From the Proceedings of the 2001 IEEE-EURASIP Workshop On Nonlinear Signal and Image Processing

SUPER-RESOLUTION: RECONSTRUCTION OR RECOGNITION?

Simon Baker and Takeo Kanade

The Robotics Institute
Carnegie Mellon University
Pittsburgh PA 15213

ABSTRACT

Super-resolution is usually posed as a *reconstruction* problem. The low resolution input images are assumed to be noisy, down-sampled versions of an unknown super-resolution image that is to be estimated. A common way of inverting the down-sampling process is to write down the reconstruction constraints and then solve them, often adding a smoothness prior to regularize the solution. In this paper, we present two results which both show that there is more to super-resolution than image reconstruction. We first analyze the reconstruction constraints and show that they provide less and less useful information as the magnification factor increases. Afterwards, we describe a "hallucination" algorithm, incorporating the *recognition* of local features in the low resolution images, which outperforms existing reconstruction-based algorithms.

1. INTRODUCTION

Super-resolution is the process of combining several low resolution images to form a higher resolution version. Numerous super-resolution algorithms have been proposed in the literature, dating back to the frequency domain approach of Huang and Tsai [10]. The first step is to register the images; i.e. compute the motion of pixels from one image to the others. We assume that registration has already been performed and focus on the second step; fusing the (aligned) low resolution images into a super-resolution image.

Fusion is normally based on the constraints that the super-resolution image, when appropriately warped and down-sampled to model the image formation process, should yield the low resolution inputs. Super-resolution is therefore posed as a *reconstruction* problem. These reconstruction constraints have been used by numerous authors since first studied by Peleg *et al.* [11] and can easily be embedded in a Bayesian framework incorporating a (smoothness) prior on the super-resolution image [4, 14, 9, 7]. Numerous other refinements are also possible. See [2] for examples.

Is that all there is to super-resolution, pose it as a reconstruction problem, write down the reconstruction constraints, add a smoothness prior, and then solve? In this paper we present two complementary results which show that there is far more. First we analyze the reconstruction constraints. How much information do they really provide? If there is no noise, and we have as many low resolution inputs as we want, could we not recover a super-resolution image with arbitrary resolution? We show that this is not the case. If the input images are quantized to 8-bit values (grey-levels in the range $0-255$), there are limits on how well we can reconstruct a super-resolution image [2]. (Ironically, noise can theoretically be used to increase the number of bits per pixel by "averaging," but the analysis of the noise free case still yields important insights into super-resolution.) In particular, we show that for large magnification factors, there is a huge "volume" of solutions to the reconstruction constraints, including numerous very smooth solutions. The smoothness prior that is typically added to resolve the ambiguity in the large solution space simply ensures that it is one of the overly smooth solutions that is chosen.

How, then, can high-magnification super-resolution be performed? Our theoretical results hold for an arbitrary number of images. Using more low resolution images therefore does not help. Suppose, however, that the input images contain text. Moreover, suppose that it is possible to perform optical character recognition (OCR) and *recognize* the text. If the font can also be determined, it would then be easy to be perform super-resolution. The text could be reproduced at any resolution by simply rendering it from the recognized text and the definition of the font. In the second half of this paper, we describe a super-resolution algorithm based on this principle, which we call *hallucination* [1]. Our hallucination algorithm, however, is based on the recognition of generic local "features" (rather than the characters detected by OCR) so that it can be applied to other phenomena. We have trained our hallucination algorithm separately on both frontal images of faces and computer generated text. We obtained significantly better results than traditional reconstruction-based super-resolution algorithms, both visually and in terms of RMS pixel intensity error.

2. THE RECONSTRUCTION CONSTRAINTS

Denote the low resolution images by $\mathrm{Lo}_i(\mathbf{m})$ where $i = 1, \ldots, N$ and $\mathbf{m} = (m, n) \in \mathbf{Z}^2$ is a vector containing the pixel coordinates. The continuous image formation equation is then:

$$\mathrm{Lo}_i(\mathbf{m}) \;=\; (E_i * \mathrm{PSF}_i)(\mathbf{m}) \;=\; \int_{\mathrm{Lo}_i} E_i(\mathbf{x}) \cdot \mathrm{PSF}_i(\mathbf{x}-\mathbf{m})\,\mathrm{d}\mathbf{x} \tag{1}$$

where $E_i(\cdot)$ is the continuous irradiance light-field that would have reached the image plane of Lo_i under the pinhole model, $\mathrm{PSF}_i(\cdot)$ is the point spread function of the i^{th} camera, and $\mathbf{x} = (x, y) \in \mathbf{R}^2$ are coordinates in the image plane of Lo_i. We decompose the point spread function into two components:

$$\mathrm{PSF}_i(\mathbf{x}) \;=\; (\omega_i * a_i)(\mathbf{x}) \tag{2}$$

where $\omega_i(\mathbf{x})$ models the blurring caused by the optics and $a_i(\mathbf{x})$ models the spatial integration performed by the CCD sensor [3]. Although our analysis is performed for arbitrary $\omega_i(\mathbf{x})$, we do assume a parametric form for a_i:

$$a_i(\mathbf{x}) \;=\; \begin{cases} \frac{1}{S_i^2} & \text{if } |x| \le \frac{S_i}{2} \text{ and } |y| \le \frac{S_i}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Research supported by US DOD Grant MDA-904-98-C-A915.

where $S_i \in [0,1]$ (since the photosensitive area is not always the entire pixel because space is needed to read out the charge [3].)

We wish to estimate a super-resolution image $\mathrm{Su}(\mathbf{p})$ where $\mathbf{p} = (p,q) \in \mathbf{Z}^2$. The coordinate frame of Su is typically defined by that of one of the low resolution input images, say $\mathrm{Lo}_1(\mathbf{m})$. In the introduction we said that we would assume that the input images Lo_i have been registered with each other, and therefore with the coordinate frame of Su. Then, denote the pixel in image Lo_i that corresponds to pixel $\mathbf{p}$ in Su by $\mathbf{r}_i(\mathbf{p})$. From now on we assume that $\mathbf{r}_i(\cdot)$ is known and constant (unlike, say, [9].)

The integration in Equation (1) is performed over the low resolution image plane. Transforming to the super-resolution image plane of Su using the registration $\mathbf{x} = \mathbf{r}_i(\mathbf{z})$ gives:

$$\mathrm{Lo}_i(\mathbf{m}) \;=\; \int_{\mathrm{Su}} E_i(\mathbf{r}_i(\mathbf{z})) \cdot \mathrm{PSF}_i(\mathbf{r}_i(\mathbf{z}) - \mathbf{m}) \cdot \left| \frac{\partial \mathbf{r}_i}{\partial \mathbf{z}} \right| \mathrm{d}\mathbf{z}. \quad (4)$$

Now, $E_i(\mathbf{r}_i(\mathbf{z}))$ is the irradiance that would have reached the image plane of the i^{th} camera under the pinhole model, transformed onto the super-resolution image plane. Assuming that $\mathbf{r}_i(\cdot)$ is correct, and that the radiance of the scene is the same for each camera, $E_i(\mathbf{r}_i(\mathbf{z}))$ should be the same for all $i = 1, \ldots N$. Denoting this value by $E(\mathbf{z})$ (the super-resolution irradiance function), we have:

$$\mathrm{Lo}_i(\mathbf{m}) \;=\; \int_{\mathrm{Su}} E(\mathbf{z}) \cdot \mathrm{PSF}_i(\mathbf{r}_i(\mathbf{z}) - \mathbf{m}) \cdot \left| \frac{\partial \mathbf{r}_i}{\partial \mathbf{z}} \right| \mathrm{d}\mathbf{z}. \quad (5)$$

The goal of super-resolution is then to recover (a representation of) $E(\mathbf{z})$. Doing this requires both increasing the resolution and "deblurring" the image; i.e. removing the effects of the convolution with the point spread function $\mathrm{PSF}_i(\cdot)$.

To proceed we need to specify which continuous function $E(\mathbf{z})$ is represented by the discrete image $\mathrm{Su}(\mathbf{p})$. The simplest case is that $\mathrm{Su}(\mathbf{p})$ represents the piecewise constant function:

$$E(\mathbf{z}) \;=\; \mathrm{Su}(\mathbf{p}) \quad (6)$$

for all $\mathbf{z} \in (p-0.5, p+0.5] \times (q-0.5, q+0.5]$ and where $\mathbf{p} = (p,q) \in \mathbf{Z}^2$ are the coordinates of a pixel in Su. Equation (4) can then be rearranged to give:

$$\mathrm{Lo}_i(\mathbf{m}) \;=\; \sum_{\mathbf{p}} \mathrm{Su}(\mathbf{p}) \cdot \int_{\mathbf{p}} \mathrm{PSF}_i(\mathbf{r}_i(\mathbf{z}) - \mathbf{m}) \cdot \left| \frac{\partial \mathbf{r}_i}{\partial \mathbf{z}} \right| \mathrm{d}\mathbf{z} \quad (7)$$

where the integration is performed over the pixel $\mathbf{p}$; i.e. over $(p-0.5, p+0.5] \times (q-0.5, q+0.5]$. The super-resolution reconstruction constraints are therefore:

$$\mathrm{Lo}_i(\mathbf{m}) \;=\; \sum_{\mathbf{p}} \mathrm{W}_i(\mathbf{m}, \mathbf{p}) \cdot \mathrm{Su}(\mathbf{p}) \quad (8)$$

where:

$$\mathrm{W}_i(\mathbf{m}, \mathbf{p}) \;=\; \int_{\mathbf{p}} \mathrm{PSF}_i(\mathbf{r}_i(\mathbf{z}) - \mathbf{m}) \cdot \left| \frac{\partial \mathbf{r}_i}{\partial \mathbf{z}} \right| \mathrm{d}\mathbf{z}; \quad (9)$$

i.e. a set of linear constraints on the unknown super-resolution pixels $\mathrm{Su}(\mathbf{p})$, in terms of the known low resolution pixels $\mathrm{Lo}_i(\mathbf{m})$. The constant coefficients $W_i(\mathbf{m}, \mathbf{p})$ in general depend on both the point spread function $\mathrm{PSF}_i(\cdot)$ and the registration $\mathbf{r}_i(\cdot)$.

3. ANALYSIS OF THE CONSTRAINTS

We now analyze the super-resolution reconstruction constraints. The equations depend upon: (1) the point spread function $\mathrm{PSF}_i(\cdot)$ and (2) the registration $\mathbf{r}_i(\cdot)$. Without any assumptions on these functions any analysis would be meaningless. If the point spread function is arbitrary, it can be chosen to simulate the "small pixels" of the super-resolution image. Similarly if $\mathbf{r}_i(\cdot)$ is arbitrary, it can be chosen (in effect) to move the camera towards the scene and thereby directly capture the super-resolution image. We therefore have to make some assumptions about the imaging conditions. First we assume that the width of the photosensitive area S is the same for all of the cameras. Second, to outlaw motions which effectively move the camera towards the scene, we assume that the registration between each pair of low resolution images is a translation. We therefore assume that the registrations take the form:

$$\mathbf{r}_i(\mathbf{z}) \;=\; \frac{1}{M}\mathbf{z} + \mathbf{c}_i \quad (10)$$

where $\mathbf{c}_i = (c_i, d_i) \in \mathbf{R}^2$ is a constant (which is different for each low resolution image Lo_i), and $M > 0$ is the *linear magnification* of the super-resolution problem. Incorporating these assumptions the continuous reconstruction constraints simplify to:

$$\mathrm{Lo}_i(\mathbf{m}) = \int_{\mathrm{Su}} \frac{\mathrm{Su}(\mathbf{z})}{M^2} \cdot \mathrm{PSF}_i\left(\frac{\mathbf{z}}{M} + \mathbf{c}_i - \mathbf{m}\right) \mathrm{d}\mathbf{z}. \quad (11)$$

Even given these assumptions, the performance of any super-resolution algorithm will depend upon the number of input images N, the exact values of $\mathbf{c}_i$, and, moreover, how well the algorithm can register the low resolution images to estimate the $\mathbf{c}_i$. Our goal is to show that super-resolution becomes fundamentally more difficult as the linear magnification M increases. We therefore assume that the conditions are as favorable as possible and perform our analysis for an arbitrary number of input images N, with arbitrary values of $\mathbf{c}_i$. Moreover, we assume that the algorithm has estimated these values perfectly. Any results derived under these conditions will only be stronger in practice.

In [2] we analyzed the reconstruction constraints in three different ways. For lack of space, we just present the most important result here, the rapid growth of the volume of solutions of the reconstruction constraints as a function of M.

Why is there a "volume of solutions" anyway? The reason is that images are always intensity quantized (typically to 8-bit values in the range 0–255 grey levels.) Suppose that $\mathrm{int}[\cdot]$ denotes the operator which takes a real-valued irradiance measurement and turns it into an integer-valued intensity. If we incorporate this quantization into our image formation model, the reconstruction constraints in Equation (11) become:

$$\mathrm{Lo}_i(\mathbf{m}) = \mathrm{int}\left[\int_{\mathrm{Su}} \frac{\mathrm{Su}(\mathbf{z})}{M^2} \cdot \mathrm{PSF}_i\left(\frac{\mathbf{z}}{M} + \mathbf{c}_i - \mathbf{m}\right) \mathrm{d}\mathbf{z}\right]. \quad (12)$$

If Su is a finite size image with n pixels, we then have:

Theorem 1 *If* $\mathrm{int}[\cdot]$ *is the standard rounding operator which replaces a real number with the nearest integer, then the volume of the set of solutions of Equation (12) grows asymptotically at least as fast as* $(M \cdot S)^{2 \cdot n}$ *(treating* n *as a constant and varying* M*.)*

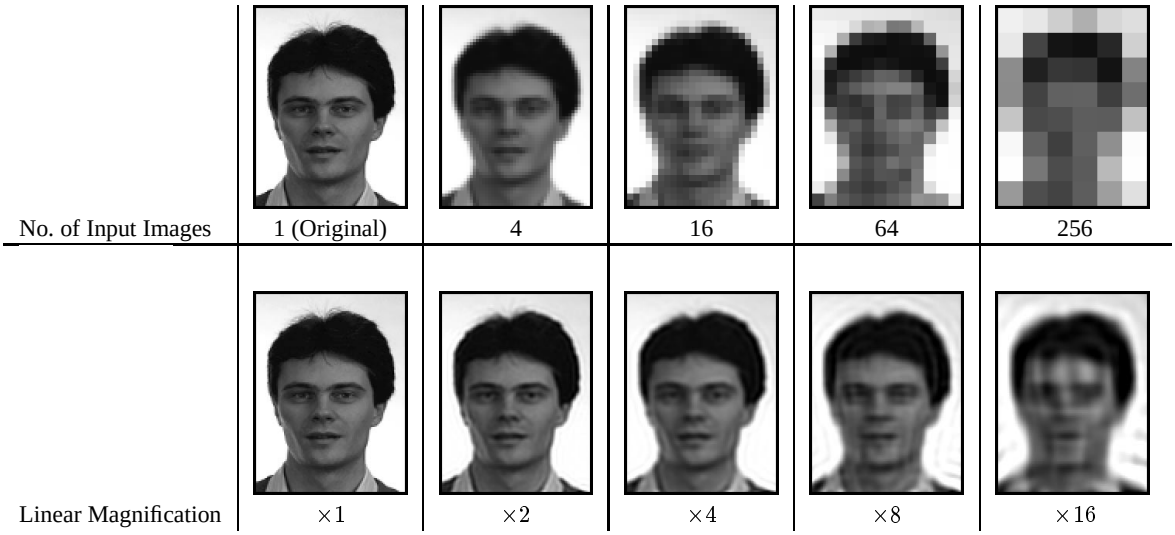


Figure 1: Results of the reconstruction-based super-resolution algorithm [9] for increasing magnification factors. The original high-resolution image (top left) is translated multiple times by random sub-pixel amounts, blurred with a Gaussian, and then down-sampled. Comparing the images in the right-most column, we see that the reconstruction algorithm does quite well given the very low resolution of the input. The degradation in performance as the magnification increases from left to right is very dramatic, however.

Proof: First note that the space of solutions is convex since integration is a linear operation. Next note that one solution of Equation (12) is the solution to:

$$\mathrm{Lo}_i(\mathbf{m}) - 0.5 = \int_{\mathrm{Su}} \frac{\mathrm{Su}(\mathbf{z})}{M^2} \cdot \mathrm{PSF}_i \left(\frac{\mathbf{z}}{M} + \mathbf{c}_i - \mathbf{m} \right) \mathrm{d}\mathbf{z}. \quad (13)$$

The definition of the point spread function as $\mathrm{PSF}_i = \omega_i * a_i$ and the properties of the convolution give $0 \leq \mathrm{PSF}_i \leq 1/S^2$. Therefore, adding $(M \cdot S)^2$ to any pixel in Su is still a solution since the right hand side of Equation (13) increases by at most 1. (The integrand is increased by less than 1 grey-level in the pixel, which only has an area of 1 unit.) The volume of solutions of Equation (12) therefore contains an n-dimensional simplex, where the angles at one vertex are all right-angles, and the sides are all $(M \cdot S)^2$ units long. The volume of such a simplex grows asymptotically like $(M \cdot S)^{2n}$ (treating n as a constant and M and S as variables). The desired result follows. □

This theorem provides an explanation of why the results obtained by most super-resolution algorithms are mixed for high magnification factors. For large magnification factors M, there is a huge volume of solutions to the quantized reconstruction constraints in Equation (12). The smoothness prior which is added to resolve this ambiguity simply ensures that it is one of the overly smooth solutions that is chosen. (Of course, without the prior any solution might be chosen which would generally be even worse.)

To illustrate this point, we conducted an experiment, the results of which are included in Figure 1. We took a high resolution image of a face (shown in the top left of the figure) and synthetically translated it by random sub-pixel amounts, blurred it with a Gaussian, and then down-sampled it. We repeated this procedure for several different down-sampling factors; $M = 2$, 4, 8, and 16. In each case, we generated multiple down-sampled images, each with a different random translation. We generated enough images so that there were as many low resolution pixels in total as pixels in the original high resolution image. For example, we generated 4 half size images, 16 quarter size images, and so on. We then applied the (reconstruction-based) algorithms of Schultz and Stevenson [14] and Hardie *et al.* [9]. The results for Hardie *et al.* [9] are shown in the figure. (The results for Schultz and Stevenson [14] were very similar and so are omitted.) We provided the algorithms with exact knowledge of both the point spread function used in the down-sampling and the random sub-pixel translations (although a minor modification of the iterative algorithm in [9] does a very good job of estimating the translation even for the very low resolution 6×8 pixel images in the right-most column.) Restricting attention to the right-most column of Figure 1, the results look very good. The reconstruction-based algorithm is able to do a decent job of reconstructing the face from tiny input images which barely resemble faces. On the other hand, the performance gets much worse as the magnification increases (from left to right.) In this noise-free synthetic experiment, the best explanation of this drop-off in performance is Theorem 1. See [2] for more details.

4. RECOGNITION-BASED SUPER-RESOLUTION

Is it possible to perform high magnification super-resolution without the results looking overly smooth? In this section, we describe an algorithm [1] that uses the information contained in a collection of recognition decisions (in addition to the reconstruction constraints.) Our approach is related to that of [8] who recently, and independently, proposed a learning framework for low-level vision, one application of which is image interpolation. Besides being applicable to an arbitrary number of images, the other major advantage of our approach is that it uses a prior that is both specific to the class of object [13] and a set of recognition decisions. The approach is also somewhat related to [6], in which the parameters of an "active-appearance" model are used for super-resolution.

We use the (standard) Bayesian formulation of super-resolution [4, 14, 9, 7] in which super-resolution is posed as finding the maximum *a posteriori* (or MAP) super-resolution image:

$$\arg\max_{\mathrm{Su}} \Pr[\mathrm{Su} \,|\, \mathrm{Lo}_i] \;=\; \arg\min_{\mathrm{Su}} \left(-\ln \Pr[\mathrm{Lo}_i \,|\, \mathrm{Su}] - \ln \Pr[\mathrm{Su}] \right). \quad (14)$$

The first term in this expression $-\ln \Pr[\mathrm{Lo}_i \,|\, \mathrm{Su}]$ is the (negative

log) probability of reconstructing the low resolution images Lo_i, given that the super-resolution image is Su. It is therefore set to be a quadratic function of the error in the reconstruction constraints.

Our approach is to embed the results of the recognition decisions in a *recognition-based prior*. Suppose that the outputs of the recognition decisions partition the set of inputs (i.e. the low resolution images Lo_i) into a set of subclasses $\{C_{i,k} \mid k = 1, 2, \ldots\}$. We then define a *recognition-based prior* as one that can be written in the following form:

$$\Pr[\mathrm{Su}] = \sum_k \Pr[\mathrm{Su} \mid \mathrm{Lo}_i \in C_{i,k}] \cdot \Pr[\mathrm{Lo}_i \in C_{i,k}]. \qquad (15)$$

In effect, there is a separate prior $\Pr[\mathrm{Su} \mid \mathrm{Lo}_i \in C_{i,k}]$ for each possible partition $C_{i,k}$. Once the low resolution input images Lo_i are available, the various recognition algorithms can be applied, and it can be determined which partition the inputs lie in. The recognition-based prior $\Pr[\mathrm{Su}]$ then reduces to the more specific prior $\Pr[\mathrm{Su} \mid \mathrm{Lo}_i \in C_{i,k}]$. This prior can be made more powerful than $\Pr[\mathrm{Su}]$ because it can be tailored to the smaller partition $C_{i,k}$.

We only have space to sketch our algorithm. The details are included in other papers [1, 2]. First, we decided to recognize generic local image features (rather than higher level concepts such as human faces or ASCII characters) because we want to apply our algorithm to a variety of phenomena. Motivated by [5] we also decided to use multi-scale features. Then, given a pixel in the low resolution image that we are performing super-resolution on, we find (i.e. recognize) a pixel in a collection of training data that is locally "similar." By similar, we mean that both the Laplacian and the image derivatives are approximately the same, at all scales. To capture this notion precisely, we used the Parent Structure vector introduced in [5]. Recognition in our hallucination algorithm therefore means finding the closest matching pixel in the training data in the sense that the Parent Structure vectors of the the two pixels are the most similar. This search is, in general, performed over all pixels in all of the images in the training data. (If we have frontal images of faces, however, we restrict this search to consider only the corresponding pixels in the training data. In this way, we treat each pixel in the input image differently, depending on its spatial location, similarly to the "class-based" approach of [13].)

For each pixel (m, n) in the input image Lo_i, we "recognize" the pixel that is the most similar in the training data. These recognition decisions partition the inputs Lo_i into a collection of subclasses, as required by the recognition-based prior. We decided to make the recognition-based prior a function of the gradient because the base, or average, intensities in the super-resolution image are defined by the reconstruction constraints. It is the high-frequency gradient information that is missing. Specifically, we define the recognition-based prior to encourage the gradient of the super-resolution image to be close to the gradient of the closest matching training samples. We therefore impose the constraint that the first derivatives of Su at each pixel should equal the derivatives of the closest matching pixel (Parent Structure) in the training data (separately for each pixel (m, n) in each input image Lo_i.)

Both the recognition-based prior (and the reconstruction constraints) are linear in the unknown pixels of Su. (They are however highly non-linear in the input images Lo_i.) Solving the constraints is therefore a high-dimensional linear least squares problem. The relative weights of these constraints are defined by a single parameter. We assume that the reconstruction constraints are the more reliable ones and so set their weighting to be much larger than the recognition-based prior to make them almost hard constraints.

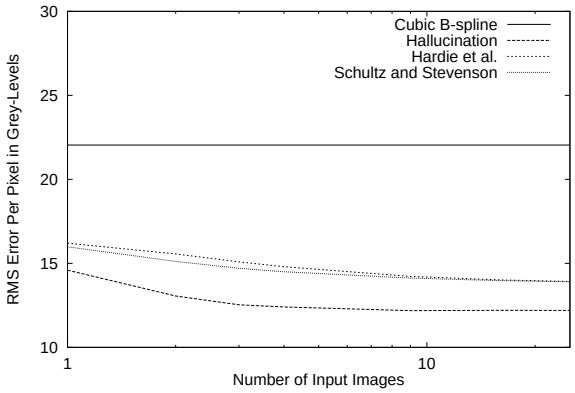


Figure 2: A comparison of our hallucination algorithm with the reconstruction-based super-resolution algorithms of Schultz and Stevenson [14] and Hardie *et al.* [9]. We plot the RMS pixel intensity error computed across the 100 image test set against the number of low resolution input images. Our algorithm outperforms the reconstruction-based super-resolution algorithms by about 20%.

4.1. Experimental Results on Human Faces

Our experiments for human faces were conducted with a training set that is a subset of the FERET dataset [12] consisting of 596 images of 278 individuals (92 women and 186 men). In the class-based approach [13], the input images (which are all frontal) need to be aligned so that we can assume that the same part of the face appears in roughly the same part of the image every time. This alignment was performed by hand marking the location of 3 points, the centers of the two eyes and the lower tip of the nose. These 3 points define an affine warp, which was used to warp the images into a canonical 96×128 pixel image.

We used a "leave-one-out" methodology to test our algorithm. To test on any particular person, we removed all occurrences of that individual from the training set. We then trained the algorithm on the reduced training set, and tested on the images of the individual that had been removed. Because this process is quite time consuming, we used a test set of 100 images of 100 different individuals rather than the entire training set. The test set was selected at random from the training set. We simulate the multiple slightly translated images required for super-resolution using the FERET database by randomly translating the original FERET images multiple times by sub-pixel amounts before down-sampling them to form the low resolution input images.

In our first set of experiments we compare our algorithm with those of Hardie *et al.* [9] and Schultz and Stevenson [14] enhancing 24×32 pixel images four-fold. In Figure 2 we plot the RMS pixel error against the number of low resolution inputs, computed over the 100 image test set. (We compute the RMS error using the original high resolution image used to synthesize the inputs from.) We also plot results for cubic B-spline interpolation for comparison. Since this algorithm is an interpolation algorithm, only one image is ever used and so the performance is independent of the number of inputs. In Figure 2 we see that our hallucination algorithm does outperform the reconstruction-based super-resolution algorithms, from 1 input image to 25. The improvement is consistent across the number of input images and is around 20%.

In Figure 3 we present some example results enhancing 12×16 pixel images eight-fold. The input images are barely recognizable as faces and the facial features such as the eyes, eye-brows,

(a) Input 12×16 (b) Hallucinated (c) Schultz and Stevenson

Figure 3: Selected results for 12×16 pixel images. (The input consists of 4 low resolution input images.) Notice how sharp the hallucinated results are compared to the results for the Schultz and Stevenson [14] algorithm. (The results for Hardie *et al.* [9] are similar to those for Schultz and Stevenson and so are omitted.)

and mouths only consist of a handful of pixels. The outputs, albeit slightly noisy, are clearly recognizable to the human eye. The facial features are also clearly discernible. The hallucinated results are also a huge improvement over the reconstruction-based algorithms of Schultz and Stevenson [14] and Hardie *et al.* [9].

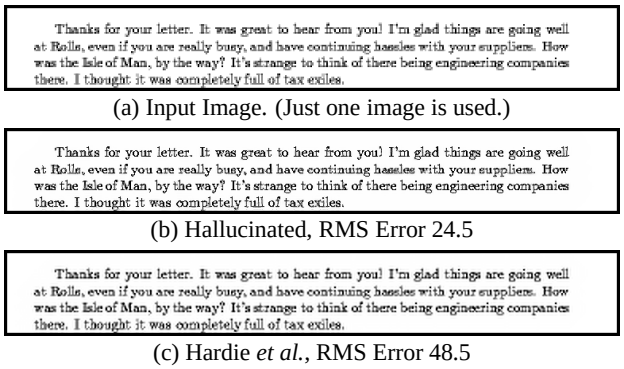
4.2. Experimental Results on Text Data

We also applied our algorithm to text data. In particular, we took the bit-map of a letter and split it into disjoint training and test samples. (The training and test data therefore contain the same font, are at the same scale, and the data is noiseless.) The results are presented in Figures 4. The input in Figure 4(a) is half the resolution of the original. The hallucinated result in Figure 4(b) is the best reconstruction of the text, both visually and in terms of the RMS intensity error. For example, compare the appearance of the word "was" in the second sentence of the text. The hallucination algorithm also has an RMS error of only 24.5 grey levels, compared to over 48 for the algorithm of Hardie *et al.* [9].

5. DISCUSSION

In the first half of this paper we showed that the super-resolution reconstruction constraints provide less and less useful information as the magnification factor increases. The major cause of this phenomenon is the spatial averaging over the photosensitive area; i.e. because S is non-zero. The underlying reason that there are limits on reconstruction-based super-resolution is therefore the simple fact that CCD sensors must have a non-zero photosensitive area in order to be able to capture a non-zero number of photons of light.

In the second half of this paper we showed that recognition decisions may provide an additional source of information for super-resolution algorithms. In particular, we developed a "hallucination" algorithm for super-resolution and demonstrated that this algorithm can obtain far better results than existing reconstruction-based super-resolution algorithms, both visually and in terms of RMS pixel intensity error. Similar approaches may possibly be useful for other reconstruction tasks in image processing.



(a) Input Image. (Just one image is used.)

(b) Hallucinated, RMS Error 24.5

(c) Hardie *et al.*, RMS Error 48.5

Figure 4: The results of enhancing the resolution of a piece of text by a factor of 2. (Just a single input image is used.) Our hallucination algorithm produces a clear, crisp image using no explicit knowledge that the input contains text. The RMS pixel intensity error is also almost a factor of 2 improvement over both of the reconstruction-based algorithms considered in this paper [14, 9].

6. REFERENCES

[1] S. Baker and T. Kanade. Hallucinating faces. In *Proc. of the Fourth ICAFGR*, 2000.

[2] S. Baker and T. Kanade. Limits on super-resolution and how to break them. In *Proc. of CVPR*, 2000.

[3] S. Baker, S.K. Nayar, and H. Murase. Parametric feature detection. *Intl. J. of Computer Vision*, 27(1):27–50, 1998.

[4] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson. Super-resolved surface reconstruction from multiple images. Technical Report FIA-94-12, NASA Ames Research Center, Moffet Field, CA, 1994.

[5] J.S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. In *SIGGRAPH '97*, pages 361–368, 1997.

[6] G.J. Edwards, C.J. Taylor, and T.F. Cootes. Learning to identify and track faces in image sequences. In *Proc. of the Third ICAFGR*, pages 260–265, 1998.

[7] M. Elad and A. Feuer. Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. *IEEE Trans. on Image Processing*, 6(12):1646–58, 1997.

[8] W.T. Freeman and E.C. Pasztor. Learning low-level vision. In *Proc. of the Seventh ICCV*, 1999.

[9] R.C. Hardie, K.J. Barnard, and E.E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. on Image Processing*, 6(12):1621–1633, 1997.

[10] T.S. Huang and R. Tsai. Multi-frame image restoration and registration. *Advances in Computer Vision and Image Processing*, 1:317–339, 1984.

[11] S. Peleg, D. Keren, and L. Schweitzer. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, pages 223–226, 1987.

[12] P.J. Philips, H. Moon, P. Rauss, and S.A. Rizvi. The FERET evaluation methodology for face-recognition algorithms. In *CVPR*, 1997.

[13] T. Riklin-Raviv and A. Shashua. The Quotient image: Class based recognition and synthesis under varying illumination. In *Proc. of CVPR*, pages 566–571, 1999.

[14] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. on Image Processing*, 5(6):996–1011, 1996.