

## **Super-Resolution Techniques Evaluated in the Context of HD Endoscopic Imaging**

M. Häfner<sup>a</sup>

M. Liedlgruber

A. Uhl

<sup>a</sup>Department for Internal Medicine, St. Elisabeth Hospital, Vienna

Technical Report 2013-04

September 2013

**Department of Computer Sciences**

Jakob-Haringer-Straße 2  
5020 Salzburg  
Austria  
[www.cosy.sbg.ac.at](http://www.cosy.sbg.ac.at)

**Technical Report Series**

# Super-Resolution Techniques Evaluated in the Context of HD Endoscopic Imaging

M. Häfner, M. Liedlgruber, and A. Uhl

## Abstract

In this work we evaluate different reconstruction-based super-resolution algorithms in order to enhance the spatial resolution of endoscopic images acquired with an HD endoscope. The main question we try to answer in this work is whether it is feasible to employ super-resolution algorithms in order to increase the resolution of our HD images to reveal new details which may have got lost due to the limited endoscope magnification inherent to the HD endoscope used (e.g. mucosal structures).

For this purpose we first briefly explain the principles of super-resolution algorithms. We then describe the evaluated algorithms in more detail, which is followed by experiments using different super-resolution algorithms. We show that for artificially generated sequences and non-endoscopic real-world sequences the SR algorithms are indeed able to deliver a higher quality as compared to a simple bicubic interpolation. In case of endoscopic sequences, however, there is no real improvement observable. But we also show that the quality measures used very frequently show a rather high disagreement upon an improvement.

## I. INTRODUCTION

**I**N image processing applications usually images with a high resolution are required or desired. This applies for example to surveillance applications where high-resolution (HR) images potentially allow to identify number plates of cars or suspicious objects and people more reliably. Medical image analysis is also a branch which benefits from high resolution images since these help medical experts to make a more reliable diagnosis (e.g. in case of MRI). But also in remote sensing applications high resolution images are always welcome in order to improve the classification of regions in such images.

The simplest approach to enhance the resolution would be to improve the sensor of the imaging device used (e.g. the CCD or CMOS chip). While reducing the pixel size and hence increasing the pixel density on such imaging chips could theoretically increase the resulting image resolution, there are practical limits. If the size of a pixel gets too small the signal-to-noise ratio (SNR) also drops, making it more difficult to clearly separate a signal from the sensor noise. The result are inevitably degraded images [1]. Hence, reducing the pixel size is no real option.

Another approach would be to increase the chip size and placing more pixels on the chip. Despite the fact that this would also increase the cost for image sensors, this would also result in an increased capacitance, making it difficult to speed up a charge transfer rate [2]. Hence, the capturing speed would suffer and therefore this approach is also considered not being effective.

As a consequence, throughout the past decade the development of super-resolution (SR) algorithms has been one of the most active research areas. Such algorithms aim at increasing the resolution of an image by employing signal processing techniques.

While the first approach dates back to 1984 [3], numerous algorithms have been developed in the past to tackle the problem of SR. These can be roughly divided into recognition-based and reconstruction-based algorithms. The first type of SR approaches tries to detect certain patterns in low-resolution (LR) images and constructs an image of higher resolution by synthesizing an image based on these patterns. Therefore such algorithms have only a limited application area since specifying such patterns is usually a non-trivial task. Hence, the vast majority of SR algorithms is reconstruction-based and exploits the information from a sequence of LR images to obtain a single HR image. The experiments shown in this paper are solely based on reconstruction-based algorithms, hence, from now on we always refer to reconstruction-based SR algorithms when we refer to SR algorithms.

The remaining part of this paper is organized as follows. In Section II we motivate the application of SR algorithms to endoscopic imagery. This is followed by a short explanation of the principles of SR algorithms in Section III. We

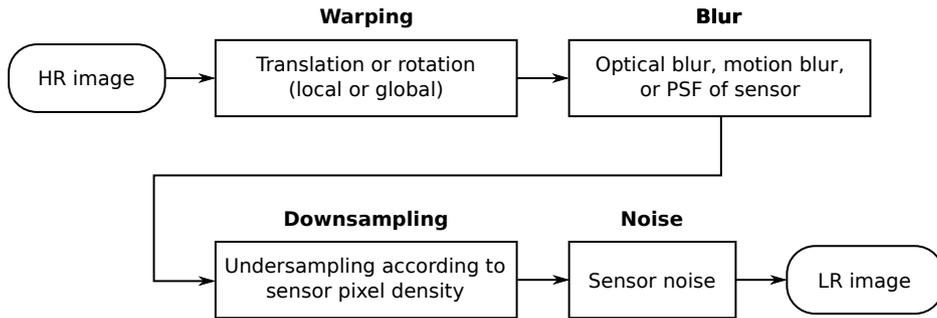


Figure 1. The observation model usually assumed in case of SR algorithms.

then briefly describe the motion estimation and SR algorithms evaluated for HR image reconstruction in Section IV and V, respectively. The termination criterion used for the SR algorithms evaluated is described in Section VI, followed by a short description of the methods used to assess the visual quality of the images produced by the different SR algorithms in Section VII. In Section VIII we describe our experimental setup, followed by the results obtained in Section IX. We conclude the paper in Section X.

## II. MOTIVATION

In the past we developed different approaches for the classification of colonic polyps (e.g. [4]–[7]). All these methods have been developed and evaluated on an image database obtained with a zoom-endoscope with a magnification factor of 150. The advantage of such endoscopes is obvious as they allow to inspect the colonic mucosa in a magnified manner, thus uncovering the fine surface structure of the mucosa as well as small lesions.

The present work, however, is solely based on endoscopic images obtained with an HD endoscope. While such an endoscope provides a roughly four times higher image resolution as compared to the previously used zoom-endoscope, the currently used HD endoscope has no zooming capabilities.

Hence, the main question we try to answer in this work is whether it is feasible to employ SR algorithms in order to increase the resolution of our HD images in order to reveal new details which may have got lost due to the limited endoscope magnification inherent to the HD endoscope used (e.g. mucosal structures). To obtain an HR image we use several LR images which are simply successive HD endoscopy video frames.

While it would be illusory to expect HR images generated by SR techniques, which are comparable to the ones obtained with the zoom-endoscope, we at least hope to be able to increase the level of detail of HD images. For this purpose we evaluate different well-known SR algorithms and measure the visual quality of the resulting HR images using different image quality metrics.

Up to our knowledge currently there exists only one work which tries to tackle the SR problem with endoscopic videos [8]. But the authors test their algorithm on LR images generated from a single video frame by shifting this frame into different directions and downscaling the shifted frames. Our work, in contrast, aims at reconstructing a HR image from several successive LR images which are not synthetically generated.

## III. PRINCIPLES AND LIMITATIONS OF SR ALGORITHMS

Before evaluating SR algorithms on our images we give a brief introduction to the principles and limitations of SR algorithms. This allows us to better understand and interpret the results we present later in Section IX. For a more thorough overview and discussion of SR algorithms we refer the reader to the excellent review articles found in [9]–[11].

### A. Principles

Fig. 1 shows the observation model which is usually assumed in case of SR algorithms. The HR image is the image which we aim to reconstruct from multiple LR images. It is the result of sampling a continuous scene into a discrete image. The HR image is then subject to a warping, which might be caused for example by camera movements or motion in the scene captured. Then, the image is usually degraded by some sort of blurring. This

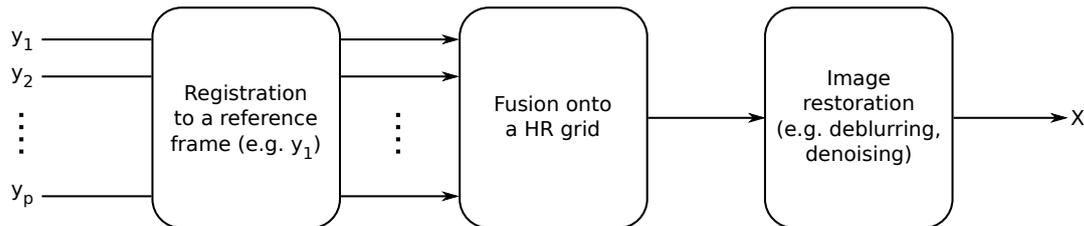


Figure 2. The reconstruction scheme common to reconstruction-based SR algorithms.

might be due to motion blur or optical blur inherent to the optics used. In addition, the image is influenced by the point spread function (PSF) of the imaging sensor. The PSF describes how several HR pixels within a certain neighborhood affect a LR pixel. This usually includes the spatial integration over the HR pixels as well as a defocus component. For the sake of simplicity the PSF is usually modeled as a simple spatial averaging over the HR pixels. Since the LR images have a lower resolution as compared to the HR image (limited by the sensor resolution), an implicit downsampling is performed. This is also the point, where aliasing artifacts are generated. Finally, depending on the sensor used, a certain amount of noise may be added to the LR image.

Multiple instances of such LR images are the basis for the reconstruction of the HR image. For this purpose SR algorithms are usually based on the scheme shown in Fig. 2. It must be noted that, while in Fig. 2 the different reconstruction stages are shown as separate steps, depending on the SR algorithm used, the steps may also be carried out simultaneously. In this figure  $y_k$  denotes an LR image,  $p$  the number of LR images available, and  $X$  the HR image we aim to reconstruct. We notice that the reconstruction basically consists of three steps:

- **Registration**

During the registration step the relative motion of each LR image with respect to a reference image (often the first LR image) is estimated with sub-pixel accuracy. As we notice from Fig. 1 this may include translation and rotation (either local or global).

- **Fusion**

Once the relative motion has been estimated, the LR images are fused onto a HR grid. In other words, the information from the different LR images is combined into an HR image, matching the resolution of the desired HR image. Since the motion between the LR images is arbitrary the respective HR image will not always match up with the regular HR grid. As a consequence non-uniform interpolation is necessary.

- **Image restoration**

Finally, to counteract blurring and noise degradations, usually an image restoration step is applied to the HR image.

In order to successfully accomplish the SR reconstruction three important conditions must be met:

- **Accurate motion estimation**

The motion estimation, based on which the registration is performed, is a very crucial step in the reconstruction process. Moreover, since the LR images are upscaled to the HR grid during the fusion step it is imperative that the motion is estimated with sub-pixel accuracy.

But achieving high accuracy during the motion estimation highly depends on the type of motion present across the LR images. If the image sequence exhibits global motion only (i.e. the motion is location-invariant and therefore constant across an entire LR image) the respective estimation can be carried out fairly easy and accurate. But as soon as the LR image sequence contains location-variant motion, the task of estimating the respective motion field can get quite complicated (e.g. by using optical flow estimation).

- **Sub-pixel shifts in LR images**

Another important preliminary for SR algorithms to work is the need for sub-pixel shifts between LR images. Fig. 3 schematically shows four different LR images. The LR image pixels denoted by the filled circles refer to the reference image all other LR images are registered to. If two LR images are shifted by integer-shifts these images are basically the same and contain the same information (except for border pixels). As a consequence an SR algorithm will not be able to recover additional information from such two frames (illustrated in Fig. 3 by the LR images made of the pixels shown as circles and diamonds, respectively).

- **Aliasing in LR images**

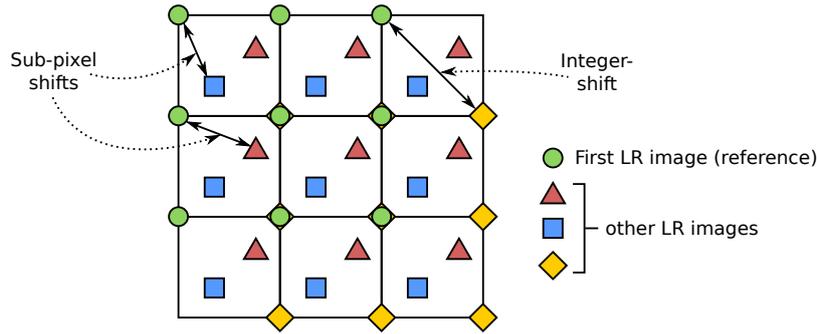
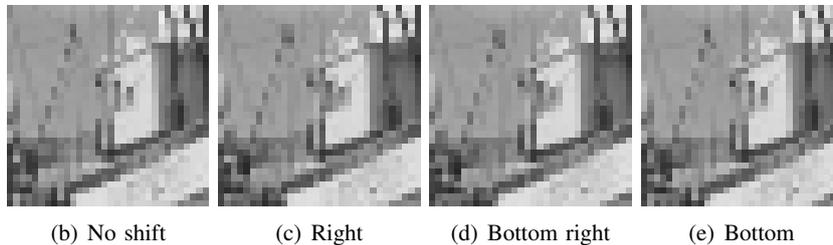


Figure 3. The difference between sub-pixel shifts and integer-shifts between a reference LR image (denoted by the filled circles) and three additional, differently shifted LR images (denoted by the triangles, the squares, and the diamonds).



(a) Example image



(b) No shift

(c) Right

(d) Bottom right

(e) Bottom

Figure 4. Illustration of aliasing artifacts necessary in order to make SR algorithms work. (a) an example image, (b)-(e) a region of the example image (denoted by a red square in (a)) after applying a blur, followed by a single pixel shift and a downscaling by a factor of four.

In order to obtain new high-frequency details for the reconstruction of the HR image we also need aliasing artifacts within the LR images. Such artifacts arise when a signal (i.e. the continuous scene, which is considered to be band-limited) is sampled below the Nyquist rate. This usually happens during the undersampling of the scene due to a limited sensor resolution.

While usually such artifacts are unwanted in signal processing, for SR reconstruction they are necessary in order to be able to obtain new image details by combining different LR images.

An example for aliasing artifacts is shown in Fig. 4. In Fig. 4(a) a test image is shown. This image is translated to the right, to the bottom right, or down by one pixel, followed by a blur (simulating the PSF) and a downsampling by a factor of four. The resulting image regions (simulated LR images) for the squared region shown in Fig. 4(a) are shown in figures 4(c)-(e). Figure 4(b) also shows a blurred and downsampled version of the squared region shown in Figure 4(a), but without any shift. The shifts by one pixel and the downsampling by a factor of four correspond to sub-pixel shifts of 0.25 pixels in the LR images.

From figures 4(b)-(e) we notice slight differences which are the result of aliasing.

#### IV. MOTION ESTIMATION

As already indicated in Section III, an accurate motion estimation is a crucial part of the SR reconstruction process. This especially accounts to the endoscopic images used throughout our experiments. Since we are facing highly complex motion patterns (e.g. position-variant transformations and parallax effects) simple motion models (e.g. perspective projections or affine transformations) are not sufficient to describe the motion between successive HD endoscopy video frames.

As a consequence we decided to use an optical flow method. While being more complex these methods are also 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 the aforementioned motion estimation methods usually assume a global motion.

Nevertheless, to be able to compare the HR reconstruction quality with simpler motion models, we also evaluate a model estimation considering a global shift and rotation only.

##### A. Global Shift and Rotation Estimation

In order to obtain an estimate for global motion which is assumed to be planar and parallel to the image plane, we use the motion estimation proposed in [12]. First, this method estimates the global rotation by investigating the correlation between the Fourier amplitudes of two images  $I_1$  and  $I_2$ . Then, image  $I_2$  is rotated such that it is aligned to the reference image  $I_1$  with respect to the rotation. Once this is done, the horizontal and vertical shifts are estimated by utilizing the Fourier shifting property.

The rotation between two images  $I_1$  and  $I_2$  is estimated by investigating the correlation between the amplitudes of the respective Fourier-transformed images  $\hat{I}_1$  and  $\hat{I}_2$ . Since shifts in the spatial domain only affect the phase angles of the Fourier-transformed images, the rotation can be estimated without knowing the actual shifts between the images. But before performing motion estimation, we multiply the LR images with a Tukey window to make the images circularly symmetric and thus obtain a more robust motion estimate.

In order to estimate the rotation, the mean of amplitudes is computed for angular slices of the frequency spectrum. Since the images are discrete we can write the mean within such a slice as:

$$M(\alpha, \hat{I}) = \frac{1}{N} \sum_{r=r_0}^R \sum_{\theta=\alpha-\Delta\omega/2}^{\alpha+\Delta\omega/2} |\hat{I}(r, \theta)|, \quad (1)$$

where  $R = \min(S_x/2, S_y/2)$  denotes the image radius with  $S_x$  and  $S_y$  being the width and height of the images, respectively.  $N$  is the number of coefficients within a slice,  $\Delta\omega$  denotes the angular width of an angular slice, measured in degrees ( $10^\circ$  degrees in our experiments), and  $r_0$  denotes a starting radius greater zero. Such a starting radius is used since within the low-frequency region the values are sampled in a very coarse manner. In addition, the low frequencies usually have rather high amplitudes, which would have a biasing effect on the mean.

By computing  $M(\alpha, \hat{I})$  for each angular slice, we obtain a sequence containing the mean of amplitudes for each slice:

$$M(\hat{I}) = \left( M(a\Delta\alpha, \hat{I}) \right)_{a=0}^{(360/\Delta\alpha)-1}, \quad (2)$$

where  $\Delta\alpha$  denotes the angular step between successive slices (i.e. the reciprocal of  $360/\Delta\alpha$  equals the number of slices) measured in degrees ( $0.1^\circ$  degrees in our experiments). It must be noted, that, since in our case  $\Delta\alpha < \Delta\omega$ , it follows that the angular slices are overlapping.

Once  $M(\hat{I})$  has been computed for  $\hat{I}_1$  and  $\hat{I}_2$ , the rotation angle (in degrees) between the images is estimated by

$$R(\hat{I}_1, \hat{I}_2) = \Delta\alpha \underset{s \in \{0, \dots, (360/\Delta\alpha)-1\}}{\operatorname{argmax}} \operatorname{corr}(M(\hat{I}_1), \operatorname{circshift}(M(\hat{I}_2), s)), \quad (3)$$

where  $\operatorname{corr}(\cdot, \cdot)$  and  $\operatorname{circshift}(\cdot, \cdot)$  denote the correlation and a circular shift operation, respectively. Having estimated  $R(\hat{I}_1, \hat{I}_2)$ ,  $I_2$  is rotated to cancel out the rotation.

To estimate the horizontal and vertical shifts between  $I_1$  and  $I_2$  the Fourier shifting property is exploited. By denoting the horizontal shift by  $\Delta x$  and the vertical shift by  $\Delta y$ , the shift between  $I_1$  and  $I_2$  can be expressed as

$$I_1(x + \Delta x, y + \Delta y) = I_2(x, y). \quad (4)$$

After transforming the images to the Fourier domain, the shift between the images can be expressed in the Fourier domain as (the Fourier shifting property)

$$\hat{I}_2(u, v) = e^{2\pi i(u\Delta x + v\Delta y)} \hat{I}_1(u, v), \quad (5)$$

which can be reformulated to

$$\frac{\hat{I}_2(u, v)}{\hat{I}_1(u, v)} = e^{2\pi i(u\Delta x + v\Delta y)}. \quad (6)$$

Now we compute the phase angle  $\angle(z) = \text{atan2}(\text{im}(z), \text{re}(z))$  on both sides. By denoting  $Z(u, v) = \frac{\hat{I}_2(u, v)}{\hat{I}_1(u, v)}$  we obtain

$$\angle(Z(u, v)) = \text{atan2}\left(\text{im}(e^{2\pi i(u\Delta x + v\Delta y)}), \text{re}(e^{2\pi i(u\Delta x + v\Delta y)})\right). \quad (7)$$

By using Euler's formula we obtain

$$\angle(Z(u, v)) = \text{atan2}(\sin(2\pi(u\Delta x + v\Delta y)), \cos(2\pi(u\Delta x + v\Delta y))), \quad (8)$$

which corresponds to

$$\begin{aligned} \angle(Z(u, v)) &= \text{atan}\left(\frac{\sin(2\pi(u\Delta x + v\Delta y))}{\cos(2\pi(u\Delta x + v\Delta y))}\right) + P \\ &= \text{atan}(\tan(2\pi(u\Delta x + v\Delta y))) + P \\ &= 2\pi(u\Delta x + v\Delta y) + P, \end{aligned}$$

with  $P = \pm\pi$  or  $P = 0$ , depending on the sign of the  $\sin(\cdot)$  and  $\cos(\cdot)$  terms. By setting  $\Delta x' = 2\pi\Delta x$  and  $\Delta y' = 2\pi\Delta y$ , we obtain a system of linear equations with one equation

$$\angle(Z(u, v)) = u\Delta x' + v\Delta y' + P \quad (9)$$

for each  $(u, v)$ . By employing, for example LSE, we can solve the system of linear equations and obtain estimates for the values  $\Delta x'$  and  $\Delta y'$ . From these we compute  $\Delta x = S_x\Delta x'/2\pi$  and  $\Delta y = S_y\Delta y'/2\pi$ .

As indicated in [12], the motion estimation described in this section does not longer work as soon as the input images are aliased. To resolve this issue the estimation of the motion must be restricted to a certain frequency band such that eventual aliasing artifacts and high frequencies do not pose a problem (i.e.  $R$  in Equ. (1) and  $u$  and  $v$  in Equ. (9) must be restricted accordingly). Especially an optimal choice for the range of  $u$  and  $v$  (width of centered window in the Fourier-spectrum being aliasing-free) in Equ. (9) is crucial to an exact shift estimation. Since the optimal values for this window width may change significantly depending on the LR images used, we test different window widths (between 3 and 25) and use the one which produces the smallest root mean square (RMS) between  $I_1$  and  $I_2$  (after applying the estimated shift to  $I_2$ ).

### B. Optical Flow Motion Estimation

The optical flow estimation we use for our experiments is based on [13] which uses the work presented in [14], [15] as the baseline model. The method works in a coarse-to-fine manner by constructing a Gauss-Laplace image pyramid for the two images the motion should be estimated between. Then the motion is estimated between the images at the coarsest level. The resulting estimate is used as a seed for the estimation at the next finer level. This is repeated until the motion has been estimated at the finest level in the pyramids.

To obtain the motion estimate at a certain pyramid level  $l$  between two images  $I_1$  and  $I_2$  (these are considered to show a scene at time instants  $t$  and  $t + 1$ , respectively), usually several constraints are introduced in order to get a stable and robust estimate (in the following we omit the pyramid level from our descriptions, since the computations are the same for all pyramid levels).

First of all, it is assumed that the gray level of a pixel is not changed during motion (gray level constancy assumption), that is

$$I_1(x + u_1, y + v_1) = I_2(x, y), \quad (10)$$

where  $(x, y)$  denotes a pixel location within the images and  $u_1$  and  $v_1$  denote the flow field from  $I_1$  to  $I_2$ . From this assumption we are able to derive an energy functional, which has to be minimized during motion estimation:

$$E_{\text{data}} = \sum_{x,y} |I_1(x + u_1(x, y), y + v_1(x, y)) - I_2(x, y)| \quad (11)$$

In [13] a symmetry constraint is introduced which not only considers the flow field from  $I_1$  to  $I_2$  but also the flow field into the other direction to obtain a more accurate motion estimate. As a consequence, based on Equ. (11), we end up with

$$E_{\text{data}}^{(1)} = \sum_{x,y} |I_1(x + u_1(x, y), y + v_1(x, y)) - I_2(x, y)| \quad (12)$$

$$E_{\text{data}}^{(2)} = \sum_{x,y} |I_1(x, y) - I_2(x + u_2(x, y), y + v_2(x, y))| \quad (13)$$

where  $E_{\text{data}}^{(1)}$  and  $E_{\text{data}}^{(2)}$  denote the energy functionals for both flow directions and  $u_2$  and  $v_2$  denote the flow field from  $I_2$  to  $I_1$ .

In addition a smoothness term is used due to the assumption of a piecewise smooth flow field:

$$E_{\text{smooth}}^{(1)} = \sum_{x,y} (|\nabla u_1(x, y)|^2 + |\nabla v_1(x, y)|^2)^\eta \quad (14)$$

$$E_{\text{smooth}}^{(2)} = \sum_{x,y} (|\nabla u_2(x, y)|^2 + |\nabla v_2(x, y)|^2)^\eta \quad (15)$$

where  $\eta$  has a value between 0 and 1. These energy terms penalize strong variations within the flow fields, hence, for the motion estimation these terms must be minimized.

Finally, the algorithm proposed in [13] uses an additional energy term, which is used to enforce symmetry between the flow fields. To obtain the respective functional we assume the following:

$$\begin{aligned} u_1(x, y) &= -u_2(x + u_1(x, y), y + v_1(x, y)) & \rightarrow & \quad u_1(x, y) + u_2(x + u_1(x, y), y + v_1(x, y)) = 0 \\ v_1(x, y) &= -v_2(x + u_1(x, y), y + v_1(x, y)) & \rightarrow & \quad v_1(x, y) + v_2(x + u_1(x, y), y + v_1(x, y)) = 0 \end{aligned}$$

Based on these assumptions we obtain the following energy functions, which again must be minimized to estimate the correct motion fields:

$$E_{\text{symmetry}}^{(1)} = \sum_{x,y} |u_1(x, y) + u_2(x + u_1(x, y), y + v_1(x, y))| + |v_1(x, y) + v_2(x + u_1(x, y), y + v_1(x, y))| \quad (16)$$

$$E_{\text{symmetry}}^{(2)} = \sum_{x,y} |u_1(x + u_2(x, y), y + v_2(x, y)) + u_2(x, y)| + |v_1(x + u_2(x, y), y + v_2(x, y)) + v_2(x, y)| \quad (17)$$

Using Eqs. (12) to (17) the objective function to be minimized is then formulated as:

$$E(u_1, v_1, u_2, v_2) = \sum_{i=1}^2 E_{\text{data}}^{(i)} + \alpha E_{\text{smooth}}^{(i)} + \beta E_{\text{symmetry}}^{(i)}, \quad (18)$$

where  $\alpha$  and  $\beta$  are parameters allowing the user to control the smoothness of the flow fields.

Although the algorithm presented in [13] is based on iterative reweighted least square (IRLS) to solve the optimization problem, the actual implementation used<sup>1</sup> is based on successive over-relaxation (SOR) since this method is faster as compared to previously used ones.

## V. ALGORITHMS EVALUATED

Once, the relative motion for all LR images has been estimated, we evaluate the approaches – described in the following in more detail – in order to reconstruct HR images.

<sup>1</sup>MATLAB package available at: <http://people.csail.mit.edu/celiu/OpticalFlow/> – downloaded on the 10<sup>th</sup> of July, 2012

### A. Iterative Back Projection (IBP)

The Iterative Back Projection method has been proposed by Irani and Peleg [16] and was chosen for our experiments due to its simplicity and intuitive nature. Although being very simple, IBP yielded excellent results on artificial image data.

Based on the observation model from Fig. 1 the formation process for an LR image  $y_k$  based on an HR image  $X$  can be formulated in matrix notation as [17]:

$$y_k = D_k B_k W_k X + \eta, \quad (19)$$

where  $D_k$  denotes the decimation matrix used for downscaling,  $B_k$  is the blur matrix modeling the PSF and all other types of blur,  $W_k$  represents the warp-matrix which represents the motion between  $X$  and  $y_k$  before downscaling (this matrix is constructed from the estimated motion information).  $\eta$  denotes a normally distributed additive noise.

Based on Equ. (19) we can formulate how to obtain an LR image  $\hat{y}_k$  based on an HR estimate  $\hat{X}$ :

$$\hat{y}_k = D_k B_k W_k \hat{X} \quad (20)$$

In our implementation the initial estimate for  $\hat{X} = \hat{X}^0$  is set to an upscaled version of  $y_1$ . Based on Equ. (20) the error  $e_k$  between an  $y_k$  and  $\hat{y}_k$  can be computed by

$$e_k = y_k - \hat{y}_k \quad (21)$$

It must be noted that, by neglecting the noise component in Equ. (20),  $e_k$  also includes the noise. Thus, by minimizing the error  $e_k$  we also minimize the noise in the HR estimate.

By using Eqs. (20) and (21) the total squared error (based on the  $L_2$ -norm) for an HR estimate  $\hat{X}$  over all  $y_k$  can therefore be written as

$$E(\hat{X}) = \frac{1}{2} \sum_{k=1}^p \|e_k\|_2^2 = \frac{1}{2} \sum_{k=1}^p \|y_k - D_k B_k W_k \hat{X}\|_2^2 \quad (22)$$

Thus, by solving the minimization problem

$$\hat{X} = \underset{X}{\operatorname{argmin}} \sum_{k=1}^p \|y_k - D_k B_k W_k X\|_2^2 \quad (23)$$

we end up with the desired HR image.

In order to obtain update weights for each pixel of  $\hat{X}$  we compute the gradient of  $E$  with respect to  $\hat{X}$ . For a single image this results in the gradient

$$G_k = W_k^T B_k^T D_k^T (D_k B_k W_k \hat{X} - y_k). \quad (24)$$

The operations  $W_k^T$ ,  $B_k^T$ , and  $D_k^T$  denote the inverse warp, a sharpening, and an upscaling, respectively.

The final values for the pixel-wise HR estimate updates are then obtained by taking the sum over all  $G_k$  in a pixel-by-pixel fashion. Simply put, the update weight for a pixel  $(x, y)$  in  $\hat{X}$  is the sum over all  $G_k$  at position  $(x, y)$ . This can be formulated as:

$$\nabla E(\hat{X}) = \sum_{k=1}^p G_k \quad (25)$$

The new HR estimate  $\hat{X}^{n+1}$  is then computed using steepest descent minimization:

$$\hat{X}^{n+1} = \hat{X}^n + \lambda \nabla E(\hat{X}^n), \quad (26)$$

where  $\lambda$  is a factor affecting the convergence speed of the iterative process.

The full iteration step can be summarized into the following equation:

$$\hat{X}^{n+1} = \hat{X}^n + \lambda \sum_{k=1}^p W_k^T B_k^T D_k^T (D_k B_k W_k \hat{X}^n - y_k) \quad (27)$$

### B. Modified IBP (ROBZ)

In [18] the authors propose a modification of the IBP approach. Instead of summing up the single gradient images  $G_k$ , they propose to compute a scaled pixel-wise median to obtain the update weight for each pixel. Thus, Equ. (25) is changed to

$$\nabla E(\hat{X}) = p \operatorname{median}(G_k)_{k=1}^p, \quad (28)$$

where the operator  $\operatorname{median}(G_k)_{k=1}^p$  denotes the pixel-wise median over all  $G_k$ .

By changing the IBP algorithm this way, outlier pixels are removed. Such outliers might arise for example due to an inaccurate motion estimation.

### C. Regularized Super-resolution (RSR)

The regularized super-resolution approach evaluated was proposed by Farsiu et al. [19]. Since basically the SR reconstruction problem is an ill-posed one [20], 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 [19] is in some way similar to the IBP method described in Section V-A. Similar to Equ. (23), the RSR approach also minimizes the error between an observed LR image  $y_k$  and a simulated LR image  $\hat{y}_k$ . But instead of the  $L_2$ -norm cost function used in Equ. (23) Farsiu et al. use the  $L_1$ -norm. Thus, the starting point for the RSR method is the following equation:

$$\hat{X} = \operatorname{argmin}_X \sum_{k=1}^p \|y_k - D_k B_k W_k X\|_1 \quad (29)$$

To compensate for the ill-posedness nature of SR reconstruction problems, an additional regularization constraint is introduced. This regularization usually contains prior information about the desired HR image solution (e.g. smoothness or preserving edges). The minimization problem then becomes

$$\hat{X} = \operatorname{argmin}_X \left[ \sum_{k=1}^p \|y_k - D_k B_k W_k X\|_1 + \lambda \Phi(X) \right], \quad (30)$$

where  $\Phi(\cdot)$  is the regularization cost function and  $\lambda$  is the regularization parameter, allowing to weight the regularization constraint against the similarity cost (i.e. the first term). The regularization function used in [19], termed as bilateral total variation (BTV), is inspired by the total variation (TV) criterion. Based on the  $L_1$  norm on the gradient magnitude, the TV criterion penalizes the total variation within an image. The BTV function follows a similar principle and looks like

$$\Phi_{\text{BTV}}(X) = \underbrace{\sum_{l=-P}^P \sum_{m=0}^P}_{l+m \geq 0} \alpha^{|m|+|l|} \|X - S_x^l S_y^m X\|_1, \quad (31)$$

where  $P$  specifies the size of the image area which is considered for the computation of the criterion. The matrices  $S_x^l$  and  $S_y^m$  shift the image  $X$  by  $l$  and  $m$  pixels horizontally and vertically, respectively. The parameter  $\alpha$  ( $0 < \alpha < 1$ ) controls the spatial decaying effect on the summation of the regularization terms.

When substituting Equ. (31) into Equ. (30), we obtain the solution of the SR problem by minimizing

$$\hat{X} = \operatorname{argmin}_X \left[ \sum_{k=1}^p \|y_k - D_k B_k W_k X\|_1 + \lambda \underbrace{\sum_{l=-P}^P \sum_{m=0}^P}_{l+m \geq 0} \alpha^{|m|+|l|} \|X - S_x^l S_y^m X\|_1 \right]. \quad (32)$$

To solve Equ. (32) steepest descent is used. For our experiments the initial estimate for  $\hat{X} = \hat{X}^0$  is set to the result of the Shift-and-Add method (see Section V-E). The full iteration step used to compute  $\hat{X}^{n+1}$  is then

$$\hat{X}^{n+1} = \hat{X}^n - \beta \left[ \sum_{k=1}^p W_k^T B_k^T D_k^T \Lambda(D_k B_k W_k \hat{X}^n - y_k) + \lambda \underbrace{\sum_{l=-P}^P \sum_{m=0}^P \alpha^{|m|+|l|}}_{l+m \geq 0} [I - S_y^{-m} S_x^{-l}] \Lambda(\hat{X}^n - S_x^l S_y^m \hat{X}^n) \right], \quad (33)$$

where  $\Lambda(\cdot)$  denotes the sign function.

Based on this iterative procedure, Farsiu et al. propose additional improvements which make the algorithm very fast by avoiding decimation and warping during the iterations. For details on these optimizations we refer the reader to [19]. The implementation used for this work is based on Equ. (33) and also allows to use optical flow motion estimation.

#### D. Projection Onto Convex Sets (POCS)

The idea of Projection onto convex sets was introduced to image processing by the work in [21]. For POCS it is assumed that all images  $X$ , represented by one-dimensional vectors, are elements of a Hilbert space  $H$ . The projection  $PX$  onto a convex set  $C \subset H$  is then defined as

$$\|X - PX\| = \underset{x \in C}{\operatorname{argmin}} \|X - x\| \quad \forall X \in H. \quad (34)$$

A set  $C$  is said to be convex if it has the following property:

$$\forall X_1, X_2 \in C : X_3 \in C \quad \text{with} \quad X_3 = \lambda X_1 + (1 - \lambda) X_2 \quad \forall \lambda \in [0, 1] \quad (35)$$

We now assume that we have  $m$  closed convex sets  $C_i, i = 1, \dots, m$  and the projection operator on  $C_i$  is denoted by  $P_i$ . It has been shown that the iteration

$$\hat{X}^{n+1} = P_m P_{m-1} \dots P_1 \hat{X}^n \quad (36)$$

will converge to a point  $\hat{X}^*$  inside  $C_0 \triangleq \cap_{i=1}^m C_i$  for an arbitrary initial image  $\hat{X}^0$ . However, while convergence is assured (as long as  $C_0$  is not the empty set), the solution  $\hat{X}^*$  is not necessarily unique.

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  $H$ . More specifically, the solution is constrained by convex sets which, according to the prior knowledge available, impose restrictions on an HR estimate in order to be a valid one. Based on such an HR estimate  $\hat{X}^n$  we then obtain a new estimate  $\hat{X}^{n+1}$  by projecting  $\hat{X}^n$  onto the convex sets according to Equ. (36). Since  $\hat{X}^{n+1}$  lies within  $C_0$  it is assured that the new estimate satisfies all constraints imposed by the convex sets. Hence,  $\hat{X}^{n+1}$  is one possible solution to the reconstruction problem.

Before the iterative SR reconstruction process can be started, two important steps are required: first, the closed convex sets, constraining the HR solution, must be defined. Then, based on the convex sets, the projection operators need to be derived from the definitions of the  $C_i$ 's.

Throughout literature several different constraint sets have been proposed (e.g. [17], [22]). In the following we list some convex set constraints which are frequently incorporated into SR methods or at least may be useful for such algorithms (along with the respective projection operators):

- **Amplitude constraint**

This constraint simply limits pixel values within images in the respective convex set to certain predefined bounds:

$$C_A = \{X : B_p^{\min} \leq X_p \leq B_p^{\max}, \forall p = 1, \dots, M\}, \quad (37)$$

where  $M$  denotes the number of pixels in  $X$  and  $B_p^{\min}$  and  $B_p^{\max}$  denote the lower and upper bound for the  $p$ -th pixel, respectively. This definition uses separate bounds for each pixel and thus allows to impose spatially-dependent limits on pixel values.

The projection of an image  $X$  onto  $C_A$  is then defined as

$$X'_p = P_A X_p = \begin{cases} B_p^{\min}, & X_p < B_p^{\min} \\ X_p, & B_p^{\min} \leq X_p \leq B_p^{\max} \\ B_p^{\max}, & X_p > B_p^{\max} \end{cases}, \quad (38)$$

where  $X'_p$  denotes the  $p$ -th pixel value within the projection of  $X$ .

- **Non-negativity constraint**

The non-negativity constraint simply prohibits pixel values below 0 and can thus be regarded as a special case of the amplitude constraint (with  $B_p^{\min} = 0$  and  $B_p^{\max} = \infty$  for all  $p = 1, \dots, M$ ). The respective convex set can be defined as

$$C_N = \{X : 0 \leq X_p, \forall p = 1, \dots, M\}, \quad (39)$$

with the respective projection

$$X'_p = P_N X_p = \begin{cases} 0, & X_p < 0 \\ X_p, & X_p \geq 0 \end{cases}. \quad (40)$$

- **Energy constraint**

The energy constraint is used to limit the energy in an image to a maximum permitted energy level:

$$C_E = \{X : \|X\|^2 \leq E\}, \quad (41)$$

where  $E$  is the maximum allowed energy within an image  $X$ . One possibility for the respective projection is then

$$X'_p = P_E X_p = \begin{cases} X_p, & \|X\|^2 \leq E \\ \frac{X_p}{\sqrt{E/\|X\|^2}}, & \|X\|^2 > E \end{cases}. \quad (42)$$

- **Data consistency constraint**

This constraint is a very important one in SR algorithms since it measures the consistency between observed LR images and a simulated LR image. In its most general form the respective convex set can be formulated as

$$C_C = \{X : \|X - g\| \leq \epsilon_R\}, \quad (43)$$

where  $g$  denotes some reference image and  $\epsilon_R$  denotes the maximum distance allowed between  $X$  and  $g$ . One possible projection for this convex set is:

$$X' = P_C X = \begin{cases} X, & \|X - g\| \leq \epsilon_R \\ X - \Delta_C(X - g), & \|X - g\| > \epsilon_R \end{cases} \quad (44)$$

where  $\Delta_C$  denotes a factor, specifying the correction strength.

In terms of the SR terminology the convex set for the  $k$ -th LR image can be formulated as

$$C_C^k = \{X : \|D_k B_k W_k X - y_k\| \leq \epsilon_R\}, \quad (45)$$

with the following projection:

$$X' = P_C^k X = \begin{cases} X, & \|D_k B_k W_k X - y_k\| \leq \epsilon_R \\ X + W_k^T B_k^T D_k^T (\Delta_C(y_k - D_k B_k W_k X)), & \|D_k B_k W_k X - y_k\| > \epsilon_R \end{cases}. \quad (46)$$

In our implementation of the POCS approach [23] we use the data consistency constraint and the amplitude constraint. Since our consistency constraint correction step also involves a sharpening of the HR estimate (induced by  $B_k^T$  in Equ. (46)), we also employ a correction amplitude constraint, which prevents over-sharpening:

$$C_{CA}^k = \{X : \|X - W_k^T D_k^T y_k\| \leq \epsilon_{CA}\}, \quad (47)$$

where  $\epsilon_{CA}$  denotes a parameter which controls the over-sharpening correction. This constraint simply states that the pixel-wise differences between the current estimate  $X$  and the LR images (after upscaling and inverse warping) must not exceed a certain limit. In fact, this constraint is a special case of the amplitude constraint with pixel-wise bounds which are based on the pixels of upscaled and warped LR images. The respective projection is

$$X' = P_{CA}^k X = \begin{cases} X, & \|X - W_k^T D_k^T y_k\| \leq \epsilon_{CA} \\ X + (\Delta_C(W_k^T D_k^T y_k - X)), & \|X - W_k^T D_k^T y_k\| > \epsilon_{CA} \end{cases}. \quad (48)$$

In our implementation the initial estimate  $\hat{X}^0$  is set to an upscaled version of  $y_1$ . In order to obtain a new HR estimate  $\hat{X}^{n+1}$ , based on a current estimate  $\hat{X}^n$ , we successively apply all projections used onto each LR image. This can be written as

$$\hat{X}^{n+1} = P_A^p P_{CA}^p P_C^p P_A^{p-1} P_{CA}^{p-1} P_C^{p-1} \dots P_A^1 P_{CA}^1 P_C^1 \hat{X}^n, \quad (49)$$

where  $P^k$  denotes the respective projection onto the  $k$ -th LR image.

As we have seen, a clear advantage of POCS is the fact that prior knowledge can be formulated conveniently in terms of convex sets and the associated projections. While it may not always be trivial to find the  $P_i$ 's, it is usually easier than finding a projection which immediately projects an arbitrary  $X$  onto the solution set  $C_0$ . Moreover, POCS is very intuitive since it allows to specify the constraints in the spatial domain, based on the observation model. However, POCS has also the drawbacks of a rather slow convergence, a high computational cost, the non-uniqueness of the solution, and strong dependence on the initial estimate.

### E. Shift-and-Add (S&A)

The Shift-and-Add method is probably one of the simplest SR algorithms, originally proposed in [24]. We nevertheless use this method for our experiments for comparison purposes.

Having estimated the motion, an initial HR estimate  $\hat{X}_0$  is computed as the pixel-wise average of the LR images, after those have been aligned properly (according to the motion estimate), upsampled, and interpolated. Using the terminology from above, this can be expressed more formally as:

$$\hat{X}_0 = \frac{1}{p} \sum_{k=1}^p W_k^T D_k^T y_k \quad (50)$$

In order to cope with degradations (noise and blur) in  $\hat{X}$  and obtain the final HR estimate  $\hat{X}$ , this method applies a deconvolution step to  $\hat{X}_0$ . For our experiments we use a regularized deconvolution [25], but any other deconvolution method might be used too.

## VI. TERMINATION OF THE ITERATIVE PROCESS

Since all SR algorithms evaluated (except for the Shift-and-Add algorithm) are iterative processes, we need to define a stopping criterion for the iterations. One possibility would be to simply fix the number of iterations carried out. But this approach has the drawback that it may be hard to find a number which works well for different kinds of images. If the number of iterations is too high this may lead to severe over-sharpening in case of certain images. If chosen too low, the resulting HR images may suffer from missing details.

Another possibility would be to measure the difference between a HR estimate  $\hat{X}^{n+1}$  and its predecessor  $\hat{X}^n$  and stop the iterative process as soon as the difference drops below a certain threshold. But again, it is not easy to find a threshold which works well for different types of images. Choosing a wrong threshold value may lead to the same problems as in case of a fixed iteration count.

For the aforementioned reasons we chose to use the adaptive termination criterion already used in [23]. Using this criterion, the iterative process can be outlined as follows:

- 1) Carry out the first three iterations and keep track of the differences between  $\hat{X}^n$  and  $\hat{X}^{n-1}$ , where  $n$  denotes the current iteration (starting at 1). The difference is measured by computing the root mean square:

$$\Delta_n = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{X}_i^n - \hat{X}_i^{n-1})^2} \quad (51)$$

- 2) Based on the differences for iterations two and three, we compute a threshold value  $\epsilon_{\text{iter}}$ , which is used later to decide upon termination of the iterative process:

$$\epsilon_{\text{iter}} = \kappa |\Delta_3 - \Delta_2|, \quad (52)$$

where  $\kappa$  is a multiplier which specifies how much changes we allow between two successive estimates to consider a solution to have converged. In our experiments we choose  $\kappa = 0.05$ .

3) Additional iterations are carried out as long as the following equation evaluates to 1:

$$\text{crit}(\Delta_n, \Delta_{n-1}) = \begin{cases} 1, & |\Delta_n - \Delta_{n-1}| \geq \epsilon_{\text{iter}} \\ 0, & |\Delta_n - \Delta_{n-1}| < \epsilon_{\text{iter}} \end{cases} \quad (53)$$

In addition we terminate the iterative process as soon as for two successive iterations the difference values are growing again.

Since the values of  $\Delta_n$  can vary significantly depending on the LR images used we do not compare these values directly against a threshold. As a consequence we compute the differences of the differences (i.e. an approximate first derivative of the differences) for the termination test in step 3. Since the value of  $\Delta_1$  highly depends on the initial HR estimate, we use the values  $\Delta_2$  and  $\Delta_3$  to compute the threshold  $\epsilon_{\text{iter}}$ . The multiplier  $\kappa = 0.05$  implies that we consider the iterative process to have converged as soon as the difference of differences falls below 5% of the difference between  $\Delta_2$  and  $\Delta_3$ . If the difference values are growing again for two successive iterations we assume that we already found a good HR estimate and stop the iterative process too.

## VII. QUALITY ASSESSMENT FOR SR ALGORITHMS

In order to be able to compare the quality of the results produced by the SR algorithms evaluated, we need some sort of metric which allows us to quantitatively measure the visual quality of the outcomes of the different SR methods. Basically we distinguish between metrics which need a reference image (e.g. the PSNR) and those, which are reference-free and hence do not rely on an reference image. Since the main focus of this work lies on applying SR algorithms to endoscopic images, we restrict our comparison to reference-free metrics as we have no reference images in case of that application scenario. The quality assessment algorithms used throughout our experiments are<sup>2</sup>:

- **Anisotropic measure [26]**

This method aims at measuring the entropy within an image for different directions. 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.

- **Metric Q [27]**

This metric measures sharpness and contrast caused by edges in an image. For this purpose Metric Q is based on the coherence computed from a singular value decomposition of local gradient matrices in tiles of an image. By summing up the product of the absolute value of the largest singular value and the coherence for all tiles, a measure for anisotropy is obtained for the whole image.

- **BRISQUE [28]**

This image quality measure is a so-called natural scene statistics-based approach. In this type of methods, certain features from a set of different images are extracted (training). Then a statistical distribution is fitted to the extracted data. In case of BRISQUE the generalized Gaussian distribution (GGD) and an asymmetric GGD (AGGD) are used. The features used operate in the spatial domain and are computed as the normalized luminance values in an image. Since it has been observed that these values exhibit Gaussian characteristics, the aforementioned distributions are fitted to the values to obtain descriptive model parameters. In addition also inter-pixel relationships are investigated in order to detect distortions (e.g. noise or blur) within an image. Once the features for training images have been obtained, support vector regression ( $\epsilon$ -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 an image with an unknown quality.

- **DIIVINE [29]**

The DIIVINE metric is very similar to BRISQUE. But instead of spatial features, this approach relies on statistical features computed in the wavelet domain. In addition, DIIVINE uses a two-stage approach for the quality assessment. In the first stage, distortions are identified, whereas in the second stage a quality score is computed for each class of distortions. Since, however, we do not aim at computing scores for each type of distortion, we modified the original implementation to perform a one-stage classification only (as done in case of BRISQUE).

<sup>2</sup>For each metric evaluated we used the MATLAB code provided by the authors of the respective work as a code-basis.

Table I  
DETAILS ON OUR ARTIFICIAL LR IMAGE SETS USED.

Training (DIIVINE and BRISQUE)					Validation (All metrics)					
Name	Color	Dimensions	Shift	Rotation	Name	ID	Color	Dimensions	Shift	Rotation
Hill	✗	256 × 256	✓	✗	Airplane	A1S	✓	256 × 256	✓	✗
	✗	256 × 256	✓	✓		A1SR	✓	256 × 256	✓	✓
House	✗	128 × 128	✓	✗	Boat	A2S	✗	256 × 256	✓	✗
	✗	128 × 128	✓	✓		A2SR	✗	256 × 256	✓	✓
Lena	✗	256 × 256	✓	✗	Elaine	A3S	✗	256 × 256	✓	✗
	✗	256 × 256	✓	✓		A3SR	✗	256 × 256	✓	✓
Text	✗	302 × 198	✓	✗	Resolution Chart	A4S	✗	128 × 128	✓	✗
	✗	302 × 198	✓	✓		A4SR	✗	128 × 128	✓	✓

Table II  
DETAILS ON THE NON-MEDICAL REAL-WORLD LR IMAGE SETS USED.

Training (DIIVINE and BRISQUE)			Validation (All metrics)			
Name	Color	Dimensions	Name	ID	Color	Dimensions
Bookcase	✓	320 × 240	Carphone	R1	✓	176 × 144
Car	✗	72 × 121	City	R2	✓	352 × 288
Foreman 1	✓	352 × 288	Container	R3	✓	352 × 288
Foreman 2	✓	352 × 288	Garden	R4	✓	352 × 288
Library	✗	350 × 286	Mobile	R5	✓	352 × 288

## VIII. EXPERIMENTAL SETUP

### A. Evaluated Image Sets

For our experiments we evaluated the different SR algorithms on a variety of LR image sequences. These sequences can be divided into artificial ones and real-world sequences. In case of the artificial sets we use well-established test images, while in case of the real-world sequences we use widely used sequences as well as endoscopic images. Two things common to all sequences are the number of LR images available for each sequence, which has been fixed to eight for our experiments, and the upscaling factor to obtain the HR results, which has been fixed to two for our experiments.

In Table I details on the 16 artificial LR image sets used are provided. We always started with a HR reference image twice as high and wide as the LR images to generate these image sets. Based on the reference image we then created eight LR frames according to the following procedure (see also the observation model in Fig. 1):

- 1) In case of the first frame we continue with step 4 since this LR frame serves as the reference frame for the motion estimation. Hence, no motion must be applied.
- 2) Apply a circular shift (one pixel horizontally and two pixels and vertically, respectively) multiplied by the frame index (starting at zero) to the image.
- 3) If enabled for the current image set, apply a counterclockwise rotation of  $0.5^\circ$  multiplied by the frame index to the image.
- 4) Apply a PSF to the image (in our case a Gaussian kernel of size  $3 \times 3$  with  $\sigma = 1$ ).
- 5) Skip every second pixel (horizontally and vertically) in order to obtain the final undersampled LR image.

As we notice, we created sequences exhibiting translational motion only and sequences also including rotation. This allows us to assess how the different motion estimation algorithms and SR algorithms perform in case of different motion types.

Table II provides details on the ten non-medical real-world sequences used. These sequences consist of the first eight frames of different popular test sequences.

The medical videos used for the experiments have been acquired during colonoscopy sessions between the years 2011 and 2012 at the Department for Internal Medicine (St. Elisabeth Hospital, Vienna) using a HD colonoscope (Pentax HiLINE HD+ 90i Colonoscope) with a video resolution of  $1280 \times 1024$ . Since our SR algorithms all work with images only, a sequence of eight consecutive frames was extracted from each video considered.

One problem, which quite frequently occurs in endoscopy videos, is a rapid movement of the tip of the endoscope, resulting in severe motion blur. As a consequence we selected the frames in a way such that the displacement between

Table III  
 DETAILS ON THE ENDOSCOPIC REAL-WORLD LR IMAGE SETS USED.

Training (DIIVINE and BRISQUE)				Validation (All metrics)			
Compressed				Compressed			
ID	Color	Dimensions	Deblocked	ID	Color	Dimensions	Deblocked
E1	✓	256 × 256	✗	E6	✓	256 × 256	✗
E1D	✓	256 × 256	✓	E6D	✓	256 × 256	✓
E2	✓	256 × 256	✗	E7	✓	256 × 256	✗
E2D	✓	256 × 256	✓	E7D	✓	256 × 256	✓
E3	✓	256 × 256	✗	E8	✓	256 × 256	✗
E3D	✓	256 × 256	✓	E8D	✓	256 × 256	✓
E4	✓	256 × 256	✗	E9	✓	256 × 256	✗
E4D	✓	256 × 256	✓	E9D	✓	256 × 256	✓
E5	✓	256 × 256	✗	E10	✓	256 × 256	✗
E5D	✓	256 × 256	✓	E10D	✓	256 × 256	✓
Uncompressed				Uncompressed			
U3	✓	256 × 256	✗	U1	✓	256 × 256	✗
				U2	✓	256 × 256	✗

two successive frames is not too large. Nevertheless, the frames selected exhibit a fairly complex motion pattern (perspective changes, zoom, translation, and rotation).

To reduce the computational demand for our methods we chose positions from which we manually extracted  $256 \times 256$ -pixel patches which serve as LR image (the position remained the same in case of a single sequence). Another issue, which might pose problems to the SR methods evaluated, are compression artifacts present in the extracted frames. Hence, in order to assess whether these artifacts affect the image quality of the reconstructed HR images, we also carry out experiments with sequences, after applying a DCT deblocking algorithm [30]. In addition, we also carried out SR experiments with three uncompressed endoscopic video sequences (the IDs of these sequences start with the letter “U”). Details on the medical image sequences used can be found in Table III.

Each of the image databases used has been tested with the SR algorithms presented in Section V. For each combination of image database and SR algorithm we carried out experiments using the two different motion estimation methods described in Section IV.

### B. Quality Assessment

Since the luminance channel of an image captures details well enough, we restrict our quality assessment to this channel. That is, prior to computing the quality scores, we convert our HR images to the CIELAB color space and apply the metrics to the luminance channel only.

The anisotropic measure and Metric Q are both tile-based. Hence, we need to specify a size for the tiles to be used. In accordance to the experiments in [26], [27] we chose a size of  $8 \times 8$  pixels for our experiments. Since the values returned by the anisotropic measure are rather small, we multiplied these scores by  $10^3$  in order to enhance the readability within the result tables.

In case of BRISQUE and DIIVINE the underlying SVR model needs to be trained. Therefore we created an additional set of image sequences (also listed in Tables I, II, and III). These sequences have been created using the same procedure as used for the validation images (already described above). A random subset of the images in the training set (50%, including the HR images if available and a bicubic interpolation of the first LR frame of each sequence) has then been rated by eight human raters. Based on these ratings, the differential mean opinion score (DMOS) was computed (we use the median instead of the mean to be resistant against outlier ratings). The DMOS values have then been used to train BRISQUE and DIIVINE. For all metrics evaluated a higher higher score means a higher quality. In case of BRISQUE and DIIVINE the score is usually in the range between 0 and 100 (there might be some outliers leaving this range).

## IX. RESULTS

An overview of the scores obtained by the different quality metrics is given in Figs. 5 to 12. These plots show the minimum and maximum quality scores for the different sequences among all SR methods evaluated (denoted

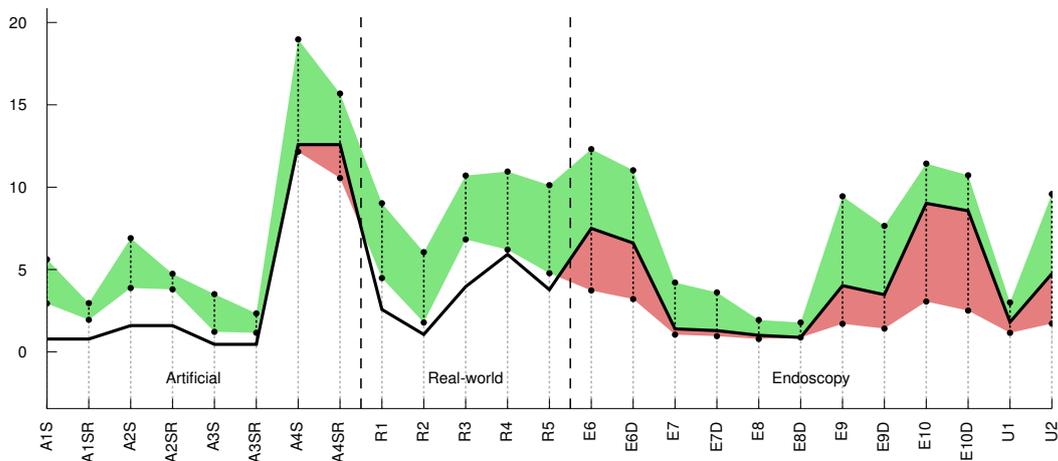


Figure 5. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (Anisotropic measure, global shift and rotation estimation).

by the black dots). The black solid line shows the scores obtained for an upscaled version of the reference frame of the respective sequence (upsampling factor of two, using bicubic interpolation). The green areas denote scores which are higher as compared to the scores for upscaling with bicubic interpolation whereas the red area shows scores below the bicubic interpolation scores.

What we immediately notice from Figs. 5 to 12 is that the anisotropic measure seems to produce rather high scores for the SR results as compared to the bicubic interpolation (see Figs. 5 and 9). However, it does not seem to make that much of a difference whether we use the simple global motion estimation or the more sophisticated local optical flow method. Only in case of the endoscopic sequences it seems that using optical flow has the effect that for almost all sequences the SR methods reach at least roughly the quality of the bicubic interpolation.

For all other metrics, the picture is quite different. In case of Metric Q (see Figs. 6 and 10) at least for the artificial and the real-world sequences we on average have rather high scores as compared to the bicubic interpolation. However, for the endoscopic sequences the SR methods seem to produce a lower quality. According to the Metric Q scores, switching from a global motion estimation to the optical flow method lowers the quality of the SR results. While this applies to all sequence types, this is especially noticeable in case of the scores for the endoscopic sequences.

In case of BRISQUE and DIIVINE we have a rather mixed picture (see Figs. 7 and 11 and Figs. 8 and 12). In case of the artificial sequences both metrics sometimes show SR result scores above and sometimes below the bicubic interpolation. This is quite similar in case of the real-world sequences. For the endoscopic sequences both metrics show dramatically low scores as compared to the score for the interpolated reference image. However, the scores get noticeably higher when using optical flow instead of a global motion estimation. While this applies to all sequence types, this can be observed especially in case of the endoscopic sequences.

A compact view on the results is provided in Table IV. This table indicates how often a SR method was able to yield a higher score as compared to the result of upsampling with bicubic interpolation. From this table we notice that, according to the anisotropic measure, the SR methods almost always deliver a higher quality score as compared to the bicubic score in case of the artificial and the real-world sequences. In case of the endoscopic sequences only IBP, ROBZ, and POCS are able to deliver higher scores in most cases. According to the remaining metrics, the overall picture is quite different. According to Metric Q, only in case of the artificial sequences in conjunction with the global motion estimation the visual quality seems to have been improved by SR methods. For all other quality metrics there are no dominant SR methods, which are consistently able to improve the visual quality by applying SR techniques. But we notice that, except for Metric Q, using optical flow seems to raise the reconstruction quality above the bicubic interpolation for endoscopic sequences at least in some cases.

Detailed scores are given in Tables V to VIII. In these tables the coloring used is the same as in the overview

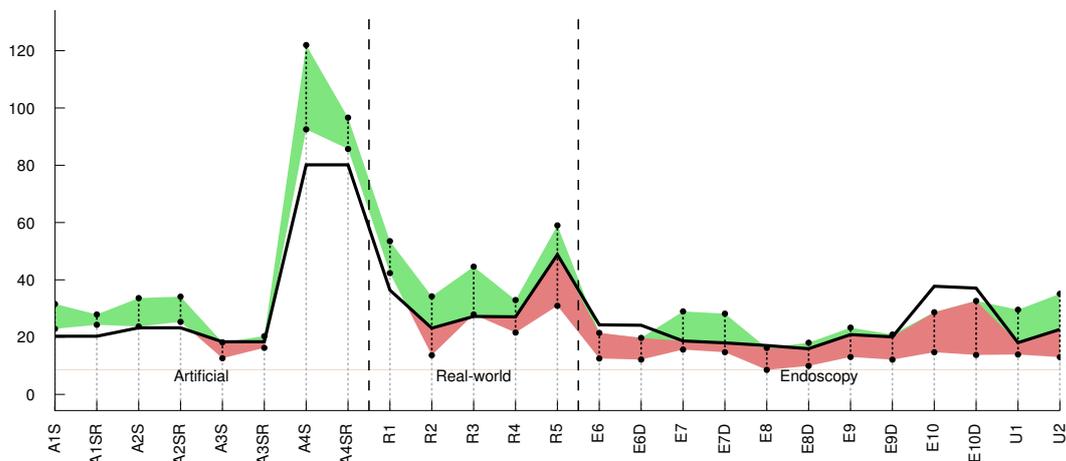


Figure 6. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (Metric Q, global shift and rotation estimation).

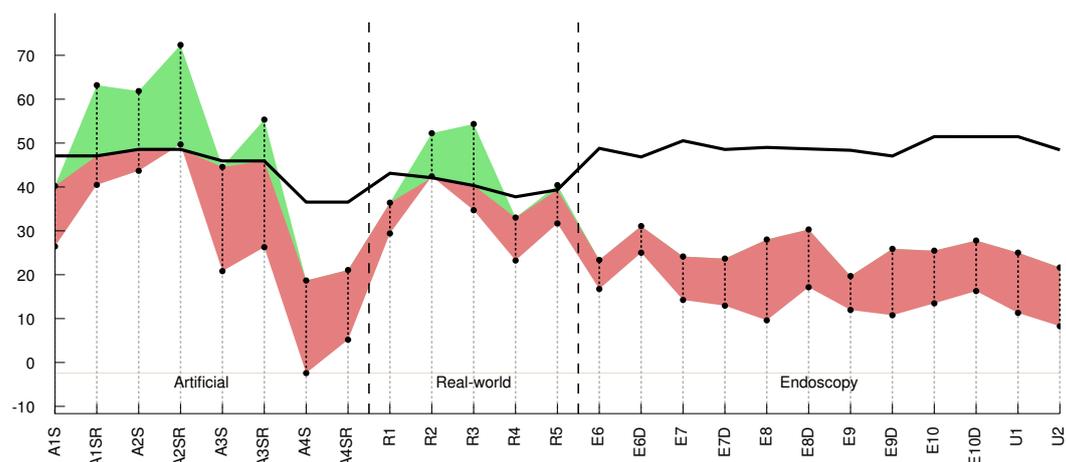


Figure 7. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (BRISQUE, global shift and rotation estimation).

plots, i.e. scores shown in red (green) are lower (higher) as compared to the scores for the result of upscaling the reference image, using bicubic interpolation. From Table V we notice that in most cases the SR methods are able to yield a higher quality score as compared to the score of bicubic interpolation. While for the real-world sequences this is always the case, there are some scores which are lower as compared to the bicubic score (in case of the endoscopic sequences noticeably more than in case of the artificial sequences). In case of Metric Q the overall picture changes significantly (see Table VI). While in case of the artificial and real-world sequences there are still quite a few improvements, the scores for the endoscopic sequences are in most cases below the bicubic scores. When using optical flow all methods seem to fail to produce a SR result of higher quality than the bicubic interpolation. In case of BRISQUE and DIIVINE (see Tables VII and VIII) things get even worse. While in case of the artificial and real-world sequences the number of scores above the bicubic score again dropped, there are almost no more improvements in case of the endoscopic images.

A different view on the results from Tables V to VIII is given in Table IX. This table indicates how often a certain SR method delivers the highest score among all SR methods, grouped by the quality metrics. From this

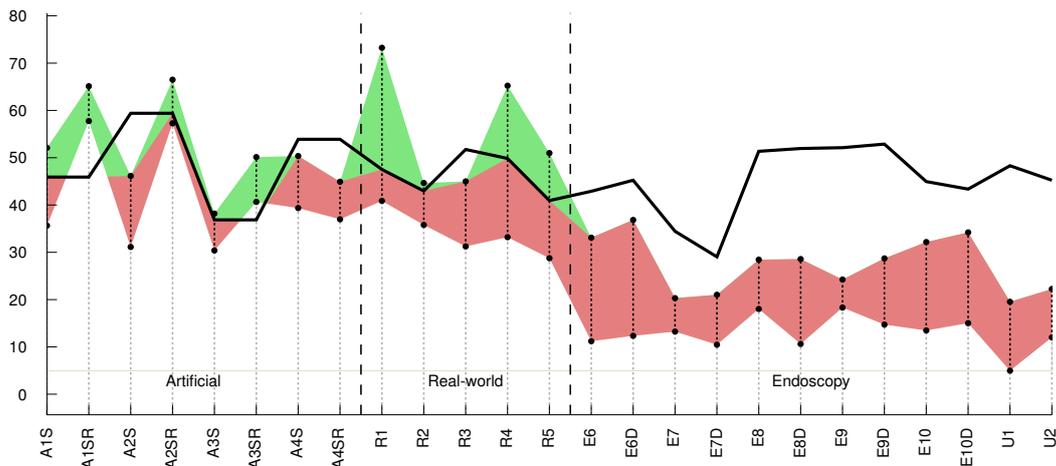


Figure 8. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (DIIVINE, global shift and rotation estimation).

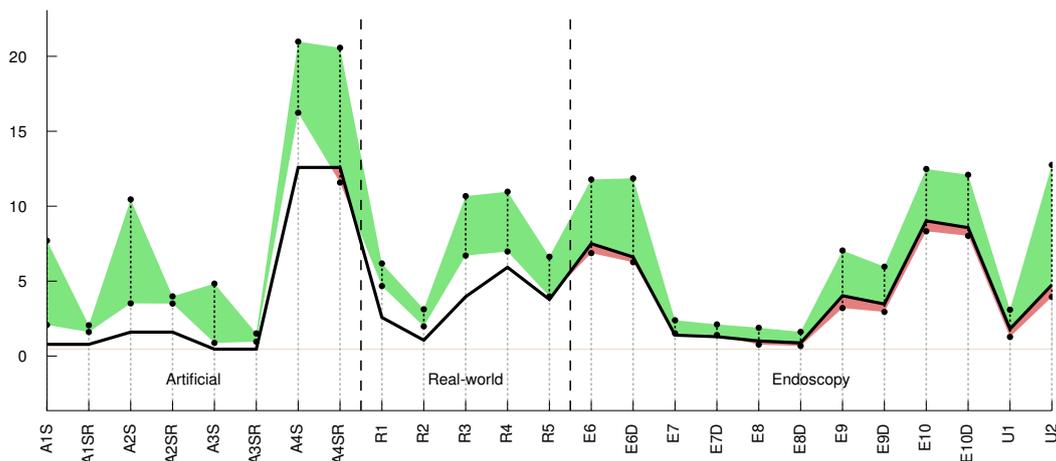


Figure 9. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (Anisotropic measure, optical flow estimation).

table we notice that when using the global motion estimation, there is a high disagreement between the different metrics on the visual quality of the SR results. The only agreement between the different metrics seems to be that the POCS method delivers the highest scores among all methods in rare cases only. When using optical flow instead, there is still a certain disagreement, but in this case all metrics agree that in most cases either S&A, IBP, or RSR deliver the highest visual quality.

#### A. Effects of Deblocking

Comparing the scores for the deblocked endoscopic sequences with the scores for the non-deblocked sequences, we notice that, according to the anisotropic measure, the visual quality of the SR results is always lower when deblocking is applied (see Table V). Metric Q and DIIVINE behave rather similar, as we can see from Tables VI and VIII. In both cases there is no clear tendency on whether applying deblocking has a positive or negative impact on the quality of the SR results. Nevertheless, for both metrics there are slightly more improvements in case of the optical flow estimation as compared to the global motion estimation. From Table VII we notice, that, compared to

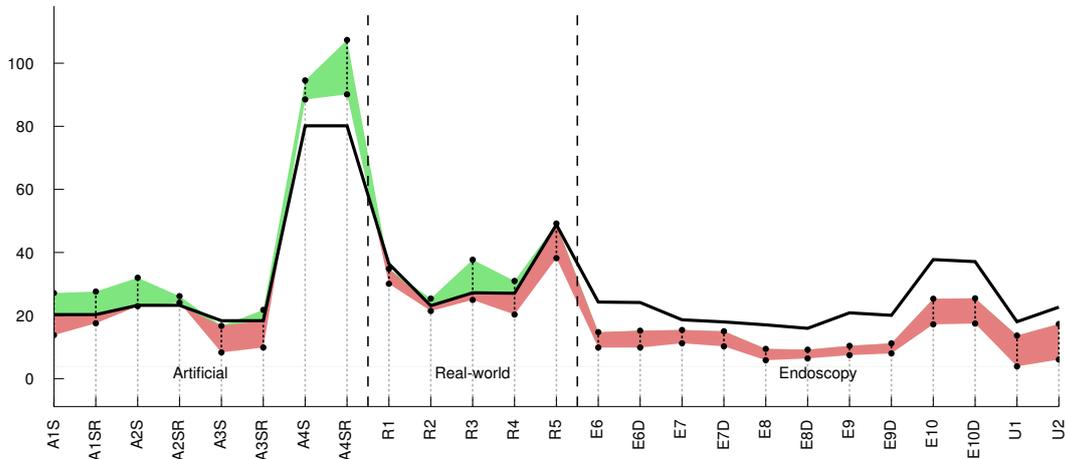


Figure 10. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (Metric Q, optical flow estimation).

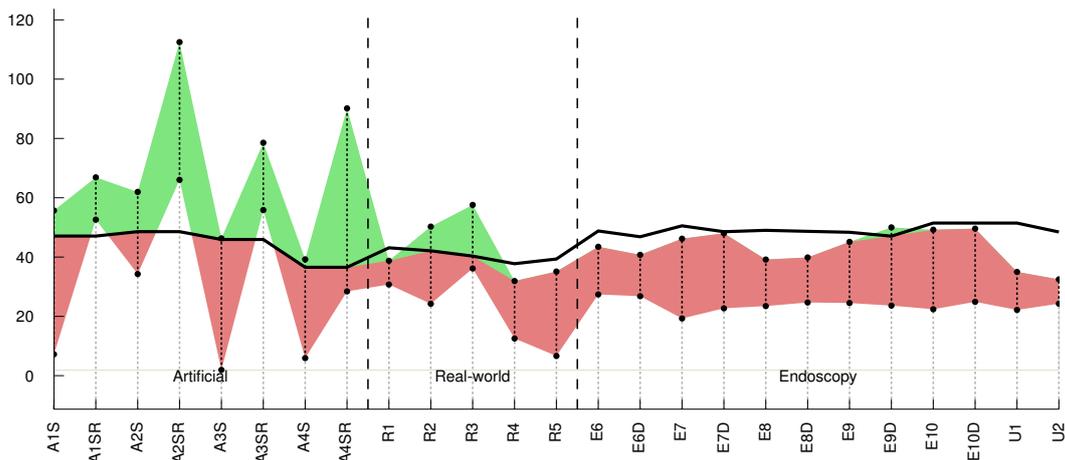


Figure 11. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (BRISQUE, optical flow estimation).

all other metrics, BRISQUE shows the highest number of quality score improvements after deblocking no matter which motion estimation method we use.

### B. Uncompressed Endoscopic Sequences

When comparing the scores for the uncompressed sequences against the scores for the other endoscopic sequences, we notice that the scores are not clearly higher (this applies to all combinations of quality metrics and motion estimation methods). However, considering the fact that we had only two sequences at hand for the evaluation process, a solid statement about whether uncompressed sequences deliver SR results of higher quality or not is not possible. In addition, the uncompressed sequences are not just uncompressed versions of other sequences in the dataset, but different sequences. This makes a comparison with the compressed sequences even more difficult.

### C. Shift vs. Shift and Rotation in Artificial Sequences

As already indicated in Section VIII-A, we used two different types of artificially generated image sequences for our experiments. Ones which exhibit translational motion only and ones which also exhibit rotation between

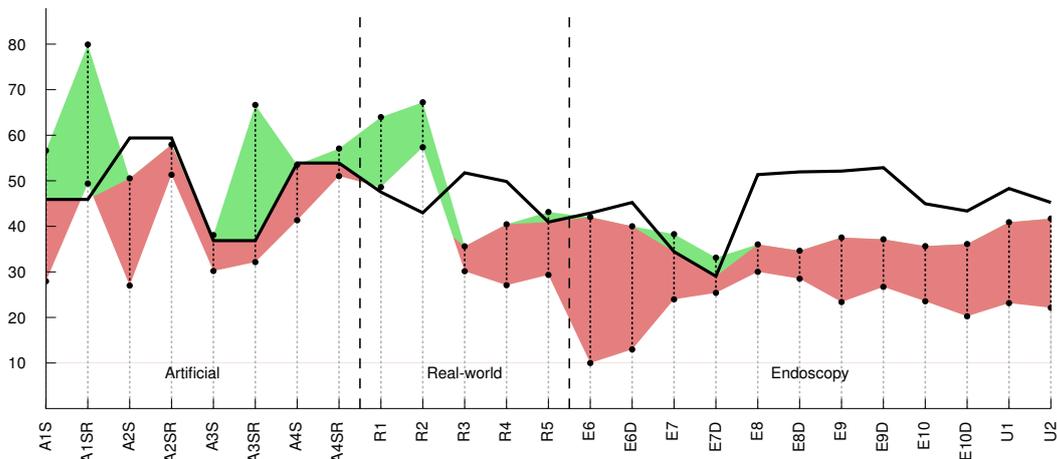


Figure 12. Minimum and maximum quality scores for the different sequences among all SR methods compared to the scores for a bicubic interpolation (DIIVINE, optical flow estimation).

Table IV

NUMBER OF CASES IN WHICH EACH OF THE SR METHODS DELIVERS A HIGHER SCORE AS COMPARED TO BICUBIC INTERPOLATION.

Metric	Sequence type (count)	Vandewalle					Optical flow				
		S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
Anisotropic	Artificial (8)	7	8	7	8	6	8	8	7	8	8
	Real-world (5)	5	5	5	5	5	5	5	5	5	
	Endoscopic (12)	2	12	0	12	6	2	12	5	12	12
Metric Q	Artificial (8)	6	7	6	6	6	6	3	7	3	3
	Real-world (5)	3	5	3	4	4	2	2	1	2	3
	Endoscopic (12)	0	7	0	4	4	0	0	0	0	0
BRISQUE	Artificial (8)	1	2	3	4	2	3	6	5	6	7
	Real-world (5)	1	3	1	2	2	1	2	0	2	2
	Endoscopic (12)	0	0	0	0	0	0	0	1	0	0
DIIVINE	Artificial (8)	4	3	4	2	4	1	3	3	3	3
	Real-world (5)	2	1	3	1	3	3	2	2	2	2
	Endoscopic (12)	0	0	0	0	0	0	1	2	1	0

the image sequence frames. From Table VI we notice that in case of Metric Q there is no clear tendency for the scores between shift-only sequences and sequences with rotation incorporated into the LR images generation process. In case of the anisotropic measure (see Table V) the scores for sequences with rotation are always lower as compared to the shift-only sequences. This is a bit more pronounced in case of the optical flow estimation. In case of BRISQUE and DIIVINE the sequences with rotation almost always produce higher scores as compared to the shift-only sequences (see Tables VII and VIII).

When comparing the scores for each metric between the two motion estimation methods evaluated there is no really big difference observable. This is quite surprising since one could expect higher scores in case of the global motion estimation since the LR sequence generation process perfectly corresponds to a global motion. In addition, the optical flow method estimates local motion which is prone to errors in case of a global motion in the data.

## X. CONCLUSION

From the results in Section IX we notice that in case of the artificial and real-world sequences there are indeed SR methods which are able to yield a higher quality in case of the SR reconstruction as compared to the simple bicubic interpolation. But we also notice that this highly depends on the quality metric used. Especially in case of the endoscopic sequences there is a high disagreement between the metrics.

While the SR algorithms seem to work quite well for artificial and non-medical real-world sequences, the overall picture is that they are not working with our endoscopic sequences. One might speculate that one possible reason for

Table V  
DETAILED SCORES OBTAINED WITH THE ANISOTROPIC METRIC.

ID	INT	HR	Vandewalle					Optical flow				
			S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
A1S	0.78	1.35	5.62	4.41	5.51	4.26	2.94	7.70	2.15	6.61	2.13	2.08
A1SR	0.78	1.35	2.63	2.96	2.62	2.94	1.95	1.92	2.06	1.60	1.96	1.97
A2S	1.60	3.55	6.90	4.67	6.61	4.55	3.88	10.45	3.65	9.21	3.58	3.52
A2SR	1.60	3.55	3.80	4.74	3.82	4.52	3.81	3.75	3.98	3.50	3.82	3.75
A3S	0.46	0.49	3.51	2.01	3.31	1.22	1.94	4.83	0.94	4.03	0.94	0.88
A3SR	0.46	0.49	2.34	1.84	2.21	1.82	1.16	1.51	1.04	1.39	0.96	1.02
A4S	12.59	17.13	15.41	17.02	15.64	18.97	12.15	19.53	20.98	16.23	20.67	19.39
A4SR	12.59	17.13	11.74	14.81	12.54	15.68	10.55	14.16	20.57	11.58	19.25	17.25
R1	2.58		4.48	9.03	4.99	8.33	7.63	5.54	6.18	4.67	5.98	6.15
R2	1.06		1.99	5.49	1.79	6.06	4.37	2.16	3.12	1.98	2.99	3.06
R3	3.96		10.70	7.51	9.89	6.83	7.50	10.67	6.72	9.77	6.70	6.77
R4	5.92		6.21	10.95	6.51	9.45	7.67	8.63	10.97	6.98	10.30	10.33
R5	3.78		4.79	9.56	4.78	10.12	6.46	4.99	6.63	3.98	6.30	5.58
E6	7.50		3.96	11.47	3.73	12.31	7.30	6.87	10.54	11.78	9.41	10.74
E6D	6.61		3.26	10.13	3.21	11.03	6.65	6.26	9.28	11.86	8.41	9.85
E7	1.40		1.34	4.20	1.05	4.21	1.67	1.58	2.39	1.52	2.24	2.22
E7D	1.29		1.20	3.60	0.96	3.61	1.56	1.44	2.11	1.41	2.05	2.00
E8	1.01		1.11	1.94	0.79	1.71	0.92	0.77	1.89	0.94	1.67	1.75
E8D	0.88		1.25	1.78	0.88	1.51	0.89	0.68	1.61	0.79	1.45	1.52
E9	4.02		2.11	7.88	1.71	9.45	2.95	3.20	6.82	3.63	5.99	7.04
E9D	3.47		1.77	6.32	1.42	7.65	2.55	2.95	5.96	3.26	5.17	5.92
E10	9.02		3.70	10.75	3.06	11.43	3.89	8.66	12.48	8.33	11.81	11.95
E10D	8.57		3.03	10.56	2.51	10.72	3.65	8.31	12.09	8.02	11.15	11.44
U1	1.81		1.38	2.88	1.16	3.00	2.53	1.28	3.08	1.79	2.72	3.09
U2	4.75		2.04	8.36	1.72	9.58	6.08	3.95	10.58	5.75	8.06	12.75

Table VI  
DETAILED SCORES OBTAINED WITH METRIC Q.

ID	INT	HR	Vandewalle					Optical flow				
			S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
A1S	20.31	24.71	22.89	31.58	23.61	29.58	27.09	27.14	13.93	21.80	13.81	14.25
A1SR	20.31	24.71	25.30	27.90	26.42	24.30	26.26	21.80	18.12	27.61	17.55	19.39
A2S	23.25	27.83	23.72	33.20	25.21	33.63	29.03	31.99	23.11	26.35	22.97	23.09
A2SR	23.25	27.83	25.33	34.12	28.15	32.38	29.04	26.16	24.70	24.58	24.10	25.22
A3S	18.38	10.95	14.30	18.17	15.61	12.62	16.59	16.74	8.37	12.47	8.33	8.93
A3SR	18.38	10.95	16.53	20.33	17.87	17.74	16.25	14.98	9.92	21.81	9.88	11.09
A4S	80.15	128.05	99.30	95.25	104.23	121.97	92.53	94.48	91.87	88.50	90.97	94.54
A4SR	80.15	128.05	87.66	85.69	89.51	96.62	87.62	101.49	99.18	90.11	91.13	107.35
R1	36.41		48.71	46.08	53.50	42.35	43.48	31.51	34.85	30.08	32.90	34.85
R2	23.13		13.67	34.23	16.62	31.81	29.63	24.08	25.39	21.46	24.41	24.46
R3	27.22		44.60	30.18	39.98	27.90	30.03	37.70	25.63	33.38	24.96	26.07
R4	27.08		29.76	30.88	32.93	21.58	28.04	26.60	30.59	20.33	27.11	30.98
R5	48.83		30.92	51.52	33.63	59.00	38.69	46.54	44.22	38.20	41.79	49.17
E6	24.29		12.55	21.42	13.20	21.47	14.00	14.75	11.88	12.19	9.86	10.92
E6D	24.14		12.15	19.72	12.83	17.47	13.24	15.21	11.94	13.17	9.87	10.99
E7	18.66		16.31	28.96	15.65	19.88	18.67	15.44	13.99	11.81	11.22	13.76
E7D	17.99		14.95	28.19	14.72	20.21	18.91	15.03	12.73	11.27	10.25	12.32
E8	17.05		12.41	16.22	12.55	8.56	11.49	7.88	9.45	5.85	6.40	8.89
E8D	15.96		13.88	18.04	13.87	9.94	13.07	7.90	9.20	6.46	6.42	8.79
E9	20.87		13.59	23.30	13.48	19.03	13.06	10.27	10.41	7.46	8.40	9.66
E9D	20.08		12.13	20.89	12.29	16.10	13.36	10.91	11.21	8.01	8.45	10.60
E10	37.75		14.74	28.69	15.74	24.27	14.72	25.34	21.91	20.62	17.22	19.77
E10D	37.10		13.77	32.64	14.75	25.46	15.77	25.46	21.56	20.88	17.47	19.90
U1	18.05		14.67	29.56	13.93	23.30	26.17	7.82	13.68	3.90	8.58	13.07
U2	22.68		13.90	35.12	13.01	28.15	27.37	10.07	17.32	6.06	12.09	14.96

this are the compression artifacts present in our sequences which cancel out important high frequency information

Table VII  
DETAILED SCORES OBTAINED WITH THE BRISQUE METRIC.

ID	INT	HR	Vandewalle					Optical flow				
			S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
A1S	47.08	62.12	29.90	26.45	40.23	38.01	27.79	7.22	55.29	17.94	55.70	50.24
A1SR	47.08	62.12	42.49	46.82	49.39	63.19	40.48	61.66	60.08	66.88	58.74	52.61
A2S	48.57	71.41	43.68	53.22	53.45	61.85	53.17	36.74	58.85	34.27	59.01	61.96
A2SR	48.57	71.41	57.72	54.93	58.91	72.36	49.69	85.36	66.03	112.48	66.23	67.87
A3S	45.93	90.60	20.82	35.29	23.65	44.57	29.44	1.94	43.80	5.05	43.96	46.29
A3SR	45.93	90.60	26.27	35.47	30.86	55.35	28.89	61.61	61.34	78.54	60.64	55.86
A4S	36.55	29.10	-0.13	4.96	-2.41	0.24	18.68	5.92	32.80	39.22	31.97	30.86
A4SR	36.55	29.10	9.46	6.71	6.39	5.19	21.07	28.42	37.73	90.14	37.07	40.07
R1	43.12		31.78	29.40	36.41	34.96	33.91	38.75	36.89	30.75	38.00	36.17
R2	42.10		43.17	43.19	52.25	48.25	42.41	37.02	50.26	24.23	49.95	48.94
R3	40.32		34.70	54.35	36.99	52.68	54.06	44.97	56.26	36.15	57.58	54.75
R4	37.76		28.64	25.49	33.02	23.24	29.26	29.57	30.00	12.54	29.36	31.91
R5	39.34		31.68	40.42	31.89	38.88	35.49	28.58	35.11	6.64	32.18	32.09
E6	48.80		22.98	19.86	23.37	16.75	21.30	43.46	27.39	30.28	31.74	27.47
E6D	46.85		31.08	25.01	29.77	26.71	29.84	40.75	26.98	28.79	31.95	26.79
E7	50.53		24.15	14.72	21.53	15.07	14.25	46.19	19.45	36.29	26.44	19.29
E7D	48.55		23.67	12.94	19.68	16.02	17.89	48.01	23.57	37.09	31.90	22.71
E8	49.01		28.02	18.21	27.81	9.61	22.32	39.16	24.03	39.15	26.62	23.46
E8D	48.68		28.54	17.15	30.31	21.55	23.25	39.13	24.75	39.86	32.37	24.67
E9	48.36		17.39	11.96	18.04	15.55	19.69	41.49	24.50	45.12	32.61	25.69
E9D	47.05		25.90	10.77	23.49	21.60	24.69	41.35	23.63	50.00	32.86	26.79
E10	51.47		25.21	17.86	21.64	13.48	25.48	49.21	23.60	35.26	26.68	22.40
E10D	51.46		26.73	18.33	23.99	16.27	27.78	49.57	25.98	36.00	30.81	24.90
U1	51.46		23.04	14.92	25.02	18.54	11.30	34.99	22.17	27.28	26.04	23.24
U2	48.45		16.58	17.21	13.49	21.64	8.26	30.83	24.27	29.24	24.55	32.47

Table VIII  
DETAILED SCORES OBTAINED WITH THE DIIVINE METRIC.

ID	INT	HR	Vandewalle					Optical flow				
			S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
A1S	45.91	73.86	47.14	35.64	46.57	45.86	52.09	27.94	56.24	28.47	56.63	45.74
A1SR	45.91	73.86	65.11	62.37	57.75	61.97	64.62	49.38	59.18	79.91	58.56	49.53
A2S	59.40	82.12	32.72	37.84	31.13	46.15	45.59	29.69	49.52	27.00	49.70	50.54
A2SR	59.40	82.12	66.50	57.27	63.03	57.89	62.26	57.93	56.11	51.32	54.86	54.92
A3S	36.86	52.78	32.04	38.19	33.70	30.39	34.14	30.20	31.41	38.08	31.49	37.11
A3SR	36.86	52.78	46.37	40.98	40.66	50.13	41.27	32.15	32.99	66.65	33.56	34.30
A4S	53.90	56.75	47.72	39.36	50.36	46.08	47.51	43.50	53.52	41.34	53.16	50.09
A4SR	53.90	56.75	41.14	36.98	44.01	44.15	44.92	52.61	57.07	51.02	55.78	54.12
R1	47.52		40.86	68.68	44.48	73.25	57.46	48.61	63.97	50.85	62.65	62.27
R2	42.97		36.38	36.71	44.66	38.92	35.78	60.06	66.64	57.35	67.23	66.54
R3	51.75		31.24	40.03	34.21	44.99	37.73	30.85	33.78	30.15	35.61	33.27
R4	49.84		56.56	39.04	65.22	33.19	51.90	40.45	27.08	27.18	27.21	29.42
R5	40.94		46.82	32.62	50.98	28.75	41.09	43.14	36.16	29.34	36.04	38.69
E6	42.90		16.15	11.20	33.09	15.21	17.79	42.03	17.36	36.60	16.74	9.99
E6D	45.24		12.34	12.51	36.84	14.71	13.89	40.01	20.13	36.47	18.62	12.99
E7	34.44		18.42	20.14	20.30	17.39	13.23	30.41	25.17	38.28	28.96	23.97
E7D	29.05		15.92	19.91	20.87	21.02	10.46	25.40	30.52	33.13	32.32	28.37
E8	51.36		18.01	26.61	28.45	27.23	20.71	32.58	30.72	36.03	30.01	31.02
E8D	51.95		10.62	24.79	27.39	28.56	15.33	28.50	31.58	34.64	32.24	31.14
E9	52.12		18.96	20.54	24.24	18.31	19.16	34.87	24.27	37.54	27.83	23.38
E9D	52.88		15.25	18.45	28.70	17.83	14.71	27.96	27.19	37.15	29.08	26.71
E10	44.95		16.79	13.47	32.16	17.05	23.28	35.65	23.55	29.35	23.86	24.16
E10D	43.37		15.02	17.62	34.20	16.24	22.45	36.12	21.63	30.95	20.25	20.73
U1	48.31		10.94	8.52	19.53	16.35	4.97	40.90	24.05	27.25	25.91	23.15
U2	45.24		12.20	12.01	22.24	19.78	13.83	41.69	22.13	22.90	23.39	22.89

important for a successful SR reconstruction.

When applying a deblocking algorithm to the endoscopic sequences, the quality scores for the SR results do not

Table IX  
NUMBER OF CASES IN WHICH EACH OF THE SR METHODS DELIVERS THE HIGHEST SCORE AMONG ALL METHODS.

Metric	Sequence type (count)	Vandewalle					Optical flow				
		S&A	IBP	RSR	ROBZ	POCS	S&A	IBP	RSR	ROBZ	POCS
Anisotropic	Artificial (8)	4	2	0	2	0	4	4	0	0	0
	Real-world (5)	1	2	0	2	0	1	4	0	0	0
	Endoscopic (12)	0	2	0	10	0	0	7	2	0	3
Metric Q	Artificial (8)	0	5	0	3	0	4	0	2	0	2
	Real-world (5)	1	1	2	1	0	1	2	0	0	2
	Endoscopic (12)	0	11	0	1	0	6	6	0	0	0
BRISQUE	Artificial (8)	0	0	1	5	2	0	0	5	1	2
	Real-world (5)	0	2	3	0	0	1	2	0	1	1
	Endoscopic (12)	5	0	3	1	3	8	0	3	0	1
DIIVINE	Artificial (8)	2	1	1	2	2	1	2	3	1	1
	Real-world (5)	0	0	3	2	0	2	1	0	2	0
	Endoscopic (12)	0	0	10	2	0	6	0	6	0	0

get consistently higher. This is most likely caused by the interpolation inherent to the deblocking, which eliminates even more details. As a consequence, a meaningful SR reconstruction gets even more challenging, if not impossible.

In case of the uncompressed endoscopy sequences the scores are also consistently lower as compared to the bicubic scores. One reason might be that even in case of the uncompressed sequences we notice a slight blur in the images. As a result there might be not enough high frequency content present for a meaningful SR reconstruction.

The bottom-line of our experimental results is that at least in case of our endoscopic sequences, applying the evaluated reconstruction-based SR techniques is not meaningful. But especially in case of uncompressed videos future work should include additional experiments on more sequences to be able to make more robust statements on the usability of case of such sequences. In addition, this work was limited to a selection of reconstruction-based algorithms. Since there also exist other types of SR algorithms, future work should also include an evaluation of other types (e.g. learning-based approaches).

## XI. ACKNOWLEDGMENTS

This work has been supported by the Austrian Science Fund (FWF) under Project No. TRP-206.

## REFERENCES

- [1] M. G. Kang and S. Chaudhuri, "Super-resolution image reconstruction," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 19–20, May 2003.
- [2] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito, "Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis," *IEEE Communications, Speech and Vision*, vol. 140, no. 1, pp. 19–25, Feb. 1993.
- [3] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, T. S. Huang, Ed. JAI Press Inc., London, 1984, pp. 317–339.
- [4] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei, "Computer-assisted pit-pattern classification in different wavelet domains for supporting dignity assessment of colonic polyps," *Pattern Recognition*, vol. 42, no. 6, pp. 1180–1191, Sep. 2008.
- [5] —, "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.
- [6] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Color treatment in endoscopic image classification using multi-scale local color vector patterns," *Medical Image Analysis*, vol. 16, no. 1, pp. 75–86, Jan. 2012.
- [7] —, "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.
- [8] K. Duda, T. Zieliński, and M. Duplaga, "Computationally simple super-resolution algorithm for video from endoscopic capsule," in *Proceedings of the International Conference on Signals and Electronic Systems (ICSES'08)*, 2008, pp. 197–200.
- [9] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.
- [10] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, Sep. 2011.
- [11] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 47–57, Aug. 2004.
- [12] P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on Applied Signal Processing (special issue on Super-resolution)*, vol. 2006, pp. 1–14, 2006.
- [13] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, May 2009.

- [14] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proceedings of the European Conference on Computer Vision (ECCV'04)*, 2004, pp. 25–36.
- [15] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunk: combining local and global optical flow methods," *International Journal of Computer Vision (IJCV)*, vol. 61, no. 3, pp. 211–231, 2005.
- [16] 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.
- [17] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 9, pp. 694–702, 1978.
- [22] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A*, vol. 6, no. 11, pp. 1715–1726, Nov. 1989.
- [23] M. Häfner, M. Liedlgruber, and A. Uhl, "POCS-based super-resolution for HD endoscopy video frames," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems (CBMS'13)*, Jun. 2013.
- [24] M. Elad, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1187–1193, 2001.
- [25] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.
- [26] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, pp. B42–B51, 2007.
- [27] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3116–3132, 2010.
- [28] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [29] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
- [30] P. List, A. Joch, J. Lainema, G. Bjøntegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 614–619, 2003.