shape. In the space spanned by (C,S), all shapes are mapped onto a strip of infinite length in the C- direction. As the value of the shape index S varies from -1 to 1, the corresponding surface shape transforms from concave pit, through trough, saddle, and ridge to convex peak. Figure 3 shows histograms of the shape index and curvedness obtained from surface regions of interest on eight polyps (two from each of pit patterns III-S, III-L, IV, and V). This gives some indication of the inter- and intra-class variability of these descriptors.

3 Pit Pattern Assignment

Experiments described in what follows used 28 OPT polyp images from 28 patients acquired using ultraviolet light and Cy3 dye. The size of each image was $1024 \times 1024 \times 1024$ voxels with aspect ratio of $1:1:1$. We included only polyps initially considered to be adenomatous polyps (pit patterns III-S, III-L, IV and V). Seven images were selected for each of the four pit patterns considered. The pit pattern type of each of these polyps was assigned by a trained pathologist at which time no secondary pit patterns were identified in these polyps. Their OPT images included some normal tissue and stalks; as such regions were not relevant for this study, they were excluded: a largest region of interest was selected through visual examination of each polyp volumetric image. The size of the extracted regions varied from $50 \times 300 \times 250$ to $350 \times 400 \times 500$.

As evidence of the reliability level of the assigned pit pattern labels, the same pathologist was asked at a later date to assign pit pattern labels to the 28 extracted regions. The

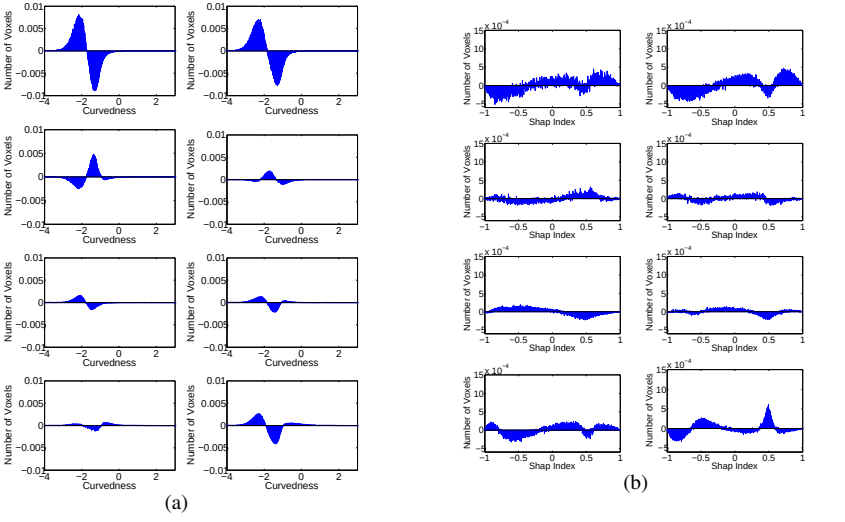


Figure 3: Histograms of (a) curvedness and (b) shape index from regions of interest on eight polyps. Each row represents one pit pattern (from top to bottom: III-S, III-L, IV and V). In each row, the first and third images are from the same polyp, and the second and the fourth are from another polyp. (Histograms have 1024 bins).

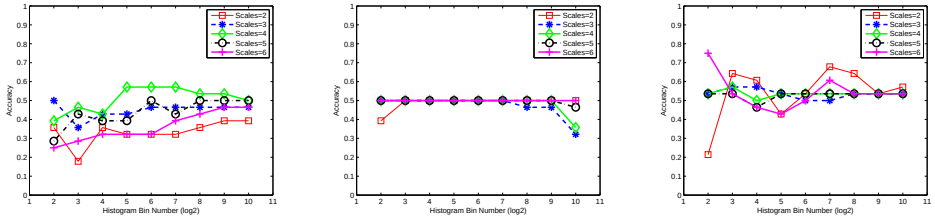
Table 1: Contingency table for pit pattern assignment by one observer. Initial session (whole polyp) vs. subsequent session (extracted region)

	III-S	III-L	IV	V
III-S	**3**	0	0	0
III-L	2	**3**	1	0
IV	2	2	**6**	4
V	0	0	0	**3**
I	0	2	0	0

regions were inspected in randomised order and the pathologist was blinded to any diagnostic information and to the previously assigned pit pattern labels. It was expected that the region extraction would have minimal effect on the pit pattern categorsiation, particularly as no secondary patterns had been noted. Nevertheless, the pit pattern assignment exhibited intra-observer variation; Table 1 is the resulting contingency table. The column represents the initial label assigned to the whole polyp while the row refers to the label subsequently assigned to the region. It can be seen that annotations of pit pattern IV are consistent over 6 samples, but for pit patterns III-S, III-L and V, only 3 in each category are consistent. In the region annotation, half of all the samples were categorized as pit pattern IV. In total, 15 polyps (54%) were annotated with the same label in both sessions.

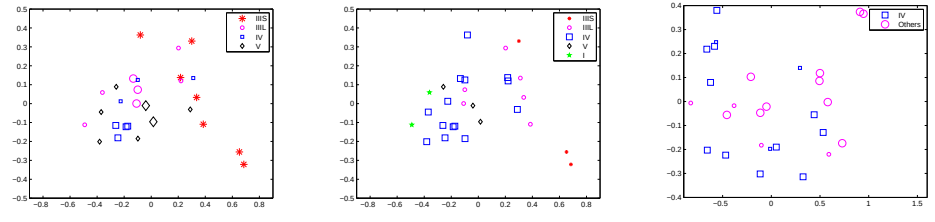
4 Pit Pattern Prediction Experiments

Support vector machines were trained to predict pit pattern labels. They were implemented using LIBSVM [2] and the 1-vs-rest method. As the number of polyps is relatively small,



(a) (b) (c)

Figure 4: SVM classification accuracy when predicting labels from (a) first session, (b) second session, and (c) in binary setting (IV vs. rest)



(a) (b) (c)

Figure 5: PCA visualisations. Correctly classified samples are larger. (a) First session labels, (b) second session labels, (c) binary labels

we report leave-one-out (LOO) cross validation results. To investigate the effect of scale we extracted features at scales of $\sigma = 2^k, k = 0,1,2,3,4,5$. To investigate the effect of the number of histogram bins, we used 2^k bins for k ranging from 2 to 10. Thus the smallest number of bins was 4, and the largest was 1024.

Figure 4(a) shows accuracy at predicting pit pattern labels assigned in the first session. An accuracy of 57% was achieved with 32 bins and 4 scales ($\sigma = 1,2,4,8$). In this case, all 7 samples in pit pattern III-S were correctly classified while only 2 samples in pit pattern III-L, 4 samples in pit pattern IV and 2 samples in pit pattern V were correctly classified. This performance is in line with the variability of the pathologist (see Table 1).

We also trained SVMs to predict the pit pattern labels assigned by the pathologist in the second session; the pathologist had assigned the 28 samples to five pit patterns III-S, III-L, IV, V and I with 3, 6, 14, 3, 2 samples respectively (Table 1), resulting in an imbalanced dataset. SVM classification results under this setting are reported in Fig. 4(b). The highest classification accuracy obtained was 50% in which case all the samples of pit pattern IV were classified in agreement with the pathologist.

Finally, SVMs were trained on labels from the second session in a balanced, binary setting: Pit pattern IV vs other patterns (14 samples per class). Figure 4(c) shows the classification results; 75% accuracy was obtained (4 bins and $\sigma = 1,2,4,8,16,32$). Three samples in pit pattern IV and four samples in non-pit pattern IV were misclassified.

Figure 5 visualises the distribution of the samples in feature space after projection onto two dimensions using PCA. Correctly classified samples are depicted with markers of larger size than the incorrectly classified samples.

5 Conclusions and Future Work

We presented a preliminary investigation of the use of differential geometry features (shape index and curvedness) to characterise the surface morphology of colorectal polyps from OPT images. In particular we evaluated the ability to predict a pathologist's pit pattern assignments based on distributions of these features. It was noted during label assignment that some patterns observed may not fit well into Kudo's pit pattern categorisation. Results were broadly in line with intra-observer disagreement although further experiments using larger data sets are needed to confirm this and to understand the extent to which pit patterns capture clinically relevant surface morphological characteristics.

References

[1] Brian Avants and James Gee. The shape operator for differential analysis of images. In *Information Processing in Medical Imaging*, pages 101–113. Springer, 2003.

[2] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[3] J. Ferlay, I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray. Globocan 2012 v1.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 11 [internet]., 2013. URL http://globocan.iarc.fr. accessed on 25/03/2014.

[4] Jan J. Koenderink. *Solid Shape*, volume 2. Cambridge University Press, 1990.

[5] S. Kudo, S. Hirota, T. Nakajima, S. Hosobe, H. Kusaka, T. Kobayashi, M. Himori, and A. Yagyuu. Colorectal tumours and pit pattern. *J. Clin. Path.*, 47(10):880–885, Oct 1994.

[6] W. Li, J. Zhang, S. J. McKenna, M. Coats, and F. A. Carey. Classification of colorectal polyp regions in optical projection tomography. In *ISBI*, pages 736–739. IEEE, 2013.

[7] J. S. Mandel, T. R. Church, F. Ederer, and J. H. Bond. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *Journal of the National Cancer Institute*, 91(5):434–437, Mar 3 1999.

[8] G. Nusko, U. Mansmann, U. Partzsch, A. Altendorf-Hofmann, H. Groitl, C. Wittekind, C. Ell, and EG Hahn. Invasive carcinoma in colorectal adenomas: multivariate analysis of patient and adenoma characteristics. *Endoscopy*, 29(07):626–631, 1997.

[9] Gabriel Peyre. Toolbox fast marching (matlab software), 2009.

[10] J. A. Sethian. *Level set methods and fast marching methods*. C. U. P., 1999.

[11] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sorensen, R. Baldock, and D. Davidson. Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science*, 296(5567):541–545, Apr 19 2002.