

Visual Features for Endoscopic Bleeding Detection

ABSTRACT

Aims: To define a set of high-level visual features of endoscopic bleeding and evaluate their capabilities for potential use in automatic bleeding detection.

Study design: Experimental study.

Place and Duration of Study: Department of Computer Architecture, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, between March 2014 and May 2014.

Methodology: The features have been established and evaluated on images extracted from 135 different endoscopic examinations. A set of measures were proposed in order to evaluate the quality of the features. The potential contribution of the features for bleeding detection is evaluated by utilizing an SVM classifier on the vectors constructed by manually assigning the features values to the endoscopic images. Potential limitations of the features' implementations are simulated by introducing noise to the assigned values. Finally, the recognition efficiency of the method is compared to implementations of two bleeding detection algorithms from the literature.

Results: The experiments show a potential of bleeding detection efficiency of 95%, assuming perfect features' implementations, and 90% when features' evaluation efficiency of 95% is assumed.

Conclusion: By introducing the visual features, the challenging problem of detecting bleeding in endoscopic images is split into a set of smaller image processing problems. The divide-and-conquer approach based on the proposed visual features is capable of achieving high detection efficiency; at the same time providing additional benefits of a more apparent processing scheme, and a possibility of justifying the decisions in a human-readable form.

Keywords: computer vision; object detection; automatic classification; decision support systems; artificial intelligence

1. INTRODUCTION

Despite continuous research in the field of supporting endoscopic examinations by means of automatic analysis and interpretation of the videos, still no single method capable of accurate bleeding detection was commonly accepted and included into medical diagnostic procedures. The application of image processing techniques is currently limited to identification of red color areas in capsule endoscopy, by an algorithm referred to as Suspected Blood Indicator. The algorithm is integrated with the software provided with the capsules. Despite its availability and ease of use, it was reported to achieve low accuracy of analysis [1] and have limited clinical use [2].

Therefore, alternative solutions are being developed by researchers. Some successful systems were already presented [3-5]; however, the bleeding detection efficiency and overall quality is still not sufficient. One of the issues is that the majority of designed systems do not

attempt to explain nor justify the established diagnosis recommendation. Moreover, as earlier investigated [6], the existing methods mostly rely on computing sparse, low-level image features and interpreting them using trained classifiers. Initial segmentation steps are rarely performed, as well as contextual analysis of the image. The typical drawbacks of the common approach can be therefore described as:

- Weak relevance of low-level features to the actual appearance of bleeding
- High complexity of the features and classifiers, resulting in the decision process falling out of human control and understanding
- No possibility of finding causes of misclassifications
- Consequently, there are no means of identifying particular features that should be included to the system in order to improve the efficiency

In order to address the difficulties, this work attempts to reflect the process of analyzing the image by a human expert. It is here assumed, that it is a process of following numerous clues, which were earlier learned during the training. The clues are inferred from identifying visual differences between images from different classes, or given by the supervisor in the form of a description. In both cases, the clues can be verbally described by using colors, shapes, texture characteristics, context information, etc. Additionally, during the learning process, the trainee establishes how strongly given clues lead to a particular decision. Finally, after the training is completed, classifying an unknown image is a process of establishing which clues are applicable for given images and estimating the overall probability of an image belonging to a particular class.

In this paper the clues leading to the decision in the cognitive process are modeled by introducing a set of high-level visual features of blood, these values are expected to imply presence or absence of bleeding by a regular classification. The features were identified and defined by observing endoscopic images presenting bleeding. In the following sections of the article the considered set of endoscopic images is firstly described, followed by a detailed description of the proposed visual features, and the process of their development. Next, a simulation of an algorithm based on the features is evaluated in terms of efficiency, and finally the conclusions are made.

2. METHODOLOGY

2.1 Endoscopic Images Set

The endoscopic images database used in the paper was acquired with the assistance of physicians from a medical university. The set consists of 74 bleeding images and 51 non-bleeding images. Each of the images has been extracted from a different endoscopic examination. Although the size may be considered small, it is related to the specificity of the endoscopic area, where it is difficult to collect large sets of images [7]. Also, collecting only a single image from one examination significantly decreases the number of images, but is required for the correctness of the analysis. Therefore, considering the aforementioned factors, the collected images are assumed to be a representative set of bleeding and non-bleeding images. For conducting the experiments, the data set was split into two subsets: set A, consisting of 36 bleeding images and 25 non-bleeding images, and set B with 38 and 26 images, respectively. The A set was later used as a training set, while the B set was used as a validation set.

2.2 The Visual Features of Bleeding

Several requirements were introduced regarding the visual features. Similarly to the clues guiding a doctor in making a decision, the features were intended to be of a high level, understandable and, possible to be described verbally. They had to include contextual dependencies based on segmentation of the image, for example describing objects with reference to the background. For the sake of simplicity the features were assumed to be boolean, meaning that the feature can be either present or absent, without any intermediate values. The features were expected to reflect concrete visual properties. Also, the implementation of a descriptor capable of verifying the presence of a particular feature could not be too complicated, and definitely simpler than verifying the presence of bleeding itself. For that reason, the features were also expected to be definable in terms of image processing techniques, providing a possibility of successful implementation. Thereby, the emphasis was put on choosing the right features with the consideration of the implementation difficulty.

Furthermore, the process of developing the features was accompanied by an estimation of the completeness of the set of features by evaluating the following statistics considering the features and the training images:

For each feature:

- Fraction of bleeding images with the feature
- Fraction of non-bleeding image with the feature

For each blood image:

- Fraction of undiscriminated bleeding images; that is, having exactly the same features as the bleeding image

Observation of the statistics significantly improved the process of choosing and defining the features. Firstly, the measures indicated whether given features are typical just for bleeding occurrences or, undesirably, for both bleeding and non-bleeding images. Secondly, the process enabled identification of particular bleeding images, which still required redefining or introducing new features. Also, it was possible to identify the bleeding occurrences being still confused with non-bleeding images, and consider that fact in the features defining process.

Finally, following the assumptions mentioned above, a set of 10 features of endoscopic bleeding were developed and precisely defined. For evaluating the statistics during the development of the features, the values of the considered features were being manually established on the set A images. The set of features is presented below along with the visual properties to be reflected and the expected implementation details. Some of the features are dependent on others, meaning that their values are to be evaluated only if the parent feature is true, otherwise they are automatically false. In order to better illustrate the idea of the features, exemplary endoscopic images for each of the features were presented in the appendix. Also, the entire process of developing the features was visualized in figure 1.

The blood color region [BIReg] Most of the bleeding cases appear as regions of characteristic shades of red color typical for intensive bleeding or fresh blood. Therefore the color information is the strongest clue for detecting bleeding. The blood color can be defined as a narrow range in RGB or HSV color space that is often present in bleeding regions and at the same time rarely occurring in non-bleeding images. The detected area must be of significant size, exceeding approximately 1% of the clearly visible part of the image. The areas of the image are to be detected by dividing the image into blocks, evaluating the average value for the block (color) and assigning it a positive or negative state of the considered feature. The adjacent positive blocks can be merged into regions, which finally imply the presence or absence of the feature.

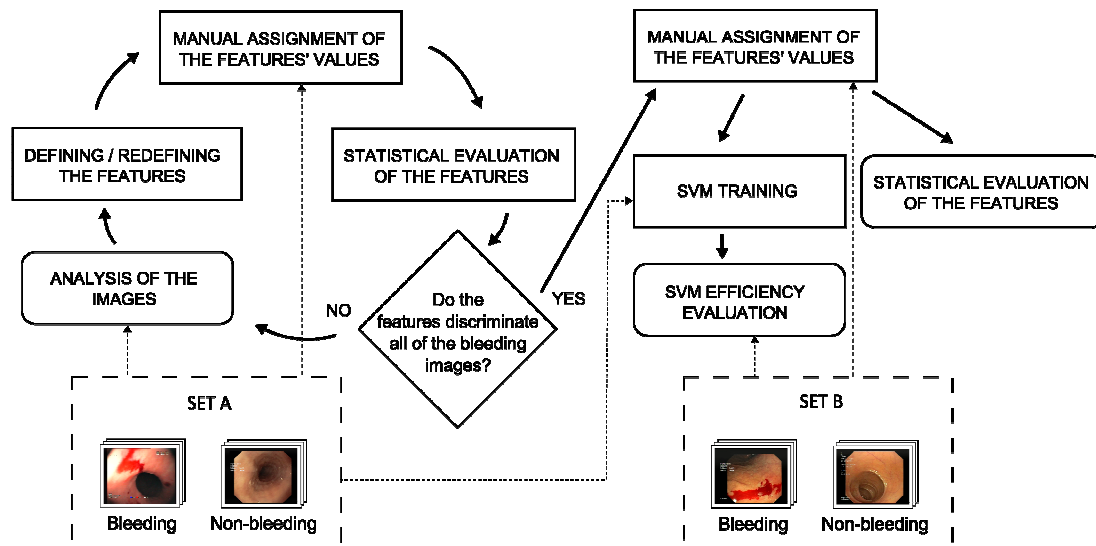


Fig. 1. Flowchart of the process of defining and evaluating the proposed features

It is important that the blood color regions feature is definitely not equivalent to the actual presence of bleeding. There exist bleeding images where the typical blood color is absent, as well as non-bleeding images, where blood color can be found. Therefore, the inclusion of additional features is necessary.

Blood color region with a smooth surface [BIReg-S] In case the blood region feature is present, occurrence of a significant amount of blood is probable. In that case, blood can fully cover the surface of the organ. This results in the occurrence of a compact, smooth blood surface. The feature value is set to true if at least one of the regions detected by the [BIReg] feature has a smooth surface. The smoothness can be measured using texture analysis techniques and compared against an experimentally tuned threshold.

The blood region has a clear boundary with an area of a different color [BIReg-C]

A common characteristic of bleeding regions is a clear, sharp boundary separating the blood area from the normal tissue. The feature value is set to true if at least a certain part of the blood region's boundary is a clear edge separating it from the adjacent area. However an additional condition is that the adjacent area, or at least its region close to the considered edge, must be of similar lighting conditions to the neighboring blood color region. The purpose is to ensure that the bleeding area and the adjacent area lie on the same organ surface, thus eliminating the edges resulting from corrugation of the organ, which is obviously not related to the occurrence of bleeding. The lighting conditions can be evaluated by analysis of the lightness component of HSL color space, then the edges can be detected and measured in the valid area.

The red color region [RdReg] Some of the bleeding cases appear as regions of a shade of red, which is less typical for blood and also appears commonly in non-bleeding images. In order to capture those cases, the feature is supposed to identify red regions similarly to the [BIReg] feature, but with a wider range of shades of red being accepted.

Red color region with a smooth surface[RdReg-S] The feature is related to the previous feature and has a similar motivation as the [BIReg-S] feature. The parameters of the processing might be, however, changed as a result of tuning towards specificity of red regions.

165 **The red region has a clear boundary with a different color [RdReg-C]** The feature is
166 [BlReg-C] equivalent for [RdReg] feature. Similarly as for the previous feature, some
167 parameters may differ due to adjustment to the red regions.

168
169 **Dispersed blood color [DspBl]** Small amounts of blood can appear as small, spread areas,
170 spots or thin blood streams. In contrast to the [BlReg] feature, large regions are ignored here
171 and the color is to be validated on a pixel level. However, some larger areas of blood
172 contaminated by other fluids or bubbles can be also detected, as they do not form
173 continuous areas of the blood color. Again, a narrow range of red color is to be defined to
174 describe blood.

175
176 **The dispersed blood lays on a blood vessel [DspBl-V]** The areas detected by feature
177 [DspBl] may be related to blood vessels, which can have the color of blood, but since they
178 are a normal endoscopic finding, they should not be reported as requiring attention.
179 Therefore it is necessary to consider this case. The feature is supposed to be true, when all
180 of the [DspBl] areas are lying on blood vessels. The implementation can be based on
181 detecting potential vessels near the considered areas by applying edge detection
182 techniques. When a net of connected edges are detected with a tree structure it can be
183 considered as a blood vessel network.

184
185 **Red color domination [RdDom]** An additional clue for analyzing the image is provided by
186 assessment of a dominant color in the image. The feature value is set to true if the dominant
187 color of the image is red, that is red color covers more than 50% of the image, excluding
188 unclear parts of the image resulting from highlights or darkness. Red color is considered as
189 a domination of red component in RGB color space, allowing a wide range of shades of red.

190
191 **Image is blurred [Blurd]** Blurriness significantly affects image analysis, hindering the
192 identification of bleeding areas. It also affects the evaluation of the remaining features.
193 Therefore the blurriness of the image is also evaluated as a feature. Blurriness can be
194 evaluated by detecting edges in the image and measuring the variance of the colors in the
195 image. High variance with a low number of edges is an indication of possible blurriness.

196 197 **3. RESULTS AND DISCUSSION**

198 199 **3.1 Features Statistics**

200
201 As mentioned before, the set of features was designed basing on the medical clues on
202 detecting bleeding and the observation of statistics over the set A images, without
203 considering the B set, which was reserved for the purpose of validation. Therefore, the B set
204 was not included into the research until the definitions of features were completed. Then,
205 exactly as for the A set, the values of the features were manually assigned to the set B
206 images.

207 In tables 1 and 2 the final values of the feature statistics over the images set were
208 presented, representing the A set and the B set, respectively. The evaluated statistics
209 include the measures introduced in section 2.2 and are denoted as follows:

210

211	$B(F_i=1)$	fraction of bleeding images with the feature present
212	$N(F_i=1)$	fraction of non-bleeding images with the feature present
213	TB	number of bleeding images with the feature present
214	FB	number of non-bleeding images with the feature present
215	TN	number of bleeding images with the feature absent
216	FN	number of non-bleeding images with the feature absent

217

Table 1. Statistics acquired for the training set

Feature	B($F_i=1$)	N($F_i=1$)	TB	FB	TN	FN
[BIReg]	0.92	0.04	33	3	1	24
[BIReg-S]	0.83	0.04	30	6	1	24
[BIReg-C]	0.67	0.00	24	12	0	25
[RdReg]	1.00	0.40	36	0	10	15
[RdReg-S]	0.69	0.36	25	11	9	16
[RdReg-C]	0.61	0.08	22	14	2	23
[DspBI]	0.58	0.12	21	15	3	22
[DspBI-V]	0.00	0.12	0	36	3	22
[RdDom]	0.83	0.64	30	6	16	9
[Blurd]	0.17	0.20	6	30	5	20

Table 2. Statistics acquired for the test set

Feature	B($F_i=1$)	N($F_i=1$)	TB	FB	TN	FN
[BIReg]	0.89	0.04	34	4	1	25
[BIReg-S]	0.87	0.04	33	5	1	25
[BIReg-C]	0.71	0.00	27	11	0	26
[RdReg]	1.00	0.54	38	0	14	12
[RdReg-S]	0.97	0.50	37	1	13	13
[RdReg-C]	0.89	0.00	34	4	0	26
[DspBI]	0.37	0.31	14	24	8	18
[DspBI-V]	0.00	0.31	0	38	8	18
[RdDom]	0.71	0.65	27	11	17	9
[Blurd]	0.24	0.23	9	29	6	20

In addition to evaluating the statistics, also undiscriminated bleeding cases in the images sets were identified. For the A set, the final values of the features obviously resulted in no undiscriminated images, since this statistic was constantly observed during designing the features, and it affected the time of finalization of the features development. Therefore, more important is the number of undiscriminated bleeding images in the B and A+B sets, where also no bleeding images were left undiscriminated.

The results show that the most crucial is the blood region feature, which was found to be present in 67 out of 74 bleeding images, and at the same time in only 2 out of 49 non-bleeding images. Therefore, the implementation of this particular feature can remain a challenge, but definitely is not as difficult as bleeding recognition itself.

3.2 Classification Efficiency Evaluation

The second phase of evaluation aimed to estimate the potential bleeding detection efficiency of the proposed set of features. For this purpose the features were put into a regular classification scheme by treating them as a feature vector of 0/1 values and subjecting them to a classification process using an SVM classifier. Since utilizing manually evaluated values of the features can be treated only as a simulation of the algorithm operation, as in this case the values of the features are perfectly assigned, also a simulation of the real implementation's limitations was performed by introducing a certain amount of random noise into the feature values. Thus, except for the algorithm based on originally assigned values of the features, also variants with 5, 10, 15 and 20 percent of randomly distorted values were evaluated. For each amount of noise 10 test were performed and an average was evaluated. Furthermore, we also compared the algorithm to two reference algorithms, which are implementations of existing bleeding detection algorithms from the literature. The reference algorithms are described in more detail in section 3.2.2. The results are presented in table 3.

3.2.1 Training and testing

The first step consisted of training and evaluation of the classifiers on the A set. (by a leave-one-out cross-validation.) This included tuning of the classifiers' parameters and, in the case of the reference algorithms, other parameters of the algorithms. The parameters optimization was achieved by performing multiple training and evaluating of the algorithm over the set A. Each training and evaluation process over the set A was performed by a leave-one-out cross-validation scheme. Therefore, since the data set consisted of approximately one image per patient, the cross-validation procedure is relevant to the most restrictive leave-one-patient-out variant [8,9]. Next, evaluation on the B set was conducted, using the classifiers trained on the A set. In both steps efficiency was evaluated as a harmonic mean of sensitivity and specificity achieved over the considered images set, which is a close equivalent of the f-measure (f-score) commonly used in automatic classification.

Table 3. Bleeding detection efficiency results of the reference algorithms, the proposed algorithm with original values of the features (F) and the algorithm variants with given amounts of noise (F(05) - F(20))

Algorithm	Set A	Set B
BD01	0.82	0.74
BD02	0.76	0.57
F	0.97	0.95
F(05)	0.93	0.90
F(10)	0.88	0.82
F(15)	0.84	0.80
F(20)	0.81	0.74

3.2.2 Reference bleeding detection algorithms

Two bleeding detection algorithms designed to support endoscopic examinations of the gastrointestinal tract were implemented as the reference algorithms. The algorithms are based on the research carried out by Penna et al. [10] and Li and Meng [4]. The algorithms were denoted as BD01 and BD02, respectively. Both algorithms are capable of marking

bleeding regions in endoscopic images. The major difference between them is the presence of the artificial neural network classifier in the BD02 algorithm. Therefore, the algorithm requires a learning phase before it can be evaluated. It is also slightly less precise, since it divides the image into small blocks to be classified, while the BD01 algorithm makes decisions for every single pixel. The authors of the BD01 algorithm reported sensitivity of 92% and specificity of 88%. The BD02 algorithm was reported to achieve sensitivity of 93,4% and specificity of 94,7%.

Both algorithms have a set of parameters influencing their efficiency. However, not all of the parameters' values were provided by the authors, which prevents perfectly accurate implementation of the algorithms. To address this problem, optimal ranges of values were assigned to the parameters and an optimization procedure was performed in order to find optimal values. 11 parameters were tuned for BD01 and 4 parameters for BD02. In the case of BD01, the assumed number of parameters was quite high, however 8 of the parameters were in fact provided by the authors and they were included into the optimization only for slight tuning.

The optimization was conducted in an automatic manner using the CRS2-LM algorithm (Controlled Random Search with local mutation) [11] from the NLOpt library [12]. Parameters ranges were treated as a parameter space to be searched for an optimal solution. The efficiency measure described in section 3.2.1 was used as an objective function to be maximized. The optimization process lasted 12 hours for each of the algorithms. After numerous iterations of the optimization algorithm, optimal parameter values were found.

Finally, both algorithms mark blood in the image instead of deciding on the presence or absence of bleeding in the image. In practice, the algorithms tend to incorrectly mark small regions in areas where no blood is present, which occurs in most images. Therefore an additional tuning was performed to set optimal threshold values determining the required size of the marked bleeding regions for classifying an image as containing bleeding.

3.2.3 Discussion

The results show that the expected efficiency of the algorithms is strongly dependent on the efficiency of the features implementation. The overall result is typically lower by 0.05 than the features efficiency. Notably high algorithm efficiency is achievable, if the features are precisely evaluated. However, even when applying 20% of noise into the features' values, the efficiency was not lower than the results of the reference algorithms. Also, if the features can be implemented with accuracy of 85%-90%, a significant advancement in bleeding detection efficiency will be achieved.

Despite putting effort into optimization of the reference algorithms, none of them reached the efficiency originally reported by the authors. The reason probably lies in the difference in the datasets used, particularly in their sizes. The authors of the BD01 and BD02 algorithms used only 11 and 10 different endoscopic examinations, respectively. These numbers might imply small diversity of the images in the datasets, which could enable achieving high results. The dataset used in the presented experiment was more than 10 times larger in terms of the diversity of the endoscopic examinations. Therefore, high efficiency rates were harder to achieve.

4. CONCLUSIONS

The presented work can be summarized as a methodical discovery and definition of the characteristic visual features of the objects to be recognized, which is the occurrence of bleeding in this case. The manual identification of the features was a required step, assuming that no sufficient methods of automatic discovery of the crucial visual features of objects exist, especially when the features are expected to be meaningful and understandable, for the sake of clarity and control over the processing. The introduction of

high-level features potentially solves the investigated limitations of the common approach to bleeding detection. The features are closely related to the visual properties of bleeding in endoscopic images. Presented statistics justify the high level of completeness of the set of features. Also the simulation of a recognition algorithm based on the features achieves high efficiency rates, showing the potential contribution of the features for automatic detection of bleeding.

By introducing the features, the problem of bleeding detection was simplified in a divide-and-conquer manner. Furthermore, the possibility of verifying and testing the values of features makes the operation of an algorithm more clear and understandable, which enables explaining the decisions and identifying causes of misclassifications. Since the meaning of the features is clear for the user, their values can serve as justification of the decisions. The user can easily verify whether the values were correctly evaluated by the system, and therefore estimate the credibility of the decision. Furthermore, the system can be easily and safely extended with new features.

What is important, the approach based on high-level features can be broadened not only to recognition of other endoscopic diseases, but also to more general image processing domains, where interpreting of visual features is to be applied. The possible disadvantages, however, include the cost of features implementation, the limitations of the boolean nature of the features and finally no consideration of the combinations of the features, which can lead to designing redundant features. Future work will focus on the implementation of the presented high-level visual features and evaluation of the actual efficiency of the algorithm. It will be also compared with implementations of chosen bleeding detection algorithms described in the literature.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Signorelli C, Villa F, Rondonotti E, Abbiati C, Beccari G, de Franchis R. Sensitivity and specificity of the suspected blood identification system in video capsule enteroscopy. *Endoscopy*. 2005;37(12):1170-3.
2. D'Halluin PN, Delvaux M, Lapalus MG, Sacher-Huvelin S, Ben Soussan E, Heyries L, et al. Does the "Suspected Blood Indicator" improve the detection of bleeding lesions by capsule endoscopy?. *Gastrointestinal Endoscopy*. 2005;61(2):243-9.
3. Lee Y, Yoon G. Bleeding detection algorithm for capsule endoscopy. *World Academy of Science, Engineering and Technology*. 2011;81:672-7.
4. Li B, Meng M. Computer-aided detection of bleeding regions for capsule endoscopy images, *Biomedical Engineering, IEEE Transactions on* 2009;56(4):1032-9.
5. Pan G, Fang Xu J, A novel algorithm for color similarity measurement and the application for bleeding detection in WCE. *International Journal of Image, Graphics and Signal Processing*. 2011;3(5):1-7.
6. Brzeski A, Blokus A, Cychnerski J, An overview of image analysis techniques in endoscopic bleeding detection., *International Journal of Innovative Research in Computer and Communication Engineering*. 2013;1(6):1350-7.
7. Liedlgruber M, Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. *IEEE Reviews in Biomedical Engineering*. 2011;4:73-88.
8. Hafner M, Liedlgruber M, Maimone S, Uhl A, Vecsei A, Wrba F. Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps. *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*. 2012;1-6.

386 9. Hegenbart S, Uhl A, Vécsei A. Systematic assessment of performance prediction
 387 techniques in medical image classification: A case study on celiac disease. Proceedings of
 388 the 22nd International Conference on Information Processing in Medical Imaging, IPMI'11.
 389 2011;498–508.
 390 10. Penna B, Tillo T, Grangetto M, Magli E, Olmo G. A technique for blood detection in
 391 wireless capsule endoscopy images. 17th European Signal Processing Conference
 392 (EUSIPCO 2009). 2009;1864-1868.
 393 11. Kaelo P, Ali M. Some variants of the controlled random search algorithm for global
 394 optimization. Journal of Optimization Theory and Applications. 2006;130:253–264.
 395 12. Johnson S. The NLOpt nonlinear-optimization package. <http://ab-initio.mit.edu/nlopt>
 396
 397
 398
 399
 400

401 **APPENDIX**

402 The table presents endoscopic images that illustrate the proposed features. Both bleeding
 403 and non-bleeding cases were provided where possible.
 404

Feature	Bleeding example	Non-bleeding example
[BReg]		
[BReg-S]		
[BReg-C]		n/a

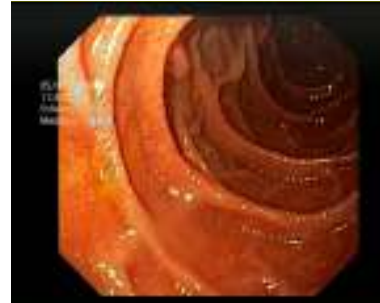
[RdReg]



[RdReg-S]



[RdReg-C]



[DspBI]



[DspBI-V]



n/a

[RdDom]



[Blurd]

