Variational local structure estimation for image super-resolution

Heng Lian

Division of Applied Mathematics, Brown University
Providence, RI 02912

May 3, 2017

Abstract

Super-resolution is an important but difficult problem in image/video processing. If a video sequence or some training set other than the given low-resolution image is available, this kind of extra information can greatly aid in the reconstruction of the high-resolution image. The problem is substantially more difficult with only a single low-resolution image on hand. The image reconstruction methods designed primarily for denoising is insufficient for super-resolution problem in the sense that it tends to oversmooth images with essentially no noise. We propose a new adaptive linear interpolation method based on variational method and inspired by local linear embedding (LLE). The experimental result shows that our method avoids the problem of oversmoothing and preserves image structures well.

Keywords: Interpolation, Variational methods

1 Introduction

Super-resolution is an image processing technique to obtain high-resolution image from its low-resolution counterpart, either using information from a single image or image sequence. Most of the literature focus on the latter [5][8]. With only one image, it is generally difficult to interpolate the missing pixels with little prior information about the image.

Many approaches have been proposed in recent years. For video/image sequence, information on subpixel motion between temporally adjacent images can be utilized. Due to the ill-posedness of the super-resolution problem, regularization[5] or a Bayesian framework[9] usually needs to be set up, the two approaches being closely related.

For only one single image, traditionally the most often used method is bilinear or bicubic interpolation, which works under the assumption that images are smooth for the most part. One recent advance on single image super-resolution

is the work by Freeman et al[4]. Using a large database of image patches consisting of low-resolution and high-resolution pairs, the algorithm will search the database for the corresponding high-resolution patch, for each of the image patches in the given low-resolution image. One obvious disadvantage of this approach is that it requires the existence of a database in the first place.

We propose a novel algorithm for single image super-resolution, without the need for a database or other training images. Our algorithm is inspired by Local Linear Embedding (LLE) used for dimension reduction. By trying to express each pixel value as a linear combination of neighboring pixel values, we can estimate the local structure of the low-resolution image which we can then use to construct a filter for interpolation in the second step.

One work similar in spirit to ours is [3]. There the authors also try to express the center pixel as a linear combination of its neighbors. Similar to Freeman[4], their algorithm also requires a training set of pairs of patches in order to estimate those linear coefficients. And the coefficients are learned in the regression framework, giving as a result a spatially adaptive filter. There algorithm can be regarded as a complement to [4] which works well on cartoon images.

One approach to single image super-resolution without training, which is also related to ours, is mentioned in [2] using digital TV filter[1], which is the discrete counterpart of the variational functional method with a total variation (TV) norm. The digital TV filter is originally designed with the goal of image denoising in mind. Although the authors demonstrated that using a TV smoothing term can largely preserve edges and other important features, the result still looks blurred and blocky (using higher order smoothness term can alleviate the second problem with the cost of exacerbating the first problem). We think that directly using denoising algorithm on the super-resolution problem is somewhat unnatural, with no special attention to the problem at hand. Our algorithm uses variational approach on the linear coefficients instead, so the smoothing term has an effect on the local filter, not directly on the image.

2 local structure estimation

Denoting the given image by u, defined on $\Omega = [0, 1] \times [0, 1]$. We start by assuming that each point $x \in \Omega$ can be approximately expressed as an weighted combination of its neighbors (eight neighbors are used in this paper):

$$u(x) \approx \sum_{i=1}^{8} w^{(i)}(x)u(x + \delta_i) \tag{1}$$

, where $x + \delta_i$ are the neighbors of x, with $\delta_1 = (-h, -h)$ etc. For grayscale images, for example, the above equality is trivially satisfied for an infinite number of combinations of $w^{(i)}, i = 1..8$, given the image u. So trying to estimate w separately for each point x is futile without further constraint. Here we use the

variational approach with a TV norm proposed in [7]:

$$\min_w \int \left(\sum_i w^{(i)} u(\cdot + \delta_i) - u \right)^2 dx + \lambda \sum_{i=1}^{8} \int |\nabla w^{(i)}| dx \qquad (2)$$

,where λ is a parameter that must be set to balance the fidelity term and the smoothness term. This minimization problem can be solved iteratively using the corresponding evolution PDE:

$$w_t^{(i)} = div \left(\frac{\nabla w^{(i)}}{|\nabla w^{(i)}|} \right) - \lambda \left(\sum_i w^{(i)} u(\cdot + \delta_i) - u \right) u(\cdot + \delta_i) \qquad (3)$$

,which can then be discretized and solved numerically.

2.1 Connections to LLE

Given a set of N vectors $\{X_i\}$ in high dimensions. The LLE solves the following minimization problem:

$$\min_W \sum_i (X_i - \sum_j W_{ij} X_j)^2 \qquad (4)$$

with the constraint

$$\sum_j W_{ij} = 1, W_{ij} \geq 0 \qquad (5)$$

where the index j runs over those of X_j that is among the k nearest neighbors of X_i. with nearest neighbors determined by some metric $d(X_i, X_j)$. This minimization problem has a non-trivial solution since it is usually assumed that the dimension of X_i is much bigger than k.

To generalize LLE, we first assume that the data come in with two components $X_i = (Y_i, Z_i)$ (think of Y_i as grid position, and Z_i as intensity value). Now we can minimize the following:

$$\min_W \sum_i (Z_i - \sum_j W_{ij} Z_j)^2 \qquad (6)$$

the index j still runs over k nearest neighbors of X_i. but now with nearest neighbors determined by some metric $d(Y_i, Y_j)$ depending on the other component of X.

If dimension of X_i is small compared to k (as in the case of an image), we must add regularization term to make the problem well-posed. And we will recover the discrete counterpart of (2) after ignoring the convexity constraint(5).

2.2 Comparison with Digital TV Filter

The digital TV filter also implicitly estimates some kind of local linear structure. It repeatedly applies the following filtering step on the current image until

(a) $w^{(i)}$ computed in our method

(b) $h^{(i)}$ for TV filter after convergence

Figure 1: The linear coefficients computed in our method and digital TV filter is quite different for a simple image

convergence (see equation (9) in [1]):

$$u^{t+1}(x) = \sum_{i=1}^{8} h^{(i)}(x)u^t(x + \delta_i) + h^{(0)}(x)u^0(x) \tag{7}$$

, where u^t is the reconstructed image at the t-th iteration, and the coefficients h are based on local image gradient and must be recomputed for each

iteration. Note one difference in form between (1) and (7) is that we don't have a corresponding $w^{(0)}$ in (1). The reason is that if $w^{(0)}$ were added, (2) would have a trivial solution, with $w^{(0)} \equiv 1$, $w^{(i)} \equiv 0, i = 1 \ldots 8$. The coefficients for digital TV filter after convergence is quite different from the coefficients estimated from our method (Fig. 1). This is due to the fact that we estimate the coefficients using a global functional, and the TV filter computes the coefficients using local information only. Another reason for the difference is that our approach estimates orientation-specific weight, the digital TV only estimates local gradient magnitude.

3   super-resolution

Suppose now we are given an $n \times n$ low-resolution image u^{low}, and have also obtained an initial reconstructed $zn \times zn$ high-resolution image $\hat{u}_0$ (by nearest neighbor interpolation, for example), where z is the magnification ratio. Using variational method outlined in the last section on u^{low}, we get eight $n \times n$ weight matrices $w^{(i)}$ representing roughly the local structure in the low-resolution image. These weight matrices cannot be directly used as filters for high-resolution image because it is smaller in size. To obtain suitable $zn \times zn$ weight matrices $\tilde{w}^{(i)}$, we can use linear interpolation on the low-resolution weight matrices to obtain the high-resolution weight matrices. Finally, the weight matrices can act as adaptive filters for interpolation:

$$u^{high}(x) = \sum_{i=1}^{8} \tilde{w}^{(i)}(x)\hat{u}_0(x + \delta_i) \tag{8}$$

4   experimental result

We compare the proposed method with both bicubic interpolation and digital TV filter using grayscale images. We use $z = 3$ in the following experiments. The smoothing parameter used is $\lambda = 0.1$ in all our experiments. It is observed that the result is quite similar for all different smoothing parameters we tried. The reason is probably that the smoothness parameter affect the smoothness of the weights, and only has indirect effect on image itself through the weights. We first apply the algorithm to the house image (Fig. 2). Both bicubic interpolation and digital TV filter oversmooth the image. To make this point clearer, the Sobel edge detector is applied on the results using the same threshold. The edge detector returns less edges on both bicubic interpolated and TV filtered images, which shows the oversmoothing problem of both methods. Another advantage of our method over bicubic interpolation is that the edges are less wiggled, as can be seen from the roof of the house image in (c) of Fig 2, for example.

   As a second example, the fingerprint image is used to see how well our method works for texture images. It is well known that TV filter does not work

on texture images as can be seen from Fig 3. Our method also gives reasonable result in this case.

5   Conclusion

In this work, we proposed a new algorithm for single-image super-resolution problem using variational method. Instead of working on the image space as in the previous work utilizing variational method, we use variational formulation to estimate the local structure of an image. The resulting adaptive filter reflects both local pixel variance and global image information. The experimental result shows some advantage of our method over some previous approaches. A future research direction might be to explore other applications of the variational estimation of the local image structure.

References

[1] Tony Chan, Stanley Osher and Jianhong Shen, "The digital TV filter and nonlinear denoising," *IEEE Transactions on Image Processing,* vol. 10, no. 2, pp. 231-241, 2001

[2] Tony Chan and Jianhong Shen, "Mathematical models for local nontexture inpaintings," *SIAM Journal on Applied Mathematics,* vol. 62, no. 3, pp. 1019-1043, 2002

[3] Adrian Corduneanu and John C. Platt, "Learning spatially-variable filters for super-resolution of text," *IEEE International Conference on Image Processing,* vol. 1, pp. 849-852, 2005

[4] William T. Freeman, Thouis R. Jones, and Egon C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications,* vol. 22, no. 2, pp. 56-65, March/April 2002.

[5] Hu He and Lisimachos P. Kondi, "A regularization framework for joint blur estimation and super-resolution of video sequences," *IEEE International Conference on Image Processing,* vol. 3, pp. 329-332, 2005

[6] Sam T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science,* vol. 290, December 2000

[7] Leonid I. Rudin, Stanley Osher and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D,* vol. 60, pp. 259-268, 1992

[8] Richard R. Schultz and Robert L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing,* vol. 4, no. 6, pp. 996-1011, 1996
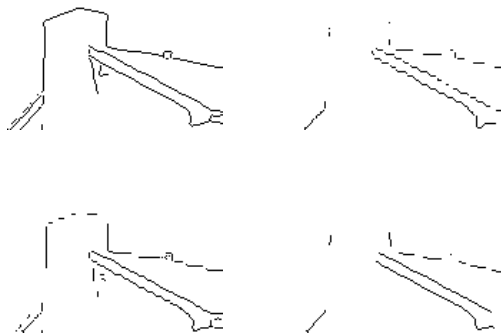
[9] Nathan A. Woods and Nikolas P. Galatsanos, "Non-stationary approximate Bayesian super-resolution using a hierarchical prior model," *IEEE International Conference on Image Processing,* vol. 1, pp. 37-40, 2005
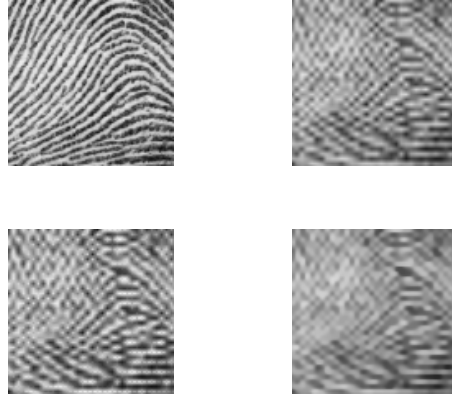
(a) interpolation results
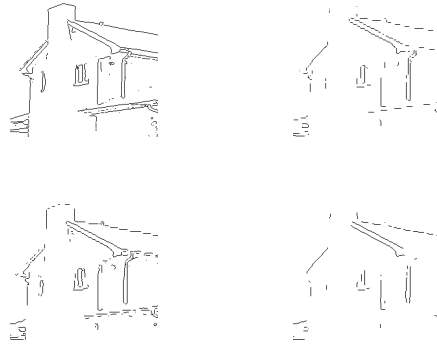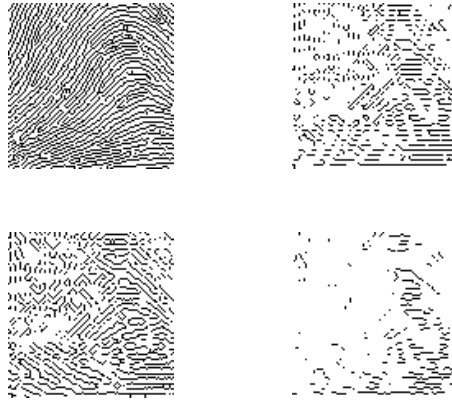

(b) response to Sobel edge detector

(c) zoom-in of part in (b)

Figure 2: Comparison of interpolation results. Upper Left: original high-resolution image. Upper Right: bicubic interpolation, PSNR=26.9. Lower Left: our method, PSNR=27.6. Lower Right: digital TV filter, PSNR=26.9.

(a) interpolation results

(b) response to Sobel edge detector

Figure 3: Comparison of results on fingerprint image. Upper Left: original high-resolution image. Upper Right: bicubic interpolation, PSNR=16.3. Lower Left: our method, PSNR=16.8. Lower Right: digital TV filter, PSNR=16.2.