

Multi-frame super-resolution with no explicit motion estimation

Mehran Ebrahimi and Edward R. Vrscay

Department of Applied Mathematics
Faculty of Mathematics, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
m2ebrahi@uwaterloo.ca, ervrscay@uwaterloo.ca
[IPCV 2008 Submission]

keywords: super-resolution, image zooming, non-local-means, inverse problems

Abstract. We introduce a novel super-resolution scheme for multi-frame image sequences. Our method is closely associated with the recently developed “non-local-means denoising filter”. In the proposed algorithm, no explicit motion estimation is performed, unlike in many other methods. Our results are comparable, if not superior, to many existing approaches, especially in the case of low signal-to noise ratio.

1 Introduction

Naturally, there is always a demand for higher quality and higher resolution images. The level of image detail is crucial for the performance of many computer vision algorithms. Current imaging devices typically consist of arrays of light detectors. A detector determines pixel intensity values depending upon the amount of light detected from its assigned area in the scene. The spatial resolution of images produced is proportional to the density of the detector array: the greater the number of pixels in the image, the higher the spatial resolution. In many applications, however, the imaging sensors have poor resolution output. When resolution can not be improved by replacing sensors, either because of cost or hardware physical limits, one can resort to resolution enhancement algorithms. Even when superior equipment is available, such algorithms provide an inexpensive alternative.

The process of producing a high-resolution (HR) image given a single low-resolution (LR) image is called *single-frame image zooming*. Re-sizing of the image does not translate into an increase in resolution. In fact, re-sizing should be accompanied by approximations to frequencies higher than those representable at original size, and at a higher signal to noise ratio. Many interpolation methods tend to smoothen and blur image detail as well as performing inefficiently in the presence of noise. Such recovery is normally performed using *a-priori* information about the image. Another possibility is to take advantage of information from several observations rather than from a single image. The problem of recovering a high quality HR image from a set of distorted (e.g., warped, blurred, noisy) and LR images is known as *super-resolution*. Fusion of the information from the observations is a fundamental challenge in the recovery process.

2 Multi-frame super-resolution problem

2.1 The classical inverse problem

Formally, we review a general model for super-resolution widely used in the literature [15, 8, 9, 5, 7]. Assume that the LR grid is

$$\Omega = [1, \dots, M] \times [1, \dots, N], \quad (1)$$

and given the positive integer z the HR grid is defined as

$$\Psi = [1, \dots, Mz] \times [1, \dots, Nz]. \quad (2)$$

Forward model. We consider a forward degradation model that converts an ideal HR image \underline{f} to degraded LR frames \underline{u}_i ,

$$\underline{u}_i = \mathcal{H}_i \underline{f} + \underline{n}_i, \quad 1 \leq i \leq k. \quad (3)$$

Here, the operator $\mathcal{H}_i = \mathcal{S}\mathcal{B}\mathcal{W}_i$ is the composition of a warping operator $\mathcal{W}_i : l^2(\Psi) \rightarrow l^2(\Psi)$ which maps the HR grid coordinate to the LR grid, a blurring operator $\mathcal{B} : l^2(\Psi) \rightarrow l^2(\Psi)$, and a downsampling operator $\mathcal{S} : l^2(\Psi) \rightarrow l^2(\Omega)$ of factor z in each direction. Also, $\underline{n}_i \in l^2(\Omega)$ denotes additive white independent Gaussian noise with zero-mean and variance σ^2 . The inverse problem of multi-frame super-resolution can be stated as follows,

Invsere problem: Given a set of k LR observed frames $\{\underline{u}_i\}_{i=1, \dots, k} \in l^2(\Omega)$ of size $M \times N$, reconstruct the HR image $\underline{f} \in l^2(\Psi)$ of size $Mz \times Nz$.

Traditionally the equations in (3) are stacked to represent a large algebraic linear equation $\underline{u} = \mathcal{H}\underline{f} + \underline{n}$. The inverse problem corresponding to such system is typically ill-posed, i.e., does not possess a unique solution that depends continuously on the measurements. A great deal of research in the area of super-resolution in the past decades has been focussed on defining effective regularization functionals to address such ill-posed inverse problem.

2.2 A word on motion estimation

Accurate motion estimation has been a very important aspect of super-resolution schemes. In the forward process, described in Equation (3), the motion parameters are represented by \mathcal{W}_i 's. In many existing super-resolution approaches, the motion is computed directly from the LR frames, while many other super-resolution algorithms unrealistically assume that motion parameters are precisely known. In general, however, accurate motion estimation of subpixel accuracy remains a fundamental challenge in super-resolution reconstruction algorithms.

In our opinion, however, it seems reasonable to assume that the motion can be relaxed from a strict grid mapping to a multi-pixel-pair intensity relation. In this view, pixel-pairs in different frames may be relevant to each other with some measured probability of confidence. In the method we propose below, instead of estimating the motion vectors explicitly, a framework is provided in which such confidence measures are evaluated and employed in the HR image reconstruction.

3 Super-resolution with no explicit motion estimation

3.1 Changing the order of blur and warp to isolate blur

It is well known that the order of \mathcal{B} and \mathcal{W}_i may be changed in the case that blur is linear spatially invariant. Hence, the equations may be written as

$$\underline{u}_i = \mathcal{S}\mathcal{W}_i\mathcal{B}\underline{f} + \underline{n}_i, \quad 1 \leq i \leq k. \quad (4)$$

If we define $\underline{v} = \mathcal{B}\underline{f}$, then the measurements become

$$\underline{u}_i = \mathcal{S}\mathcal{W}_i\underline{v} + \underline{n}_i, \quad 1 \leq i \leq k. \quad (5)$$

In our algorithm, we simply focus on solving \underline{v} , noting that any existing deblurring algorithm can be applied to reconstruct \underline{f} after we obtain a solution for \underline{v} . In order to reduce notation, we define $\underline{v}_i = \mathcal{W}_i\underline{v}$, so that the above equations become

$$\underline{u}_i = \mathcal{S}\underline{v}_i + \underline{n}_i, \quad 1 \leq i \leq k. \quad (6)$$

Furthermore, in many multi-frame super-resolution algorithms it is customary to employ one of the LR frames as a reference frame. The algorithm proposed in the next section provides an estimation of \underline{v}_i denoted by $SR(\underline{v}_i)$ for any $1 \leq i \leq k$.

3.2 HR reconstruction of the i -th frame given the j -th frame

In what follows, we let \widehat{u}_i denote the interpolation \underline{u}_i from Ω to Ψ yielded by some standard technique, e.g., bilinear interpolation. Therefore, $\widehat{\cdot}$ is a mapping from $l^2(\Omega)$ to $l^2(\Psi)$. Note that, the interpolation \widehat{u}_i provides an approximation of \underline{v}_i , i.e., $\underline{v}_i \approx \widehat{u}_i$ for $1 \leq i \leq k$. Here, however, we seek a superior approximation to \underline{v}_i , which will be denoted as $SR(\underline{v}_i)$.

The following scheme is inspired by the work on image and image-sequence denoising in [2–4]. For any $x \in \Psi$, we evaluate the conditional expectation of $SR(\underline{v}_i)(x)$ given the observed LR image \underline{u}_j by

$$E[SR(\underline{v}_i)(x) | \underline{u}_j] = \frac{1}{W(x, i, j)} \sum_{y \in \Psi} w(x, y, i, j) \widehat{u}_j(y), \quad \text{such that} \quad (7)$$

$$w(x, y, i, j) = \left[\exp \left(- \frac{\| \widehat{u}_i(\mathcal{N}^d\{x\}) - \widehat{u}_j(\mathcal{N}^d\{y\}) \|_2^2}{h^2} \right) \right], \quad \text{and} \quad (8)$$

$$W(x, i, j) = \sum_{y \in \Psi} w(x, y, i, j), \quad (9)$$

where $\mathcal{N}^d\{x\}$ denotes a square neighborhood of length $(2d+1) \times (2d+1)$ centered at x . The confidence measures mentioned earlier are expressed in terms of $w(x, y, i, j)$'s, where W is a normalization factor. Note that in evaluation $w(x, y, i, j)$ we have used the interpolated copies of \underline{u}_i and \underline{u}_j , respectively denoted by \widehat{u}_i and \widehat{u}_j . Employing such notion automatically takes into account translations of \underline{u}_i and \underline{u}_j by sub-pixel accuracy. Also, h is a regularization parameter that can be adjusted to control the smoothness of the output.

3.3 HR reconstruction of the i -th frame given the whole image sequence

Eventually, we evaluate the conditional expectation of $SR(\underline{v}_i)(x)$ for any $x \in \Psi$, given the information of all of the frames $\{\underline{u}_j\}$ for $1 \leq j \leq k$:

$$E[SR(\underline{v}_i)(x) | \{\underline{u}_j\}_{1 \leq j \leq k}] = \frac{1}{G(i)} \sum_{1 \leq j \leq k} g(|i - j|) \times E[SR(\underline{v}_i)(x) | \underline{u}_j], \quad (10)$$

$$G(i) = \sum_{1 \leq j \leq k} g(|i - j|), \quad (11)$$

where g is a decaying function of $|i - j|$, and G is a normalization factor. The expression $g(|i - j|)$ in this equation represents the temporal confidence on the expectations computed for each of the various frames, j , which has been taken into account in reconstructing the HR image $SR(\underline{v}_i)$. In the experiments reported below, we have assumed that each of the frames in hand are equally likely useful in producing the HR details of the i -th frame. Hence, we have taken g to be a box-function of large enough support which yields $g(|i - j|) = 1$. As a result,

$$E[SR(\underline{v}_i)(x) | \{\underline{u}_j\}_{1 \leq j \leq k}] = \frac{1}{k} \sum_{1 \leq j \leq k} E[SR(\underline{v}_i)(x) | \underline{u}_j]. \quad (12)$$

4 Computational experiments

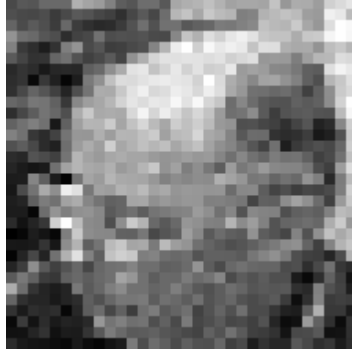
As in the case of the NL-means denoising algorithm, the algorithm described above is computationally intensive. The major computational burden exists in the complexity of computing the weights $w(x, y, i, j)$. A primary scheme to overcome this complexity, introduced in [2–4], is to restrict the search window by restricting $y \in \Psi \cap \mathcal{N}^r\{x\}$ i.e., y lies in a square neighborhood of x with size $(2r + 1) \times (2r + 1)$, as opposed to the entire field of Ω .

Figure 1 shows the result of evaluating $E[SR(\underline{v}_i)(x) | \{\underline{u}_j\}_{1 \leq j \leq k}]$, on an image sequence taken from the data-set library of MDSP at U. California Santa Cruz (<http://www.soe.ucsc.edu/milanfar/software/sr-datasets.html>), originally obtained from the Adyoron Intelligent Systems Ltd., Tel Aviv, Israel. We have taken the first 20 frames, i.e., $k = 20$, of size 32×32 from this sequence, i.e., $M = N = 32$, and have added independent additive white Gaussian noise of standard deviation $\sigma = 0.05$ to the data set. Figure 1(a) shows the result of applying nearest neighborhood interpolation on the second frame of this sequence, i.e., when $i = 2$ for a zooming factor of $z = 3$. In Figure 1(b), the result of bilinear interpolation on the same frame is shown. We have plotted the result of $E[SR(\underline{v}_2)(x) | \{\underline{u}_j\}_{1 \leq j \leq 20}]$ in Figure 1(c). In this experiment, we have taken $d = 4$ (corresponding to a neighborhood of size 9×9), a restricted search window of radius $r = 13$, a zooming factor of $z = 3$, and a smoothness parameter $h = 0.08$.

5 Conclusions

In this paper, we have introduced a novel multi-frame super-resolution technique which does not require explicit motion estimation. Our algorithm was inspired by the non-local-means denoising algorithms introduced in [2–4]. The computational burden of

(a) Nearest neighbor interpolation



(b) Bilinear interpolation



(c) Proposed multi-frame super-resolution



Fig. 1. Experimental results: (a) Nearest neighborhood interpolation, (b) Bilinear interpolation, (c) Proposed algorithm with the following parameters: The original image sequence is of size 32×32 , i.e., $M = N = 32$, of a $k = 20$ frames sequence. The HR counterpart of second frame, i.e., $i = 2$ is desired. Additive white Gaussian noise of $\sigma = 0.05$ is added. Neighborhood of radius $d = 4$, search window radius $r = 13$, zooming parameter $z = 3$, and smoothness parameter $h = 0.08$.

the scheme is a formidable challenge, which precludes any iteration scheme to improve the results. Since there are many parameters in our algorithms, it seems that a fair comparison with other super-resolution algorithms cannot be made. As a result, no comparisons were presented in this paper. That being said, experiments with a number of sets of parameters suggest that our algorithm yields results which are quite comparable if not superior to some of the algorithms in [8, 9] especially when the image sequence is of very low signal-to-noise ratio.

Acknowledgments

This research has been supported in part by the Natural Sciences and Engineering Research Council of Canada, in the form of a Discovery Grant (ERV). ME has been supported primarily by an Ontario Graduate Scholarship.

References

1. S. Borman, *Topics in multiframe superresolution restoration*, PhD thesis, Graduate Program in Electrical Engineering, University of Notre Dame, Indiana, April 2004.
2. A. Buades, B. Coll, and J.M. Morel, *A nonlocal algorithm for image denoising*, IEEE International conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 6065, San-Diego, California, June 20-25, 2005.
3. A. Buades, B. Coll, and J.M. Morel, *A review of image denoising algorithms, with a new one*, SIAM Journal on Multiscale Modeling and Simulation (MMS), Vol. 4, No. 2, pp. 490-530, 2005.
4. A. Buades, B. Coll, J.M. Morel, *Denoising image sequences does not require motion estimation* IEEE Conference on Advanced Video and Signal Based Surveillance, 2005, pp. 70-74
5. S. Chaudhuri, *Super-resolution imaging*, Boston, MA, Kluwer, 2001.
6. M. Ebrahimi and E.R. Vrscay, *Solving the Inverse Problem of Image Zooming using Self-Examples*, Lecture Notes in Computer Science, Volume 4633, Book: Image analysis and Recognition, pp. 117-130, Springer Berlin / Heidelberg, 2007.
7. M. Elad and A. Feuer. *Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images*, IEEE Trans. on Image Processing, vol. 6, no. 12, pp. 1646-1658, 1997.
8. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, *Advances and Challenges in Super-Resolution*, International Journal of Imaging Systems and Technology, August 2004, Volume 14, no 2, pp. 47-57.
9. S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, *Fast and Robust Multi-Frame Super-Resolution*, IEEE Transactions on Image Processing, October 2004, Volume 13, no 10, pp. 1327- 1344.
10. W.T. Freeman, T.R. Jones, and E.C. Pasztor, *Example-based super-resolution*, IEEE Comp. Graphics And Appl., Vol. 22, No. 2, pp. 56-65, 2002.
11. R. Hardie, K. Barnard, and E. Armstrong. *Joint MAP registration and high-resolution image estimation using a sequence of undersampled images*, IEEE Trans. on Image Processing, vol. 6, no 12, pp. 1621-1633, December 1997.
12. M. Irani and S. Peleg, *Improving resolution by image registration*, CVGIP: Graph. Models Image Processing, vol. 53, pp. 324-335, December 1993.
13. IEEE signal processing magazine, *Special issue on super-resolution image reconstruction*, vol. 20, no. 3, pp. 19-86, May 2003.

14. S. Mallat, *A wavelet tour of signal processing*, San Diego, CA: Academic, 1998.
15. N. X. Nguyen, *Numerical algorithms for image superresolution*, PhD thesis, Graduate Program in scientific computation and computational mathematics, Stanford University, July 2000.
16. R. Schultz and R. Stevenson. *Extraction of high-resolution frames from video sequences*, IEEE Trans. on Image Processing, vol. 5, pp. 996-1011, June 1996.
17. A. Tekalp, M. Ozkan and M. Sezan. *High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration*, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. III, pp. 169-172, San Francisco, CA, 1992.
18. R. Tsai and T. Huang. *Multi-frame image restoration and registration*, In Advances in Computer Vision and Image Processing, vol. 1, Greenwich, CT, 1984.
19. H. Ur and D. Gross. *Improved resolution from subpixel shifted pictures*, CVGIP: Graphical Models and Image Processing, vol. 54, no. 2, pp. 181-186, March 1992.
20. S.C. Zhu and D. Mumford, *Prior learning and Gibbs reaction-diffusion*, IEEE Trans. on Patt. Analysis and Machine Intel., Vol. 19, No. 11, pp. 1236-1250, 1997.