Improving super-resolution image reconstruction by in-plane camera rotation

Stefan Bonchev
Faculty of mathematics and informatics
Sofia university
Sofia, Bulgaria.
stefan_b@fmi.uni-sofia.bg

Kiril Alexiev
Institute of Parallel Information Processing
Bulgarian Academy of Sciences
Sofia, Bulgaria
alexiev@bas.bg

Abstract – *In a digital optical imaging system, image resolution is constrained by several factors, including focus plane array pitch and optics. Super-resolution approaches aim at overcoming some of these limits by incorporating additional information of the object and/or combining several pictures of the same object, taken with some displacements between each other. This paper considers the second class of methods. The obtainable resolution improvement in this case has an upper limit, determined by the signal-to-noise ratio of the image taken. Moreover, some fourier spectrum frequencies below this limit are unrestorable. Here an approach is introduced to overcome this by active control of the camera movements. An experiment, verifying the approach is presented.*

Keywords: Super-resolution, Image restoration, Image processing, Fourier spectrum

1 Introduction

Super-resolution presents set of methods for increasing the resolution of an imaging system beyond some of its inherent limits. Their aim is to reconstruct high-resolution (HR) image of some object given a set of its low-resolution (LR) images. Publications on the subject describe several approaches to the problem.

The simpler ones employ augmented interpolation of a single input image only. In contrast to 'pure interpolation', in them, the values of the pixels of the resultant higher-resolution image are generated adding some more information about its semantics. This information could be derived manually or automatically. In [3], [19] and [10] several sets of hand-crafted rules targeting edge preservation based on human intuition are suggested. Another approach is to incorporate additional information coming from real data. As an example, in [18], edge preservation is done by a neural network, trained on pairs of LR and HR images.

When employing these approaches, the additional information in the resultant image is pre-determined in some way. They may work well with a given class of images (e.g. cartoons) but lack universality. Also, most often, they lack efforts to get physically valid resultant image.

Another set of approaches reconstruct an HR image by fusing the information from several LR images by using some model describing physically the imaging process. This could be achieved in several ways. The most widespread way is to 'invert' the process of obtaining a couple of LR images from a single HR one [5]. Here each LR image serves as a constraint to the HR one. This method has an advantage, as it could be naturally extended by employing additional (statistically derived or obtained in another way) prior information about the desired image in the form of additional constraints [2]. Two ways of expressing the constraints for the HR image are most commonly used, leading to the projection onto convex sets (POCS) and maximum a posteriori (MAP) approaches. In the POCS approach, each LR image restricts the HR image within the limits of convex set of the possible ones. Additional information about the HR image is included easily, if it could be expressed as such belonging-to-a-convex-set restriciton. The convexity of the sets servers as a guarantee that subsequent projections will converge to an image belonging to the sets intersection. The MAP (or Bayesian) approach, expresses each restriction as a functional on the HR image. This enables wider types of additional information to be included and works naturally with prior information gathered statistically. The resultant image in the MAP approach is derived by minimising these functionals, usually in mean square sense. Also, some other approaches exist - e.g the authors in [1] use non-uniform interpolation followed by deconvolution, but they lay beyond the scope of this paper.

An approach that fuses several LR images is harder to implement and is computationally more expensive than working with a single image only, but it has the principle advantage of using the information about the object coming from several of its images and hopefully

can result in better HR image.

In this paper, an improvement to a method from the second class is suggested by controlling camera position when taking frames. Similar successfull attempts are done already, described in [4], where the optical sensor is shifted using micro-actuators.

Here we suggest to take several sets of frames, rotating the camera between them. As it is discussed later, this gives an advantage over using only sensor shift motions.

The orgainsation of this paper is as follows: In section 2 some theoretical results about the limitations of SR reconstruction are presented and discussed. A not widely covered in the literature on the subject limitation is pinpointed, that could not be overcome by in-plane camera shifts only. In section 3 a mathematical model of the imaging process is presented. Using it, the advantage of translation-and-rotation over purely translational motion is motivated. In section 4 some experimental results are shown, confirming the validity of the approach. Conclusion and some possible directions for future work are presented in section 5.

2 Limitations of super-resolution reconstruction

The question of limitations of the super-resolution process is not entirely covered in the literature. Using space domain analysis, the authors of [2] show, that the problem does not have a unique solution when the zoom factor (the ratio between the LR and HR pixel size) is integer. Also, they prove the problem is ill-posed in other cases and its condition numer increases rapidly with the increase of the zoom. In [11] a more in-depth analysis is done and a conclusion is given that the maximal theoretically attainable zoom factor is 5.7.

Both publications share the consideration that low-resolution images are directly derived from the high-resolution one. This consideration, despite useful in its particular context, does not reflect the fact, that the irradiation of the focal plane array presents a continuous function (at the scale we are working at). Taking this into account gives a more precise view of the process and enables the proposed way to improve super-resolution reconstruction.

In [20] a dependence of the maximum obtainable zoom factor on the ratio of pixel size to pixel step is derived. The criterion for determining this maximum is based on the signal-to-noise ratio deterioration which is due to attenuation of the high frequencies when integrating the irradiance over the pixel area.

In all publications the authors implicitly assume that the boundary of the super-resolution zoom is a single one, and that resolution could be achieved on all coarser scales. Unfortunately, as the real data comes from integrating the incoming irradiance over non-zero sized pixels, some frequencies of it's spectrum are lost, resulting in 'gaps' in resolution under this absolute limit. This effect could not be overcome by acquiring more images under in-plane camera translations, as the set of non-recoverable frequencies remains the same. However, as later shown, it could be avoided or mitigated by in-plane camera rotation, as this kind of movement reduces or eliminates the unrecoverable frequency set under the presented imaging model.

3 Imaging model

3.1 Image formation and restoration

In the subsequent sections we will use the following multi-frame imaging model: let's have N images of a single scene, and for the k-th one, the focus plane array (FPA) registers some irradiance, which is integrated over every pixel and results in a raster image represented by the $r \times s$ matrix of values $y_k[m,n]$, $m = 1 \dots r$, $n = 1 \dots s$, $k = 1 \dots N$. The sensor resolution is determined by its dimensions divided by its pixel count in horizontal and vertical direction, resulting in the pixel size T (without losing generality, we consider square pixels only). This presents a good approximation of modern sensor matrices. If we imagine that we have integrated $I = I_1(\cdot)$ (the index is unimportant and is arbitrary chosen) over a sensor with smaller pixels we may get another, possibly bigger $p \times q$ matrix of values $x[i,j]$ representing in an image of the scene with higher-resolution.

In the case, when the object's lightfield is almost constant between the frames and the change of camera's focal point position is small, we could assume, that

$$I_k(u,v) = I(\varphi_k(u,v), \psi_k(u,v)) \tag{1}$$

where φ_k and ψ_k present an invertible map between the HR and k-th LR image coordinates and have continuous partial derivatives on the whole image or on a set of 'image patches'. The question of finding φ_k and ψ_k is the subject of many research efforts, but goes beyond the scope of this paper. Some approaches to acquiring and using the model of motion we consider could be found in [17] and [6].

In this case, as shown in [5], the super-resolution task could be expressed as

$$\overline{y} = A\overline{x} \tag{2}$$

in the space domain, or [12] as

$$\mathcal{F}[\overline{y}] = A\mathcal{F}[\overline{x}] \tag{3}$$

in the frequency domain. Here $\overline{y}$ denotes a vector of all LR pixel values, $\overline{x}$ - a vector of all HR pixel values, and $\mathcal{F}[a]$ denotes the Fourier spectrum of a. As mentioned in [20], by solving (2) or (3) one does - explicitly or implicitly - de-aliasing followed by de-convolution.

It's very common these tasks to be ill-posed or underdetermined. The reason for this emerges from the nature of the imaging process [2] and, when dealing with real data, from sensor noise, quantization error and error in determining φ_k and ψ_k.

Some authors ignore the ill-posedness [9], some take care of it. As an example, [13] and [14] employ special set-theoretic approach for solving (2). Other authors choose to use regularization, using smoothness [16, 7] or domain speciffic functional [8].

The approach in this paper seeks improvement in HR image reconstruction by influencing φ_k and ψ_k(thus also A) by using active camera rotation.

3.2 Assumptions

The considerations in the following secions are made under a number of assumptions about the particular conditions under which the imaging process is done. Some of them are inherent in the optical imaging process, or are easily achievable. Other come from the speciffic use of active camera concerned in this article.

Due to the non-ideal optical system, I could be considered a band-limited function. Also, as we choose the sample count in $\overline{x}$, we could think that it has enough samples for the perfect linear reconstruction of I. As mentioned above, we'll consider I to be constant between frames. A way to extend the imaging model used in this article to global linear changes of I is implemented in [15]; non-global changes of I due to bigger camera or object movements would shift the task from restoring an image to restoring some properties of the photographed object.

Also, we assume that the coordinate maps φ_k and ψ_k represent a translation followed by a rotation. This is an acceptable approximation for the situation when the real camera is subjet to a combination of the following kind of movements: small translational movement parallel to the focal plane, small pan/tilt movement and/or arbitraty rotation arround the camera principal optical axis.

3.3 Imaging in the space domain

Under the assumptions mentioned above, the imaging process could be modeled as a three stage one. First, the coordinates of the intensity function I are substituted leading to $I_k(u,v) = I(\varphi_k(u,v), \psi_k(u,v))$. Then I_k is convolved with a 2D rectangle function and finally sampled. This presents in another way the process, when for the k-th image we get the intensity $y_k[m,n]$ of the pixel with coordinates m, n by integrating I over its corresponding area.

Collecting the values of the pixels $y_k[m,n]$ for all k, m and n into the vector $\overline{y}$ we could write the equation

$$\overline{y} = \mathcal{A}[I] \tag{4}$$

where $\mathcal{A}$ presents a linear map from the space of I to $\mathbb{R}^{srN}$.

Under our assumption that I is perfectly linearly reconstructible given HR image pixel values, (4) is equivalent to the system of equations

$$A\overline{x} = \overline{y} \tag{5}$$

where A is an $u \times (pq)$ matrix, $u = srN$, and $\overline{x}$ is a vector of length pq presenting the collection of HR image pixel values [5].

Given $\overline{y}$, (4) has (unique) solution for I iff (5) has (unique) solution for $\overline{x}$. This also means, that the null spaces of $\mathcal{A}$ and A are isomorphic and consequently, reducing the dimensionality of the null space of $\mathcal{A}$ should extend the set of recovarable images by reducing the null space of A.

We can also write the similar to (5) equation for each LR image:

$$A_k\overline{x} = \overline{y}_l \tag{6}$$

Up to row reordering, A equals the vertical concatenation of A_1 through A_N :

$$A\bar{x} = \begin{pmatrix} A_1 \\ \dots \\ A_N \end{pmatrix} \bar{x} = \begin{pmatrix} \bar{y}_1 \\ \dots \\ \bar{y}_N \end{pmatrix} = \overline{y} \tag{7}$$

Also, under the assumptions mentioned above, there exist similar linear maps $\mathcal{A}_1$ through $\mathcal{A}_N$ (corresponding to A_1 through A_N respectively) such that

$$\mathcal{A}_1[I] = \overline{y}_1, \ \dots, \ \mathcal{A}_N[I] = \overline{y}_N \ , \ \mathcal{A}[I] = \begin{pmatrix} \overline{y}_1 \\ \dots \\ \overline{y}_N \end{pmatrix} \tag{8}$$

Taking into account that $\overline{y} = \mathcal{A}[I]$ presents the concatenation of $\overline{y}_l = \mathcal{A}_l[I]$ for $l = 1 \ \dots \ N$, it is obvious that the null space of $\mathcal{A}$ equals the intersection of the null spaces of $\mathcal{A}_1$ through $\mathcal{A}_N$.

3.4 Imaging in the frequency domain

In the text below we denote the Fourier spectrum of g by $\mathcal{F}[g]$. As $\mathcal{F}$ presents linear invertible map, the imaging process could be uniquely presented in the frequency domain by the linear maps $\mathcal{B}_k$, $k = 1 \dots N$ such that

$$\mathcal{F}[\mathcal{A}_k I] = \mathcal{B}_k(\mathcal{F}[I]), k = 1 \dots M \tag{9}$$

Also, there exists a map $\mathcal{B}$ such that

$$\mathcal{F}[\mathcal{A}I] = \mathcal{B}(\mathcal{F}[I]), k = 1 \dots M \tag{10}$$

As $\mathcal{F}$ is invertible and the null space of $\mathcal{A}$ equals the intersection of the null spaces of $\mathcal{A}_1$ through $\mathcal{A}_N$, the null space of $\mathcal{B}$ presents the intersection of the null spaces of $\mathcal{B}_1$ through $\mathcal{B}_N$.

For brevity, we'll denote $\mathcal{F}[I]$ by $\widehat{I}$. When the maps φ_k and ψ_k between the HR and the LR images' coordinates represent translation, followed by rotation, we could easily decompose $\mathcal{B}_k$ $(k = 1 \dots N)$ as $\mathcal{B}_k = \mathcal{S}_k.\mathcal{V}_k.\mathcal{W}_k$, where $\mathcal{S}_k$ reflects the sampling process, $\mathcal{V}_k$ - the process of integration of the irradiance over the pixel area and $\mathcal{W}_k$ - the coordinate transformation expressed by φ_k and ψ_k.

Under the assumptions in 3.2, $\mathcal{W}_k$ presents is a composition of image spectrum phase change, followed by image spectrum rotation:

$$[\mathcal{W}_k(\widehat{I})](u,v) = \widehat{I}(\xi_l(u,v), \eta_l(u,v)).Q_l(u,v) \qquad (11)$$

Here ξ_k and η_k represent the effect of sensor rotation (which results spectrum rotation) and Q_k - the effect of sensor translation (which results in spectrum phase shift). The values of Q_k are nonzero as they lay on the unit circle in the complex plane, the coordinate rotaional map determined by ξ_k and η_k is invertible, so the linear operator $\mathcal{W}_k$ is non-singular. The null space of $\mathcal{S}_k$ depends on translations between frames and under our assumptions is a subspace of the null space of $\mathcal{V}_k$. We'll ignore it and concentrate on the latter.

3.5 Motivation for in-plane camera rotation approach

The effect of the integration over a square pixel of size $T \times T$ on the spectrum $\mathcal{W}_k(\mathcal{F}[I])$, expressed by $\mathcal{V}_k$, is:

$$[\mathcal{V}_k.W_k](\widehat{I}))](u,v) = [\mathcal{W}_k(\widehat{I})](u,v).L(u,v) \qquad (12)$$

where

$$L(u,v) = \operatorname{sinc}(u/T).\operatorname{sinc}(v/T) \qquad (13)$$

Changing the multiplication order and substituting coordinates, we can express the effect of integration on $\widehat{I}$ for the k-th image as multiplying $\widehat{I}(w,t)$ by

$$L_k(w,t) = L(\xi_k^{-1}(w,t), \eta_k^{-1}(w,t)) \qquad (14)$$

In this proces, the frequency of $\widehat{I}$ with coordinates (w,t) is lost if either $u = \xi_k^{-1}(w,t)$ or $v = \eta_k^{-1}(w,t)$ equals to exact multiple of π as it has zero image. However, when there is a rotation between the frames, some or all of the zeros of $L_k(\cdot,\cdot)$, $k = 1 \dots N$ will not overlap in the band where $\widehat{I}$ is non-zero. This leads to reduction (in the sence of subset) of $\mathcal{B}_1$'s through $\mathcal{B}_N$'s null spaces intersection and therefore to reduction in the dimensionality of the null space of $\mathcal{B}$. As under our assumptions the later is isomorphic to the null space of A this could bring better image reconstruction possibility.

4 Experimental results

Here we present the results of two experiments, one with simulated low-resolution images, derived from a real-world photo and the other using a synthetic resolution test target. In the first experiment, two sets of low-resolution images were generated from the 256 by 256 pixel high-resolution image (fig.2). The first set contained ten 81 by 81 pixel images simulating translational camera motion; the second - ten 81 by 81 pixel ones, with the camera making two in-plane rotations (fig.1). Then, the image was restored by iterated back-projection leading to LMS solution of 5. The absence of noise in the simulated set-up made regularisation unneeded. The resulting PSNR ratios between the original and the restored image are 35,06 and 37,45 db for the translational and translation/rotation cases. Note the lack of the 'net' artifact (fig.3) in the case when the camera makes rotations (fig.4); in this case, the visible artifacts in the lower right corner are due to the area non being covered by all LR images.

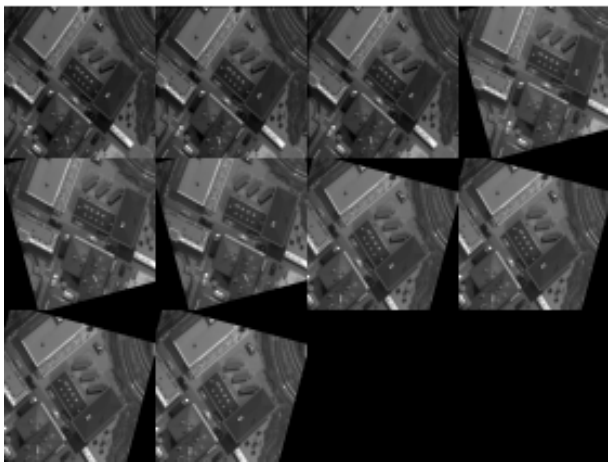

Figure 1: LR image set generated simulating camera translation and in-plane rotation

Figure 2: portion of the original 256x256 pixel image

Figure 3: portion of the same image, restored from 10 81x81 images under translational motion

Figure 4: portion of the same image, restored from 10 81x81 images with two rotations

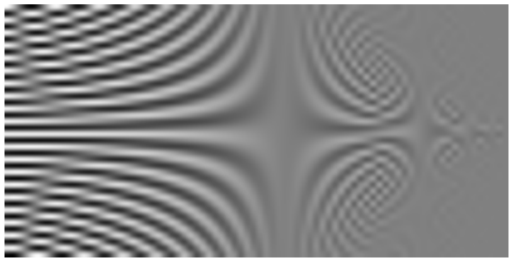

Figure 5: An 80x40 simulated LR resolution test target image

In the second experiment two sets of 30 resolution test target 80x40 pixel images were synthesized again

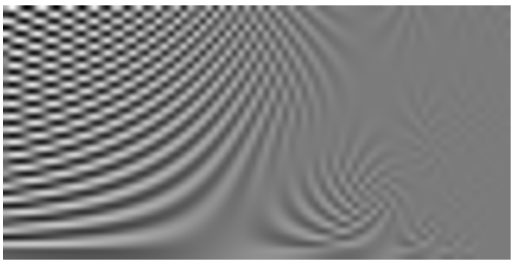

Figure 6: Another of the 80x40 simulated LR resolution test target images set

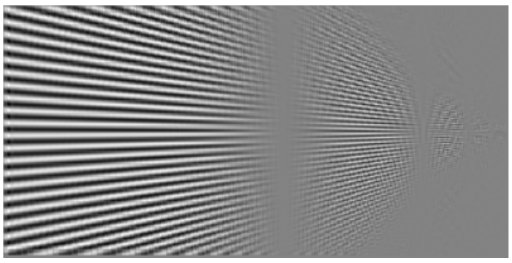

Figure 7: Restored 400x200 HR resolution test target image, no camera rotation

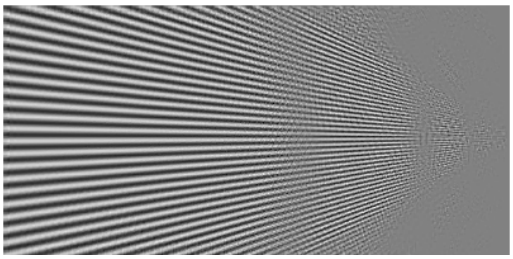

Figure 8: Restored 400x200 HR resolution test target image with camera rotation

simulating only camera translation for the first and translation-and-rotation for the second. One of the LR images is shown on fig.5. From them, 400x200 pixel HR image was restored; the usage of this large maginification factor is to avoid aliasing effects in the restored image. In the case of purerly translational camera motion, shown on fig.7 the 'gaps' in resolution, stemming from the null space of A could be easily seen. The mitigation of this effect is visible on fig.8, verifying the correctness of the proposition in the previous section.

5 Conclusions and future work

5.1 Conclusions

In this paper, we have considered the task of super-resolution by fusing the information from several low-resolution images. The model used considered that the array of optical sensors is irradiated by a continuous op-

tical flux $I(\cdot,\cdot)$. We have identified an obstacle in the super-resolution reconstruction of images, namely the loss of certain fourier spectrum frequencies. This obstacle is brought by non-zero pixel size and cannot be overcome by taking more snapshots under (local) translation only. A way to solve it by in-plane rotating the camera between taking images was proposed. Some experimental results with synthetic data were presented, showing the viability of the method.

5.2 Future work

It would be interesting to investigate the case of other camera motions, corresponding to φ_k and $\psi_k$representing locally affine transform. The question of finding optimum rotation angle(s) and number of LR images is interesting as well. Also, by finding the attenuation of some frequencies that are part of the null spaces of high to low resolution transition operators for some of the images will give the possibility for better high resolution image reconstruction.

6 Acknowledgments

This paper is partially supported by the Bulgarian Ministry of Education and Science under grants VU-MI-204/06

References

[1] Simon Baker and Takeo Kanade. Super resolution optical flow. Technical Report CMU-RI-TR-99-36, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, October 1999.

[2] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on PAMI*, 24(9):1167–1183, September 2002.

[3] S. Battiato, G. Gallo, and F. Stanco. A locally adaptive zooming algorithm for digital images. *Image Vision and Computing Journal*, 20(11):805–812, September 2002.

[4] M. Ben-Ezra, A. Zomet, and S.K. Nayar. Jitter Camera: High Resolution Video from a Low Resolution Detector. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 135–142, Jun 2004.

[5] Sean Borman and Robert L. Stevenson. Linear models for multi-frame super-resolution restoration under non-affine registration and spatially varying psf. In *Computational Imaging II*, volume 5299 of *Proc. SPIE*, pages 234–245, San Jose, CA, USA, 2004.

[6] Michael M. Chang, A. Murat Tekalp, and M. Ibrahim Sezan. Simultaneous motion estimation and segmentation. *IEEE Transactions on Image Processing*, 6(9):1326–1333, 1997.

[7] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multi-frame super-resolution. *IEEE Transactions on Image Processing*, 13:1327–1344, 2003.

[8] Takeo Kanade Goksel Dedeoglu and Jonas August. High-zoom video hallucination by exploiting spatio-temporal regularities. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, volume 2, pages 151–158, June 2004.

[9] Michal Irani and Shmuel Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4:324–335, 1993.

[10] Xin Li and Michael T. Orchard. New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10(1):1521–1527, October 2001.

[11] Zhouchen Lin and Heung-Yeung Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Transations on PAMI*, 26(1), January 2004.

[12] Shmuel Peleg Michal Irani. Super resolution from image sequences. In *Proc. 10thInternational Conferenc on Pattern Recognition*, volume ii, pages 115–120, June 1990.

[13] A.J. Patti, M.I. Sezan, and A.M. Tekalp. High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur. In *Proceedings of the IEEE International Conference on Image Processing*, volume I, pages 343–347, 1994.

[14] Andrew Patti, M. Ibrahim Sezan, and A. Murat Tekalp. Super resolution video reconstruction with arbitrary sampling lattices and non-zero aperture time. *IEEE Transactions on Image Processing*, 6:1064–1076, 1997.

[15] Lyndsey C. Pickup, David P. Capel, and Stephen J. Roberts Andrew Zisserman. Bayesian image super-resolution, continued. In *Advances in Neural Information Processing Systems*, pages 1089–1096. MIT Press, 2006.

[16] Robert L. Stevenson Richard R. Schultz. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5:996–1011, 1996.

[17] Steven M. Seitz and Charles R. Dyer. Physically-valid view synthesis by image interpolation. In *In Proc. IEEE Workshop on Representations of Visual Scenes*, pages 18–25, 1995.

[18] Carl Staelin, Carl Staelin, Darryl Greig, Darryl Greig, Mani Fischer, Mani Fischer, Ron Maurer, and Ron Maurer. Neural network image scaling using spatial errors. In *Technical Reports*. HP Laboratories Israel, October 2003.

[19] Battiato Gallo Stanco. Smart interpolation by anisotropic diffusion. In *International Conference on Image Analysis and Processing*, pages 572–577, Los Alamitos, CA, USA, 2003. IEEE Computer Society.

[20] Klamer Schutte Thuan Q. Pham, Lucas J. van Vliet. Influence of signal-to-noise ratio and point spread function on limits of super-resolution. In *Image Processing: Algorithms and systems IV*, Proc. SPIE IS & T, pages 169–180, 2005.