Colonoscopy Image Processing using the Structural Entropy

Szilvia Nagy, Brigita Sziová, Levente Solecki and László T. Kóczy
Széchenyi István University
H-9026 GyQr
Egyetem tér 1, Hungary
{nagysz@sze.hu}{szibr@sze.hu, koczy@sze.hu}{solecki@sze.hu}

ABSTRACT: *The fine surface of the bowel and the colorectal polyps reflect the colonoscopy images. This is analogous to the case of combustion engine surface scans, where the grooving and wear can be detected from the fine pattern superposed to a cylinder curvature. While comparing we can found that the colonoscopy images and to have many reflections, whereas the roughness scanners detect small dust particles and well as the vibrations from the environment. We bring a model in this work to take care of both the problems using histogram stretching together with a special type of filtering. Besides we have used the masks to control the effect of the operators. In the testing we have measured the effects of the processing steps on the structural entropy of the image because the structural entropies are used in characterization of the images.*

Keywords: Colonoscopy, Roughness, Background Subtraction, Reflection, Outliers, Entropy

Received: 13 January 2021, Revised 7 June 2021, Accepted 19 May 2021

DOI: 10.6025/jmpt/2021/12/3/74-80

1. Introduction

In measured 2D data, such as images or surface scans there can be characteristic lengths that carry the information which is needed, while larger or smaller scale patterns are not necessary. Also, in measured data such points can arise, which do not fit into the trend of the other measured points. In the case of images or scanned surfaces such outliers can also form larger domains. In the following considerations we give a method for eliminating the outliers as well as removing larger scale background.

We apply our method for two kinds of data, the first one is microgeometrical scan of combustion engine cylinder surface, where the data consist of height values measured on a regular 2D grid, the second is colonoscopic images, where the data consist of 3 colour intensity values at points of regular grid, again. The background that should be neglected is the curvature of the cylinder surface and the bowel wall respectively, while the outliers are the measurement noise, and the impurities in the first case and the light reflections from the wet surface in the second one.

Colonoscopy is mainly used in diagnosis of colorectal cancer, which is one of the most frequently occurring cancer types. It develops from special types of colorectal polyps [1,2], thus diagnosing in the pre-cancerous, polyp phase helps survival

greatly. Not all the polyp types have the tendency to develop into malign object, these polyps can remain in the bowel without causing any complications, moreover, it is better to leave them intact, as all the surgeries can potentially cause damage. The polyps can be classified as benign or malign according to the pattern on their surface [3-6]. Of course, the first task is to find the polyp, which is done by qualified medical experts nowadays, however there are several groups [7-9] studying the possibility of an automated detection. Previously, we developed a method that can classify an image segment to containing or not containing polyp classes, which works well for several types of images [10-11], however, for low quality pictures, the method fails. This was the reason we tried to preprocess the images.

In combustion engines the inner surface of the cylinder must be smooth enough to let the piston move, and at the same time it must store oil and the tiny particles arising as wear byproducts. The cylinder lining's pattern depends on the process it is made by. One of the most traditional and widespread method is honing, which results in straight grooves with a slight slope compared to the surface of the piston. The movement of the piston, however, causes fine scratches parallel to its movement. These patterns can be detected by surface scanners.

In order to build reliable classification scheme, we used Rényi entropy combinations and tested, whether there was difference between the new and the worn surfaces. However, in this classification, the large-scale pattern of the background dominates, and the fine pattern's own entropy is hard to distinguish. Outlier bring even more unwanted information to the system, thus we decided to remove them from the pictures.

In this article, after describing the image properties in Section 2, we give the histogram stretching and mask acquiring method together with the preparation of the outlier pattern to be subtracted. Section 3 describes the treatment of the large-scale background with polynomial fitting and large-scale averaging and give the results. The last section contains the conclusion.

2. Image Properties

In the case of the colonoscopy mages, we used image databases from multiple sources [7-9]. These images are taken in non-ideal circumstances, the bowel is always dark inside, so the endoscope has a light source, which is point-like. At the same time the inner surface is wet, moving, pink (ideally), i.e. it is very far from ideal image source. They are almost monotonic, with lots of reflections, multiple types of patterns from veins to polyp pits to the more-or-less distorted cylindrical shape of the bowel and the polyp protruding into the bowel space.

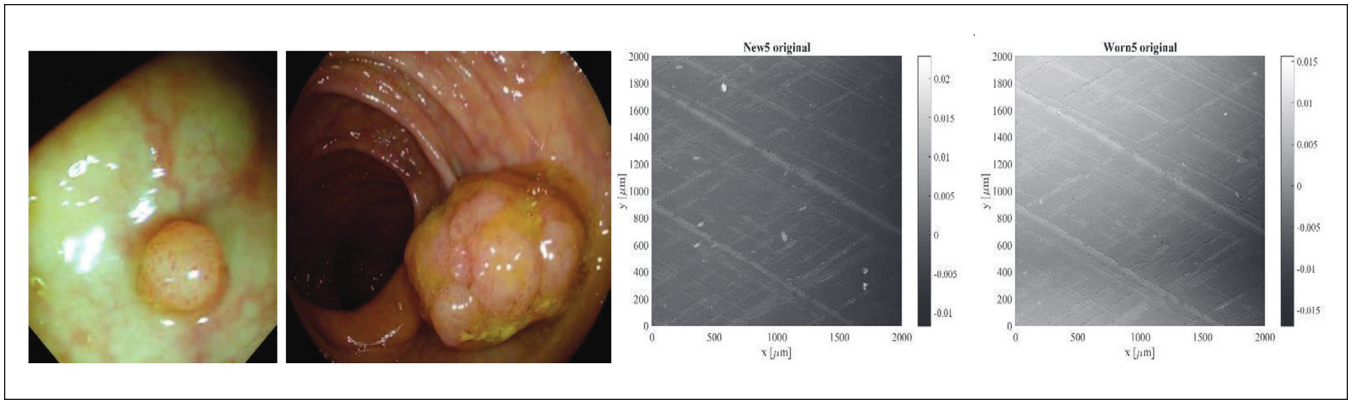

Figure 1. First two photographs: endoscopy images (picture No. 219 from database CVC_Colon [7] and No. 116 from ETIS-Larib [8]) with visible reflections, blood vessels, polyp patterns, bubbles and impurities. 3rd and 4th subplot: new and worn surface scans at measuring point 5 of the combustion engine cylinder surface, with visible curving and outliers (the outliers are trimmed down to make the other details visible)

In the case of the combustion engines, we took silicone replicas from the new engine, and of the same part of the surface after 500 hours polycyclic endurance test. For preparing the samples Struers RepliSet-F5 replica kit was used. The replicas measured at 5 different points along their axes (each to different environments: the first point was between the first two piston rings, the 5th location a bit over the lowest turning point of the piston) with TalysurfCLI2000 surface scanner's optical sensor based on chromatic length aberration (CLA). These images are much easier to treat than the colonoscopy ones, as they have

quite uniform pattern, a regular background and few, very small-sized outliers.

3. Contrast Enhancement, Mask Preparation and Reflection Filtering

After separating the colour channels of the colonoscopy images, we had to treat 3 separate 2D data, similarly to the case of the surface scans. Studying the intensity histograms of the images, in both cases the outliers – the reflections and the impurities – form a large peak around the maximum intensity value. In the case of the colonoscopy images there is a large peak at the minimum intensity value as well, because all the images have black frame. In the case of the scanned surfaces there are usually some measurement error toward the low levels as well, this means however, only a small peak at the minimum intensity value.

As a first step, we stretched the histogram so, that the peaks around the highest and lowest values would be cut off. Instead of these points we used the maximum and the minimum values of the new histogram, respectively. During this process we also defined a mask both for those points where the intensity was below the lower threshold, and for those above the upper threshold. The low-masks consisted of 0s in those points where the intensity was below the lower limit, and 1 in the other points, while the high-masks were 1 everywhere, except for the points where the intensity was above the higher threshold.

As a next step we extended the masks to those neighbouring points, where the intensity value was a by a certain level higher than its environment excluding the already masked points.

With the new masks we calculated such a matrix-to-besubtracted which was 0 everywhere, except for the masked points, where the values were calculated the following way. For each of the points we determined the average intensity of its surroundings of size 41 by 41 but did not count the points with 0 mask values into the surroundings. We put the difference between the average and the actual pixel intensity to the matrixto-be-subtracted and used it for eliminating the outliers.

The masks were used differently in the case of the two types of data. For surface scans, both the high and the low mask covered valuable domains of the matrix, while in the case of the colonoscopy images the black mask denoted those parts of the image where the colonoscope could not provide data. As a result, for colonoscopy, these black points had to be excluded from the points of the matrix-to-be-subtracted, while for engine cylinder surfaces, not.

As a last step we applied a mean filtering around the edges of the masks, if it was necessary.

As an example, the matrix-to-be-subtracted and the resulting picture of the second subplot of Figure1 can be seen in Figure 2 and Figure 3. (in the case of the cylinder surfaces the matrices to be subtracted are hardly visible due to the large size of the scan (2000 by 2000 points) and the small sizes of the measurement errors, especially for the low outliers.

4. Background Subtraction

As in the case of the combustion engine cylinder surfaces the background is a cylinder of very large diameter compared to both the honing pattern and the size of the scanned surface element, as a first approach we fitted a second order polynomial to the matrix of the scanned surface. This approach was effective in some cases, sometimes, higher order polynomials were necessary. There were more surface scans, however, where this method resulted in even more complicated background pattern. Clearly, polynomial fitting could not be used in the case of the bowel pictures, as the background is way too complicated for a polynomial.

The simplest solutions for determining the background pattern seemed to be a large-scale averaging throughout the complete picture. Of course, in the case of the colonoscopy pictures the black frame should not be counted into the average, as it does not belong to the valuable part of the image, thus we applied the low-mask again. The resulting background is surprisingly near the ideal one, its effect can be seen in the first subplot of Figure 4. In the case of the combustion engines, the changing of the order of the process made the detection of the outliers simpler. The resulting picture can be seen in the second subplot of Figure 4.

5. Structural Entropy as a Basis of Classification

The purpose of this image processing was to determine, whether it can improve the classification efficiency of our scheme

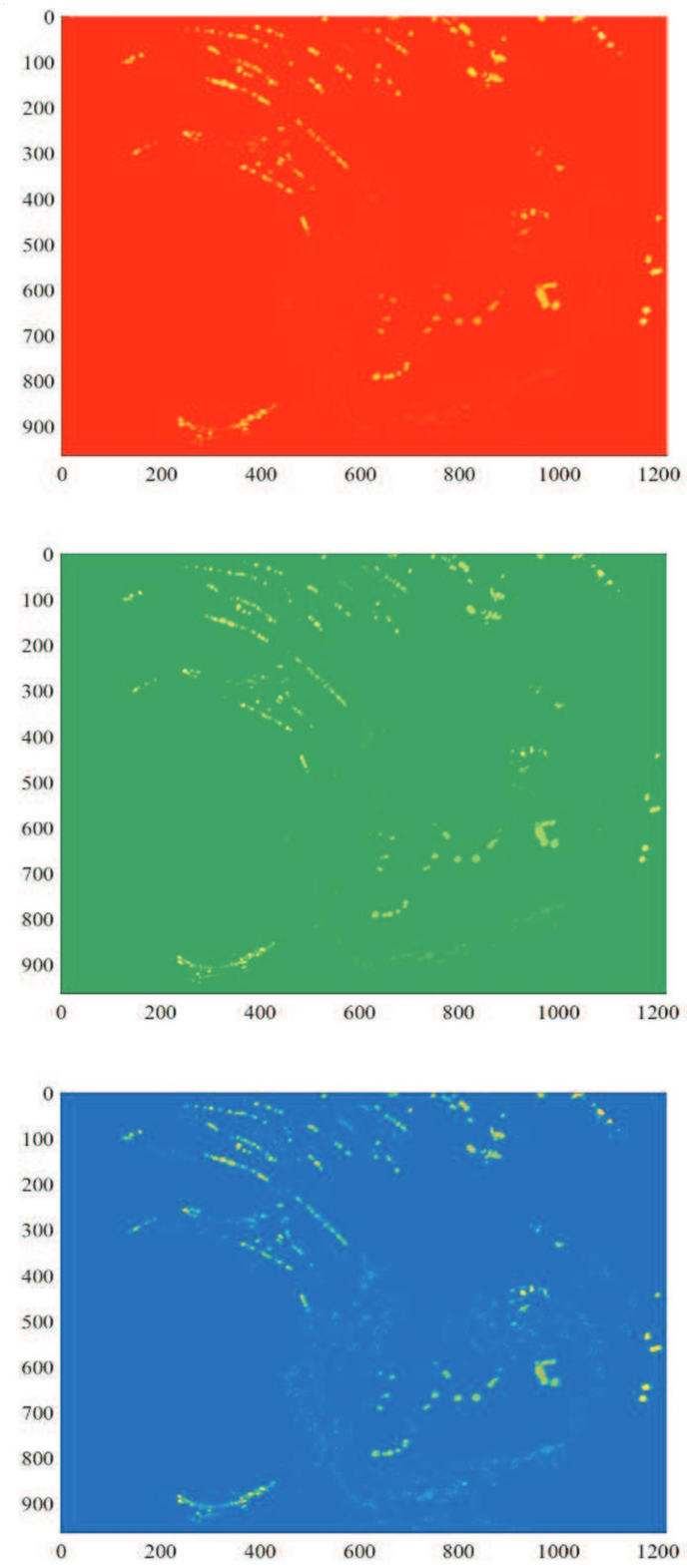

Figure 2. Matrix-to-be-subtracted of the second picture of Figure 1 (picture 116 from ETIS-Larib [8]) fo the three colour channels R, G, and B

Figure 3. The corrected picture after reflection subtraction for a colonoscopy image, for the 2nd picture of Figure 1

[10,11]. For the engine surfaces we started to develop a classification [12], where the wear of the surface could be determined from the scan, while for the colonoscopy images determining the polyp content was the goal. In the case of both data types we used a fuzzy classification scheme with structural entropy as one of the antecedents.

Structural entropy-based characterization was first used in quantum mechanics [13], [9-11], then it was adapted scanning microscopy images of semiconductor device parts [12]. Later it was found to characterize the surface roughness of gold catalyzers well [13].

These structural entropy quantities are defined using the generated entropies of Rényi. The *n*th entropy can be written as

$$S_n = \frac{1}{1-n} \ln \left(\sum_{i=1}^{N} I_i^{\,n} \right) \tag{1}$$

Here I_i means the normalized pixel intensity, and N the number of pixels. The pixel intensities must fulfill the requirements of being a probability distribution, i.e., all intensity should be nonnegative and its sum should be 1.

Pipek and Varga studied two of Rényi entropy differences

$$S_{str} = S_1 - S_2, \text{ and } -\ln q = S_0 - S_2, \tag{2}$$

and found that these quantities together describe the shape of the distribution. The quantity q is known in quantum mechanics, it is the spatial filling factor.

If $S_{str}(q)$ is plotted for an image, it can be told from the plot, what type of localization does the intensity have, i.e., if the picture contains mainly exponentially decreasing components, its $S_{str}(q)$ point will be near the curve belonging to exponential distributions.

6. The Processing Steps' Effects on the Structural Entropy Maps

The structural entropies moved to very low (10^{-4}) values in the case of the combustion engines but remained along the same path for both the new and the worn surfaces. The newworn classification is not improved, moreover, in many locations it became less effective.

In the case of the colonoscopy images, the structural entropies also decrease in magnitude, however, the processed pictures usually perform slightly better than the original pictures.

In both cases the localization types remain very slow, Gaussian, or even slower.

7. Conclusion

A method for subtracting the outliers and the background for two types of 2D data is presented. Using the histogram of the images, the lowest and highest value points, that are sparsely populated are cut off, and marked as outliers on masks. The masks are used for excluding these outliers from the averages when the new values corresponding to the masked points are calculated.

Background subtraction is based on large-scale averaging of the data proved to be more effective than polynomial fitting even in the case of regular surfaces such as combustion engine cylinder surfaces. The processing makes the image features more distinguishable to the naked eye, however, for computational intelligence-based classification methods, its effect is not always positive.

Acknowledgement

The authors would like to thank the financial support of the project EFOP-3.6.2-16-2017-00015, GINOP-2.3.4-15-2016-00003, and the ÚNKP-17-4 New National Excellence Programme of the Ministry of Human Capacities of Hungary. This work was supported by National Research, Development and Innovation Office (NKFIH) K108405 and K124055.

References

[1] Søreide, K., Nedrebø, B. S., Reite, A., et al. (2009). Endoscopy Morphology, Morphometry and Molecular Markers: Predicting Cancer Risk in Colorectal Adenoma, *Expert Rev. Mol. Diagn*, 9, 125-137.

[2] Rácz, I., Jánoki, M., Saleh, H. (2010). Colon Cancer Detection by 'Rendezvous Colonoscopy': Successful Removal of Stuck Colon Capsule by Conventional Colonoscopy, *Case Rep. Gastroenterol*., Volume 4, Karger, 2010, p 19–24.

[3] Jass, J. R. (2006). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features, *Histopathology*, Volume 50, Wiley, 113–130.

[4] Kudo, S., Hirota, S., Nakajima, T., et al. (1994). Colorectal tumours and pit pattern. *J Clin Pathol*, 47, *p* 880-885.

[5] Kudo, S., Tamura, S., Nakajima, T., et al. (1996). Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointest Endosc*, 44, 8-14.

[6] Kudo, S., Rubio, C.A., Teixeira, C.R., et al. (2001). Pit pattern in colorectal neoplasia: endoscopic magnifying view. *Endoscopy*, 33, 367-373.

[7] Bernal, J., Sanchez, F. J., Vilariño. F. (2012). Towards Automatic Polyp Detection with a Polyp Appearance Model, *Pattern Recognition*, 45, 3166-3182.

[8] Silva, J. S., Histace, A., Romain, O., Dray, X., Granado, B. (2014). Towards embedded detection of polyps in WCE images for early diagnosis of colorectal cancer, *Int J Comput Assisted Radiology and Surgery*, 9, 283-293.

[9] Bernal, J., Sanchez, F. J., Fernández-Esparrach, G., Gil, D., Rodrígez, C., Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, *Computerized Medical Imaging and Graphics*, 43, 99-111.

[10] Nagy, Sz., Lilik, F., Kóczy, L. (2017). Entropy based fuzzy classification and detection aid for colorectal polyps, *IEEE Africon* 2017, Cape Town, South Africa, 15-17, September.

[11] Nagy, Sz., Sziová, B., Kóczy, L. T. (2018). The effect of image feature qualifiers on fuzzy colorectal polyp detection schemes using KH interpolation - towards hierarchical fuzzy classification of coloscopic still images, accepted for publication at Fuzz IEEE 2018, Rio de Janeiro.

[12] Solecki, L., Nagy, Sz. (2016). Wavelet Analysis and Structural Entropy Based Intelligent Classification Method for Combustion Engine Cylinder Surfaces, *In*: Proceedings of the 8th European Symposium on Computational Intelligence and Mathematics, ESCIM, 5-8th October 2016, Sofia, 115-120.

[13] Pipek, J., Varga, I., Universal classification scheme for the spatial localization properties of one-particle states in finite dimensional systems, *Phys. Rev. A*, Volume 46, APS, Ridge NY-Washington DC, 1992, 3148—3164.

[14] Varga, I., Pipek, J. (2003). Rényi entropies characterizing the shape and the extension of the phase space representation of quantum wave functions in disordered systems, *Phys. Rev. E*, Volume 68, APS, Ridge NY-Washington DC, 026202.

[15] Molnár, L. M., Nagy, Sz., Mojzes, I. (2010). Structural entropy in detecting background patterns of AFM images, Vacuum, Volume 84, Elsevier, Amsterdam, 2010, 179-183.

[16] Bonyár, A., Molnár, L. M., Harsányi, G. Localization factor: a new parameter for the quantitative characterization of surface structure with atomic force microscopy (AFM), *MICRON*, Volume 43, Elsevier, Amsterdam, 2012, 305-310.