

Super-resolution using neural networks based on the optimal recovery theory

Yizhen Huang · Yangjing Long

Published online: 18 January 2007
© Springer Science + Business Media, LLC 2007

Abstract An optimal recovery based neural-network Super Resolution algorithm is developed. The proposed method is computationally less expensive and outputs images with high subjective quality, compared with previous neural-network or optimal recovery algorithms. It is evaluated on classical SR test images with both generic and specialized training sets, and compared with other state-of-the-art methods. Results show that our algorithm is among the state-of-the-art, both in quality and efficiency.

Keywords Super resolution · Neural-network · Optimal recovery

1 Introduction

With the popularization of various image-capturing devices such as commercial digital cameras, camphones and camcorders, digital images can be obtained more easily and conveniently than ever before. However, High-Resolution (HR) sensors are still costly. This is because, from the hardware perspective, high spatial resolution requires small pixel size

i.e. large number of pixels per unit area, and hence limits the amount of light available for each pixel, which exaggerates the effect of shot noise and degrades the image quality evidently. To avoid suffering from shot noise, there exists the limitation of the pixel size reduction, and the optimally limited pixel size is estimated at about $40 \mu\text{m}^2$ for a $0.35 \mu\text{m}$ CMOS process. The current image sensor technology has almost reached this level.

This urges and expedites the development of software solution to increase resolution level. Such resolution enhancement techniques may be generally referred to as Super Resolution (SR) image reconstruction, which primarily consists of three categories: (a) Multiple-frame SR by extracting a single HR image from a sequence of Low-Resolution (LR) images aligned in sub-pixel accuracy (see [1–3] for excellent surveys) (b) Single-frame example-based SR that increases the resolution of a single image with training based on analogous images [4–11, 29–31] (c) Single-frame SR that does not depend on any other images [12–27]. The latter category of single-frame SR is more commonly referred to as image interpolation or digital zooming. In this article, we focus on single-frame SR.

A finer taxonomy can be made for single-frame SR according to branches of their algorithms. Due the limited space, only a few are listed: (a) Linear methods [12–16] are relatively fast once the filter coefficients are found. This is especially the case for short kernel methods [12]. But the visual quality of the resulting images is often quite unsatisfactory. Shorter kernel filters usually result in blurred images or aliasing effects [13, 14], and with longer filters the images often exhibit ringing along sharp edges [15, 16]. (b) Images interpolated with edge-directed approaches [17, 18] look sharper but often seem too much like drawings, especially for large zoom factors. (c) The ideal filter response approaches [19, 20] are with theoretical perfection and poor practical

Y. Long: This work was supported in part by the National 211/985 Project of Shanghai Jiaotong University under Subgrant PRP[S03009009].

Y. Huang (✉) · Y. Long
Department of Computer Science and Engineering, School of Electronics Information & Electric Engineering, Shanghai Jiaotong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, PR China
e-mail: hyz12345678@sjtu.edu.cn

Y. Long
Department of Mathematics, Shanghai Jiaotong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, PR China, Digital Media and Data Reconstruction Lab., Shanghai Jiaotong University

performance. (d) Projection-Onto-Convex-Set (POCS) algorithms produce images with high Peak Signal-to-Noise Ratio (PSNR) but are computationally expensive [21, 22]. (e) Partial Differential Equation (PDE) models [23, 24] establish inter-pixel correlation from a more abstract point of view, yet they usually yield images with jagged artifacts and are vulnerable to noises. (f) Optimal recovery interpolators [23, 25–27], and (g) Neural-network based methods [29–31], which are to be detailed in Section 2.

The branch, to which each SR algorithm belongs, affects the output image quality and computing expense intrinsically. Based on this observation, one way to find potentially useful SR algorithms is by introducing new category of methodology. However, this is rather difficult considering the fact that SR has been intensively studied for decades (see the seminal work by Tsai and Huang [33] for multiple-frame SR and early textbook [34] for single-frame SR).

Another way is by combining several existing SR algorithms to take their respective advantages. Driven by this, a hybrid method of optimal recovery and neural-network is proposed in this paper. The rest of the manuscript is arranged as follows: In Section 2, the two branches of algorithms, from which our hybrid method evolves, together with the main idea of our method are presented; Section 3 details our proposed method; Section 4 is experimental results; Finally conclusion is provided in Section 5.

2 Related work and main idea

Like other image reconstruction problems such as Color Filter Array interpolation [28, 32], SR is an inherently ill-posed inverse problem since there may exist a virtually infinite set of magnified images from the original given data. The problem is usually placed into an inter-pixel correlation model [6, 7, 11, 12, 15–17, 23, 24, 29–31], which has been acknowledged to be highly non-linear. Back propagation neural networks are capable of learning complex nonlinear functions. This naturally motivates the development of neural-network approaches [29–32] that produces better results in high-frequency regions.

One problem using neural-networks is that, there exists large information redundancy in the functional relation between input and output neurons, which affects efficiency greatly. This is especially the case if neural-networks are applied directly in a brute force way [29, 30, 32]. Another problem is that, the training step for neural-networks is considerably time-consuming. Fortunately, this can be done as an initialization and does not occupy any computing resource for real-time processing. On the other hand, the training step is also an opportunity to incorporate priori information. Such information is often referred to as example images, and statistics show that, example images help to improve

both the objective quality (commonly measured in terms of PSNR, Mean Square Error (MSE), or Normalized Color Difference (NCD)) and subjective quality (namely visual experience by our Human Vision System (HVS)) of output images [4, 6, 8, 11].

A lot of image reconstruction algorithms based on the optimal recovery theory arise recently [23, 25–28]. Although quite different from each other, they benefit from the common strengths intrinsic to the optimal recovery representation: the subjective quality of output images is satisfactory [27]. However, for each image patch, not only does it require computing matrix multiplication but also inverse matrices, which is inefficient and induces relatively large errors. Specifically, if the matrix to be inverted is singular or approximately singular, the errors induced are considerable. Therefore optimal recovery algorithms are slow and produce images with relatively inferior subjective quality [25, 27] among the state-of-the-art methods.

The success of these algorithms motivates us to ask that whether it is possible to integrate an optimal recovery based approach within a neural-network framework and, if so, will the two different branches of algorithms complement each other to offer a better algorithm? We believe that the answers should be positive. An important observation is that, optimal recovery condenses and extracts local image patch information and enables neural-networks to be trained and simulate in more effective way.

Guided by these ideas, an optimal recovery based neural-network SR algorithm is developed, which is computationally less expensive and outputs images with high subjective quality, compared with previous neural-network or optimal recovery algorithms. It is evaluated on classical SR test images and compared with other state-of-the-art methods. Results show that our algorithm is among the state-of-the-art, both in quality and efficiency.

3 Proposed algorithm

The proposed algorithm consists of the following 3 portions:

3.1 Optimal recovery model

The first key point of using optimal recovery for image interpolation is determining the quadratic signal class K :

$$K = \{x \in R^n : x^T Q x \leq \varepsilon\}$$

from a set of training data. The training data is usually taken from the local features of the image and selecting a proper training set is discussed at the end of this subsection. For now assume that a training set of patches $S = x_1, \dots, x_m$ representative of the local data is given for estimating the local quadratic signal class. The Q for which the ellipsoid

$$x^T Qx \leq \varepsilon \tag{1}$$

for some constant ε must be representative of the training set S . In other words Q must be a matrix such that when an image patch y is similar to the vectors in S then Eq. (1) holds for y . Let matrix S be formed by arranging the image patches in S as columns:

$$S = (x_1, \dots, x_m) \tag{2}$$

and consider the equation relating the image patch y to the training set S using a column of m weights, a :

$$Sa = y \tag{3}$$

Vector y is similar to the vectors in S when a has small energy. Using standard notation for singular value decomposition of S , Eq. (3) can be rewritten as:

$$U \Lambda V^T a = y$$

The weight vector a is given by

$$a = V \Lambda^{-1} U^T y$$

and the sum of the squares, or the energy of a is:

$$a^T a = y^T U \Lambda^{-2} U^T y$$

Since

$$SS^T = U \Lambda^2 U^T$$

it follows that

$$\begin{aligned} a^T a &= y^T (SS^T)^{-1} y \\ &= y^T Qy \end{aligned}$$

where Q is the pseudo inverse of SS^T . If y is very similar to the training set S then

$$a^T a \leq \varepsilon$$

and with the new Q it follows that

$$y^T Qy \leq \varepsilon$$

This is the desired form for our ellipsoidal signal class. It results from an ‘‘Occam’s razor’’ type of assumption that small weights are used to represent vectors that are similar to our training set S . For more theoretical basis, refer to [36] or the appendix of [27].

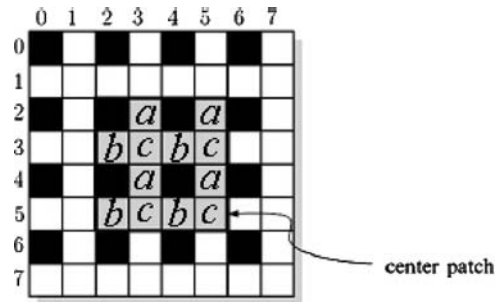


Fig. 1 Local HR image used for selecting S to estimate the quadratic class for the center 4×4 patch (dark pixels are part of the decimated image)

Given the formulation of the quadratic signal class from a training set S , or equivalently the formulation of Q , the next key point is determining the training set S . One commonly used approach is based on the proximity of their locations to the position of the vector being modeled. Patches are generated from the local neighborhood.

For example, in Fig. 1 to model the quadratic signal class that the center patch

$$x = [x_{2,2}, x_{2,3}, x_{2,4}, x_{2,5}, x_{3,2}, \dots, x_{5,5}]^T \tag{4}$$

belongs to, let

$$S = \left\{ \begin{pmatrix} x_{0,0} \\ x_{0,1} \\ \vdots \\ x_{3,3} \end{pmatrix}, \dots, \begin{pmatrix} x_{4,4} \\ x_{4,5} \\ \vdots \\ x_{7,7} \end{pmatrix} \right\} \tag{5}$$

where S is formed by choose all the possible 4×4 image blocks in the 8×8 region of Fig. 1.

However, only darks pixels belong to the decimated image and are known, other pixels are missing samples to be interpolated. To resolve this problem, an initial interpolation should be performed before applying our proposed interpolator. Section 3.1 discusses this issue specifically.

Once the quadratic signal class K is obtained, the known representors are products between Q^{-1} and vectors with one at the location of known samples and zero everywhere else. The optimal recovery solution is a linear combination of the known representors.

For example, to achieve 2X magnification, the known representors (denoted as $F_{x,y}$ where (x, y) is its location) in Fig. 1 are the dark pixels for known samples. The known representor $F_{2,2}$ at the location (2,2) is the product between Q^{-1} and the vector with 1 for $x_{2,2}$ and 0 for all other vector components.

The grey pixels labeled with ‘‘a’’ are computed by the average of the known representors at its immediate left and right neighborhood e.g. the intensity of the pixel (2,3) is

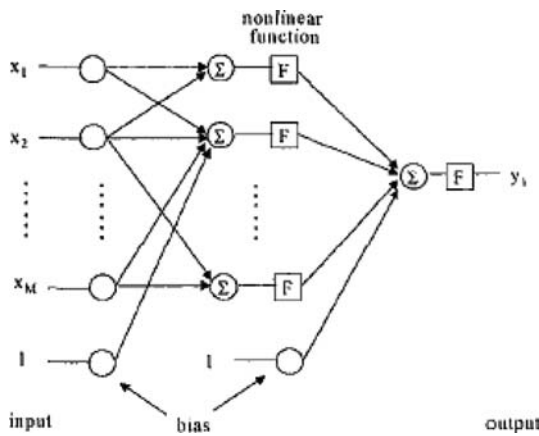


Fig. 2 Schematic of a 3-layer feed-forward neural network

estimated and updated to be $(F_{2,2} + F_{2,4})/2$. Similarly, the grey pixels labeled with “*b*” are computed by the average of the known representors of its immediate top and bottom neighbors, and the “*c*” labeled pixels is the average of the known representors of its four immediate neighbors.

3.2 Neural-network design

We use the 3-layer feed-forward neural network as shown in Fig. 2: The input layer has 64 neurons, corresponding to the 8×8 image patch, the hidden layer has 50 neurons and the output layer has 9 neurons indicating the 9 used known representors. The sigmoid transfer function of the hidden layer is $\text{tansig}(x) = 2/(1 + \exp(-2x)) - 1$ and that of the output layer is $\text{logsig}(x) = 1/(1 + \exp(-x))$. The training algorithm is selected to be the Scaled Conjugate Gradient method [35].

In the training step, the neural network is fed with 8×8 image patches arranged in vector form for the 64 input neurons and the 9 known representors computed by the optimal recovery algorithm in Section 3.1 as target values for the output neurons; in the simulation step, the known representors are generated by the neural network rather than by the optimal recovery theory.



Fig. 4 Twenty 768×512 example images for the generic training set. The images are numbered from left to right, top to bottom, of which the 7, 8, 9, 10 and 14th images are selected to form the specialized training set to SR the image Boat

3.3 Two-pass iterative mechanism

As stated in Section 3.1, an initial interpolation is need before our proposed algorithm can work. From another perspective, our algorithm can be considered as a process of updating a rough interpolated SR image to be a more precise one. Hence it can be utilized as an iterative process with the output image of the previous iteration used as the input of the current iteration.

The simplest initial interpolation may be pixel replication. Bilinear or bicubic initial interpolation can slightly improve the output image quality if using our algorithm in a one-pass way, which indicates that, the output image quality depends on the initial interpolation step. Interestingly, according to experience from experiments, using our algorithm in a two-pass way generates visual results that are very similar regardless of the initial interpolation step, and more times of iteration only waste the computing resource but yield negligible performance gain. So a two-pass iterative mechanism is adopted in the experiments part in Section 4. Its block diagram is shown in Fig. 3.

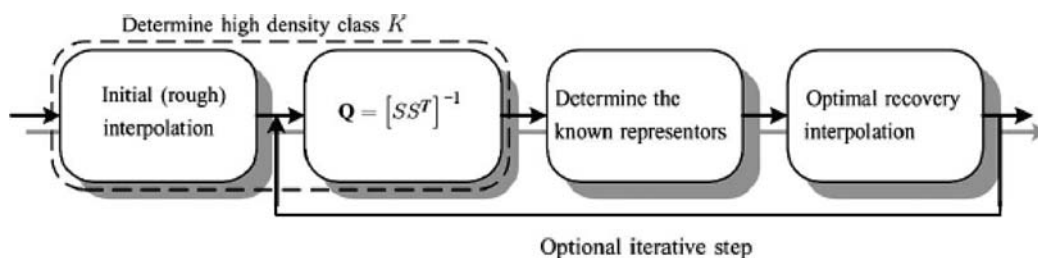


Fig. 3 A block diagram of our proposed algorithm

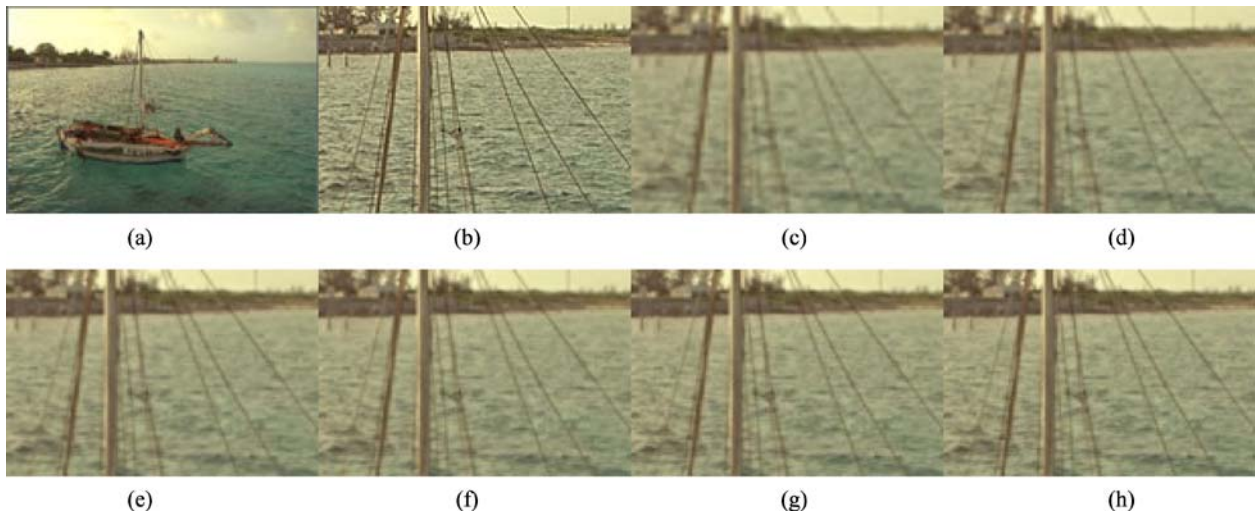


Fig. 5 Boat's mast 2X magnification (from left to right, top to bottom): (a) The entire true HR image (5X downsampled), (b) The true HR image cropped from (a), (c) Input (2X magnified), (d) The bicubic spline in Adobe Photoshop, (e) The optimal recovery algorithm, (f) The neural-network algorithm, (g) The proposed algorithm with generic training set, (h) The proposed algorithm with specialized training set

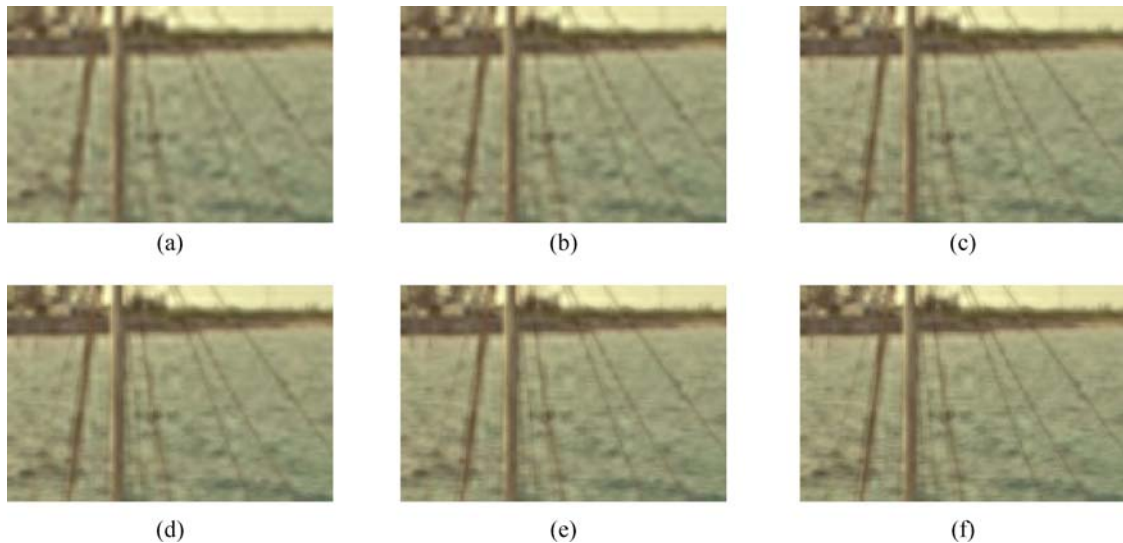


Fig. 6 Boat's mast 3X magnification (from left to right, top to bottom): (a) Input (3X magnified), (b) Bicubic spline in Adobe Photoshop, (c) The optimal recovery algorithm, (d) The neural-network algorithm, (e) The proposed algorithm with generic training set, (f) The proposed algorithm with specialized training set

4 Experiments and results

For performance evaluation, our method is compared with the traditional bicubic interpolation (the bicubic resampling tool in Adobe Photoshop), the state-of-the-art optimal recovery algorithm (our own implementation of the algorithm described in [27]) and the neural-network algorithm (obtained directly from the author [31]). All codes run on an AMD Athlon 64 3000 MHz. Twenty 768×512 example images displayed in Fig. 4 are used to generate 8×8 image patches for training the neural network.

In what follows, both subjective visual evaluation and objective quantitative measurement are presented. The Peak Signal-to-Noise Ratio (PSNR) is used as an objective indicator to measure image quality, which is defined by

$$PSNR = 10 \times \log \left(\frac{255^2}{\frac{1}{cols \times rows} \sum_{i=1}^{cols} \sum_{j=1}^{rows} e^2(i, j)} \right)$$

Where “cols” and “rows” are the column and row numbers of the image, respectively, 255 is the maximum value for pixel

Table 1 Average PSNR over the 3 channels of each method for noise-free 2X magnification (in dB)

	Bicubic	Optimal recovery	Neural-network	Proposed
Lena	39.72	34.19	38.35	39.63
Peppers	40.20	40.32	39.77	40.09
Boy	34.41	35.93	36.26	35.87
Boat	34.58	36.43	37.05	38.21

Table 2 Computation time corresponding to Table 1

	Optimal recovery	Neural-network	Proposed
Lena	2.81	13.42	1
Peppers	2.57	19.24	1
Boy	2.89	14.37	1
Boat	2.92	12.53	1

intensity, and $e(i, j)$ is the difference between the real and computed images at pixel location (i, j) .

The PSNR results and corresponding computation time of each method are listed in Tables 1 and 2. The bicubic spline is waived from comparing computation time because its source code remains unavailable to us. Our algorithm is used as the baseline 1, to which the ratios of all other methods are computed. We can see that, our algorithm is computationally less expensive while offering an excellent PSNR in most cases. Figures 5 and 6 depict in close-up the results of the image Boat by each method for 2X and 3X magnification respectively. Due to the limited space, only the mask part of the boat is cropped from the entire HR image and presented in full size. It is shown that, our proposed method produces visually pleasing HR images with fine details at edges while synthesizing plausible texture and shine at the water ripple region.

5 Conclusion

An optimal recovery based neural-network Super Resolution algorithm is developed. The proposed method is computationally less expensive and outputs images with high subjective quality, compared with previous neural-network or optimal recovery algorithms. It is evaluated on classical SR test images and compared with other state-of-the-art methods. Results show that our algorithm is among the state-of-the-art, both in quality and efficiency.

Besides, we believe extensions of our algorithm, such as clustering and classification of example image patches, using multiple neural-networks for specialized learning and simulation of image patches with different gradients and directions, etc., may bring the performance increase. We will pursue this interesting direction in our future research.

References

1. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. *IEEE Signal Proc. Mag.* **20**(3), 21–36 (2003)
2. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(9), 1167–1183 (2002)
3. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Advances and challenges in super-resolution. *Int. J. Imag. Sys. Technol.* **14**(2), 47–57 (2004)
4. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. *Int. J. Comp. Vis.* **40**(1), 25–47 (2000)
5. Qiu, G.: Inter-resolution look-up table for improved spatial magnification of images. *J. Vis. Commun. Image Represent.* **11**, 360–373 (2000)
6. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. *IEEE Comp. Graph. Appl.* **22**(2), 56–65 (2002)
7. Freeman, W.T., Pasztor, E.C.: Learning to estimate scenes from images. In: *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*, pp. 775–781 (1998)
8. Sun, J., Zheng, N., Tao, H., Shum, H.: Image hallucination with primal sketch priors. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. II, pp. 729–736 (2003)
9. Freeman, W.T., Pasztor, E.C.: Markov networks for super-resolution. In: *Proceedings of the 34th Annual Conference on Information Sciences and Systems* (2000)
10. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image Analogies. In: *Proceedings of SIGGRAPH 2001*, pp. 327–340 (2001)
11. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: *Proceedings of the IEEE International Conference on Computer Vision*, vol. I, pp. 275–282 (2004)
12. Keys, R.G.: Cubic convolution interpolation for digital image processing. *IEEE Trans. Acous., Speech and Signal Proc.* **29**, 1153–1160 (1981)
13. Ramstad, T., Wang, Y., Mitra, S.K.: An efficient image interpolation scheme using hybrid IIR Nyquist filters. *Opt. Eng.* **31**, 1277–1283 (1992)
14. Chan, A.K., Chui, C.K., Zha, J., Liu, Q.: Local cardinal spline interpolation and its application to image processing. In: *Proceedings of the Conference on Curves and Surfaces in Computer Vision and Graphics II*, SPIE Proceedings, vol. 1610, pp. 272–283 (1991)
15. Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing: part I-theory. *IEEE Trans. Sign. Proc.* **41**, 821–833 (1993)
16. Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing: part II-efficient design and applications. *IEEE Trans. Sign. Proc.* **41**, 834–848 (1993)
17. Thurnhofer, S., Mitra, S.: Edge-enhanced image zooming. *Opt. Eng.* **35**(7), 1862–1870 (1996)
18. Li, X., Orchard, M.T.: New edge-directed interpolation. *IEEE Trans. Image Proc.* **10**(10), 1521–1527 (2001)
19. Chan, S.C., Ho, K.L., Kok, C.W.: Interpolation of 2-D signal by subsequence FFT. *IEEE Trans. Circuits Sys. Part II* **40**, 115–118 (1993)
20. Kim, S.P., Su, W.: Direct image resampling using block transform coefficients. *Sign. Proc.: Image Commun.* **5**, 259–272 (1993)
21. Stasinski, R., Konrad, J.: POCS-based image reconstruction from irregularly-spaced samples. In: *Proceedings of the 7th IEEE International Conference on Image Processing Vol. 2*, pp. 315–318 (2000)
22. Stasinski, R., Konrad, J.: Improved POCS-based image reconstruction from irregularly-spaced samples. In: *Proceedings of the 7th European Signal Processing Conference, Vol. 2* pp. 461–464 (2002)

23. Chen, G., de Figueiredo, R.J.P.: A unified approach to optimal image interpolation problems based on linear partial differential equation models. *IEEE Trans. Image Proc.* **2**, 41–49 (1993)
24. Karayiannis, N.B., Venetsanopoulos, A.N.: Image interpolation based on variational principles. *Sign. Proc.* **25**, 259–288 (1991)
25. Muresan, D.D., Parks, T.W.: Optimal recovery approach to image interpolation. In: *Proceedings of the 8th IEEE International Conference on Image Processing*, vol. III, pp. 7–10 (2001)
26. Shenoy, R.G., Parks, T.W.: An optimal recovery approach to interpolation. *IEEE Trans. Sign. Proc.* **40**(8), 1987–1996 (1992)
27. Muresan, D.D., Parks, T.W.: Adaptively quadratic (AQua) image interpolation. *IEEE Trans. Image Proc.* **13**(5), 690–698 (2004)
28. Muresan, D.D., Parks, T.W.: Demosaicing using optimal recovery. *IEEE Trans. Image Proc.* **14**(2), 267–278 (2005)
29. Swanston, D.J., Bishop, J.M., Mitchell, R.J.: simple adaptive momentum: New algorithm for training multilayer perceptrons. *Elect. Lett.* **30**, 1498–1500 (1994)
30. Scalero, R.S., Tepedelenlioglu, N.: A fast new algorithm for training feedforward neural networks. *IEEE Trans. Sign. Proc.* **40**, 203–210 (1992)
31. Plaziac, N.: Image interpolation using neural networks. *IEEE Trans. Image Proc.* **8**, 1647–1651 (1999)
32. Go, J., Sohn, K., Lee, C.: Interpolation using neural networks for digital still cameras. *IEEE Trans. Consum. Elect.* **46**(3), 610–616 (2000)
33. Tsai, R.Y., Huang, T.S.: Multiframe image restoration and registration. *Adv. Comp. Vis. Image Proc.* **1**, 317–339 (1984)
34. Anderws, H.C., Hunt, B.R.: *Digital Image Restoration*. Prentice-Hall NJ, (1977)
35. Moller, A.F.: A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* **6**(4), 525–533 (1993)
36. Michelli, C.A., Rivlin, T.J.: *Optimal Estimation in Approximation Theory*, Plenum NY, (1976)