Multi-observer survey on quality enhancement techniques for still image acquisition in colonoscopy

Sebastian GROSS[1,2], Martin SCHINK[1], Thomas STEHLE[1], Alexander BEHRENS[1]

[1]Institute of Imaging & Computer Vision, RWTH Aachen University, 52056 Aachen, Germany
[2]Internal Medicine III, Universittsklinkum Aachen, 52057 Aachen, Germany

sebastian.gross@lfb.rwth-aachen.de

Abstract. *Colonoscopy, endoscopic examination of the colon, is a crucial part of the diagnostic and therapeutical process recommended for various diseases of the colon. During examination, still images are frequently taken for documentary purposes and later computer-based analysis. Still images without blur caused by movement and focus variation are difficult to capture because of limited endoscope manoeuvring control and colon peristalsis. This problem can be engaged by capturing a video sequence rather than a single image. A final sharp frame can subsequently be selected from the sequence via sharpness metrics. Three sharpness measures were evaluated: Variance, Sum Modulus Difference, and Perceptual Sharpness Metric. Alternatively, an approach called Extended Depth of Field aims at combining a refined image from the sharpness information of multiple frames. In this paper, a multi observer survey comparing the different approaches is detailed. Analysis of the survey results yields that sharpness considerations paired with performance reasons make Sum Modulus Difference the preferred algorithm for the selection of sharp images from colonoscopic video streams.*

Keywords

Colonoscopy, endoscopy, sharpness metrics, Extended Depth of Field, sharpness, blur, survey, variance, .Sum Modulus Difference, Perceptual Sharpness Metric.

1. Introduction

Colonoscopy, endoscopic examination of the colon, is a crucial part of the diagnostic and therapeutical process recommended for various diseases of the colon. A special illumination techniques called Narrow Band Imaging (NBI) [1] has been identified to possess great potential in reducing the complexity of colon polyp classification [2]. It is an exhaustive replacement for chromoendoscopy which requires laborious tasks like spraying and removing dye from polyp surfaces in its workflow. Furthermore, NBI images have been proven to provide a solid base for computer-based polyp classification [3].

However, still images without blur caused by movement and focus variation are difficult to capture because of limited endoscope manoeuvring control and colon peristalsis. Four frames taken from a scene of 0.32 seconds (eight images) are shown in Fig. 1. There is considerable movement between the four frames and the fourth image is clearly the sharpest for human observers.

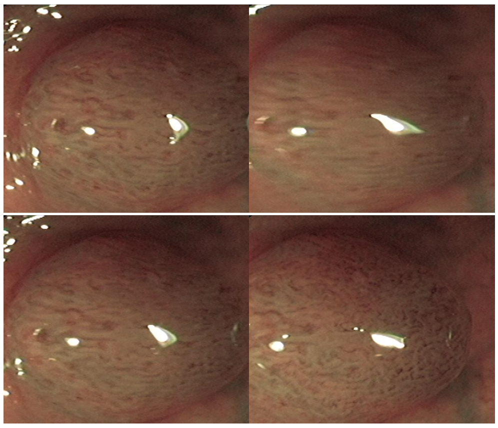

Fig. 1. Four frames from a 0.32 seconds images sequence: Top row left: frame 1, second sharpest images. Top row right: Strongly blurred image, least sharpest. Bottom row left: Third sharpest image. Bottom row right: Sharpest images, most detail.

The problem of acquiring sharp stills can be engaged by capturing an image sequence rather than a single image. A final sharp image can subsequently be selected from the sequence or be calculated by combining sharpness information from multiple images. In this paper we will introduce a multi-observer survey comparing different approaches presented in literature and discuss the results.

The rest of the paper is organized as follows. Section 2 sums up the algorithms investigated in the survey and offers references for further information. Section 3 highlights the survey setup and execution. Section 4 illustrates the survey results. Section 5 discusses the results and draws conclusions.

2. Still image acquisition

Sharpness describes the property of an image to represent edges and corners. High frequency components are necessary for such details. There is, however, no exact measurement for sharpness. Nevertheless, several metrics were published in literature in an effort to quantify sharpness. Three of those sharpness measures [4] were investigated for the purpose of enhancing still images in NBI colon endoscopy: Variance, Sum Modulus Difference, and Perceptual Sharpness Metric. Additionally, a method for combining sharpness information from multiple images into a single sharp still which is called Extended Depth of Field (EDoF) has been investigated.

2.1. Variance

Erasmus et al. [5] suggested using the sum of the image variance as a reference for image sharpness. The image is converted into gray-scale. Calculation of the sharpness metric value is performed by

$$S_{variance} = \frac{1}{NM}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x_{i,j} - \frac{1}{NM}\sum_{i=1}^{M}\sum_{j=1}^{N}x_{i,j}\right)^2 \quad (1)$$

where N is the width and M the height of the image in pixels. Single pixels are addressed as (i,j) and gray values are denominated $x_{i,j}$.

2.2. Sum Modulus Difference

Sum Modulus Difference (SMD) [6] evaluates the partial derivatives in x and y by summation of the magnitudes of amplitude differences of horizontal an vertical neighbour pixels. The sharpness metric is calculated by

$$S_{SMD} = \sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\left(|x_{i,j} - x_{i-1,j}| + |x_{i,j} - x_{i,j-1}|\right). \quad (2)$$

2.3. Perceptual Sharpness Metric

Yang et al. [7] suggested an algorithm called Perceptual Sharpness Metric (PSM) in 2006. The image is divided into blocks of 8×8 pixels in size. Each block is subjected to the Discrete Cosine Transformation (DCT)

$$DCT_{m,n}(k_1,k_2) = \sum_{i=0}^{7}\sum_{j=0}^{7} x_{i,j}\cos\left[\frac{\pi}{8}(i+\frac{1}{2})k_1\right]\cos\left[\frac{\pi}{8}(j+\frac{1}{2})k_2\right] \quad (3)$$

where $k_1 = 0..7$ and $k_2 = 0...7$ represent the indices for the coefficient of the DCT of the 8×8 sized block with the position (m,n). The coefficients in the first row and first column of the transformed block are weighted with the Sharpness Sensitivity Function $SSF(d) = d^{0.269}(-3.533 + 3.533d)\exp(-0.548d)$ which is based on studies analyzing the sensitivity of the human eye [7]. The results for each block are summed up to

$$\tilde{S}_{m,n} = \sum_{d=1}^{8} SSF(d) \cdot [DCT_{m,n}(0,d-1) + DCT_{m,n}(d-1,0)] \quad (4)$$

and the final sum for all blocks is calculated by

$$S_{PSM} = \frac{1}{M \cdot N}\sum_{m=1}^{M/8}\sum_{n=1}^{N/8}\tilde{S}_{m,n} \quad (5)$$

where the final sum S_{PSM} is normalized to the size of the image in pixels MN.

During the evaluation of PSM, poor performance was observed. However, a close look at Eqn. 4 revealed a possible cause. The weighed coefficients of the DCT are summed up, but potentially cancel each other out if the underlying cosine alters the sign. The intuitive solution to this problem is

$$\tilde{S}_{m,n} = \sum_{d=1}^{8} SSF(d) \cdot [|DCT_{m,n}(0,d-1)| + |DCT_{m,n}(d-1,0)|] \quad (6)$$

where the magnitude is calculated for each coefficient before summation. The change was implemented and both versions of the sharpness measure were evaluated. The version with the changes as indicated in Eqn. 6 is referred to as modified Perceptual Sharpness Metric (mPSM).

2.4. Extended Depth of Field

EDoF refers to the combination of depth information from multiple frames into a single image. The goal is to use pixel information from different stills to sharpen the image as much as possible. As a first step, motion of the camera and the object in between the shots has to be compensated. To this end, we apply a transformation to deform the images based on [8]. Subsequently, images are decomposed into different sub-bands using Discrete Wavelet Transform (DWT) and Stationary Wavelet Transform (SWT) [9]. The sub-band decomposition is illustrated in Fig. 2 for DWT and Fig. 3 for SWT. The information for each pixel in the corresponding sub-bands of all frames is combined. Subsequently, the inverse transforms are used to recompose the final image. Bradley and Bamford proposed a fast one-path algorithm for EDoF in 2004 [9]. A chart of the complete algorithm adapted

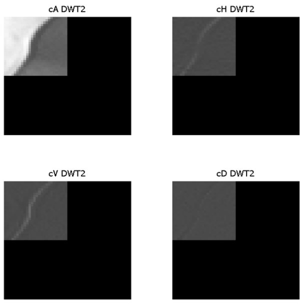

Fig. 2. Sub-band decomposition for Discrete Wavelet Transform (DWT)

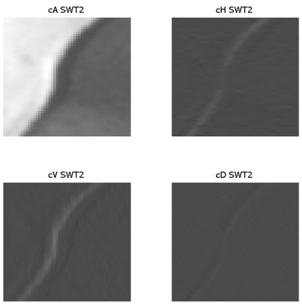

Fig. 3. Sub-band decomposition for Stationary Wavelet Transform (SWT)

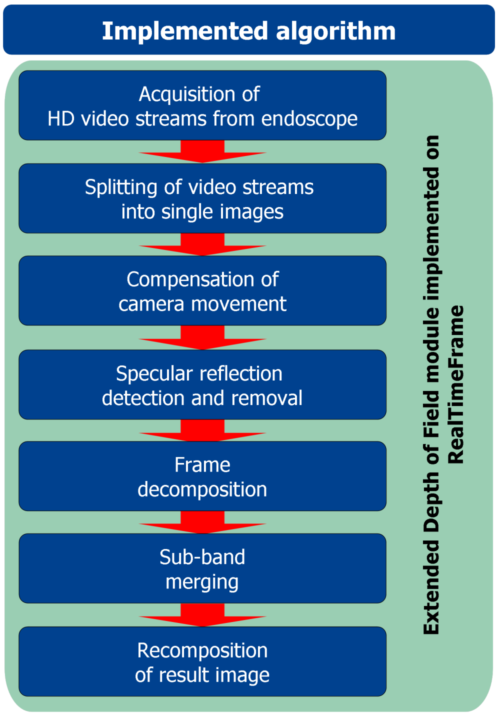

Fig. 4. Graphical representation of the Extended Depth of Field (EDoF) algorithm for colonoscopic NBI video streams used in the survey.

for the use with colonoscopic video streams [10] is depicted in Fig. 4.

Fig. 5. Frames from an image sequence generated for Extended Depth of Field algorithms

An image sequence created by Evans [11] for the purpose of applying EDoF algorithms is presented in Fig. 5. The focus varies in each of the stills. The complete series consists of eight images. The EDoF algorithm was applied using two setups. In the first setup one level of DWT and two levels of SWT were applied. The result can be seen in Fig. 6. The second setup makes use of three levels of SWT. No DWT is applied in this case. The result image is shown in Fig. 7.

Fig. 6. Results for the EDoF algorithm employing one level of Discrete Wavelet Transform (DWT) and two levels of Stationary Wavelet Transform (SWT)

Fig. 7. Results for the EDoF algorithm employing three levels of Stationary Wavelet Transform (SWT)

3. Survey setup

A survey study was designed to compare the different approaches introduced in Sec. 2. Six video scenes showing varying forms and types of tumours under NBI illumination were captured during actual colonoscopic procedures. They were each decomposed into a frame sequence with a length of 50 to 100 frames and a size of 1440×540 pixels each. Representative images are displayed in Fig. 8.

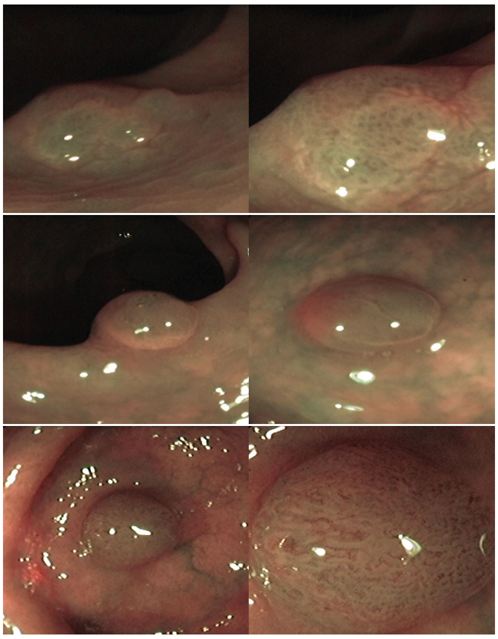

Fig. 8. Representative images taken from the six scenes used in the survey.

Videos taken during colonoscopic examinations of the colon regularly show specular reflections on the mucosa surface. These reflections do not contain any information and - even worse - may cause unpredictable corruption of sharpness metric calculations and movement compensation. Therefore, specular reflections were removed applying the algorithm described in [12].

The three sharpness metrics and the Extended Depth of Field approach as well as the specular reflection removal algorithm were implemented as plug-ins for RealTimeFrame which is a framework for rapid prototyping of video processing algorithms [13]. It offers templates for plug-ins and provides easy to use access to video streams from several configurable sources.

Subsequently, the sharpness metrics and EDoF were applied to a central image area of 576×468 pixels of the images. Each approach determined a single still image to be the sharpest of the sequence. The resulting images were mixed with two images which where selected by the survey team from the image sequence. This process was repeated for each of the six scenes.

The procedure results in seven sharp images per sequence which were offered to seven human observers during the survey. The observers were asked to rate two images in their relative sharpness at a time. Possible answers were 'Left image is sharper.', 'Both image are equal in sharpness.', and 'Right image is sharper.'. A sharper image in this direct comparison was awarded one point. In case of a draw, the points were split. For each image sequence, the points were summed up and, subsequently, normalized to the sum of all points for the sequence. Furthermore, the average and standard deviation were calculated for each set of images.

4. Results

The survey described in Sec. 3 was run and the results are listed in Tab. 1. The highest results for each sequence and the average are highlighted by underlining. The overall performance of SMD, mPSM, and EDoF averaged over all six sequences is identical.

	Variance	SMD	PSM	mPSM	EDoF
Sequence 1	<u>0.73</u>	0.54	0.60	0.41	0.14
Sequence 2	0.46	0.80	0.04	0.58	<u>0.86</u>
Sequence 3	0.00	0.63	0.46	0.38	<u>0.95</u>
Sequence 4	0.13	<u>0.64</u>	0.14	0.51	0.63
Sequence 5	0.63	<u>0.84</u>	0.03	<u>0.84</u>	0.20
Sequence 6	0.46	0.11	0.31	<u>0.81</u>	0.74
Average	0.40	<u>0.59</u>	0.26	<u>0.59</u>	<u>0.59</u>
Std. deviation	0.28	0.26	0.23	0.20	0.34

Tab. 1. Percentage of achievable points for each approach. Second lowest row: Average percentage. Lowest row: Standard deviation.

Variance is strongest for sequence 1 and above average for sequence 5. SMD and gives the best results in sequence 4 and sequence 5. Additionally, it performs well in sequence 43 and very well in sequence 2. PSM does not perform best in any of the sequences. However, the results in sequence 1 are above average. mPSM is strongest in sequence 5 and sequence 6. EDoF wins in sequence 2 and sequence 3. It also returns excellent results for sequence 6 and above average results for sequence 4 where it is a close runner-up.

Furthermore, the computational demands of the approaches were investigated. After their performance in the survey, especially SMD, (m)PSM, and EDoF were optimized for performance. The results of the performance evaluations are presented in Tabl. 2.

And indeed, the performance of the algorithms offers a different perspective. Variance and SMD are equally fast. PSM is a lot slower, but still fast enough to work as a realtime sharpness metric for video streams of 25 frames per second. EDoF is, as expected, nowhere near real-time capability. Images were acquired from the video stream and processed after all images were acquired.

5. Conclusion

The sharpness results presented in Sec. 4 collected by the survey highlighted in Sec. 3 yield no straightforward conclusion. The average performance of SMD, mPSM, and

	Variance	SMD	(m)PSM	EDoF
Sequence 1	3.77	3.59	19.21	1259.60
Sequence 2	4.06	2.82	19.84	1250.32
Sequence 3	3.92	3.88	20.28	1251.00
Sequence 4	3.65	4.67	21.37	1159.78
Sequence 5	3.78	4.06	19.68	1500.68
Sequence 6	3.92	3.35	20.05	1768.83
Average	3.85	3.73	20.07	1365.03
Std. deviation	0.15	0.63	0.73	228.41

Tab. 2. Average runtime for a single frame for each approach and each sequence. Second lowest row: Average run time over all sequences. Lowest row: Standard deviation of run times.

EDoF is identical. However, taking the performance results into account, SMD offers the sharpest results and demands least processing power. Therefore, SMD is clearly to be preferred for the task of selecting a sharp frames from an endoscopic video streams.

However, EDoF shows great potential and offers, as it is the most complex approach, huge potential for further optimization in both resulting images sharpness and processing times reduction. To this end, a possible approach based on the results of [10] would be the reduction of the number of input images for the EDoF algorithm by preselecting sharp images. This would reduce both processing time and errors uncorrected by the movement compensation.

Acknowledgements

The project was supported by the excellence initiative of the German federal and state governments.

References

[1] GONO, K., OBI, T., YAMAGUCHI, M., MACHIDA, H., SANO, Y., YOSHIDA, S., HAMAMOTO, Y., ENDO, T. Appearance of enhanced tissue features in narrow-band endoscopic imaging. In *Journal of Biomedical Optics*, 2004, vol. 9, no. 3, p. 568-577.

[2] TISCHENDORF, J. J. W., WASMUTH, H. E., KOCH A., HECKER, H., TRAUTWEIN, C., WINOGRAD, R. Value of magnifying chromoendoscopy and narrow band imaging (NBI) in classifying colorectal polyps: A prospective controlled study. In *Endoscopy*, 2007, vol. 39, no. 12, p. 1092-1096.

[3] STEHLE, T., AUER, R., GROSS, S., BEHRENS, A., WULFF, J., AACH, T., WINOGRAD, R., TRAUTWEIN, C., TISCHENDORF, J. Classification of colon polyps in NBI endoscopy using vascularization features. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2009, vol. 7260, p. 72602S 1 - 12.

[4] GROSS, S., SCHINK, M., STEHLE, T., BEHRENS, A., TISCHENDORF, J., TRAUTWEIN, C., AACH, T. Echtzeitfähige Extraktion scharfer Standbilder in der Video-Koloskopie. In *Bildverarbeitung für die Medizin 2010*, Aachen (Germany), 2010, p. accepted.

[5] ERASMUS, S. J., SMITH, K. C. A. An automatic focusing and astigmatism correction system for the SEM and CTEM. In *Journal of Microscopy*, 1982, vol. 127, p. 185-199.

[6] BRENNER, J. F., DEW, B. S., HORTON, J. B., NEIRATH, P. W., SELLERS, W. D. An automated microscope for cytologic research. In *J. Histochem. Cytochem.*, 1971, vol. 24, no. 1, p. 100-101.

[7] YANG, K., GUEST, C. C., DAS, P. Perceptual Sharpness Metric (PSM) for Compressed Video. In *Proc. IEEE International Conference on Multimedia and Expo*, 2006, p. 777-780.

[8] BEHRENS, A. Creating Panoramic Images for Bladder Fluorescense Endoscopy. In *Acta Polytechnica: Journal of Advanced Engineering*, 2008, vol. 48, no. 3, p. 50 - 54.

[9] BRADLEY, A. P., BAMMFORD, P. C. A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform, In *Image and Vision Computing '04 New Zealand (IVCNZ'04)*, 2004, pp. 279-284.

[10] GROSS, S., SCHINK, M., BEHRENS, A., STEHLE, T., AACH, T. Extraction of sharp stills during endoscopic procedures using Exteneded Depth of Field. In *Biomedica 2010*, Aachen (Germany), 2010, p. accepted.

[11] EVANS, L. Beaded Deer Project. In *http://www.lin-evans.net*, 2004, 05.02.2010.

[12] STEHLE, T. Removal of Specular Reflections in Endoscopic Images. In *Acta Polytechnica: Journal of Advanced Engineering*, 2006, vol. 46, no. 4, p. 32 - 36.

[13] GROSS, S., STEHLE, T. RealTimeFrame - A Real Time Processing Framework for Medical Video Sequences. In *Acta Polytechnica: Journal of Advanced Engineering*, 2008, vol. 48, no. 3, p. 15 - 19.

About Authors...

Sebastian GROSS was born in 1978. He studied Electrical Engineering with a focus on Information and Communication Technology at the RWTH Aachen University. Since his graduation in 2007 he has been working towards a PhD degree and is a research scientist at the 'Institute of Imaging & Computer Vision' at RWTH Aachen University. His research interests are in medical image processing, pattern recognition, and computer vision.

Martin SCHINK studies Electrical Engineering at RWTH Aachen University. The topic of his diploma thesis supervised by the Institute of Imaging & Computer Vision, RWTH Aachen University, is 'Automated Extraction of Sharp Stills from Endoscopic Real-Time Video Streams'.

Thomas STEHLE studied Computer Science at Karlsruhe Institute of Technology, Germany, and received his Diploma degree in 2005. Since then he has been working as research scientist towards his PhD degree at the Institute of Imaging and Computer Vision, RWTH Aachen University, Germany. His research interests are medical image analysis and processing, motion estimation, camera calibration and computer vision.

Alexander BEHRENS was born in 1980. He studied Electrical Engineering with a focus on Information Processing at the University of Hanover, Germany. He received his Dipl.-Ing. degree with honors in 2006. He is currently working as a research scientist toward a PhD degree at the 'Institute of Imaging & Computer Vision' at RWTH Aachen University, Germany. His research interests are in medical and industrial image processing, signal processing, pattern recognition, and computer vision.