Comparison of Different Super-Resolution Methods for HD Video Endoscopy

M. Häfner[1], M. Liedlgruber[2], A. Uhl[2]

Department for Internal Medicine, St. Elisabeth Hospital, Vienna
Department of Computer Sciences, University of Salzburg, Austria
mliedl@cosy.sbg.ac.at

Abstract. The main question we try to answer in this work is whether it is feasible to employ super-resolution (SR) algorithms to increase the spatial resolution of endoscopic high-definition (HD) images in order to reveal new details which may have got lost due to the limited endoscope magnification of the HD endoscope used (e.g. mucosal structures).
For this purpose we compare the quality achieved of different SR methods. This is done on standard test images as well as on images obtained from endoscopic video frames. We also investigate whether compression artifacts have a noticeable effect on the SR results.
We show that, due to several limitations in case of endoscopic videos, we are not consistently able to achieve a higher visual quality when using SR algorithms instead of bicubic interpolation.

1 Introduction

Throughout the past years different approaches aiming at the classification of colonic polyps have been developed. Most of these works are based on traditional endoscopes. But there also exists work based on zoom-endoscopes equipped with an optical zoom (e.g. [1]). Such endoscopes are advantageous as they allow to inspect the colonic mucosa in a magnified manner, revealing the fine surface structure of the mucosa and small lesions. However, throughout the past few years high-definition (HD) endoscopes got more and more popular. While providing a roughly four times higher image resolution as compared to many zoom-endoscopes, they are often not equipped with an optical zoom.

One possible way to unveil more details from such HD images is to use super-resolution (SR) algorithms. There already exists a work in which an SR method is applied to wireless capsule endoscopy video frames [2]. In this work the authors test their algorithm on low-resolution (LR) images generated from a single video frame by shifting it into different directions and downscaling the shifted frames. But this does not reflect a realistic application scenario.

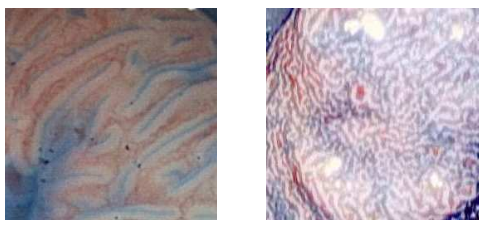
Hence, the main question we try to answer in this work is whether it is feasible to use SR algorithms to increase the resolution of endoscopic HD images and reveal new details. Figure 1 shows two tubulovillous adenoma, one captured with a zoom-endoscope and one captured with an HD endoscope without zoom. The dramatic difference between these images in terms of visible details is obvious.

While we can not expect high-resolution (HR) images generated by SR methods to be comparable to the ones obtained with zoom-endoscopes, we at least hope to be able to increase the level of detail in HD images. For this purpose we evaluate a set of SR methods on real-world HD endoscopy videos. Since these are compressed, they also suffer from compression artifacts. Thus, we also investigate the impact of such artifacts on the SR quality of the algorithms evaluated using deblocking on the video sequences. We also carry out experiments using uncompressed videos.

Throughout literature many SR algorithms are evaluated on artificially generated LR images only. That is, although real-world video test sequences are available, the respective sequences are subject to blur and downsampling to generate LR frames (e.g. [2]). These frames are then used to reconstruct an HR image. While this is a practical way to assess if an algorithm works (i.e. an accurate quality assessment is possible since the HR ground truth is available), this hardly matches real-world scenarios.

Since our aim is to reveal new details in endoscopy videos, it is not meaningful to evaluate SR algorithms on artificially generated LR frames. We thus apply SR algorithms to HD videos frames. Doing so, we face different problems:

- *Compression artifacts:* Although there exists work which specifically aims at SR for videos (e.g. [2]), these algorithms are quite often evaluated only on uncompressed sequences. Since HD videos would require a fairly high amount of storage if stored uncompressed, they are usually compressed. This comes at the price of sometimes clearly noticeable compression artifacts (e.g. 8×8 DCT blocks). To get an idea of how these artifacts influence the SR reconstruction quality, we also carry out experiments with sequences which have been deblocked using the method described in [3].
- *Lack of aliasing artifacts:* The images we used in this work show a lack of aliasing artifacts. One reason for this is that the videos are compressed. To investigate whether SR algorithms produce better results on uncompressed videos we will also examine different algorithms on endoscopic videos sequences which have not been compressed.
  In addition, since in endoscopic images there are usually no sharp edges (i.e. high frequency content) which may result in aliasing artifacts due to undersampling, it is clear that our images do not expose clearly noticeable



(a) Zoom (b) HD

Fig. 1. Illustration of the difference between two different imaging modalities.

aliasing. Another cause may be the presence of noise and blur, caused by the sensor and small camera movements.

- *Complex motion:* Since in endoscopic videos we are facing highly complex motion (e.g. position-variant transformations) simple motion models fail to describe the motion between successive HD endoscopy video frames.
  As a consequence we use an optical flow method [4]. Such methods are more versatile when it comes to the estimation of arbitrary complex motion between images. This is mainly due to the fact that optical flow methods allow to estimate local motion, while simpler methods usually work well only with global motion. In our case this task is hindered to some extent by compression artifacts and a lack of aliasing artifacts.

2 Materials and methods

To be able to assess whether SR with endoscopy images is feasible, we carried out experiments using different SR methods. These are Shift-and-Add (S&A) [5], regularized super-resolution (ROB) [6], iterated back-projection (IBP) [7], and robust super-resolution (ROBZ) [8]. Although the LR images used are color images, we apply the SR algorithms only to the intensity component in the CIELAB color space since this channel usually contributes most to textural features. The color components of the HR images are obtained by a bicubic upscaling of the first frame of the respective LR sequence.

For our experiments we evaluated the SR algorithms on two different sets of LR image sequences: widely used sequences and sequences extracted from endoscopic videos. For all sequences we always use eight LR frames and an upscaling factor of two for the SR algorithms. The endoscopic sequences are based on successive frames of videos acquired during colonoscopy sessions between the years 2011 and 2012 at the Department for Internal Medicine (St. Elisabeth Hospital, Vienna) using an HD colonoscope (Pentax HiLINE HD+ 90i Colonoscope) with a resolution of 1280×1024 pixels. To reduce the computational demand for the SR methods we chose positions from which we manually extracted 256×256-pixel patches which serve as LR images (the position remained the same in case of a single sequence). Details on the LR sequences used can be found in Table 1. As we notice from this table, experiments with two endoscopy videos without compression have also been conducted (U1 and U2). Furthermore, the sequences D1-D5 have been obtained by a deblocking of E1-E5.

Due to the lack of reference HR images in case of endoscopic images, we use two reference-free quality metrics. The first metric used is called BRISQUE [9]. It is a so-called natural scene statistics-based approach, which computes statistical features based on edge responses at different scales. Based on the features for the training set, support vector regression (ϵ-SVR) is used to learn a mapping from feature space to quality scores. The learned model is then used to predict the quality score for a new image with an unknown quality. For the training we generated a different set of sequences (similar to the ones in Table 1). This set has been rated by eight human raters. Based on these ratings, the differential

Table 1. Details on our LR image sets used (the columns "Compr." and "Debl." indicate the sequences which are compressed and subject to deblocking, respectively).

Name	ID	Color	Compr.	Debl.	Name	ID	Color	Compr.	Debl.
Carphone	R1	Y	N	N	Endoscopy 5	E5	Y	Y	N
City	R2	Y	N	N	Endoscopy 6	D1	Y	Y	Y
Container	R3	Y	N	N	Endoscopy 7	D2	Y	Y	Y
Garden	R4	Y	N	N	Endoscopy 8	D3	Y	Y	Y
Mobile	R5	Y	N	N	Endoscopy 9	D4	Y	Y	Y
Endoscopy 1	E1	Y	Y	N	Endoscopy 10	D5	Y	Y	Y
Endoscopy 2	E2	Y	Y	N	Endoscopy 11	U1	Y	N	N
Endoscopy 3	E3	Y	Y	N	Endoscopy 12	U2	Y	N	N
Endoscopy 4	E4	Y	Y	N					

mean opinion score (DMOS) was computed (using the median instead of the mean to be resistant against outlier ratings), which is used for the training.

The second metric measures the entropy within an image for different directions [10]. This is done by first computing the discrete Pseudo-Wigner distribution for an image. Then, an approximate PDF is computed, for which the pixel-wise Rény-entropy is computed. By repeating these computations for different directions and taking the mean over all entropy values for each direction considered, an anisotropic entropy measure for an image is obtained.

For both metrics a higher score indicates a higher quality. For BRISQUE the score is usually in the range 0 to 100 (some scores may leave this range due to SR reconstruction results not represented appropriately in the training set).

To assess whether the SR methods produce useful results, we also create an upscaled image from the first image of each sequence using bicubic interpolation, for which we also compute quality scores.

3 Results

Table 2 shows the results obtained. The anisotropic scores have been multiplied by 10^4 due to the small values originally yielded. According to these scores, the methods IBP and ROBZ almost always produce results of higher quality than INT. For sequences R1-R5 all SR methods are able to improve the visual quality. In case of sequences E1-E5, the methods S&A and ROB sometimes yield a score below the INT score. The overall picture is very similar for the deblocked sequences D1-D5. The combinations of SR methods and sequences, showing an improvement over INT, highly correlate with the respective combinations in case of sequences E1-E5. Interestingly, deblocking almost always lowers the scores (in case of INT as well as in case of the SR methods). For the uncompressed videos the SR algorithms sometimes yield a higher score than INT, but not always.

The BRISQUE scores are quite different. Comparing the scores for the SR results with the INT scores shows that SR rarely improves the image quality.

Table 2. Detailed metric results for the SR algorithms (INT denotes interpolation).

	Anisotropic					BRISQUE				
ID	INT	S&A	IBP	ROB	ROBZ	INT	S&A	IBP	ROB	ROBZ
R1	25.8	55.4	61.8	46.7	59.8	51.3	45.0	62.0	52.4	65.5
R2	10.6	21.6	31.2	19.8	29.9	32.6	47.1	40.3	26.5	42.8
R3	39.6	106.7	67.2	97.7	67.0	28.6	12.8	36.0	-9.3	37.5
R4	59.2	86.3	109.7	69.8	103.0	52.4	48.6	23.5	14.6	30.3
R5	37.8	49.9	66.3	39.8	63.0	38.2	52.1	23.4	-13.7	25.3
E1	75.0	68.7	105.4	117.8	94.1	49.2	45.9	31.1	32.9	41.9
E2	14.0	15.8	23.9	15.2	22.4	48.4	35.5	24.3	37.5	36.1
E3	10.1	7.7	18.9	9.4	16.7	49.8	49.0	25.1	51.4	29.3
E4	40.2	32.0	68.2	36.3	59.9	49.3	52.7	26.6	56.5	37.0
E5	90.2	86.6	124.8	83.3	118.1	51.2	36.8	28.9	41.3	37.4
D1	66.1	62.6	92.8	118.6	84.1	49.1	46.0	31.9	23.8	41.9
D2	12.9	14.4	21.1	14.1	20.5	50.2	36.0	26.3	36.8	40.6
D3	8.8	6.8	16.1	7.9	14.5	47.9	47.8	24.9	49.3	32.0
D4	34.7	29.5	59.6	32.6	51.7	48.7	51.4	28.1	56.6	39.1
D5	85.7	83.1	120.9	80.2	111.5	52.9	38.8	30.4	41.2	41.8
U1	18.1	12.8	30.8	17.9	27.2	50.6	46.3	28.9	15.6	30.6
U2	47.5	39.5	105.8	57.5	80.6	53.7	48.3	37.6	27.1	41.6

This accounts to all image sequences. Regarding E1-E5, we observe a higher score for SR in three cases only. These are the same combinations of SR methods and sequences for which a higher score is yielded by BRISQUE as compared to INT in case of the sequences D1-D5. While deblocking sometimes lowers the scores, this happens not as often as in case of the anisotropic metric. The SR scores for sequences U1-U2 are always below the scores for INT.

The bottom line is that the scores yielded by the metrics are quite different. While the anisotropic measure seems to detect an improvement by SR methods as compared to INT quite often, this is not the case for BRISQUE. According to the anisotropic measure, at least the IBP and ROBZ methods are always able to produce high quality SR results. For BRISQUE such a trend is not observable. When using deblocking the metrics behave slightly different. While the anisotropic metric almost always shows a lowered score, in case of BRISQUE about 30% of the combinations yield a higher score. Moreover, the ROBZ method always yields a higher BRISQUE score after deblocking.

4 Discussion

Our experimental results indicate that it highly depends on the quality metric used for image quality assessment whether SR algorithms seem to deliver a higher quality than a bicubic interpolation. We have also shown that a deblocking in case of endoscopic videos – at least for the sequences used in this work – has no positive impact on the visual quality of the SR outcomes. In contrast, we have shown, that the deblocking consistently lowers the quality. This may be attributed to a lack of aliasing artifacts in endoscopic images, which gets even worse when deblocking is applied (due to an additional smoothing of details). This is also supported by the fact that even in case of uncompressed videos SR algorithms seldom deliver a higher quality as compared to an interpolation.

In future work we therefore plan to conduct a study with physicians, routinely performing endoscopy, to perform subjective tests. This will provide a solid basis to assess the real quality of the SR results obtained. We will also investigate how SR algorithms affect diagnostic performance.

Acknowledgments

This work is partially funded the Austrian Science Fund (FWF) under Project No. TRP-206.

References

1. Tischendorf JJW, Gross S, Winograd R, Hecker H, Auer R, Behrens A, et al. Computer-aided classification of colorectal polyps based on vascular patterns: a pilot study. Endoscopy. 2010 Mar;42(3):203–207.
2. Duda K, Zieliński T, Duplaga M. Computationally Simple Super-Resolution Algorithm for Video from Endoscopic Capsule. In: Proceedings of the International Conference on Signals and Electronic Systems (ICSES'08); 2008. p. 197–200.
3. List P, Joch A, Lainema J, Bjontegaard G, Karczewicz M. Adaptive Deblocking Filter. IEEE Transactions on Circuits and Systems for Video Technology. 2003;13(7):614–619.
4. Liu C. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis [dissertation]. Massachusetts Institute of Technology; 2009.
5. Elad M. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. IEEE Transactions on Image Processing. 2001;10(8):1187–1193.
6. Farsiu S, Robinson D, Elad M, Milanfar P. Advances and Challenges in Super-Resolution. International Journal of Imaging Systems and Technology. 2004 Aug;14(2):47–57.
7. Irani M, Peleg S. Improving resolution by image registration. CVGIP: Graphical Models and Image Processing. 1991 Apr;53(3):231–239.
8. Zomet A, Rav-Acha A, Peleg S. Robust super-resolution. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR'01). 1; 2001. p. 645–650.
9. Mittal A, Moorthy AK, Bovik AC. No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing. 2012;21(12):4695–4708.
10. Gabarda S, Cristóbal G. Blind image quality assessment through anisotropy. Journal of the Optical Society of America A. 2007;24(12):B42–B51.