POCS-based Super-Resolution for HD Endoscopy Video Frames

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

M. Liedlgruber and A. Uhl
Multimedia Signal Processing and Security Lab
Department of Computer Sciences
University of Salzburg, Austria
{mliedl, uhl}@cosy.sbg.ac.at

Abstract

The main question we try to answer in this work is whether it is feasible to employ super-resolution (SR) algorithms to increase the spatial resolution of endoscopic high-definition (HD) images in order to reveal new details which may have got lost due to the limited endoscope magnification inherent to the HD endoscope used (e.g. mucosal structures). For this purpose we propose a SR algorithm, which is based on the Projection onto convex sets (POCS) approach. This algorithm is able to avoid over-sharpening, which is often seen with other methods. Since POCS-based approaches are iterative ones, we also propose an adaptive iteration scheme.

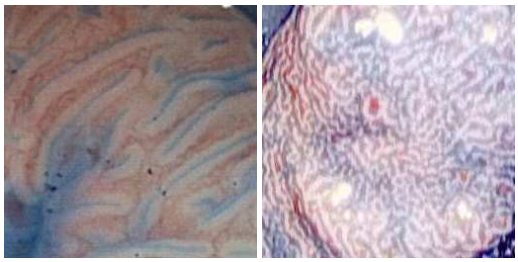
We compare the quality of the reconstruction of our method against the quality achieved by other SR methods. This is done on standard test images as well as on images obtained from endoscopic video frames.

We show that, while our approach produces competitive results on standard test images, we are not able to reveal new details in endoscopic images for various reasons.

1 Introduction

In the past we developed different approaches for the classification of colonic polyps (e.g. [6, 5, 7, 8]). 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 a HD endoscope. While such an endoscope provides a roughly four times higher image resolution as compared to the previously used zoom-endoscope, it provides 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). Fig. 1 shows two tubulovillous adenoma, one captured with a zoom-endoscope (not HD, but zooming-capabilities) and one captured with a HD endoscope (no zooming possible). We immediately notice the dramatic difference between these images in terms of the details visible.



(a) Zoom (b) HD

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

While it would be illusory to expect high-resolution (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.

Up to our knowledge currently there exists only one work which tries to tackle the SR problem with endoscopic images [1]. But in this work the authors test their algorithm on low-resolution (LR) images generated from a single video frame by shifting it into different directions and downscaling the shifted frames. Our work, in contrast, aims at reconstructing a HR image from several successive LR

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

images which are not synthetically generated.

The remaining part of this work is organized as follows: in Section 2 we give a brief introduction to super-resolution and POCS, followed by a description of our proposed method in Section 3. In Section 4 we describe the experimental setup used and present the results obtained. We conclude the paper in Section 5.

2 Super-resolution with POCS

2.1 Principles of Super-resolution

Fig. 2 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 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. 3. 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. 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. 2 this may include translation and rotation (either local or global).

- **Fusion**
  Based on the estimated motion the LR images are fused onto a HR grid. That is, the information from the different LR images is combined into an HR image, matching the resolution of the desired HR image.

- **Image restoration**
  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 the estimation of the motion is a crucial step. It is even imperative that the motion is estimated with sub-pixel accuracy. However, this can get quite complicated, depending on the motion present in the LR images. Apart from that, sub-pixel shifts between LR images are needed. 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. 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, which 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.

Based on the observation model from Fig. 2 the formation process for an LR image y_k based on an HR image X can be formulated in matrix notation as [3]:

$$y_k = D_k B_k W_k X + \eta, \tag{1}$$

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. η denotes a normally distributed additive noise. The matrices D_k^T, B_k^T, and W_k^T represent the reverse operations, i.e. upscaling, sharpening, and inverse motion, respectively.

2.2 POCS

The idea of POCS was introduced to image processing by the work in [13]. 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

$$||X - PX|| = \underset{x \in C}{\operatorname{argmin}} ||X - x|| \qquad \forall X \in H. \tag{2}$$

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

$$\forall X_1, X_2 \in C : X_3 \in C \tag{3}$$

with

$$X_3 = \lambda X_1 + (1 - \lambda) X_2 \qquad \forall \lambda \in [0, 1]$$

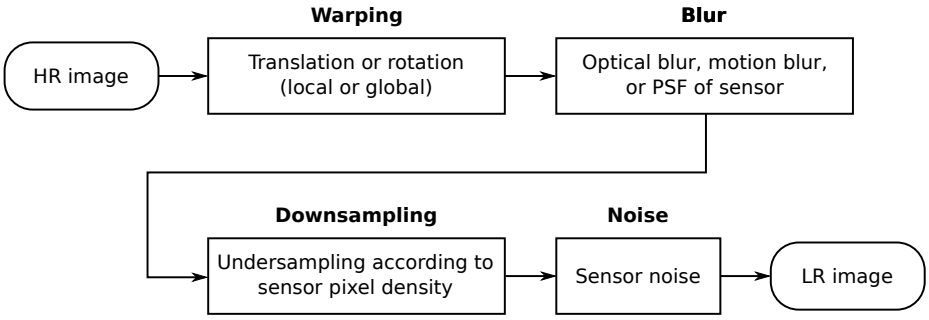


Figure 2. The observation model usually assumed in case of SR algorithms.

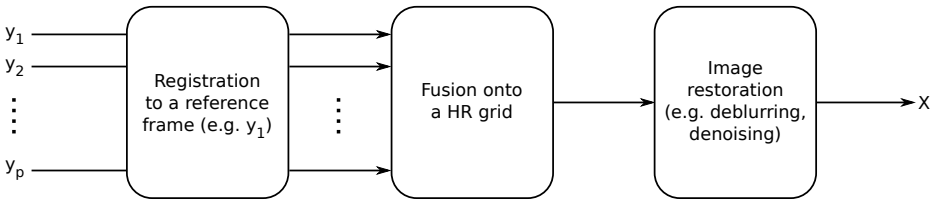


Figure 3. The reconstruction scheme common to reconstruction-based SR algorithms.

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

$$\hat{X}^{n+1} = P_m P_{m-1} \ldots P_1 \hat{X}^n \tag{4}$$

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. (4). 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. [12, 3]). In the following we list the convex set constraints which are used throughout this work (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:

$$C_A = \{X : B_p^{\min} \leq X_p \leq B_p^{\max}, \forall p = 1, \ldots, M\}, \tag{5}$$

  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

$$X'_p = P_A X_p = \begin{cases} B_p^{\min}, & X_p < B_p^{\min} \\ X_p, & B_p^{\max} \leq X_p \leq B_p^{\max} \\ B_p^{\max}, & X_p > B_p^{\max} \end{cases}, \tag{6}$$

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

- **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

$$C_C = \{X : ||X - g|| \leq \epsilon_R\}, \tag{7}$$

  where g denotes some reference image and ϵ_R denotes the maximum distance allowed between X and g. One

possible projection for this convex set is:

$$X' = P_C X = \begin{cases} X, & ||X-g|| \leq \epsilon_R \\ X - \Delta_C(X-g), & ||X-g|| > \epsilon_R \end{cases} \quad (8)$$

where Δ_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

$$C_C^k = \{X : ||D_k B_k W_k X - y_k|| \leq \epsilon_R\}, \quad (9)$$

with the projection given in Equ. (10).

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.

3 Proposed Method

Since the motion, usually observed in endoscopy videos, is quite complex (e.g. a combination of rotation, zoom, and non-uniform shifts), we decided to use an optical flow method to estimate the motion between two images. The optical flow method we use for our experiments is based on [10]. This 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 at the coarsest pyramid 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.

In our implementation of the POCS approach 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. (10)), we also employ a correction amplitude constraint, which prevents over-sharpening:

$$C_{CA}^k = \{X : ||X - W_k^T D_k^T y_k|| \leq \epsilon_{CA}\}, \quad (11)$$

where ϵ_{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 given in Equ. (12).

In our implementation the initial estimate $\hat{X}^0$ is set to an upscaled version of y_1. Based on Equ. (4) we end up with

$$\hat{X}^{n+1} = P_A^p P_{CA}^p P_C^p \cdots P_A^1 P_{CA}^1 P_C^1 \hat{X}^n, \quad (13)$$

where $P_{(\cdot)}^k$ denotes the respective projection onto the k-th LR image. It must be noted that in our implementation projections affect only those pixels which need correction (according to the respective thresholds).

Since the proposed method is an iterative process, 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 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 an adaptive termination criterion in this work. 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:

$$\Delta_n = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{X}_i^n - \hat{X}_i^{n-1}\right)^2} \quad (14)$$

2. Based on the differences for iterations two and three, we compute a threshold value ϵ_{iter}, which is used later to decide upon termination of the iterative process:

$$\epsilon_{\text{iter}} = \kappa|\Delta_3 - \Delta_2|, \quad (15)$$

where κ 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:

$$\text{crit}(\Delta_n, \Delta_{n-1}) = \begin{cases} 1, & |\Delta_n - \Delta_{n-1}| \geq \epsilon_{\text{iter}} \\ 0, & |\Delta_n - \Delta_{n-1}| < \epsilon_{\text{iter}} \end{cases} \quad (16)$$

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

$$X' = P_C^k X = \begin{cases} X, & ||D_k B_k W_k X - y_k|| \leq \epsilon_R \\ X + W_k^T B_k^T D_k^T \left(\Delta_C (y_k - D_k B_k W_k X)\right), & ||D_k B_k W_k X - y_k|| > \epsilon_R \end{cases} \quad (10)$$

$$X' = P_{CA}^k X = \begin{cases} X, & ||X - W_k^T D_k^T y_k|| \leq \epsilon_{CA} \\ X + \left(\Delta_C (W_k^T D_k^T y_k - X)\right), & ||X - W_k^T D_k^T y_k|| > \epsilon_{CA} \end{cases} \quad (12)$$

Since the value of Δ_1 highly depends on the initial HR estimate, we use the values Δ_2 and Δ_3 to compute the threshold ϵ_{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 Δ_2 and Δ_3.

4 Experimental Setup & Results

4.1 Experimental Setup

To be able to compare the visual results of our SR method, we carried out additional experiments using other methods as well. These are Shift-and-Add (S&A) [2], regularized super-resolution (ROB) [4], iterated back-projection (IBP) [9], and robust super-resolution (ROBZ) [14]. In case of color images we apply the SR algorithms to the intensity component of the images only (in this case the color components are simply upscaled using bicubic interpolation).

For our experiments we evaluated the SR algorithms on different 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 sequences extracted from endoscopic videos. 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 case of the artificial sequences we generated eight LR frames based on a HR image according to the observation model in Equ. (1) (the sequences are subject to shift and rotation). The endoscopic sequences are based on successive frames of videos acquired during colonoscopy sessions between the years 2011 and 2012 at the Department for Internal Medicine (St. Elisabeth Hospital, Vienna) using a HD colonoscope (Pentax HiLINE HD+ 90i Colonoscope) with a video resolution of 1280×1024. To reduce the computational demand for the SR methods we chose positions from which we manually extracted 256×256-pixel patches which serve as LR image (the position remained the same in case of a single sequence). Details on the LR sequences used can be found in Table 1. In this table the column "Synthetic" indicates whether the respective LR sequence has been generated from a single HR image.

Name	ID	Color	Synthetic
Airplane	A1	✓	✓
Boat	A2	✗	✓
Elaine	A3	✗	✓
Chart	A4	✗	✓
Carphone	R1	✓	✗
City	R2	✓	✗
Container	R3	✓	✗
Garden	R4	✓	✗
Endoscopy 1	E1	✓	✗
Endoscopy 2	E2	✓	✗
Endoscopy 3	E3	✓	✗
Endoscopy 4	E4	✓	✗

Table 1. Details on our LR image sets used.

Since we have no reference HR image in case of endoscopic images, we use a reference-free quality metric. The metric used is called BRISQUE [11]. It is a so-called natural scene statistics-based approach. It fits a statistical distribution to features extracted from different images. Once the features for training images have been obtained, support vector regression (ϵ-SVR) is used to learn a mapping from feature space to quality scores. The learned model is then used to predict the quality score for an image with an unknown quality.

Since the quality metric relies on training data for SVR, we generated a different set of sequences (similar to the one in Table 1). This set has 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 the metric. A higher score returned by BRISQUE means a higher quality, where the score is usually in the range between 0 and 100 (there might be some outliers leaving this range).

4.2 Results

From the results presented in Table 2 we notice that our method is in many cases able to produce a higher score as compared to the interpolation result (the first LR frame after upsampling, using bicubic interpolation). But in case of the endoscopy sequences our method fails to produce a higher

ID	INT	HR	S&A	IBP	ROB	ROBZ	Proposed
A1	46.8	67.3	45.2	60.5	5.8	60.6	47.5
A2	50.6	81.6	60.2	57.1	44.3	58.6	62.1
A3	47.6	101.4	50.0	78.5	59.8	77.3	66.3
A4	19.8	2.6	8.8	21.5	45.4	16.3	24.4
R1	51.3	-	45.0	62.0	52.4	65.5	63.0
R2	32.6	-	47.1	40.3	26.5	42.8	43.2
R3	28.6	-	12.8	36.0	-9.3	37.5	34.9
R4	52.4	-	48.6	23.5	14.6	30.3	34.2
E1	49.2	-	45.9	31.3	32.9	41.9	30.1
E2	48.4	-	35.5	24.3	37.5	36.1	23.8
E3	49.8	-	49.0	25.1	51.4	29.3	25.1
E4	49.3	-	52.7	26.6	56.5	37.0	27.1

Table 2. Detailed metric results for the SR algorithms (INT denotes interpolation and HR denotes the ground truth).

score than interpolation. The picture is similar for the other SR algorithms: while in quite a few cases these methods are able to produce an image with a higher score in case of artificial and real-world sequences, they fail almost always in case of the endoscopy sequences. It is also no surprise that almost all methods fail to produce a higher quality as compared to the HR images. Only in case of the "Chart"-sequence (A4), all methods produce a higher quality than the original HR image. This may be due to the fact that this sequence is made of monochrome images, resulting in very sharp SR results.

The proposed approach is able to outperform all other methods in case of the "Boat"-sequence (A2). In case of some other sequences (i.e. Chart, Carphone, City, and Garden) our method outperforms quite a few other methods. But in case of the endoscopy sequences it is almost always inferior as compared to the other methods.

5 Conclusion

In this paper we proposed a POCS-based algorithm for super-resolution. We showed that the results are in many cases better or at least competitive as compared to other SR algorithms. But we also showed that all SR algorithms (including the proposed one) almost always fail to improve the visual quality of endoscopic images. This may be due to the fact that these sequences are extracted from compressed videos which show a high amount of compression artifacts (i.e. blocking). Apart from this, the endoscopy frames do not contain very much high frequency content.

In future work we will therefore evaluate the SR algorithms also on uncompressed endoscopy video frames. In addition we will investigate whether applying deblocking algorithms allows to improve the quality of the SR results. Since in this work we relied on a single metric only for the assessment of the visual quality of the SR results, we will also investigate additional metrics.

References

[1] 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)*, pages 197–200, 2008.

[2] M. Elad. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. *IEEE Transactions on Image Processing*, 10(8):1187–1193, 2001.

[3] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Transactions on Image Processing*, 6(12):1646–1658, Dec. 1997.

[4] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, Aug. 2004.

[5] 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*, 42(6):1180–1191, Sept. 2008.

[6] 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*, 12(4):407–413, Dec. 2009.

[7] 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*, 16(1):75–86, Jan. 2012.

[8] 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*, 107(3):565–581, Sept. 2012.

[9] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing*, 53(3):231–239, Apr. 1991.

[10] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, Massachusetts Institute of Technology, May 2009.

[11] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[12] H. Stark and P. Oskoui. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, 6(11):1715–1726, Nov. 1989.

[13] D. C. Youla. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Transactions on Circuits and Systems*, 25(9):694–702, 1978.

[14] 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)*, number 1, pages 645–650, 2001.