Single Image Super Resolution

Andrew Burkard, Hui Fang, Alayna Ruberg

4 August 2018

Abstract

Our project seeks to reproduce the results of image super resolution using sparse representation shown by Yang, Wright, Huang, and Ma[YWHM10] using satellite imagery. We discuss super-resolution techniques and Yang's method, then show that satellite images can be reconstructed using Yang's algorithm. We show that the existing dictionary in the Yang algorithm performs similarly to a new dictionary trained on satellite images.

1. Background

Super-resolution (SR) image reconstruction is an area of research that deals with creating a high-resolution (HR) image from one or multiple low-resolution (LR) images. Initial forays into this research required multiple LR images of the same scene or subject to construct an HR image of the same scene or subject, but this approach is difficult to apply due to the lack of availability of multiple LR images. Instead, recent developments in super-resolution techniques have made it possible to construct a HR image from a single LR image.

Yang, Wright, Huang, and Ma [YWHM10] have shown that such reconstruction is possible with the help of sparse representations of low-resolution image patches. Extending the work in [FREM04], which uses ℓ_1 minimization to obtain HR images, they show that using LR image patches that are well-represented with linear combinations of elements from an over-complete dictionary, it's possible to reconstruct an HR image by creating a sparse representation of each LR image patch, and then use the coefficients of the representation to generate the HR output. They compare various methods of SR image reconstruction, showing that their algorithm performs best for both generic images and face images.

There have been continued efforts to build upon the work in [YWHM10] seeking to improve the computation time without compromising the results. In [ZEP12], the authors suggest several modifications of the algorithm to improve computation speed, and even propose a bootstrapping scheme to eliminate the need for a training set at all. [PE14] tackles the underlying assumption in [YWHM10] and [ZEP12] that image patches will have the same exact sparse representation in the LR and HR dictionaries (which implies both must be the same large size), using a statistical model instead to describe the relationship between the two, mainly seeking to decrease the size of the LR dictionary and improve the scale-up performance of the algorithm.

While these extensions are interesting contributions, [YWHM10] remains the main focus for our paper. We will use those techniques on a more specialized set of data: satellite images.

2. Data

We collected 100 square satellite images via the Google Maps Static API [goo18]. Each image is located at the geographic center of one of the world's 100 most populous urban areas. We chose this so that the images would be certain to contain interesting features rather than empty oceans or deserts, as would likely occur when using random coordinates. The images were taken at the API's zoom level of 15, which is enough to clearly make out details in building shapes and streets. Each image is available as a 128x128 low-resolution version and a 256x256 high-resolution version. Because the downscaling is done by the image provider, and the algorithm used is not readily apparent (or necessarily the same for each image), it presents a compelling test case for super-resolution.

We split the images into a set of 90 for training a dictionary and 10 for evaluation. For comparison, we also retain the dictionary trained on images of flowers and everyday scenes from [YWHM10].

A peculiarity of note in the data is the presence of Google watermarks in the imagery. Both the high and low-resolution images contain a Google logo in the bottom left corner and various watermarks appearing in random locations. While the logo is always in the same location in the high and low-resolution versions of an image, the watermarks can appear at different locations with different sizes between the images. Figure 1 shows an example of this. This will have an effect on quantitative comparisons between the expected and the reconstructed images. However, the effect will be present across all reconstruction algorithms used.

(a) A 128x128 low-resolution satellite image

(b) A 256x256 low-resolution satellite image

Figure 1: The same map area in low and high resolution versions. Notice that the logo is in the same pixel-alignment in the bottom-left, but the watermarks are not.

3. Methodology

The work in [YWHM10] has two parts: first training the two dictionaries and then utilizing them to upscale a test set of simulated LR images. The premise of this technique is that an image can be thought of as a signal that has the same sparse representation in a LR and HR dictionary, $\mathcal{D}_L$ and $\mathcal{D}_H$ respectively, made up of atoms of image patches. The HR image $\mathbf{X}$ is made up of image patches $\mathbf{x}$ which have a sparse representation in $\mathcal{D}_H$, as is the LR image $\mathbf{Y}$ is made up of images patches $\mathbf{y}$ with a sparse representation in $\mathcal{D}_L$. If these dictionaries are furthermore isometric (or close to being so), then you can recover $\mathbf{X}$ from $\mathbf{Y}$.

The two dictionaries can be found by training simultaneously via linear programming and convex optimization rather than machine learning algorithms requiring massive training sets, using sparse coding techniques and ℓ_2 minimization. The joint training is possible as the two dictionaries are concatenated together, so that the objective function is

$$\min_{\mathcal{D}_H,\mathcal{D}_L,Z} \frac{1}{N}||X_C - \mathcal{D}_C Z||_2^2 + \hat{\lambda}||Z||_1 \tag{1}$$

where

$$X_C = \begin{bmatrix} \frac{1}{\sqrt{N}} X_H \\ \frac{1}{\sqrt{M}} Y^L \end{bmatrix}, \mathcal{D}_C = \begin{bmatrix} \frac{1}{\sqrt{N}} \mathcal{D}_H \\ \frac{1}{\sqrt{M}} \mathcal{D}_L \end{bmatrix} \tag{2}$$

and the ℓ_1 norm $||Z||_1$ is used to enforce sparsity. N and M represent the dimension of the HR and LR image patches in vector form. Extensive research cited in [YWHM10] has indicated that while this minimization is not convex on $\mathcal{D}_C$ and Z simultaneously, it is convex when either are fixed. Thus a training algorithm that alternates between updating $\mathcal{D}_C$ and Z can reach convergence through linear programming (when updating Z) and quadratically constrained quadratic programming (when updating $\mathcal{D}_C$).

Once the dictionaries are determined from the training set, $\mathcal{D}_L$ is used to recover the sparse representation α from the LR patches $\mathbf{y}$ with ℓ_1 minimization. The objective function of this optimization is

$$\min ||\alpha||_1 \text{ s.t. } ||F\mathcal{D}_L\alpha - F\mathbf{y}||_2^2 \tag{3}$$

where F is the linear feature extraction operator (typically a high-pass filter to capture on textures for more visually pleasing results) and α is the vector of coefficients approximating the LR patch $\mathbf{y}$. One can also use Lagrange multipliers to formulate this optimization as

$$\min_{\alpha} ||F\mathcal{D}_L\alpha - F\mathbf{y}||_2^2 + \lambda||\alpha||_1 \tag{4}$$

which is a familiar regularized Lasso linear regression. The work in [Don] suggests that as long as the vector α is sufficiently sparse, ℓ_1 minimization will successfully recover the sparsest solution.

With the sparse signal recovered, $\mathcal{D}_H$ can then be used to generate the a high resolution image $\mathbf{X}_0$. As the actual HR image $\mathbf{X}$ and the LR $\mathbf{Y}$ are of the same scene, they should be related as

$$\mathbf{Y} = SH\mathbf{X} \tag{5}$$

where S is the down-sampling operator and H is the blurring filter. In other words, the images should be consistent. The optimization earlier to reconstruct $\mathbf{X}$ might produce a result that is actually inconsistent, so we use gradient descent to project the reconstruction result $\mathbf{X}_0$ onto the solution space of $\mathbf{Y} = SH\mathbf{X}$ such that down-sampling our result will get back the input $\mathbf{Y}$.

Our paper will first use the dictionaries provided with the code from [YWHM10] to upscale our own test set of ten LR satellite images, comparing them to the actual HR satellite image and the bicubic upscaling results. Then, we train a new set of dictionaries using patch samples from our our dataset of ninety satellite images, comparing how well it reconstructs the images to the results from the original pretrained dictionary.

To evaluate image quality, we use peak signal-to-noise ratio (PSNR) along with visual inspection, as these were the criterion used in Yang's paper. While PSNR appears to give a definitive, easy to calculate quantitative measure, it does not always align with perceived visual quality. This is especially true for the high frequency details we wish to capture with super-resolution [WBSS04].

4. Results

The details of exactly what locations our ten test images correspond to are found in Table 2. Of the 10 test images, we selected three to compare more closely in this paper that represent the diversity of images tested.

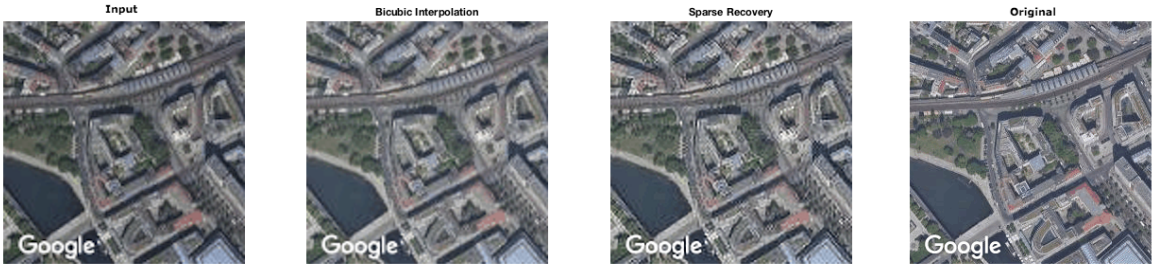

Figure 2: Comparison between sparse recovery and bicubic interpolation using our dictionary trained on 90 satellite images.

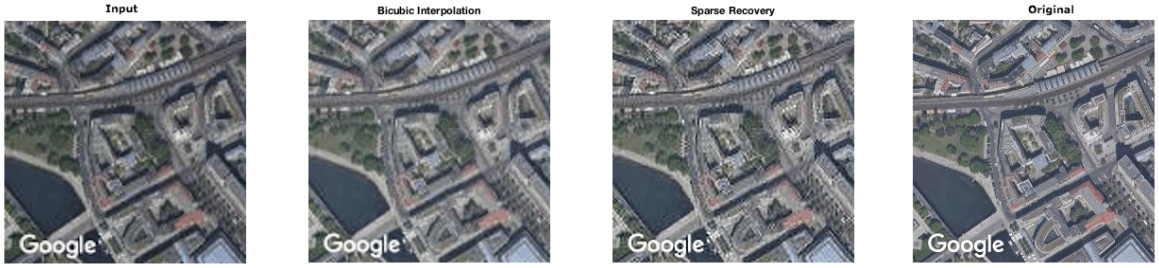

Figure 3: Comparison between sparse recovery and bicubic interpolation using the existing dictionary in the Yang algorithm.

Figure 2 compares the LR input with bicubic interpolation, sparse recovery, and the original HR image using the dictionary we trained using 90 satellite images. The image generated using sparse recovery has high detail, which arguably represents the original HR image more closely than the blurry image generated with bicubic interpolation.

Figure 3 compares the LR input with bicubic interpolation, sparse recovery, and the original HR image using the existing dictionary in the Yang algorithm (which was trained with a variety of images, including flowers, animals, and buildings, but no satellite images). Similar to Figure 2, the image generated using sparse recovery has higher detail than the blurry image generated with bicubic interpolation.

Figures 4 and 5 make the same comparisons, but with a more uniform image. The high level of detail reconstruction with sparse recovery is especially apparent in this image.

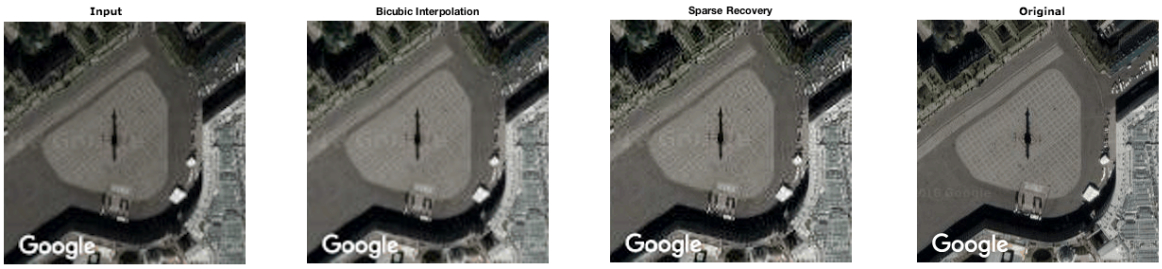

Figure 4: Comparison between sparse recovery and bicubic interpolation our dictionary trained on 90 satellite images.

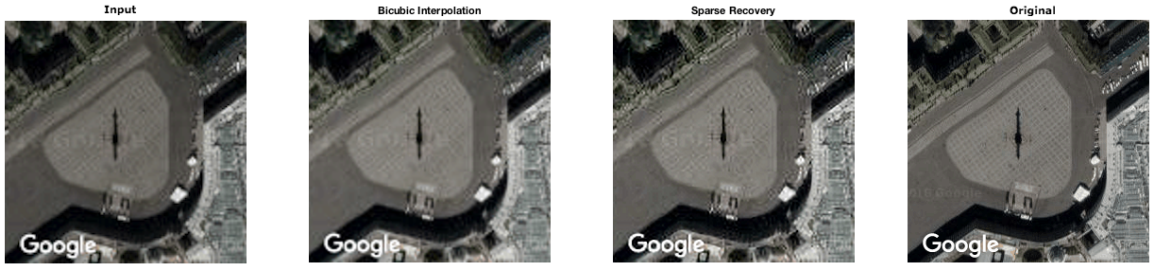

Figure 5: Comparison between sparse recovery and bicubic interpolation using the existing dictionary in the Yang algorithm.

We make the same comparisons in Figure 6 and 7, but with a denser image. The sparse recovery method performs similarly well.

Figure 6: Comparison between sparse recovery and bicubic interpolation our dictionary trained on 90 satellite images.

Figure 7: Comparison between sparse recovery and bicubic interpolation using the existing dictionary in the Yang algorithm.

When comparing images generated using the existing dictionary with the images generated using our satellite dictionary, there is no visible difference between our satellite dictionary and the existing dictionary.

(a) Original HR dictionary

(b) New satellite HR dictionary

Figure 8: There is very little visible difference between the two sparse recovery images.

Visual inspection is only one way of comparing the images, so Yang's algorithm includes the PSNR (peak signal-to-noise ratio) of each reconstructed image as a quantitative measure. A higher PSNR should indicate a better image. Bicubic interpolation consistently resulted in a higher PSNR (see Table 1) for images generated using either dictionary. Our satellite dictionary had slightly higher PSNR values compared to the original Yang dictionary.

Test Image No.	1	2 (Fig.6)	3	4	5	6	7	8 (Fig.2)	9	10 (Fig.4)
Bicubic	21.13	25.67	25.79	25.11	18.88	19.08	25.10	19.48	25.99	27.01
Sparse (our dict)	20.38	25.35	25.54	24.31	18.04	18.12	24.73	18.74	25.81	26.54
Sparse (Yang dict)	20.37	25.28	25.49	24.25	18.02	18.10	24.66	18.72	25.73	26.49

Table 1: Comparison of PSNR values for each image in the test set

Having higher PSNR values for bicubic interpolation is unexpected given our visual assessment and a different result from in [YWHM10], where the sparse recovery algorithm resulted in lower PSNR values compared to bicubic interpolation. We did expect to have higher PSNR

Test Image No.	1	2	3	4	5
Approx Coordinates	41.82, -87.75	43.69, -79.42	43.86, 125.33	45.49, -73.58	45.74, 126.64
Test Image No.	6	7	8	9	10
Approx Coordinates	48.86, 2.33	51.49, -0.11	52.52, 13.40	55.75, 37.61	59.93, 30.31

Table 2: Key for Table 1

values using our new dictionary compared to Yang's, so that result at least aligns with expectations, even if the magnitude of the change is marginal, and the visual difference undetectable to the human eye.

5. Conclusions

When we compare images generated using the existing Yang dictionary to images generated using our dictionary of satellite images, we see very little visible difference. One possible explanation for this is the training set: perhaps if it were larger, the dictionary might perform better. However, it's not altogether surprising that the images are similar, given the features of the existing dictionary images. Indeed, if you visualize the dictionary atoms side-by-side they look to be very similar, being made up of the same types of textural and directional edge details.

(a) Original HR Dictionary (Zoomed) (b) New Satellite HR Dictionary (Zoomed)

Figure 9: Both are made up of a variety of textural and edge details.

This is likely because the dictionaries are trained to pick up on high frequency data and distill to the basic components of such. It seems that no matter what the training image is, it will by and large exhibit those same basic components.

We note that it seems the algorithm sometimes applies a textural patch and creating artifacts when it seems a smooth edge path would be better. One parameter in the sparse recovery algorithm that due to time constraints we had not experimented with is the smoothing parameter λ in equation 4: ours was set at 0.2 for all sparse reconstructions. A higher λ would result in a smoother image, and a lower λ a more textural one. It is possible that if we tried a variety of λ, we would be able to find a more optimal smoothing parameter that produces better results for satellite images.

While Yang's algorithm visually looks superior to bicubic interpolation in reconstructing satellite images, the low PSNR values would suggest otherwise. This again highlights an issue we mentioned while outlining our methodology, questioning whether the the PSNR is an accurate indicator of the quality of the upscaling results. The authors in [YWHM10] also question the validity of the metric, though since it happened to align with the visual assessment that their algorithm produced better results, they did not delve much further into the matter. Since the PSNR does not align with our visual assessment, it is a more immediate concern, but beyond the scope of this paper to develop further.

Regardless, more work can be done to improve the image quality beyond sparse representation, as other research has continued to build upon this technique.

References

[Don] David L. Donoho. *Communications on Pure and Applied Mathematics*, 59(6):797–829.

[FREM04] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing*, 13(10):1327–1344, Oct 2004.

[goo18] Google maps static api, Jul 2018.

[PE14] T. Peleg and M. Elad. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Transactions on Image Processing*, 23(6):2569–2582, June 2014.

[WBSS04] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13(4):600–612, 2004.

[YWHM10] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, Nov 2010.

[ZEP12] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In Jean-Daniel Boissonnat, Patrick Chenin, Albert Cohen, Christian Gout, Tom Lyche, Marie-Laurence Mazure, and Larry Schumaker, editors, *Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

A. Code

Our code and data are available at: https://github.com/aburkard/super-resolution