Super-Resolution Techniques for Minimally Invasive Surgery

Vincent De Smet[1], Vinay P. Namboodiri[1] and Luc Van Gool[1,2]

[1] ESAT-PSI/IBBT, K.U.Leuven, Belgium.
{vincent.desmet,vinay.namboodiri,luc.vangool}@esat.kuleuven.be
[2] Computer Vision Laboratory, BIWI/ETH Zürich, Switzerland

Abstract. We propose the use of super-resolution techniques to aid visualization while carrying out minimally invasive surgical procedures. These procedures are performed using small endoscopic cameras, which inherently have limited imaging resolution. The use of higher-end cameras is technologically challenging and currently not yet cost effective. A promising alternative is to consider improving the resolution by post-processing the acquired images through the use of currently prevalent super-resolution techniques. In this paper we analyse the different methodologies that have been proposed for super-resolution and provide a comprehensive evaluation of the most significant algorithms. The methods are evaluated using challenging in-vivo real world medical datasets. We suggest that the use of a learning-based super-resolution algorithm combined with an edge-directed approach would be most suited for this application.

Keywords: super-resolution, minimally invasive surgery, endoscopy

1 Introduction

The use of video cameras in medical procedures has introduced a wide range of techniques for minimally invasive diagnosis and surgery (MIS) like laparoscopy and thoracoscopy, or more generally endoscopy, to modern medicine. These procedures, also referred to as keyhole surgery, allow a surgeon to make only a small incision through which a camera can be inserted. The surgeon can then perform the operation or diagnosis based on a video feed which is displayed on a monitor in the operating room. The cameras used for these procedures typically record in a relatively low resolution, making the magnified video on the display look blurry and making it hard for the surgeon to distinguish some details such as smaller veins. Imaging techniques like super-resolution (SR) [11] can offer a significant visual improvement in cases like these. Super-resolution is a term used to refer to image interpolation techniques which try to recover as much missing detail as possible, while often also reducing noise artifacts. In this paper we introduce both classical and state-of-the-art SR techniques to the field of medical imaging for minimally invasive surgery, and provide a comprehensive evaluation of these methods. Related research has been done by Greenspan [8], who describes the

application of classical reconstruction-based SR algorithms to different medical imaging modalities: magnetic resonance imaging (MRI), functional MRI (fMRI) and positron emission tomography (PET). Kouam and Ploquin [9] present a super-resolution algorithm for ultra-sound images based on the analysis of the point spread function. Robinson et al. [12] explore specialized reconstruction-based methods on applications in the area of X-ray digital mammography and optical coherence tomography (OCT). The use of SR for minimally invasive surgical applications has been touched upon by Lerotic and Yang [10], who apply an SR method called Projection Onto Convex Sets (POCS) in order to investigate the use of fixational movements for robotic assisted MIS. The use of SR, while promising, also has its inherent challenges that need to be specifically addressed. In this paper we advocate the use of SR for endoscopy and consider the issues involved therein.

2 Super-Resolution Methods

We start by briefly describing the different classes and the specific algorithms that we use for our evaluation of super-resolution in the context of minimally invasive surgical procedures. The selection of methods encompasses most of the currently prevalent techniques for super-resolving images.

2.1 Edge-Directed Interpolation

One class of super-resolution algorithms, commonly referred to as interpolation methods, relies on techniques for improved interpolation beyond the basic bicubic interpolation. These methods are based on the idea that interpolation of images by focusing on reducing the blurring of edges results in visually pleasing results. We explore two methods that focus on improved interpolation of edges.

Gradient Profile Prior Sun *et al.* [15] propose an approach to explicitly ensure the sharpness of gradients in super-resolution by using a Gradient Profile Prior (GPP). The authors model the distribution of gradient profiles in an image as a Generalized Gaussian Distribution (GGD), which depends on a shape parameter λ and a sharpness parameter σ. By examining 1 million gradient profiles from 1000 natural images they show that a λ value of 1.6 gives a good estimate for the shape of the GDD for natural images, independent of the image resolution. The value of σ for the GGD that models the low-resolution gradient profiles is estimated from the gradient profiles in the upsampled version of the input image. The σ for the high-resolution GGD is estimated based on the σ for low-resolution, or it can be estimated from the input image, as done in [16]. Each pixel of the upsampled input image is associated with its nearest gradient maximum. Its gradient intensity is then multiplied by the ratio of the high-resolution and low-resolution GGDs, resulting in sharper edges across the image. This procedure is fast and can be applied independently at each pixel and is therefore easily parallelizable.

Total Variation Regularization An approach proposed by Chatterjee *et al.* [3] considers the use of total variation (TV) regularization combined with spectral and spatial data fitting constraints. The use of regularization is required to address the under-constrained nature of the SR problem. This regularization is performed in a TV framework by including constraints for minimizing the sum of the L_1 norm of the gradient over the whole image. This constraint preserves edges while keeping noise to a minimum. The use of spatial data-fitting constraints with TV alone does not suffice in preserving textures. In [3], the authors use an additional spectral data constraint to solve this problem.

2.2 Reconstruction-Based SR

When multiple degraded observations are available, i.e. frames of a video sequence, reconstruction-based super-resolution methods can be used. These methods are based on the idea that each of the frames in a sequence, when interpolated and registered to the first frame with sub-pixel accuracy, can bring extra information about the scene. A backprojection constraint prevents the result from straying too far from the input when blurred and downsampled back to its original resolution. This constraint is usually accompanied by a smoothness prior, which enforces the smoothness inherent to natural images. These priors are needed because of the under-determined nature of this inverse problem. The reconstruction-based super-resolution problem is usually solved as a maximum likelihood estimation (MLE) or a maximum a posteriori (MAP) estimation [1].

For our application the most challenging step is the registration of the different frames. The related reconstruction-based works have been demonstrated to work on data with relatively straightforward motion, where global transformations suffice to model the movement. When dealing with the type of video sequences acquired from endoscopic surgery, these models fail to capture the inter-frame movement of the scene. The highly non-rigid transformations of internal tissue and organs, independent on each other and the rest of the scene, requires a more complex transformation model. In order to model these movements we opt to work with an optical flow registration algorithm here. Our implementation is based on the work by Fransens *et al.* [5]. They use a Bayesian framework to solve the super-resolution problem in a MAP sense. The inversion process, in which the super-resolved image is estimated, is interleaved with the computation of a dense optical flow field. Possible occlusions are handled with a visibility map, which is optimized in an expectation-maximization process.

2.3 Learning-Based SR

Learning-based SR methods try to learn the correlation between low-resolution and high-resolution details, based on a set of example images. This image database is constructed from low-resolution/high-resolution image pairs, created by blurring and downsampling a high-resolution image to create the low-resolution counterpart. One of the foremost papers using this approach is the work by Freeman *et al.* [6]. We give an overview of this method and a related method by Chang *et al.*, based on Locally-Linear Embedding (LLE) [2] in this section.

Exemplar-Based Learning Freeman *et al.* [6] propose the use of corresponding LR-HR image patches from a database as exemplars for super-resolution. Their method divides the images into small patches, which then function as nodes in a Markov Random Field (MRF). For each patch of an input image a match is found in the database, after which the corresponding high-resolution patch is taken from the database and used as a candidate in the MRF. The arcs between high- and low-resolution nodes represent a data compatibility function, while the arcs between high-resolution nodes can be interpreted as a continuity function. The data compatibility function is obtained by finding k-nearest neighbors for each low-resolution input patch and calculating their distance. The continuity function between neighboring high-resolution patches is obtained by evaluating the overlap between the high-resolution candidates. The MRF is then optimized by minimizing the weighted sum of these two cost functions using belief propagation.

In our implementation of this method we use patch sizes of 7×7. To find the k-nearest neighbors we use the Adaptive Locality-Sensitive Hashing method as proposed by Wang *et al.* [17]. Instead of using the patch-based continuity criterion in belief propagation, we use denser patch sampling where each pixel in the input image is used as the center for a patch. We find the nearest neighbor for each patch and the corresponding high-resolution patch is directly used in the result image. We average the contribution from all overlapping patches for each pixel in the high-resolution domain. This modification achieves results which are comparable to those attained when using belief propagation and allows the algorithm to be parallelized and executed much faster. While this method requires a database of images, we choose to use only the LR input image as a database. This ensures that we avoid hallucination effects due to spurious matches in a huge database. This concept of single image learning has been confirmed recently by Glassner *et al.* [7].

Neighborhood Embedding of Exemplars The computation time and perceptual quality of results of patch-based methods depends on the size of the learning database. The approach proposed by Chang *et al.* [2] does not depend on this size. Their approach involves learning a compact dictionary of LR-HR patches using a clustering algorithm from the database. For each patch of the LR image a set of k-nearest neighbors is found in the dictionary. k reconstruction weights that minimize the reconstruction error of each LR patch are then obtained using the LLE dimensionality reduction approach [13]. The authors postulate that these weights can be used to also obtain the corresponding HR patch by combining the related k HR patches from the dictionary. This is done for each patch with a large overlap between patches for regularization. In our implementation we learn the dictionary from the LR input image itself.

2.4 Combining Edge-Directed And Learning-Based SR

A recent method by Tai *et al.* [16] proposes a combination of the GPP method with the exemplar-based learning method. The authors observe that while the

GPP method enhances the edges, the textures are not suitably enhanced by this method. Therefore, they propose a technique in which the texture is also enhanced by using a single exemplar image. The combination is done using a simple way that ensures that the details are maximally enhanced. This is done by noting that the gradient profile prior does not enhance the texture suitably but does enhance the edges more than the exemplar technique alone. Therefore a combination of the maximal gradient from the exemplar and gradient profile prior technique ensures that both sharp edges and visually pleasing textures are obtained. A back-projection constraint ensures that the result is true to the original input image. The authors have provided a partial implementation of their method which we have fully implemented and evaluated for the application described in this paper.

3 Evaluation

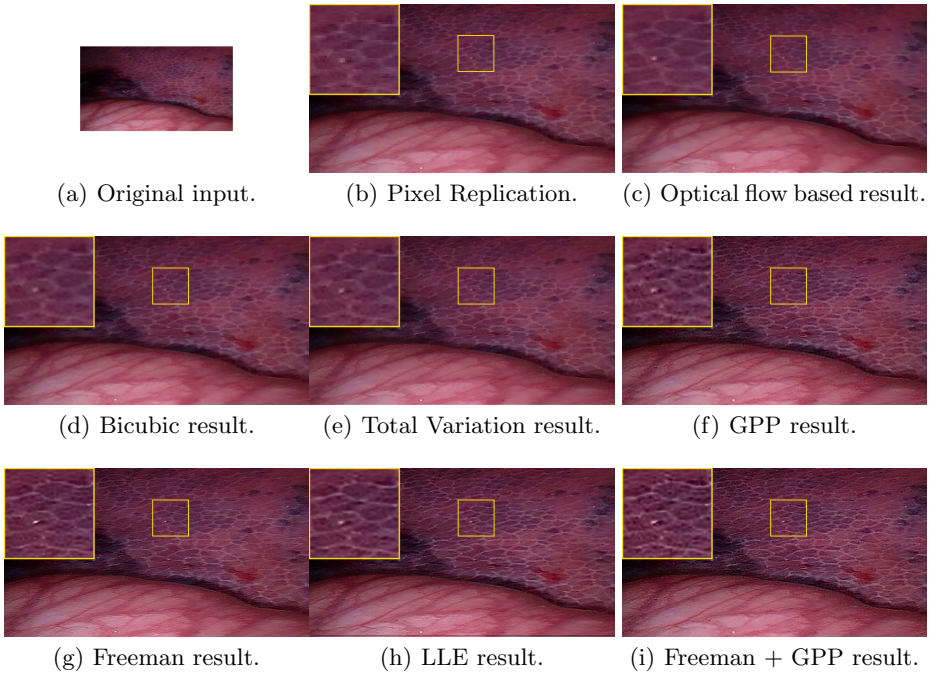
We evaluate the techniques discussed in the previous section on several laparoscopic/endoscopic datasets, and on a cardiovascular scene. We present visual results for a limited set of examples in the paper. Further visual results are available at our website[3]. We also provide quantitative results for the various datasets in subsection 3.4.

3.1 In-vivo Porcine Liver

The results for the various super-resolution techniques applied on an image of an in-vivo laparoscopic procedure on a porcine liver are shown in Fig. 1. The results are shown for a 2× upsampling of the original image. The original image has textured regions as well as smooth regions with veins as can be seen in Fig. 1(a). We have processed this image with several techniques which use only this image as input. Only the multiframe optical flow-based SR method, discussed in subsection 2.2, uses several (10) frames from the original video. The dataset is drawn from the "VIP Laparoscopic / Endoscopic Video Dataset"[4] [14]. For each result we show a closer zoom of a part of the image for clearer comparison between the various super-resolution algorithms. As can be seen from the results, the TV method adds some slight sharpness to the edges. The GPP method shows more sharpening, but also enhances noise/artefacts, resulting in a sharper but slightly noisier result. The learning-based methods (Freeman and LLE) hallucinate more well-defined edges but, because their image database consists only of the LR input image, also hallucinate small scale texture from the noise. This is slightly more noticable in the Freeman result because the LLE-based algorithm uses a linear combination of patches rather than the actual database patches themselves. Freeman's algorithm on the other hand shows slightly sharper edges, because its resulting patches are not smoothed by the averaging that occurs

[3] http://homes.esat.kuleuven.be/~vdesmet/endoscopy/
[4] http://www.doc.ic.ac.uk/~pmountne/vision/

when using a linear combination of patches. The combination algorithm joins the advantages of the GPP and Freeman algorithms, and adds the most detail. The reconstruction-based result gives a slight sharpening, albeit not as much as the learning-based methods. An interesting feature of this result is that it additionally performs a denoising, because its result is created from multiple images. It needs to be noted however that the movement between video frames in this example consisted mainly of the textured region in the upper half of the image and the region in the lower half moving independently of each other. Thus the registration of the frames with optical flow works relatively well here. Larger and more complex movements can lead to failed registration and thus very unreliable results, as will be shown in subsection 3.3.



(a) Original input. (b) Pixel Replication. (c) Optical flow based result.

(d) Bicubic result. (e) Total Variation result. (f) GPP result.

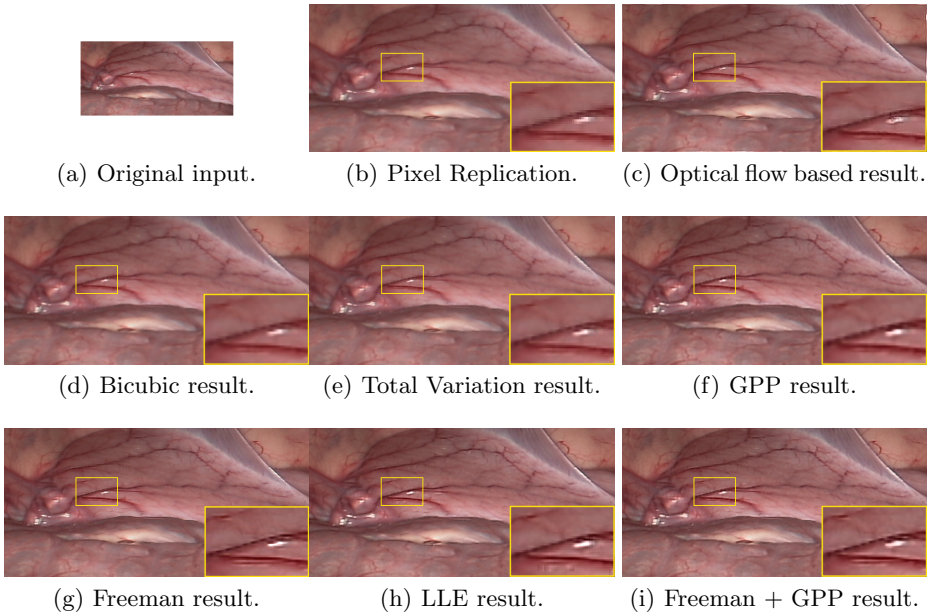
(g) Freeman result. (h) LLE result. (i) Freeman + GPP result.

Fig. 1. Results for in-vivo porcine liver.

3.2 In-vivo Porcine Cardiac Surface

A second example from the same video dataset collection is shown in Fig. 2. This example is taken from an in-vivo laparoscopic procedure exploring a porcine cardiac surface. Again a 2× upsampling factor was used. Similar conclusions can be drawn here, the combined edge- and exemplar-based technique results in the sharpest image with a clear contrast at the edges. Note here that in the bicubic and the edge-directed results the lower veins in the zoomed in part of the image

could be mistaken for a single vein, while the learning-based results make the distinction clearer. Again the reconstruction-based algorithm also gives a sharp result with high fidelity to the scene (using 10 inputframes). The movement in this example consisted of the beating of the cardiac surface, which is modelled well by the optical flow algorithm.



(a) Original input. (b) Pixel Replication. (c) Optical flow based result.

(d) Bicubic result. (e) Total Variation result. (f) GPP result.

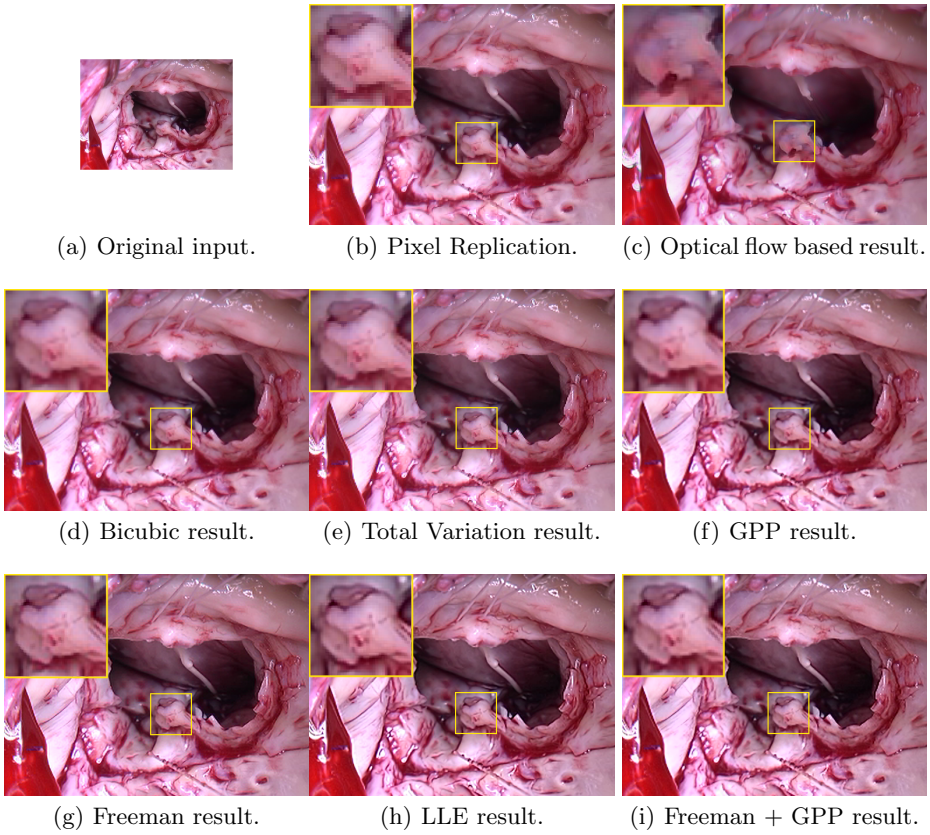
(g) Freeman result. (h) LLE result. (i) Freeman + GPP result.

Fig. 2. Results for in-vivo porcine cardiac surface.

3.3 In-vivo Human Cardiovascular Procedure

The previous examples showed results for laparoscopic procedures. In order to show that the super-resolution techniques do not limit themselves to laparoscopic and endoscopic data, we show results for a cardiovascular scene. The original image can be seen in Fig. 3(a) and the results for the various techniques are shown in Fig. 3(b-g). An immediately observable fact is that the reconstruction-based method performs poorly on this kind of dataset, even though it has more data available (12 frames) than the other techniques. The reason is that this dataset has very complex deformable motion with non-uniform and variable texture information in the scenes. There is also more movement than in the previous examples. Therefore, the accurate sub-pixel registration that is required for the multiframe methods to work is not available. The other methods perform well. The edge-based methods preserve the strong edges as can be seen in the zoomed in view Fig 3(e-f). However, results from the learning-based methods shown in Fig. 3(g-h) show more improvement due to the introduction of more high-frequency information. The combined edge- and learning-based method performs

similarly well and has slightly stronger edges than the learning-based method alone.



(a) Original input. (b) Pixel Replication. (c) Optical flow based result.

(d) Bicubic result. (e) Total Variation result. (f) GPP result.

(g) Freeman result. (h) LLE result. (i) Freeman + GPP result.

Fig. 3. Results for in-vivo human cardiovascular procedure.

3.4 Quantitative Results

We obtain quantitative results using Peak Signal to Noise Ratio (PSNR) by downsampling the input images by a factor of two and then upsampling them using the various techniques presented in section 2. The results are provided in table 1. The quantitative results however do not correspond well to the perceived visual quality of the results. For instance, the bicubic interpolation method, which generates substantial blurring, provides the highest PSNR result. This disconnect between PSNR and perceived visual quality is well known and has been explored by Eskicioglu and Fisher [4]. While PSNR when used for measuring substantial noise does provide quantitative results that correspond to visual quality, this metric does not suit the comparison of improvements that

are obtained using super-resolution algorithms. We are currently investigating appropriate metrics that can be used for evaluating super-resolution algorithms.

Method	Porcine Liver	Porcine Cardiac Surface	Human Cardiovascular
Bicubic	36.2205	38.7317	27.6883
TV	31.4599	31.3720	23.2105
LLE	33.0907	33.7748	25.3247
Freeman	33.3718	35.7921	27.2585
GPP	34.4335	35.0422	25.9193
GPP + Freeman	32.3606	33.1568	25.3031
Optical Flow	33.0403	31.7700	22.6763

Table 1. PSNR values for all methods.

4 Discussion

In the previous section, we provided qualitative and quantitative results for various super-resolution algorithms. We are interested in ensuring their use for critical minimally invasive surgical applications. This necessitates consideration of several other factors.

The hallucination-based algorithms show the ability to introduce high-frequency detail. However, one concern in using these algorithms is that they could potentially hallucinate non-existent image features that are artefacts rather than actual image features. The other class of methods that shows good results is edge-based methods. These do not introduce artefacts and they enhance the edges. However, they do not enhance the texture suitably. Therefore, the combination of edge and exemplar-based learning algorithms seems to be the most promising for surgical applications. The edge-based method can be used to attenuate the introduction of artefacts from the learning-based method. Currently, we used the method for combination as proposed by Tai *et al.* [16]. It would be interesting to explore the combination of these techniques to further improve the results and to make it more robust for medical applications. The reconstruction-based method relies heavily on accurate sub-pixel registration of images and is therefore unreliable and best avoided for medical applications, unless the inter-frame scene movement is not too complex.

Another concern for medical applications is the need for real time processing of the algorithms. The edge-based algorithm is highly efficient and can potentially be implemented in real time by exploiting the high degree of parallelizability. The learning-based algorithms, despite also having potential for parallelization, are more complicated. However, the neighborhood embedding method [2] is more scalable as it depends on learning a dictionary of fixed size. For medical applications, a dictionary can be learned offline that is most suited for a specific kind of operation. These methods, when implemented efficiently, can potentially run in real time by using either a learned dictionary or a small database and by exploiting the potential for parallelization. The reconstruction-based method is again at a disadvantage here, as it requires a high computational effort for both the calculation of the dense optical flow field and the actual super-resolution.

To conclude, we propose the use of super-resolution for minimally invasive surgical applications. As has been shown here, this appears feasible and promising. We argue that the use of a combined edge- and learning-based technique would be most suited for the application.

References

1. Capel, D., Zisserman, A.: Computer vision applied to super resolution. IEEE Signal Processing Magazine 20(3), 75–86 (May 2003)
2. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: IEEE Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 275–282 (2004)
3. Chatterjee, P., Namboodiri, V.P., Chaudhuri, S.: Super-resolution using sub-band constrained total variation. In: First International Conference on Scale Space and Variational Methods in Computer Vision, (SSVM). pp. 616–627 (2007)
4. Eskcioglu, A.M., Fisher, P.S.: Image quality measures and their performance. Communications, IEEE Transactions on 43(12), 2959–2965 (1995)
5. Fransens, R., Strecha, C., Van Gool, L.J.: Optical flow based super-resolution: A probabilistic approach. Computer Vision and Image Understanding 106(1), 106–115 (2007)
6. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. International Journal of Computer Vision 40(1), 25–47 (2000)
7. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: IEEE International Conference on Computer Vision (ICCV) (2009)
8. Greenspan, H.: Super-resolution in medical imaging. The Computer Journal 52(1), 43–63 (2009)
9. Kouame, D., Ploquin, M.: Super-resolution in medical imaging: An illustrative approach through ultrasound. In: Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 249–252 (2009)
10. Lerotic, M., Yang, G.Z.: The use of super resolution in robotic assisted minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention. pp. 462–469 (2006)
11. Milanfar, P.: Super-Resolution Imaging. CRC Press (2010)
12. Robinson, D., Chiu, S., Lo, J., Toth, C., Izatt, J., Farsiu, S.: Novel applications of super-resolution in medical imaging. In: Milanfar, P. (ed.) Super-Resolution Imaging, chap. 13, pp. 383–412. CRC Press (2010)
13. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
14. Stoyanov, D., Scarzanella, M., Pratt, P., Yang, G.Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention 2010, vol. 6361, pp. 275–282 (2010)
15. Sun, J., Sun, J., Xu, Z., Shum, H.Y.: Gradient profile prior and its applications in image super-resolution and enhancement. IEEE Transactions on Image Processing 20(6), 1529–1542 (2011)
16. Tai, Y.W., Liu, S., Brown, M.S., Lin, S.: Super resolution using edge prior and singale image detail synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
17. Wang, Q., Tang, X., Shum, H.: Patch based blind image super resolution. In: IEEE Conference on Computer Vision. pp. 709–716 (2005)