Bridging the Resolution Gap Between Endoscope Types for a Colonic Polyp Classification

M. Häfner
Department for Internal Medicine
St. Elisabeth Hospital
Vienna, Austria

M. Liedlgruber, A. Uhl, and G. Wimmer
Department of Computer Sciences
University of Salzburg, Austria
Salzburg, Austria
Email: {mliedl,uhl, gwimmer}@cosy.sbg.ac.at

***Abstract*—In this work we investigate whether cross endoscopic modality classification of colonic polyps is feasible, i.e. images of a high-magnification endoscope are used as training set to classify images of a high-definition endoscope.**

In order to compensate the scale differences between the images acquired with the different imaging modalities we apply different super-resolution methods to endoscopic high-definition sequences. We then use a set of feature extraction methods for the classification of the super-resolution reconstruction results.

To be able to assess whether super-resolution algorithms are helpful in this scenario, we also compare the results obtained from these experiments against the classification results based on original high-definition frames and against classification rates based on upscaled versions of high-definition frames.

We show that classifying images acquired with a high-definition endoscope and training the underlying classifier with images acquired with a high-magnification endoscope is feasible, but the improvements by using super-resolution algorithms are highly feature-dependent.

I. Introduction

Throughout the past years a variety of different methods for an automated classification of colonic polyps based on endoscopic images has been developed. The majority of these works is based on traditional endoscopes. But there also exists work which is based on imagery acquired using an endoscope with high magnification capabilities (e.g. [1]). One advantage of such endoscopes is that they allow to inspect the colonic mucosa in a magnified manner, thus revealing the fine surface structure of the mucosa as well as small lesions. However, throughout the last few years high-definition (HD) endoscopes got more and more popular. While this type of endoscopes provides a roughly four times higher image resolution as compared to many zoom-endoscopes, they are often not providing optical magnification.

While using HD endoscopes has many benefits, large scale databases containing this type of imagery enriched with histopathologic ground truth as required for computer supported decision support systems based on classification or retrieval are still missing. On the other hand, such databases exist for high magnification endoscopes due to their long-lasting availability. Thus, one interesting question is whether it is in general possible to classify HD images, while using high-magnification images for the training process (i.e. mixing images from different endoscope types). This is especially interesting since it is a common problem that endoscopy image databases are quite often rather limited in terms of the images available [2]. However, one problem with this approach are the scale differences between the two imaging modalities.

One possible way to deal with this problem would be to simply downscale the images in a high-magnification database by a certain factor to obtain more similar resolutions between the two modalities. This, however, would most likely result in a substantial loss of details, which is not desirable. Another possibility is to use super-resolution (SR) algorithms to bridge the resolution gap between the two modalities to some extent. For our application scenario we can potentially use multiple successive low resolution frames from a video to construct a high resolution image. While applying super-resolution algorithms successfully to endoscopic videos poses different problems [3], we chose the latter option in favor of loosing potentially important details by simply downscaling.

In recent work, SR algorithms have been evaluated on endoscopic images [4], [5]. Based on patches, extracted from HD video frames (termed low- resolution (LR) in our context), high-resolution (HR) images were created, using a set of different SR algorithms. The quality of the SR reconstruction has then been assessed using different image quality metrics. The outcome of these studies was that, at least for the quality metrics and SR methods evaluated, there is no real benefit from applying SR algorithms to endoscopic images in order to reveal new details.

While, similar to [4], [5], in this work we apply SR algorithms to HD images, the main aim of the present study is different. We compare the classification results of the SR images using a classifier trained on high-magnification images with the classification results of the original HD images using a classifier trained on the original HD images or on high-magnification images. This way we are able to determine whether it is possible and meaningful to train classifiers with images of a different endoscopic imaging modality exhibiting more mucosal detail compared to the images to be classified and if it makes sense to additionally bridge the resolution gap by means of SR. We also investigate to which extent scale invariant methods are able to lower the effect of scale differences and how these features perform in combination with SR methods (in terms of overall classification rates).

Figure 1 shows two tubulovillous adenoma, one captured with a zoom-endoscope and one captured with an HD endoscope without optical zoom. We immediately notice the dramatic difference in terms of the details visible.

(a) Zoom (b) HD (c) SR method

Fig. 1. Illustration of the difference between two different imaging modalities in (a) and (b). And a region of the same size extracted from the outcome of an SR method applied to (b) in (c).

The remaining part of this work is organized as follows: In Section II we briefly describe the SR algorithms and feature extractions methods evaluated. In Section III we describe the experimental setup used and present the results obtained. We then conclude the paper in Section IV.

II. Methods Evaluated

In the following we briefly describe the SR algorithms and the feature extraction methods which have been employed for the experiments in this work.

A. SR Algorithms

The set of SR algorithms chosen for our experiments is the same as evaluated in [3]. In addition, as indicated in [3], we face complex motion in endoscopy videos. We therefore use the optical flow estimation by Black and Anandan [6] (part of the implementation available for [7]).

In the following y_k denotes the k-th LR image from the input sequence and $\hat{X}_n$ denotes the HR estimate after the n-th iteration of the respective iteration.

- **Iterative Back Projection (IBP):** The Iterative Back Projection [8] was chosen for our experiments due to its simplicity and intuitive nature. Simply stated, this method computes the pixel-wise difference between y_i and $\hat{X}_n$ after applying the respective warp, smoothing, and downsampling. The difference image is then upsampled, followed by computing the gradient image, and warping back the gradient image to the image space of $\hat{X}_n$. The final update for $\hat{X}_n$ is obtained by summing up the gradient images pixel-wise for all y_i in and adding the resulting image, multiplied by a constant factor, to $\hat{X}_n$.
- **Robust Super-Resolution (ROBZ):** This method, proposed in [9], is basically a modification to the IBP method. Instead of summing up the single gradient images, the authors propose to compute a pixel-wise median to obtain the update weight for each pixel. By changing the IBP algorithm this way, outlier pixels are removed. Such outliers might arise, for example, due to an inaccurate motion estimation.
- **Projection Onto Convex Sets (POCS):** The key idea of POCS-based SR algorithms is to express every piece of prior knowledge about the solution as a constraint in image space. More specifically, the solution is constrained by convex sets which, according to the prior knowledge available, impose restrictions on a HR estimate in order to be a valid one. The experiments in this work are based on the POCS-approach proposed in [4] as this method has been developed in the context of endoscopic imaging.
- **Regularized Super-Resolution (RSR):** The RSR method used in this work was proposed in [10]. Since the SR reconstruction problem is an ill-posed one [11], regularized approaches aim at finding the desired HR image in the space of possible solutions by imposing one or more constraints on the SR reconstruction. The algorithm proposed in [10] is in some way similar to the IBP method described, as it also aims at minimizing the error between an observed LR image y_k and a simulated LR image. But besides a different cost function, the approach in [10] uses an additional regularization constraint to compensate for the ill-posedness nature of SR reconstruction problems. The constraint used is termed as bilateral total variation (BTV), which penalizes the total variation within an image with a spatial decaying effect.

For our experiments the initial HR estimate $\hat{X}_0$ is set to an upscaled version of y_1 in case of IBP, ROBZ, and POCS. For RSR, $\hat{X}_0$ is set to the pixel-wise mean of all LR images after registration and upscaling. In addition, in case of RSR, we use a regularized deconvolution on $\hat{X}_0$ to cope with noise and blur.

Since all the SR methods evaluated work in an iterative manner, we employ the adaptive termination criterion proposed in [4] to decide upon termination of the iterative process.

B. Feature Extraction

In this work we evaluated the following set of feature extraction methods for a subsequent classification. These methods have been chosen because of their quite different characteristics to analyze their different behaviors.

- **Local Binary Patterns (LBP):** Based on a grayscale image, this operator generates a binary sequence for each pixel by thresholding the neighbors of that pixel by the center pixel value. The binary sequences are then treated as numbers (i.e. the LBP numbers). In our case the $LBP_{8,1}$ operator has been used (i.e. eight neighbors, radius of one). Once all LBP numbers for an image are computed, a histogram based on these numbers is generated and used as feature vector.
- **Dual-Tree Complex Wavelet Transform (DT-CWT):** The DT-CWT is used with six scales and six orientations. Based on the absolute values of the detail subband coefficients, the statistical features mean and standard deviation are computed for each subband [12]. This process is repeated for each color channel. The resulting values are concatenated to obtain the final vector.
- **FRACTAL:** The method proposed in [13] is based on the computation of fractal features. After a conversion of an image to grayscale, the image is filtered using the MR8 filter bank. Subsequently the local fractal dimension for each pixel of the eight filter responses is computed. Then a bag-of-visual-words approach is

used to generate a histogram for an image. These histograms are then used as features for the classification.

- **Segmented Shape Features (SSF):** This method has been specifically designed for the classification of colonic polyps. SSF analyzes the shape of connected components (blobs) from images (after a conversion to grayscale) segmented by a variation of the fast level lines transform. But in contrast to the algorithm proposed in [14], we use a slightly modified algorithm, which prevents merging of blobs. The final feature vector of an image consists of the histograms computed from three shape features (convex hull feature, skeletonization feature, and perimeter feature) and a contrast feature extracted from the blobs.
- **MBFSI:** In order to derive multiscale blob features (MBF) [15], a series of flexible threshold planes (i.e. blurred versions of the image itself) are applied to a textured image (after a conversion to grayscale) and then the topological and geometrical attributes of the blobs in the obtained binary images are used to describe image texture. Two features are used to describe an image, the number of blobs and the shapes of the blobs. The shape features are invariant to spatial scaling within a small range, but the number of blobs changes to some extent. Hence, in order to exploit the scale invariance as a stand-alone feature, we use the scale invariant shape features only (MBFSI).
- **Edge Features (EF):** After a grayscale conversion of the input image, this method aims at finding regions which correspond to pits, as typically observed on a colonic mucosa. Based on these regions, different features are extracted [16]. In accordance to the work in [16], the experiments in this work are also based on a feature selection (the feature selection has been carried out on the high-magnification images). The feature selection yielded two different features to be used in our experiments: the mean image intensity across all pits detected and the mean irregularity across all pits detected (the irregularity for one pit is computed by dividing the maximum radius of the pit by the minimum radius).

While the DT-CWT method operates in a multi-resolution fashion, the LBP operator works on small pixel neighborhoods only. The methods EF, MBFSI, and SSF, in contrast, are specifically designed to analyze shapes. In addition, to support bridging the resolution gap, three of our methods are designed to be scale invariant (i.e. FRACTAL, MBFSI, and SSF). While, in general, the EF method is not scale invariant, the subset of features yielded by the feature selection is.

III. Experimental Setup and Results

A. Experimental Setup

Each LR sequence used in this work is based on four successive frames taken from 62 videos acquired during colonoscopy sessions between the years 2011 and 2013 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. In order to acquire the videos, 37 patients underwent endoscopy.

TABLE I. Ground truth information for the LR sequences used in our experiments.

	Non-neoplastic	**Neoplastic**	**Total**
LR sequences	19	43	62
Patients	13	35	48

TABLE II. Ground truth information for high-magnification image database used for training in our experiments.

	Non-neoplastic	**Neoplastic**	**Total**
Images	198	518	716
Patients	14	32	46

The high-magnification images, serving as training set, are based on 327 endoscopic color images (either of size 624×533 pixels or 586×502 pixels) acquired between the years 2005 and 2009 at the Department of Gastroenterology and Hepatology (Medical University of Vienna) using a zoom-colonoscope (Olympus Evis Exera CF-Q160ZI/L) with a magnification factor of 150. In order to acquire the images, 40 patients underwent colonoscopy. To obtain a larger set of images, we manually extracted subimages (regions of interest) with a size of 256×256 pixels from the original images. This resulted in an extended image set containing 716 images in total.

Lesions found during colonoscopy have been examined after application of dye-spraying with indigocarmine, as routinely performed in colonoscopy. Biopsies or mucosal resection have been performed in order to get a histopathological diagnosis. In case of the HD video sequences, the Pentax i-SCAN image enhancement has been enabled in addition to the topical staining (i.e. i-SCAN mode 3, which enhances the visibility of pit pattern and vascular features).

The ground truth for the LR sequences and, as a consequence, for the SR reconstruction results is given in Table I. One notices that we carry out a classification between non-neoplastic and neoplastic polyps. A more fine-grained classification would theoretically be possible, but this would potentially lead to rather unstable results due to the quite limited number of LR sequences available. Inside the colon of a single patient different types of lesions may develop. Since such a patient appears in more than one class, the total number of patients in Table I is higher (48) as compared to the number of patients who underwent colonoscopy (37).

Table II shows the ground truth information used for the high-magnification image set. Similar to the LR sequences ground truth, again the number of patients shown in this table is slightly higher (46) as compared to the total number of patients who actually underwent colonoscopy (40).

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 simple bicubic upscaling of the first frame from the respective LR sequence. For the classification we employ the k-NN classifier with different choices for k (i.e. $k = 1,\ldots,10$). The classification setup (i.e. different image sets for training and evaluation) will be referred to as distinct sets classification in the remaining part of this work (DS classification).

We carry out three different types of classification scenarios, based on different evaluation image sets:

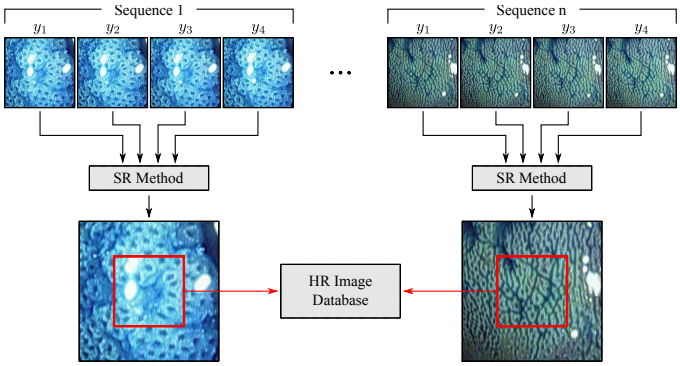

Fig. 2. Illustration of the process behind creating the HR image databases.

- **SR scenario:** We apply the SR algorithms to the LR sequences. To reduce the computational demand for the SR methods we chose positions in the original HD frames 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). For the SR reconstruction we use an upscaling factor of two (resulting in 512×512-pixel HR images). We then extract a 256×256-pixel center patches from the resulting HR images for the classification (red square in 2).
- **"Normal" scenario:** The first frame of each LR sequence is used to construct the image set for evaluation. Choosing the first LR frame can be justified by the fact that the first frame is also used as the reference frame when applying the SR algorithms.
- **"Bicubic" scenario:** While being similar to "Normal", the first LR frame of each sequence is subject to upscaling using an upscaling factor of two and bicubic interpolation. The 256×256-pixel center patches of the upscaled images are then used to construct the evaluation set (red square in 2).

It must be pointed out that, by applying the four different SR methods and by considering the "Normal" and "Bicubic" scenarios, we end up with a total of six different image databases for validation. Each of these databases is then subject to a separate classification. The process of creating a single HR image database from the video sequences available is illustrated in Fig. 2. In case of a distinct set scenario, the training set consists of the high-magnification images.

To assess whether differences in classification results are statistically significant we employ McNemar's test [17]. For two methods M_1 and M_2 this test statistic keeps track of the number of images which are misclassified by method M_1 but classified correctly by method M_2 and vice versa. Throughout this work we chose a significance level of $\alpha = 0.05$. This implies that, if M_1 and M_2 are significantly different, there is a confidence level of 95% that the differences between the outcomes of the methods are not caused by random variation.

B. Results

By investigating the results obtained, we want to answer different questions. We thus split up the result discussion in different parts in which we analyze and discuss different aspects of the results obtained.

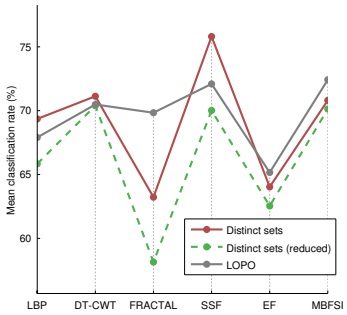

Fig. 3. Comparison of the mean overall classification rates obtained for DS, reduced DS, and LOPO (in case of "Normal").

TABLE III. MEAN CLASSIFICATION RESULTS AND THE STANDARD DEVIATIONS FOR THE DIFFERENT EVALUATION SETUPS.

	DS	DS (reduced)	LOPO	DS vs. LOPO
LBP	69.4 ± 0.0	65.8 ± 1.0	67.9 ± 2.8	✓
DT-CWT	71.1 ± 2.8	70.4 ± 2.4	70.5 ± 1.1	✗
FRACTAL	63.2 ± 2.5	58.1 ± 5.1	69.8 ± 1.9	✗
SSF	75.8 ± 1.3	70.0 ± 1.9	72.1 ± 2.4	✗
EF	64.0 ± 2.7	62.5 ± 2.9	65.2 ± 3.0	✗
MBFSI	70.8 ± 2.9	70.1 ± 6.4	72.4 ± 1.2	✓ (-)

1) Is a Classification based on Different Endoscope Types Feasible?: Figure 3 shows a comparison of the mean overall classification rates (over all choices for k) for a distinct sets classification (one time with the original high-magnification set used as training set, and one time with a reduced high-magnification image set) for "Normal" (i.e. the original LR frames have been used for evaluation). To obtain the reduced high-magnification set we selected 100 random subsets of the original images in such a way that the number of images per class corresponds to the LR image set. These subsets have then been used as training sets for 100 classification runs. The classification rates for the reduced set shown in Fig. 3 correspond to the mean rates over all 100 classification runs. In addition we carried out experiments using leave-one-patient-out cross-validation (LOPO). Both, the training and evaluation images are taken from the LR image set (with the restriction, that the training set must not contain an image from the patient, currently under classification).

It is important to recall that the number of images available in the high-magnification image set is much higher as compared to the images in the LR images set. This might also have an effect on the resulting classification rates in case of the DS classification. This is the reason why additional experiments have been conducted with the reduced high-magnification set.

From the results in Fig. 3 we notice that for some features extraction methods the mean classification rates in case of LOPO are higher as compared to the DS classification, while other features seem to yield higher classification rates in case of distinct sets classification. But in most cases the differences seem to be quite small. As expected, we also notice that the rates in case of a reduced training set are consistently lower as compared to the DS rates.

Table III shows the classification results in a more detailed fashion. The results shown in this table are the mean overall classification rates over all choices for k, along with the respec-

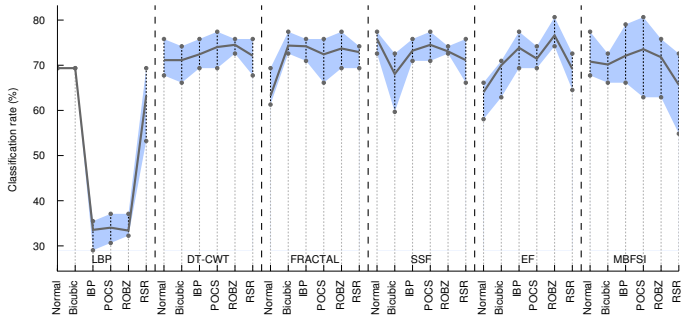

Fig. 4. Overview of the results from our experiments.

tive standard deviations. In case of the reduced image set the mean and standard deviation values have been computed from all 1000 overall classification rates (i.e. 10 choices for k times 100 classification runs). The last column in this table indicates whether, according to McNemar's test, the differences between the results produced by the different evaluation setups are statistically significant. For McNemar's test we fixed the value for k to 4, since for this value the overall classification rates are in most cases the highest. A check mark indicates a significant difference between the results of DS and LOPO and the sign given in brackets shows whether a DS result is significantly higher (+) or significantly lower (-) as compared to the respective LOPO result.

The most important thing we notice from Table III is that, despite the fact that there are differences in the results between DS and LOPO, these are not statistically significant for the majority of the features. Only in case of MBFSI and LBP we observe statistically significant differences. While in case of LBP the overall rates are nevertheless equal, MBFSI yields slightly lower results in case of DS. As a consequence, we can conclude that classifying the LR images, while using the high-magnification images for the training, is indeed feasible – at least for the feature extraction methods and image databases used. But we have also seen that the number of training images available makes a difference. Hence, the outcome that the DS results are sometimes higher as compared to the LOPO results can partially be explained by the fact that in case of DS the number of training images is much higher, which is beneficial, especially when using the k-NN classifier.

2) Can SR Algorithms Help to Bridge the Resolution Gap?: As we saw, there are almost no statistically significant differences between a DS classification and a LOPO classification. We now aim to answer the question whether we are able to improve the classification results by classifying images after applying SR algorithms.

Figure 4 gives an overview of the results from the respective experiments. In this figure the solid line shows the mean over the overall classification rates obtained with the different choices for k for one specific combination of SR algorithm and feature extraction method. The shaded area indicates the range between the minimum and maximum overall rates over all values for k for one combination. "Normal" and "Bicubic" denote the cases where the original and upscaled images are used for classification, respectively (i.e. the two scenarios as described in Section III-A). Table IV shows the results in more detail (the mean rates along with the standard deviations).

TABLE IV. MEAN CLASSIFICATION RESULTS ALONG WITH THE RESPECTIVE STANDARD DEVIATIONS (GIVEN IN PERCENT).

	Normal	Bicubic	IBP	POCS	ROBZ	RSR
LBP	69.4 ± 0.0	69.4 ± 0.0	33.5 ± 2.4	34.0 ± 2.1	33.4 ± 1.5	63.2 ± 5.2
DT-CWT	71.1 ± 2.8	71.1 ± 2.3	72.4 ± 2.1	74.0 ± 2.3	74.5 ± 1.3	72.1 ± 2.3
FRACTAL	63.2 ± 2.5	74.4 ± 1.4	74.2 ± 1.7	72.4 ± 3.0	73.7 ± 2.6	72.9 ± 1.7
SSF	75.8 ± 1.3	68.1 ± 3.7	73.2 ± 1.4	74.5 ± 1.8	73.1 ± 0.8	71.1 ± 3.1
EF	64.0 ± 2.7	70.0 ± 2.5	73.9 ± 2.6	71.5 ± 1.9	76.6 ± 2.3	69.2 ± 2.6
MBFSI	70.8 ± 2.9	70.2 ± 2.0	72.1 ± 5.2	73.5 ± 5.4	71.8 ± 4.3	65.6 ± 5.3

TABLE V. SIGNIFICANCE TEST RESULTS FOR DS, COMPARING "NORMAL" AND THE REMAINING SR METHODS (WITH $k = 4$).

	Bicubic	IBP	POCS	ROBZ	RSR
LBP	✗	✓ (-)	✓ (-)	✓ (-)	✗
DT-CWT	✗	✗	✗	✗	✗
FRACTAL	✓ (+)	✓ (+)	✓ (+)	✓ (+)	✗
SSF	✗	✗	✗	✗	✗
EF	✗	✗	✗	✓ (+)	✗
MBFSI	✗	✗	✗	✗	✗

From Fig. 4 we notice that the results are considerably lower for most SR methods in case of LBP. While in case of "Normal" and "Bicubic" the rates are equal, SR algorithms yield lower results in a consistent manner. The main problem of LBP seems to be that this operator analyzes texture in a small pixel neighborhood of a fixed size only. Thus the feature is quite sensitive to scale changes. In addition, in case of "Normal" and "Bicubic" all images are classified as neoplastic, which explains the fact that the minimum and maximum mean overall rates are equal in these cases. Hence, LBP is not suited for our DS classification. The overall picture is quite different for the other feature extraction methods evaluated.

When comparing the SR rates with the "Normal" rates for the remaining features, we see that in most cases applying an SR method yields at least slightly higher mean overall classification rates. Only in case of SSF there is no gain from a higher resolution in terms of the classification rates. In case of DT-CWT, FRACTAL, EF, and MBFSI the classification rates for HR images are mostly higher as compared to "Normal". But we also notice, that the improvement of the classification rates are rather small in case of some features, when comparing the respective SR rates with "Bicubic". This is especially noticeable for the FRACTAL feature. While for this feature the SR rates are much higher as compared to the "Normal" rate, the "Bicubic" rate is even higher than the SR rates. For all features, except for FRACTAL, there are quite often SR methods which are able to increase the classification rates. The clearly higher rates using SR or bicubic upscaling compared to "Normal" in case of the two methods FRACTAL and EF indicate that these two methods are not scale invariant, although they are designed to be scale invariant.

Table V shows which combinations of feature extraction methods and SR algorithms are able to yield a statistically significant different results as compared to "Normal". Again, the results shown have been obtained by fixing k to 4.

From this table we notice that there are only three features which deliver significantly different results as compared to "Normal". As already seen in Fig. 4, the "LBP" feature in most cases yields lower mean classification rates when applying an SR method. This can also be seen from Table V, where three out of the four SR methods yield significantly lower overall classification rates. In case of the FRACTAL feature, almost all SR methods yield significantly higher overall classification rates. However, as already mentioned earlier, even "Bicubic"

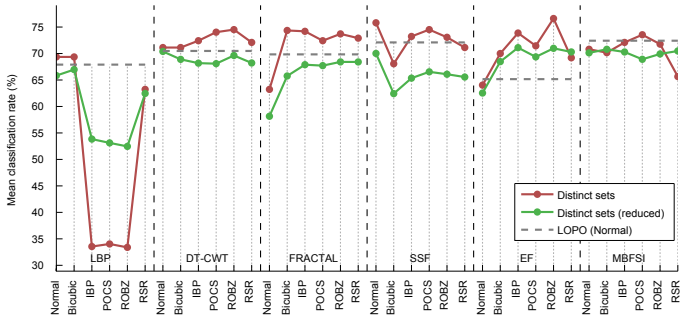

Fig. 5. Comparison of the mean overall classification rates for DS, DS reduced, and LOPO-CV for "Normal".

yields much higher mean overall rates as compared to "Normal". Thus, we can conclude for this feature that it would be sufficient to apply a bicubic upscaling and that applying an SR method in case of this feature has no additional benefit. Only the EF feature seems to really benefit from applying an SR method. Although this is only the case for the ROBZ method, we also notice from Fig. 4 that in case of this combination the mean overall classification rate is considerably higher as compared to "Normal" and "Bicubic".

3) Summary of Results: Figure 5 summarizes our experimental results. The solid lines show the mean overall classification rates obtained by DS (red) and reduced DS (green). The dashed line shows the mean overall classification rates obtained for a LOPO classification for "Normal" (i.e. when classifying LR frames without applying SR algorithms and without using a separate image set for training). We see that for DT-CWT, FRACTAL, and EF applying SR algorithms and using a DS classification provides consistently higher mean overall classification rates as compared to LOPO/"Normal". But for some of these feature extraction methods we notice that, when using a reduced DS classification, the classification rates drop below the LOPO/"Normal" rates. Only in case of EF it doesn't really matter whether we use reduced DS or DS – in both cases the rates remain above the LOPO/"Normal" rates. In case of SSF and MBFSI only a few SR methods are able to deliver higher rates as compared to LOPO/"Normal". For LBP this is never the case.

IV. Conclusion

In this work we have shown that classifying images acquired with an HD endoscope and training the underlying classifier with images acquired with a high-magnification endoscope is feasible.

We have also shown that by applying SR algorithms we are even able to increase the overall classification rates for some feature extraction methods. But from the experiments carried out it is also evident that the improvements are sometimes only marginal. Apart from that, the improvements quite often vanish as soon as the training set in case of a DS classification is reduced. In other words, the improvements observed in case of DS can partially be explained by the fact that the number of images available for training is much higher when using high-magnification images.

We have also seen that it is highly feature-dependent whether there is a benefit from bridging the resolution gap with SR algorithms or not. While LBP completely fails to benefit from SR algorithms, other methods (e.g. EF and FRACTAL) seem to benefit quite well from higher image resolutions.

Acknowledgments

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

References

[1] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 565–581, Sep. 2012.

[2] M. Liedlgruber and A. Uhl, "Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review," *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 73–88, 2012.

[3] M. Häfner, M. Liedlgruber, and A. Uhl, "Comparison of different super-resolution methods for HD video endoscopy," in *Proceedings of Bildverarbeitung für die Medizin 2014 (BVM'14)*, ser. Informatik aktuell, Aachen, Germany, Mar. 2014, accepted.

[4] ——, "POCS-based super-resolution for HD endoscopy video frames," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS'13)*, 2013, pp. 185–190.

[5] ——, "Super-resolution techniques evaluated in the context of HD endoscopic imaging," Department of Computer Sciences, University of Salzburg, Austria, http://www.cosy.sbg.ac.at/research/tr.html, Tech. Rep. 2013-04, 2013.

[6] M. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, pp. 75–104, 1996.

[7] S. Deqing, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, 2010, pp. 2432–2439.

[8] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, Apr. 1991.

[9] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR'01)*, no. 1, 2001, pp. 645–650.

[10] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[11] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[12] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei, "Feature-extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *Pattern Analysis and Applications*, vol. 12, no. 4, pp. 407–413, Dec. 2009.

[13] A. Uhl, A. Vécsei, and G. Wimmer, "Fractal analysis for the viewpoint invariant classification of celiac disease," in *Proceedings of the 7th International Symposium on Image and Signal Processing (ISPA 2011)*, Dubrovnik, Croatia, Sep. 2011, pp. 727 –732.

[14] A. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in HD colonoscopy," in *Proceedings of the 3rd International MICCAI Workshop on Medical Computer Vision (MICCAI-MCV'13)*, ser. Springer LNCS, vol. 8331, 2013.

[15] Q. Xu and Y. Q. Chen, "Multiscale blob features for gray scale, rotation and spatial scale invariant texture classification," in *Proceedings of 18th International Conference on Pattern Recognition (ICPR'06)*, vol. 4, Sep. 2006, pp. 29–32.

[16] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Endoscopic image classification using edge-based features," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR'10)*, 2010, pp. 2724–2727.

[17] B. Everitt, *The Analysis of Contingency Tables*. Chapman and Hall, 1977.