

Improving the Quality of Electronic Cleansing of Colorectal CT Images Using a Hybrid Method

¹Hossein Beigi Harchegani, ²Hamid Moghaddasi, ³Mahtab Shaebani, ⁴Hamid Beigy,
³Babak SalavatiPour

¹Department of Health Information technology, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

²Associate Professor of Health Information Management & Medical Informatics, Department of Health Information technology, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

³Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

⁴Department of Computer Engineering, Sharif University of Technology, Tehran, Iran

³Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
moghaddasi@sbmu.ac.ir

ABSTRACT

Introduction: Colorectal cancer, as one of the most important fatal cancers, is caused by the lack of timely diagnosis of colorectal polyps. Presently, because of the advancements in CT imaging of the colorectal device, the CTC-CAD is a promising method for the duly diagnosis of these appendages. In this regard, Electronic Colon Cleansing (ECC) is one of the effective factors that enhance diagnostic accuracy in the methods used in CTC-CAD. To date, various methods have been utilized for ECC (e.g., the mosaic decomposition (MD) method) that each has advantages and limitations. Therefore, the aim of this study is to combine the methods of linear computing of previous studies and also some image processing methods to improve the quality of electronic cleansing of data on residual materials existed in CT colorectal images. This proposed method is called LM_ECC.

Method: In this study, to implement ECC, the thresholding method, statistical functions, and image processing methods were combined. Then, to evaluate the proposed method, 22 images were randomly selected and ranked by seven radiologists. Regarding the extent of the interpretable, the images taken before and after ECC were collected using MD and LM_ECC methods. The concordance of concordance of all three categories of opinions was calculated based on Kendall's tau-b correlation coefficient test. Next, the average of the ranked opinions obtained for the main images and the results of the LM_ECC method, as well as for the MD and the LM_ECC method, were included in two T-tests.

Findings: The value of t-test between the mean score of radiologists' opinions for the main images and the results of the LM_ECC method ($p < 0.001$) is -9.355, while it is -5.414 between the mean score of radiologists for the MD results and the results obtained from the LM_ECC method ($p < 0.001$).

Conclusion: Based on the coefficient of concordance, it is found that there is a high agreement between the ranked opinions of the radiologists, based on which the results of the T-tests show the significant effect of the LM_ECC method on electronic cleansing compared to the main images and the results of MD method. Therefore, it can be concluded that the LM_ECC method is able to improve the quality of electronic cleansing of colorectal CT images.

Keywords: Electronic Colon Cleansing, CAD, Polypeptide diagnosis, CT Colonography

1 Introduction

Given the death rate of 57,000 per year due to colon cancer in the United States and recent advances in colorectal imaging by Computerized Tomography (CT) technology, the CT Colonography (CTC) is known a tool for detecting intestinal cancer [1-6]. CTC is a CT imaging from the abdominal cavity, which is done to detect the colorectal polyps [4,6]. CTC images are utilized in computer-aided diagnostic systems. CTC-CADs consist of three main parts: patient preparation before imaging, standard imaging, and soft computing on images for diagnosis [7-9]. All diagnostic techniques of CTC-CADs require those images that do not have confusing data [4,10,11] because the remainder of the residual materials from colon can be misinterpreted as a part of the colon. Therefore, it would result in increasing the false positives and subsequently reducing accuracy. Today, the electronic cleansing of colorectal CT images is considered a promising technique to remove the residual material in CTC images for the purification of the virtual cleansing after imaging [4,10].

The first and most basic solution proposed for identifying the confounding data is the thresholding method. The methods in this regard were introduced based on the use of statistical image features, vector quantization (in order to dimension reduction), image gradient information, and the classification of the Markov random field [12-17]. Further methods have focused on using the edge modeling during image categorization to effectively describe the labeled areas. Afterward, the image gradient was used in the later methods using a Sobel mask filtering [18-21]. Then, more complex and effective algorithms with several carefully designed steps were proposed. These efficient methods utilize the effective features of the images and the combination of several highly accurate categorization methods. For this purpose, Cai et al. (2011) presented the mosaic decomposition (MD) method. According to the report, the sensitivity of 97.1%, the specificity of 85.3%, the accuracy of 94.7%, and AUC = 0.96 can be obtained in the classification of areas containing residual material [10], which are approximately good results for ECC [4].

Many of the methods proposed for ECC employ the nonlinear computing. Using this computation method may lead to heavy processing and subsequently increasing the run time. Unlike these methods, thresholding is a very simple method with linear computing; however, it contains many challenges. Firstly, thresholding does not eliminate the Partial Volume Effect (PVE), because the voxels of air and residual material are categorized incorrectly when using this method. Therefore, it has an inconsistent effect on segmentation. Secondly, as the thresholding method is sensitive to any range of intensities, a slight change in the threshold value results in a change in the segmentation result, especially the shape of the intestinal surface. Thirdly, thresholding increases the rippling effects of the intestinal tract. Thus, a sharp boundary between the colon and the internal colon space is created, which means the removal of

the mucous membrane and the mucous membrane is the key to the discovery of the polyps, and its removal is very unsatisfactory [4].

The mentioned methods that utilized the nonlinear computing such as the MD method have obtained relatively good results, but other methods can also be presented to improve the quality of electronic cleansing image using linear computing methods such as thresholding and solving its challenges. The aim of this study is to provide a combination of linear methods and some image processing methods in order to improve the quality of electronic cleansing of CTC images.

2 Method

2.1 Procedure

The electronic cleansing method of proposed CTC images, called the linear method (LM)_ECC, can be divided into two main groups of labeling and cleansing (Fig. 1).

LM-ECC			
step	Method used	purpose	Grouping
1	Thresholding on the whole image	Determining the initial ROIs	labeling
2	Four-neighbor method	Removing bony ROIs	
3	Thresholding on the selected ROIs	Removing non-frontier residual material and identifying the margins of areas containing residual material	Cleansing
4	Eight-neighbor method	classifying the margins and removing the residual material	
5	Gaussian filter	smoothing the sharp edges of the mucus	

Fig. 1. Procedures for implementation and their purposes

Step 1 (thresholding on the whole image)

For the first part of the labeling, the thresholding method is used based on relation (1).

$$T(x,y) = \begin{cases} 1. & I(x,y) > 1500 \\ 0. & I(x,y) \leq 1500 \end{cases} \quad (\text{relation 1})$$

Step 2 (The four-neighbor method)

First, the thresholding method was implemented and then the selected neighboring sites were investigated for differentiation of the residual material from the bones having the same HU values. To show the differentiation of the contrast or residual material from the rectangular bone, for each continuous region, the values in $T(x,y)$ are plotted on $I(x,y)$, where each of them is a region of interest (ROI). After obtaining all ROIs for each of them, they all are investigated such that if one of the four ROIs illustrated in Fig. 2 shows the air data, it means that the ROI is related to the residual material; otherwise, it is the bone and therefore should be removed from the selected ROIs.

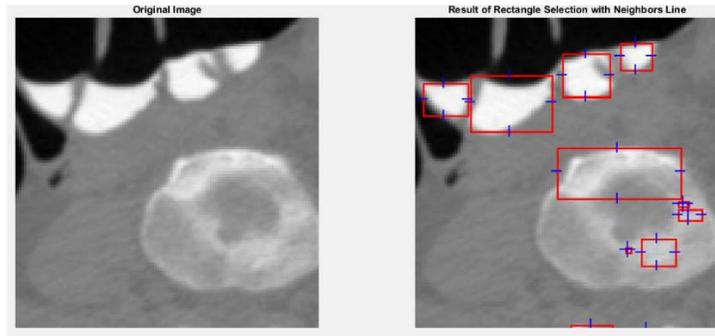


Fig. 2: ROIs and Neighborhood Indicators

In order to reduce the effect of boundary values in exploring the quadratic neighborhood, first, the outlier data (Fig. 3) are eliminated. Then, the average of other ROI values is calculated. Eventually, the average of the values inside the area of interest, after removing the outliers, was considered as the internal criterion while a value with a suitable distance for the external data was considered as the external criterion.

$$\mu = \frac{\sum_x \sum_y I(x, y) * (T(x, y) > 500)}{\sum_x \sum_y (I(x, y) = 1)} \quad (\text{relation 2})$$

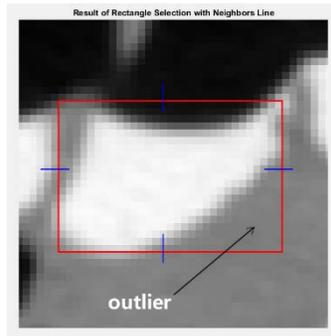


Figure 3: The outlier data in the area of interest

Step 3 (Thresholding on the selected ROIs)

After the labeling step, the values for residual materials in these areas should be properly converted to the values of the air data. To perform this correctly, the ROIs first look at the cubicles to cover the residual materials data that may not have been placed in the ROI. Next, they multiply these cubicles by a few pixels and then the data for each ROI are classified by applying thresholds based on the relation (3). Afterward, the matrix I is updated. All pixels within the ROI are investigated; a value less than 1400 is related to either air or soft tissue and thus does not require any adjustment. But, if the value is greater than 1500, the data is related to residual material, which should be converted to an air value that is close to zero. Other values between 1,400 and 1,500 represent the margins between the residual material and air, or extra material and soft tissue, which should be carefully sorted and categorized. For this reason, the matrix E is defined to hold these sides. This matrix, which has the same size as the

matrix I, is initialized with a zero number and takes a value of 1 for the values from 1400 to 1500 per ROI.

$$I(x,y) = \begin{cases} I(x,y) & I(x,y) < 1400 \\ 0 & I(x,y) > 1500 \\ I(x,y) & I(x,y) \geq 1400 \text{ and } I(x,y) \leq 1500 \end{cases} \quad (\text{relation 3})$$

Step 4 (The eight-neighbor method)

After calculating relation 3, the matrix E contains the boundaries of residual material with soft tissue or air, which should be carefully classified. To do so, the eight-neighbor method is utilized. For all values of one in the matrix E, the neighbors are examined in eight directions. If air data is encountered in the opposite directions, all data between them are converted into air, which results in eliminating the margins between air and residual material with high precision.

Step 5 (Gaussian filter)

In reviewing the eight neighborhoods, if the values in the opposite directions were not related to the air data, it means that the adjacent edge is between the residual material and the soft tissue, which is smoothed by applying the gossip filter to the sharp edges of the mucus. The results can be seen in Fig. 4.

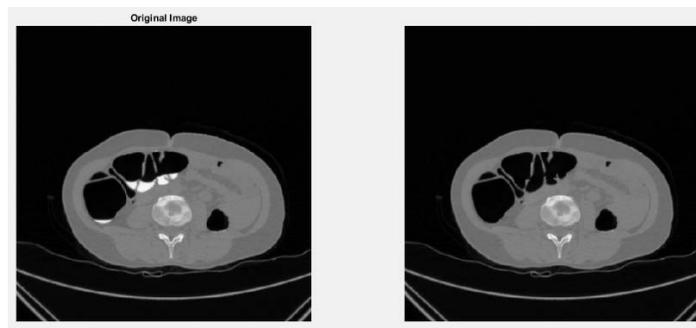


Fig. 4: Initial CTC image and the same image after doing the LM_ECC electronic cleansing method

2.2 The evaluation method

First, based on relation (4) with P-value = 0.05, 22 images were selected randomly among the 35 sets of CTC images (average annual performance) of patients who referred to an educational/therapeutic center and performed the CTC [22].

$$n = \frac{z^2_{(1-\frac{\alpha}{2})} P(1-P)}{d^2} \quad (\text{relation 4})$$

These images were taken separately by seven radiologists and then they were asked to rate the quality of these obtained images based on the following values.

1. Cannot be interpreted, due to the presence of residual material
2. Can be interpreted, but heavily influenced by residual material
3. Can be interpreted, on average influenced by residual material
4. Can be interpreted, very little under the influence of residual material
5. Can be interpreted, unaffected by residual material

Then, both LM_ECC and MD methods were implemented in MATLAB R2017a and executed on the same images, the images from these two methods, along with the main images. By keeping the name of the applied method secret, they were provided separately to seven radiologists and then they were asked to re-rank the cleansed images according to previous criteria. The results of the image ranking before and after cleansing performed by MD and LM_ECC method were imported in the SPSS software version 23. Afterward, Kendall's tau-b test was utilized to measure the agreement of radiologists.

The T-test was done to compare the effect of the LM_ECC method on the quality of the main images cleansed compared to the quality of the main images. The pre-test was done before the cleansing and post-test were done after cleansing. The dependent variable is defined as the mean of the opinions of the radiologists. The variable values were 1 to 5, with a larger number representing the higher quality of the images. The null hypothesis means that there is no significant difference between the mean of the scores obtained from the opinions of radiologists in the pre-test and post-test. In contrast, the alternative hypothesis means that there is a significant difference between the tests.

3 Results

Usually, CTC images contain data from residual materials that have been in the patient's colon during the imaging. To perform the electronic cleansing of CTC images, the proposed method first should be able to label those places where the residual materials are accumulated. Then, it should remove these sites so that the data related to the residual material are completely removed. Second, the data on the edges of the colorectal mucosa adjacent to these materials will not be affected by the changes because the correct values of these data are very effective in raising the accuracy of the diagnosis of colorectal challenges. As shown in Fig. 5, HU values in the CTC start at zero for air and will end with a value of about 2000 for fluid and bone.

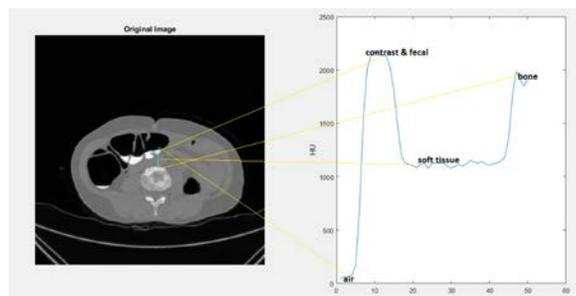


Fig. 5. HU range for air, residual material, soft tissue and bone

As shown in Fig. 5, in the use of the LM_ECC method for labeling the residual materials, the same HU values for the residual materials and bone is the biggest difficulty in separating these two types of data and applying the thresholding. Ignoring this point in the next CTC-CAD systems including the segmentation and diagnosis of polyps is the source of many errors. Thus, handling this challenge has a critical effect on the accuracy of subsequent operations. The distinction between bones and the residual materials is the non-adjacent or adjacent to air. It has to be noted that the residual materials are always in the adjacent of the air data.

In reviewing this neighborhood, there is also another problem in which the regions in the discussion do not have a completely clear boundary. Therefore, one should be considered as a value within the ROI range and then should be compared with a certain amount at the depth of the proximity, based on which the choice of these two values can be a source of some challenges for the efficiency of the proposed method.

To examine the agreement of radiologists in ranking the images, Kendall's coefficient of concordance was used based on the opinions of seven radiologists. The results of this coefficient of concordance for the main images, MD-cleansed images, and images cleansed by the LM_ECC method are presented in Tables 1, 2, and 3, respectively. For all values reported in Tables 1 and 2, the p-value is less than 0.001, while for the values of Table 3, the p-values are greater than 0.001, as represented in parentheses.

Table 1: Results of Kendall's coefficient of concordance of Radiologists' opinions for Main

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4	Radiologist 5	Radiologist 6	Radiologist 7
Radiologist 1	-	0.806	0.788	0.827	0.782	0.74	0.864
Radiologist 2	-	-	0.907	0.708	0.818	0.698	0.797
Radiologist 3	-	-	-	0.696	0.757	0.715	0.766
Radiologist 4	-	-	-	-	0.793	0.61	0.784
Radiologist 5	-	-	-	-	-	0.593	0.784
Radiologist 6	-	-	-	-	-	-	0.649
Radiologist 7	-	-	-	-	-	-	-

Table 2: Results of Kendall's coefficient of concordance of Radiologists' opinions for cleansed Images from MD

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4	Radiologist 5	Radiologist 6	Radiologist 7
Radiologist 1	-	0.88	0.728	0.784	0.779	0.784	0.766
Radiologist 2	-	-	0.874	0.791	0.829	0.853	0.839
Radiologist 3	-	-	-	0.6	0.783	0.694	0.656
Radiologist 4	-	-	-	-	0.584	0.632	0.696
Radiologist 5	-	-	-	-	-	0.741	0.651
Radiologist 6	-	-	-	-	-	-	0.649
Radiologist 7	-	-	-	-	-	-	-

Table 3: Results from Kendall’s coefficient of concordance of radiologists’ opinions for cleansed images by LM_ECC

	Radiologist 1	Radiologist 2	Radiologist 3	Radiologist 4	Radiologist 5	Radiologist 6	Radiologist 7
Radiologist 1	-	0.708	0.578(0.011)	0.454(0.102)	0.71	0.639(0.007)	0.552(0.064)
Radiologist 2	-	-	0.84	0.815(0.001)	0.777	0.914	0.723(0.010)
Radiologist 3	-	-	-	0.692(0.003)	0.591(0.003)	0.746	0.620(0.013)
Radiologist 4	-	-	-	-	0.644(0.003)	0.748(0.002)	0.881(0.008)
Radiologist 5	-	-	-	-	-	0.678	0.579(0.016)
Radiologist 6	-	-	-	-	-	-	0.667(0.012)
Radiologist 7	-	-	-	-	-	-	-

The results of the Paired Samples T-test are reported in Table 4 to evaluate the effect of the LM_ECC method on the quality of the main images cleansed as compared to the quality of the main images based on the average of the radiologists' opinions.

$$t(21) = -9.355 . p < 0.001 \quad \text{(relation 5)}$$

Table 4: Paired samples T-test for the main images and the LM_ECC method
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Original Image & LM_ECC Method	-2.140727	1.073373	.228844	-2.616634	-1.664820	-9.355	21	.000

The results of the paired samples T-test (Table 5) were used to compare the effect of the LM_ECC method on the quality of the main images and compare the effect of the MD method on the quality of the same images, based on the average of the radiologists' opinions.

$$t(21) = -5.414 . p < 0.001 \quad \text{(relation 6)}$$

Table 5: Paired samples T-test results for MD and LM_ECC method
Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 2	MD Method & LM_ECC Method	-.584273	.506199	.107922	-.808709	-.359837	-5.414	21	.000

4 Discussion and Conclusion

An important feature of the LM_ECC method is to take advantages of the various previous methods and to combine them with image processing techniques. For instance, the simplicity of linear computing based on statistical features, enhanced accuracy of the proposed method in the diagnosis and labeling of areas with residual material using image gradient information and linear methods, the edge detection methods for removing the labeled areas, and image processing filters are used to smoothen the boundary between residual material and soft tissue.

In reviewing the agreement between the radiologists' opinions using Kendall's tau-b coefficient of concordance, the obtained results for their opinions about 22 images before and after the cleansing by the MD method and the method LM_ECC are presented in Tables 1, 2, and 3 respectively. It can be seen that the minimum coefficient of concordance is 0.593, 0.584, and 0.454, respectively, while the maximum coefficient of concordance is 0.907, 0.88 and 0.881, respectively. Clearly, in all contrasting views of radiologists in Tables 1 and 2, p-value <0.001 and the Kendall value is greater than 0.5, indicating the agreement of the seven radiologists in rating the images. However, in Table 3, only two items obtained a p-value > 0.05, which is close to 10% with regard to the 21 Kendall values. It indicates the agreement of most radiologists in ranking the cleansed images using LM_ECC.

According to the T-test results presented in Table 4, to investigate the effect of the LM_ECC on the quality of the main images cleaned up compared to the quality of the original images, there is a significant difference between the pre-test and the post-test. The results of this test, as shown in relation (5), with a degree of freedom of 21, reveal that the scores of electronic cleansing are significantly higher than before the cleansing. Therefore, according to test results, the null hypothesis is rejected. This means that the LM_ECC method has a significant effect on the quality of images cleansing than the main images.

Based on the results of the T-test to compare the effect of the LM_ECC method on the quality of the main images compared to the MD method on the quality of the same images (Table 5), it is clear that there is a significant difference between the pre-test and the post-test. The results in this test (relation 6), with the degree of freedom of 21, show that the scores of electronic cleansing data are significantly higher than the MD method using the LM_ECC method. Therefore, according to test results, the null hypothesis is rejected. This means that LM_ECC has a significant effect on the quality of images cleansing as compared to the MD method.

Finally, based on the computational results of the tests, it can be concluded that the LM_ECC method, which benefits the advantages of linear methods and methods of image processing neighborhood analysis to improve the quality of electronic cleansing images, is able to improve the quality of these images in practice compared to main images and the MD method.

ACKNOWLEDGMENTS

This research is a part of the Ph.D. thesis at Shahid Beheshti University of Medical Sciences and has been approved by the ethics committee by the code of IR.SBMU.REC.1396.164.

Conflict of Interest

There are no conflicts of interest.

REFERENCES

- [1] Moghaddasi et al., "Application of Virtual Reality in Colonoscopy", *Journal of Health Informatics & Management*, 2018. 2(1).1000109
- [2] Bellini, D., et al., Bowel preparation in CT colonography: Is diet restriction necessary? A randomised trial (DIETSAN). *European radiology*, 2018. 28(1): p. 382-389.
- [3] "Cancer Facts and Figures," A. C. Society, Ed., ed, 2003.
- [4] *Virtual Colonoscopy - Technical Aspects: InTech China*, 2011.
- [5] Y. H. e. al., "Computer-aided Diagnosis Scheme for Detection of Polyps at CT Colonography," *Radiographics*, vol. 22, pp. 963- 979, 2002.
- [6] L. Li, et al., "An image segmentation approach to extract colon lumen through colonic material tagging and hidden Markov random field model for virtual colonoscopy."
- [7] R. SL, et al., "Patient preferences for CT colonography, conventional colonoscopy and bowel preparation," *Am J Gastroenterol*, pp. 578–585, 2003.
- [8] L. PA, et al., " Dietary fecal tagging as a cleansing method before CT colonography 2002;224. ," *Radiology*, vol. 224, pp. 393–403, 2002.
- [9] D. J. Vining, et al., "Technical feasibility of colon imaging with helical CT and virtual reality," *Ann. Meeting Amer. Roentgen Ray. Soc*, p. 104, 1994.
- [10] e. a. Cai W., "An Electronic cleansing method for inhomogeneously tagged regions in noncathartic CT colonography.," vol. 30, pp. 559-574, 2010.
- [11] e. a. Wang Z., "An improved electronic colon cleansing method for detection of colonic polyps by virtual colonoscopy," *IEEE Transaction on Biomedical Engineering*, vol. 53, pp. 1635-1646, 2006.
- [12] S. Lakare, et al., "Electronic colon cleansing using segmentation rays for virtual colonoscopy."
- [13] Y. Zhang, et al., "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Medical Imaging*, vol. 20, pp. 45-57, 2001.
- [14] D. Chen, et al., "A novel approach to extract colon lumen from CT images for virtual colonoscopy," *IEEE Trans. Medical Imaging*, vol. 19, pp. 1220-1226, 2000.
- [15] R. Leahy, et al., "Applications of Markov random fields in medical imaging," *Information Processing in Medical Imaging*, pp. 1-14, 1991.

- [16] K. Held, et al., "Markov random field segmentation of brain MR images," IEEE Trans. Medical Imaging, vol. 16, pp. 878-886, 1997.
- [17] L. Li, et al., "Segmentation of MR brain images: a self-adaptive online vector quantization approach," SPIE Medical Imaging, vol. 4322, pp. 1431-1438, 2001.
- [18] Z. Liang, et al., "Parameter estimation and tissue segmentation from multispectral MR images," IEEE Trans. Medical Imaging, vol. 13, pp. 441-449, 1994.
- [19] S. Lakare et al., "3D Digital Cleansing Using Segmentation Rays", IEEE Visualization, pp. 37-44, 2000.
- [20] e. a. Skalski A., "Colon cleansing for virtual colonoscopy using non-linear transfer function and morphological operations.," IEEE international workshop on Imaging systems and techniques, pp. 1-5, May 2007.
- [21] Z. M.E., et al., "Digital subtraction bowel cleansing for CT colonography using morphological and linear filtration methods.," IEEE Transactions on Medical Imaging, vol. 23, pp. 1335-1343, 2004.
- [22] Cowan, G., Statistical data analysis. 1998: Oxford university press.