

Open Access

Recent Development of Computer Vision Technology to Improve Capsule Endoscopy

Junseok Park^{1*}, Youngbae Hwang^{2*}, Ju-Hong Yoon², Min-Gyu Park², Jungho Kim², Yun Jeong Lim³ and Hoon Jai Chun⁴

¹Digestive Disease Center, Institute for Digestive Research, Department of Internal Medicine, Soonchunhyang University College of Medicine, Seoul, ²Intelligent Image Processing Research Center, Korea Electronics Technology Institute (KETI), Seongnam, ³Department of Internal Medicine, Dongguk University Ilsan Hospital, Dongguk University College of Medicine, Goyang, ⁴Division of Gastroenterology and Hepatology, Department of Internal Medicine, Institute of Gastrointestinal Medical Instrument Research, Korea University College of Medicine, Seoul, Korea

Capsule endoscopy (CE) is a preferred diagnostic method for analyzing small bowel diseases. However, capsule endoscopes capture a sparse number of images because of their mechanical limitations. Post-procedural management using computational methods can enhance image quality. Additional information, including depth, can be obtained by using recently developed computer vision techniques. It is possible to measure the size of lesions and track the trajectory of capsule endoscopes using the computer vision technology, without requiring additional equipment. Moreover, the computational analysis of CE images can help detect lesions more accurately within a shorter time. Newly introduced deep learning-based methods have shown more remarkable results over traditional computerized approaches. A large-scale standard dataset should be prepared to develop an optimal algorithms for improving the diagnostic yield of CE. The close collaboration between information technology and medical professionals is needed. **Clin Endosc 2019;52:328-333**

Key Words: Capsule endoscopy; Computer vision technology; Deep learning

INTRODUCTION

Since the introduction of capsule endoscopy (CE) in the year 2000, it has become the preferred diagnostic method for small bowel diseases because of its low invasiveness. However, the diagnostic yield of CE can be influenced by many factors. Several quality indicators have been suggested to standardize the methods of CE and reduce interpretation-related errors.¹ Generally, CE video sequences are reviewed after post-proce-

dural reconstruction. This process is time-consuming. In addition, there is the possibility of misinterpretation due to the limitation of human concentration.

The evolution of computer vision technology can ameliorate the diagnostic abilities of CE. Computational methods regarding modifying and interpreting CE images may reduce the image review time and error rates significantly.² Moreover, the introduction of deep learning to computer vision has resulted in the outstanding improvement of lesion recognition.^{3,4}

Since capsule endoscopes remain passive moving devices, only limited information can be obtained from their images. Many mechanical improvements to endoscopes have been studied. For safety reasons, the immediate clinical application is difficult. Advances in computer vision have allowed us to gain more details regarding the current generation of capsule endoscopes. It is possible to measure the size of lesions and predict their location more accurately than with other methods. Subsequent therapeutic procedures can be performed more systemically with this information.

This review focuses on important advances in computer

Received: October 8, 2018 **Revised:** November 18, 2018
Accepted: November 25, 2018

Correspondence: Yun Jeong Lim
Department of Internal Medicine, Dongguk University Ilsan Hospital, Dongguk University College of Medicine, 27 Dongguk-ro, Ilsandong-gu, Goyang 10326, Korea
Tel: +82-31-961-7133, **Fax:** +82-31-961-9339, **E-mail:** limyj@dumc.or.kr
ORCID: <https://orcid.org/0000-0002-3279-332X>

*These authors contributed equally to this study.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

vision technology that can be applied to CE in the deep-learning era. These advances are organized into four categories: image enhancement for improved visual quality, depth sensing for three-dimensional image interpretation, Simultaneous Localization and Mapping (SLAM) for the exact localization of capsule endoscope, and automated lesion detection for reducing review time. Technical information in this review will be explained with scenarios familiar to clinicians. This review is expected to promote active communications between medical and information technology (IT) experts.

IMAGE ENHANCEMENT FOR BETTER VISUAL QUALITY

Capsule endoscopes capture images with low lighting and limited power. Videos with low resolutions and low frame rates are transmitted wirelessly to a recorder installed outside of the human body. In addition, blur images are often captured due to capsule endoscopes' short depths. These degrada-

tions in image quality can increase the difficulty of providing accurate diagnoses. The computational processing of these images can correct the fundamental problem of CE (Fig. 1).

Noise is an inevitable problem in imaging systems. The hardware limitations of commercially available capsule endoscopes can produce noisy images that need to be fixed by post-procedural corrections. Classical noise suppression methods, including the use of bilateral filters and Gaussian blur filters, may produce erroneous and unusual CE results.⁵ The ability to reduce noise while maintaining the details of images is required for CE. Non-local means filters, adaptive median (AM) filters, block-matching and 3D filtering, and K-nearest neighbor filters have been compared in terms of their endoscopy-image correcting abilities. The AM filter, particularly, showed better results in reducing impulse noise while preserving image details than other 3 methods. Gopi et al. have proposed double density dual-tree complex wavelet transform (DDDT-CWT) methods for reducing noise of images (Table 1).⁶ These authors first converted images into YCbCr color spaces. They then applied a DDDT-CWT-based grayscale

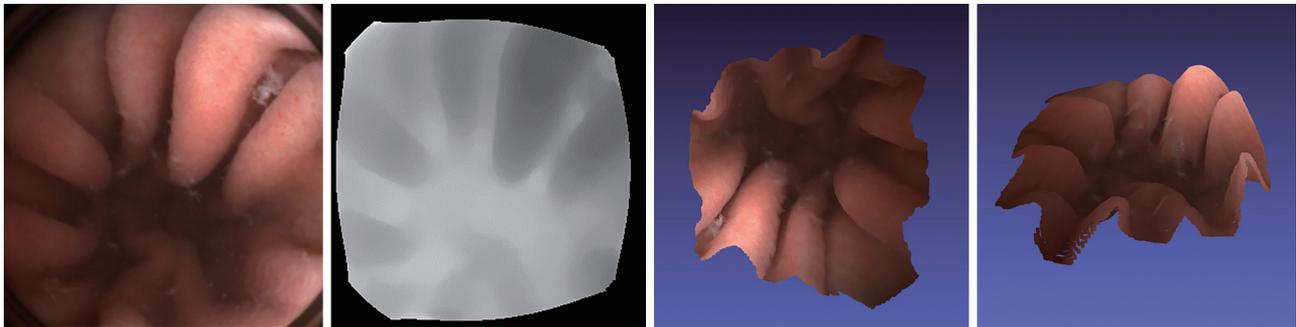


Fig. 1. Examples of deblurred capsule endoscopy images using computer vision technology. The two images on the left side with blur were obtained directly from one of the capsule endoscope's cameras. The blur of images is corrected, as shown on the right side, using depth information measured with two cameras at different angles.

Table 1. Computer Vision Technologies for the Enhancement of Capsule Endoscopy Images

Study	Suggested algorithm	Purpose	Outcome
Gopi et al. ⁶	DDDT-CWT	Noise reduction	Improved PSNR and SSIM than other three algorithms
Liu et al. ⁷	TV minimization on MFISTA/FGP framework	De-blurring	Improved PSNR for the simulation results of CE images
Peng et al. ⁸	Synthesis from DPM with aligned nearby sharp frames	De-blurring	Improved SSD errors, showing experimental result on video sample
Duda et al. ⁹	Average of upsampled and registered low-resolution images	De-blurring	Improved PSNR
Singh et al. ¹²	Interpolation function using DWT	De-blurring	Improved PSNR, MSE, and ME
Wang et al. ¹³	Adaptive dictionary pair learning	De-blurring	Improved PSNR for the dataset of CE images

CE, capsule endoscopy; DDDT-CWT, double density dual-tree complex wavelet transform; DPM, direct patch matching; DWT, discrete wavelet transform; FGP, fast gradient projection; ME, maximum error; MFISTA, monotone fast iterative shrinkage/thresholding algorithm; MSE, mean square error; PSNR, peak signal-to-noise ratio; SSD, sum of squared differences; SSIM, structural similarity index.

noise reduction method separately for each color spaces. They demonstrated the performance of DDDT-CWT by comparing the DDDT-CWT method to three other methods.

Capsule endoscopes are usually equipped with fisheye lenses that have small depths of field. Blurred images may be obtained due to fast camera motions with low frame rates and the use of the wrong lens focus. Liu et al. have introduced a deblurring method that uses total variation minimization framework and the monotone fast iterative shrinkage/thresholding technique combined with a fast gradient projection algorithm (Table 1).⁷ They demonstrated the effectiveness of this algorithm by presenting the simulation results of images that had noise and blur experimentally added to them. Furthermore, blurry video frames can be corrected by using synthesized images with references to nearby sharp frames. Peng et al. have proposed a synthesis method that follows a non-parametric mesh-based motion model to align sharp frames with blurry frames.⁸ Various endoscopic video samples with blurred frames can be sufficiently corrected with their method.

Capsule endoscopes are restricted in terms of size and data transmission bandwidth. There is a limit to applying better optical or imaging sensors to capture high resolution images. The computational resolution enhancement of images after

transmission is an efficient method for obtaining accurate diagnoses. The algorithm proposed by Duda et al. was simpler than other methods.⁹ It can be calculated in real-time.⁹ They averaged upsampling and registered low-resolution image sequences. Häfner et al. have introduced a method to prevent the over-sharpening problem that occurs in the super-resolution process and evaluated their method in the context of colonic polyp classification.^{10,11} In addition, Singh et al. have introduced a method of interpolation function using discrete wavelet transform.¹² Their algorithm showed superior results in enhancing endoscopic images over other traditional image super-resolution techniques.¹² Wang et al. have also proposed an adaptive dictionary pair learning technique.¹³ They formed the dictionary pair by selecting relevant normalized patches of high-resolution images and low-resolution images. Their method can restore the textures and edges of CE images effectively.

DEPTH SENSING FOR THREE-DIMENSIONAL INTERPRETATION

The depth of images can provide additional information about a subject. However, commonly used endoscopic imag-

Table 2. Computer Vision Technologies for Depth Sensing and Capsule Endoscope Localization

Study	Suggested algorithm	Purpose	Outcome
Karargyris et al. ¹⁴	Shape-from-shading	Depth sensing	Create three dimensional-surfaced CE videos
Fan et al. ¹⁵	SIFT, epipolar geometry	Depth sensing	Three-dimensional reconstruction of the GI tract's inner surfaces from CE images
Park et al. ¹⁶	Stereo-type capsule endoscope, direct attenuation model	Depth sensing	Create three-dimensional depth map, size estimation for lesions observed in stereo-type CE images
Turan et al. ²⁴	Vision-based SLAM, Shape-from-shading	Capsule localization	Improved RMSE for the three-dimensional reconstruction of stomach model and capsule trajectory length

CE, capsule endoscopy; GI, gastrointestinal; RMSE, root mean square error; SIFT, scale invariant feature transform; SLAM, simultaneous localization and mapping.



Fig. 2. Depth map and three-dimensional reconstruction sample of a capsule endoscope with a stereo-camera. Depth maps are calculated with capsule endoscope stereo-cameras. Bright pixels on the second image from left indicate that farther than dark ones. The depth information allows us to construct three-dimensional models of the structure, as shown the two images on the right side.

ing systems produce flat images without depth information. Depth information can be obtained by the computerized analysis of endoscopic images. Various signals, including focus, shading, and motion, can be used for depth estimation. The Shape-from-X technique is named after how it can use various types of signals for the purpose. Karargyris et al. have developed a Shape-from-Shading technique for CE (Table 2).¹⁴ They reconstructed the three dimensional-surfaced video frames of protruding features. Moreover, the Shape-from-Motion technique can take a video sequence as input and recover camera motion and geometric structures and Fan et al. have adopted this technique for constructing three-dimensional meshes through Delaunay triangulation.¹⁵

Recently developed capsule endoscopes with stereo-vision

can accurately and robustly estimate depth maps from the gastrointestinal tract. Park et al. have used a novel capsule endoscope consisting of two cameras for depth-sensing and three-dimensionally rendering intestinal structures (Fig. 2).¹⁶ They can also measure the size of lesions in a large bowel phantom model accurately.

SLAM FOR THE EXACT LOCALIZATION OF CAPSULE ENDOSCOPES

The exact location of lesions is important for determining the subsequent interventions of CE. The three-dimensional position of capsule endoscopes in the abdominal cavity can

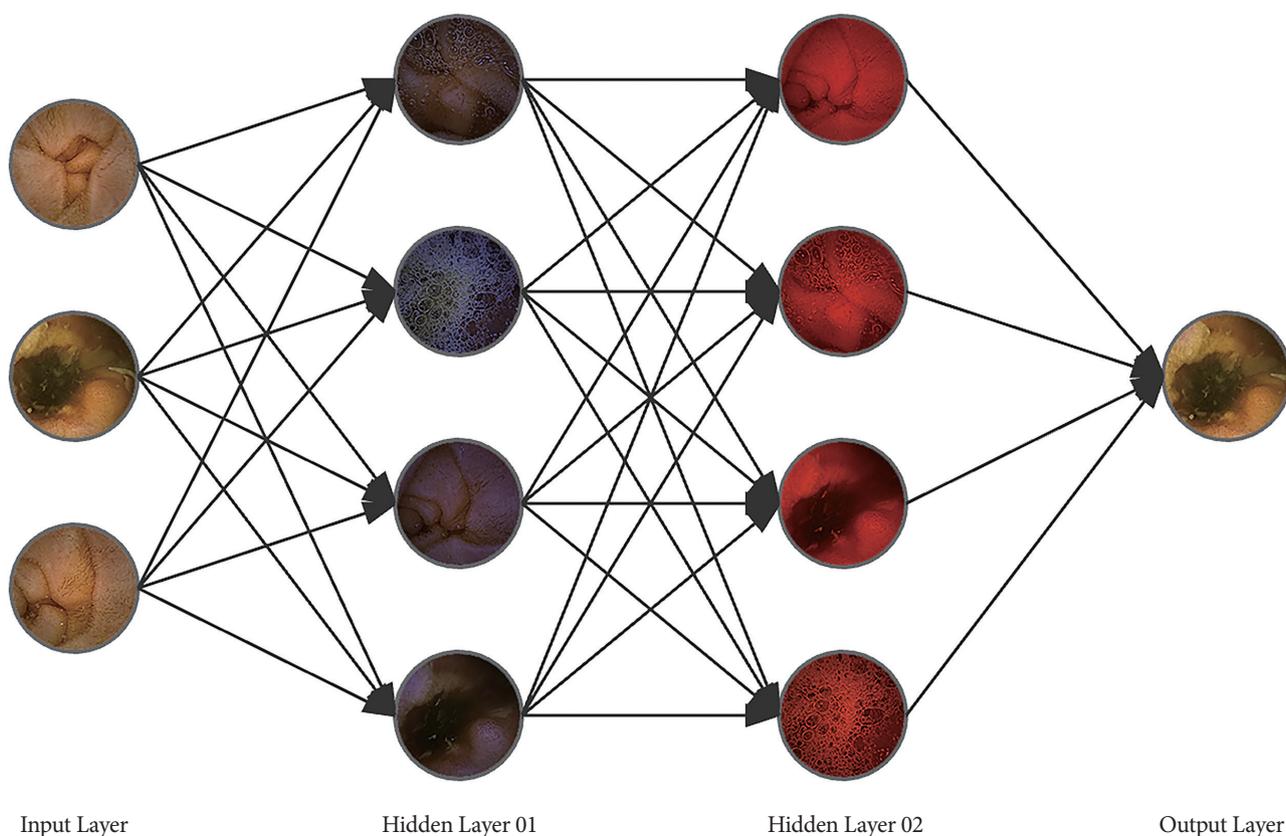


Fig. 3. Scheme of automated lesion detection for capsule endoscopy images using Deep-learning. The input images are numerically weighted via the hidden layers of large datasets. The image with the most weight is selected on the output layer.

Table 3. Deep Learning-Based Computer Vision Technologies for Analyzing Capsule Endoscopy Images

Study	No. of images for training	No. of images for testing	Outcome
Zou et al. ³⁷	60,000	15,000	Classify CE images according to the organ of origin, accuracy: 95%
Jia et al. ³⁸	8,200 (2,050 positives)	1,800 (800 positives)	Bleeding detection for annotated CE images, F1 score: 0.9955 ^{a)}

CE, capsule endoscopy.

^{a)}The harmonic average of the precision and recall, $F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$.

be obtained with the external sensor arrays of a CE system.¹⁷ However, the three-dimensional spatial position of capsule endoscopes does not represent its intraluminal location in the gastrointestinal tract. It is necessary to track the trajectory of capsule endoscopes and measure their distance from specific landmarks of the intestine in order to determine their intraluminal location. The analysis of the color and texture of images can help divide CE videos into specific regions and estimate the motion of capsule endoscopes, including their rotation and displacement.¹⁸⁻²¹

The intestine has a dynamic environment due to continuous peristalsis. Its internal surface also has many textureless regions. To overcome such circumstantial disadvantages, SLAM technology that can simultaneously perform camera position estimations and three-dimensional reconstructions can be applied. Mahmoud et al. have tracked the specific points of organs using epipolar geometry.^{22,23} The information from specific points using two different perspectives may be successfully used to reconstruct a semi-dense map of organs.^{22,23} Moreover, a recent non-rigid map fusion-based direct SLAM method has achieved high accuracy for the extensive evaluation of pose estimation and map reconstruction (Table 2).²⁴ By analyzing shapes and shades, vision-based SLAM methods can add depth information for CE images. Furthermore, the experimental results of image reconstruction have suggested the effectiveness of both looping the trajectory of capsule endoscopes and scanning the inner surface of organs.

AUTOMATED LESION DETECTION FOR REDUCING REVIEW TIME

CE image analysis requires long and insipid review times. In addition, only a small fraction of CE images contains clinically significant lesions.²⁵ These long review times can lead to high-lesion miss rates, even if interpretations are performed by well-trained professionals.²⁶ Choosing the appropriate images to review will shorten the review time and contribute to providing accurate diagnoses. However, the automatic detection of pathology using CE has long been a challenge. Recent studies regarding the analysis of the color and texture of images have shown adequate results in discovering hemorrhages and other representative lesions.²⁷⁻³²

Since the introduction of deep learning methods to computer vision, image recognition performance on large scale datasets has been greatly improved (Fig. 3). Deep learning-based image recognition technology has been applied to endoscopic image analysis and has shown surprising results in pathology detection.³³⁻³⁶ Zou et al. have analyzed 75,000 images with a Convolutional Neural Networks-based method to categorize

images into organs of origin (Table 3).³⁷ For detecting polyps and classifying normal CE images, a deep learning-based Stacked Sparse AutoEncoder method has shown improved pathology detection results for 10,000 images.³⁸ Recent works in deep-learning have shown better performance. However, deep-learning methods need large datasets to overcome the fundamental overfitting problem.³⁹

CONCLUSIONS

The computational analysis of images can improve the clinical yield of CE without the assistance of mechanical augmentation. Image enhancement techniques can correct errors and improve the quality of images, depth information can be used to measure lesions and track the movement of capsule endoscopes, and automated lesion recognition can reduce CE image review times. Moreover, the recently introduced stereo-vision capsule endoscope and deep-learning methods in computer vision can lead to the outstanding improvement of CE image analysis. Lastly, the close collaboration between medical and IT professionals would enable CE to achieve higher diagnostic yields.

Conflicts of Interest

The authors have no financial conflicts of interest.

REFERENCES

1. Shim KN, Jeon SR, Jang HJ, et al. Quality indicators for small bowel capsule endoscopy. *Clin Endosc* 2017;50:148-160.
2. Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat Rev Gastroenterol Hepatol* 2015;12:172-186.
3. Jia X, Meng MQ. Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017 Jul 11-15; Seogwipo, Korea. Piscataway (NY): IEEE; 2017. p. 3154-3157.
4. Yuan Y, Meng MQ. Deep learning for polyp recognition in wireless capsule endoscopy images. *Med Phys* 2017;44:1379-1389.
5. Obukhova N, Motyko A, Pozdeev A, Timofeev B. Review of noise reduction methods and estimation of their effectiveness for medical endoscopic images processing. In: 2018 22nd Conference of Open Innovations Association (FRUCT); 2018 May 15-18; Jyväskylä, Finland. Piscataway (NY): IEEE; 2018. p. 204-210.
6. Gopi VP, Palanisamy P, Niwas SI. Capsule endoscopic colour image denoising using complex wavelet transform. In: Venugopal KR, Patnaik LM, eds. *Wireless networks and computational intelligence*. Berlin: Springer-Verlag Berlin Heidelberg; 2012. p. 220-229.
7. Liu H, Lu W, Meng MQ. De-blurring wireless capsule endoscopy images by total variation minimization. In: *Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*; 2011 Aug 23-26; Victoria, Canada. Piscataway (NY): IEEE; 2011. p. 102-106.

8. Peng L, Liu S, Xie D, Zhu S, Zeng B. Endoscopic video deblurring via synthesis. In: 2017 IEEE Visual Communications and Image Processing (VCIP); 2017 Dec 10-13; St. Petersburg (FL), USA. Piscataway (NY): IEEE; 2017. p. 1-4.
9. Duda K, Zielinski T, Duplaga M. Computationally simple super-resolution algorithm for video from endoscopic capsule. In: 2008 International Conference on Signals and Electronic Systems; 2008 Sep 14-17; Krakow, Poland. Piscataway (NY): IEEE; 2008. p. 197-200.
10. Häfner M, Liedlgruber M, Uhl A. POCS-based super-resolution for HD endoscopy video frames. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems; 2013 Jun 20-22; Porto, Portugal. Piscataway (NY): IEEE; 2013. p. 185-190.
11. Häfner M, Liedlgruber M, Uhl A, Wimmer G. Evaluation of super-resolution methods in the context of colonic polyp classification. In: 2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI); 2014 Jul 18-20; Klagenfurt, Austria. Piscataway (NY): IEEE; 2014. p.1-6.
12. Singh S, Gulati T. Upscaling capsule endoscopic low resolution images. *International Journal of Advanced Research in Computer Science and Software Engineering* 2014;4:40-46.
13. Wang Y, Cai C, Zou YX. Single image super-resolution via adaptive dictionary pair learning for wireless capsule endoscopy image. In: 2015 IEEE International Conference on Digital Signal Processing (DSP); 2015 Jul 21-24; Singapore. Piscataway (NY): IEEE; 2015. p. 595-599.
14. Karargyris A, Bourbakis N. Three-dimensional reconstruction of the digestive wall in capsule endoscopy videos using elastic video interpolation. *IEEE Trans Med Imaging* 2011;30:957-971.
15. Fan Y, Meng MQ, Li B. 3D reconstruction of wireless capsule endoscopy images. *Conf Proc IEEE Eng Med Biol Soc* 2010;2010:5149-5152.
16. Park MG, Yoon JH, Hwang Y. Stereo matching for wireless capsule endoscopy using direct attenuation model. In: 4th International Workshop, Patch-MI 2018, Held in Conjunction with MICCAI 2018; 2018 Sep 20; Granada, Spain. Cham: Springer; 2018. p. 48-56.
17. Marya N, Karellas A, Foley A, Roychowdhury A, Cave D. Computerized 3-dimensional localization of a video capsule in the abdominal cavity: validation by digital radiography. *Gastrointest Endosc* 2014;79:669-674.
18. Shen Y, Guturu PP, Buckles BP. Wireless capsule endoscopy video segmentation using an unsupervised learning approach based on probabilistic latent semantic analysis with scale invariant features. *IEEE Trans Inf Technol Biomed* 2012;16:98-105.
19. Zhou R, Li B, Zhu H, Meng MQ. A novel method for capsule endoscopy video automatic segmentation. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2013 Nov 3-7; Tokyo, Japan. Piscataway (NY): IEEE; 2013. p. 3096-3101.
20. Spyrou E, Iakovidis DK. Video-based measurements for wireless capsule endoscope tracking. *Meas Sci Technol* 2014;25:015002.
21. Bao G, Mi L, Pahlavan K. Emulation on motion tracking of endoscopic capsule inside small intestine. In: 14th International Conference on Bioinformatics and Computational Biology; 2013 Jul; Las Vegas (NV), USA.
22. Mahmoud N, Cirauqui I, Hostettler A, et al. ORBSLAM-based endoscope tracking and 3D reconstruction. In: Peters T, Guang-Zhong Y, Navab N, eds. *Computer-assisted and robotic endoscopy*. Cham: Springer International Publishing; 2017. p. 72-83.
23. Mahmoud N, Hostettler A, Collins T, Soler L, Doignon C, Montiel JMM. SLAM based quasi dense reconstruction for minimally invasive surgery scenes. arXiv:1705.09107.
24. Turan M, Almalioglu Y, Araujo H, Konukoglu E, Sitti M. A non-rigid map fusion-based direct SLAM method for endoscopic capsule robots. *Int J Intell Robot Appl* 2017;1:399-409.
25. Faigel DO, Baron TH, Adler DG, et al. ASGE guideline: guidelines for credentialing and granting privileges for capsule endoscopy. *Gastrointest Endosc* 2005;61:503-505.
26. Zheng Y, Hawkins L, Wolff J, Goloubeva O, Goldberg E. Detection of lesions during capsule endoscopy: physician performance is disappointing. *Am J Gastroenterol* 2012;107:554-560.
27. Mamonov AV, Figueiredo IN, Figueiredo PN, Tsai YH. Automated polyp detection in colon capsule endoscopy. *IEEE Trans Med Imaging* 2014;33:1488-1502.
28. Kumar R, Zhao Q, Seshamani S, Mullin G, Hager G, Dassopoulos T. Assessment of Crohn's disease lesions in wireless capsule endoscopy images. *IEEE Trans Biomed Eng* 2012;59:355-362.
29. Chen G, Bui TD, Krzyzak A, Krishnan S. Small bowel image classification based on Fourier-Zernike moment features and canonical discriminant analysis. *Pattern Recognition and Image Analysis* 2013;23:211-216.
30. Szczypiński P, Klepaczko A, Pazurek M, Daniel P. Texture and color based image segmentation and pathology detection in capsule endoscopy videos. *Comput Methods Programs Biomed* 2014;113:396-411.
31. Fu Y, Zhang W, Mandal M, Meng MQ. Computer-aided bleeding detection in WCE video. *IEEE J Biomed Health Inform* 2014;18:636-642.
32. Iakovidis DK, Koulaouzidis A. Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. *Gastrointest Endosc* 2014;80:877-883.
33. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NY): IEEE; 2016. p. 779-788.
34. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology* 2018;155:1069-1078.e8.
35. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas (NV), USA. Piscataway (NY): IEEE; 2016. p. 770-778.
36. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
37. Zou Y, Li L, Wang Y, Yu J, Li Y, Deng WJ. Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: 2015 IEEE International Conference on Digital Signal Processing (DSP); 2015 Jul 21-24; Singapore. Piscataway (NY): IEEE; 2015. p. 1274-1278.
38. Jia X, Meng MQ. A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016 Aug 16-20; Orlando (FL), USA. Piscataway (NY): IEEE; 2016. p. 639-642.
39. Hwang Y, Park J, Lim YJ, Chun HJ. Application of artificial intelligence in capsule endoscopy: where are we now? *Clin Endosc* 2018;51:547-551.