

# An Enhanced Super-Resolution System with Improved Image Registration, Automatic Image Selection, and Image Enhancement

Yu-Chuan Kuo (郭又銓), Chien-Yu Chen (陳建宇), and Chiou-Shann Fuh (傅楸善)

Department of Computer Science and Information Engineering,  
National Taiwan University, Taipei, Taiwan

## Abstract

In this paper we propose a system that reconstructs high-resolution images with improved super-resolution algorithms, which is based on Irani & Peleg iterative method and employs our initial interpolation, image registration, automatic image selection, and image enhancement methods. When the target of reconstruction is a moving object with respect to a stationary camera, high-resolution images can still be reconstructed, whereas previous systems only work well when we move the camera and the displacement of the whole scene is the same.

## Keywords

image improvement, image enhancement, super resolution, image registration, interpolation

## 1. Introduction

Due to the environmental constraint and the resolution of the sensor, we can only get low quality images at times. In order to improve the image quality and resolution by human eyes, more than a single input image is required. With image sequences, a blurring scene, a dim figure, or an unclear object of poor quality can be reconstructed to a super-resolution output image and can then be easily observed and recognized. Previous research regarding super resolution is mainly divided into iterative methods [1], frequency domain methods [2], and Bayesian statistical methods [3].

In Section 2 we introduce an improved super-resolution method with particular choices of initial guess and a better image registration method. Then we propose a novel idea of image selection in Section 3 so as to make the system better and faster. In Section 4 we apply a post-processing of image enhancement to make the output image clearer. Experiments and conclusions are described in Sections 5 and 6 respectively.

## 2. Improved Irani & Peleg Iterative Method

### 2.1 Brief Description of Traditional Irani & Peleg Method

Irani [1] developed the iterative algorithm using image registration to reconstruct the super-resolution image in 1991. The method mainly consists of three phases, initial guess, imaging process, and reconstruction process.

At first, a low-resolution image is taken as reference on which we may reconstruct a “guessed” super-resolution image by interpolation techniques. That is, directly put extra pixels in between the original reference image and then infer the pixel value with respect to its neighbor intensities.

With the initial guess, imaging process is then applied according to the following formula,

$$g_k^{(n)} = (T_k(f^{(n)}) * h) \downarrow s$$

where  $g_k$  is the  $k$ th observed image frame;  $f$  is the super-resolution scene;  $h$  is the blurring operator;  $T_k$  is the transformation operator that transforms other low-resolution images to the reference frame; and  $s$  is the down-sampling operator.

The whole process represents the imaging process that takes pictures with a simulated camera.

Then, we compare the result of the imaging process with the real low-resolution image we have in hand. The differences are used to improve the reference image in the current iteration.

$$f^{(n+1)} = f^{(n)} + \frac{1}{K} \sum_{k=1}^K T_k^{-1}(((g_k - g_k^{(n)}) \uparrow s) * p)$$

where  $K$  is the total number of low-resolution images that are used;  $p$  is the de-blurring operator;  $f^{(n)}$  is the reconstruction result after  $n$ th iterations. Repeatedly apply the above process until the reference frame converges to a satisfactory result after several iterations.

## 2.2 Improved Initial Guess

When the magnification factor and reconstruction image sizes get larger, the computation time becomes longer. Typical runtime is on the order of hours and are machine-dependent. The initial guess as described above will largely affect the performance of our result, and if a better initial guess is applied, great amount of computation time will be saved.

Because initial guess is done merely once at the beginning of the process, the complexity of the whole Irani & Peleg method does not depend on the complexity of the initial guess, which is based on interpolation techniques. Here we introduce only 3<sup>rd</sup> order (cubic) interpolation that takes 4 neighboring pixels into consideration and then evaluate performances of 1st~5th order initial guess by Peak Signal-to-Noise Ratio (PSNR)<sup>1</sup>.

Third order, or cubic interpolation considers 4 unknown variables. Suppose the interpolation function is  $y = f_3(x) = ax^3 + bx^2 + cx + d$ , and known neighboring pixels include  $(-1, A)$ ,  $(0, B)$ ,  $(1, C)$ , and  $(2, D)$ ; then

$$\begin{aligned} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} &= \begin{pmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \\ \Rightarrow \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} &= \begin{pmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} \\ &= \begin{pmatrix} -0.1667 & 0.5 & -0.5 & 0.1667 \\ 0.5 & -1 & 0.5 & 0 \\ -0.3333 & -0.5 & 1 & -0.1667 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} \end{aligned}$$

Similarly, other orders of interpolation also solve for coefficients of  $f_n(x)$ . Applying  $f_n(x)$  in 2-dimensional interpolation algorithm, we can get all pixels in an integral row up-sampled first by interpolation in  $x$  direction, and then get all pixels by interpolation in  $y$  direction.

We observe that different order of interpolation results in different initial-guess images and different convergence rates of image quality as the number of iteration grows. By choosing the most appropriate order of interpolation, we will get the best results of Irani & Peleg method, since initial guess has a great influence on the performance of image registration and on the necessary number of iterations to achieve the peak image result. In most situations, 3<sup>rd</sup> order interpolation ranks the best choice of initial guess if both complexity and reconstructed image quality are concerned. We evaluate the performance of different orders of interpolation by PSNR between the original image and reconstructed images. The experimental results are

shown in Figure 1.

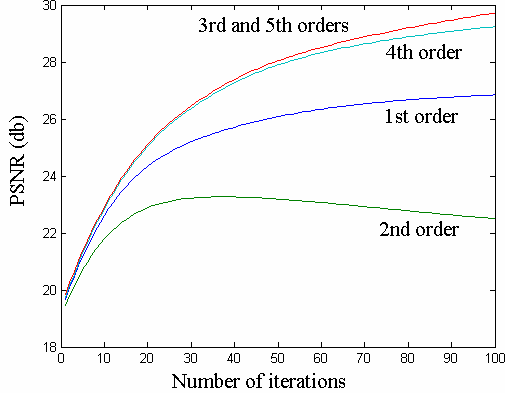


Figure 1. Performance with 1<sup>st</sup> to 5<sup>th</sup> order of interpolation applied for initial guess.

Using initial guess with different orders of interpolation has different PSNR convergence rates. Blue, green, and cyan curves represent 1<sup>st</sup>, 2<sup>nd</sup>, and 4<sup>th</sup> orders respectively. Performance with 3<sup>rd</sup> and 5<sup>th</sup> orders of interpolation achieves similar results as the red curve shows.

## 2.3 Improved Image Registration

Image registration is critical in the performance of our algorithm since each iteration refines each pixel on the high-resolution image using the information of the corresponding pixel on the low-resolution images. We introduce two methods to achieve image registration. The local matching technique looks for a set of corresponding pairs and the global matching technique looks for the corresponding position of the whole low-resolution image on the simulated high-resolution image.

### 2.3.1 Local Matching Technique

For each interesting point  $(x, y)$  on low-resolution image  $i$ , the mapping function  $LR_i(x, y)$  looks for its corresponding point  $(u, v)$  on the simulated high-resolution image. Function  $LR_i(x, y)$  minimizes absolute difference  $LAD_i(x, y; u, v)$  within a local window  $w$ . Translation  $LT_i(x, y)$  is the translation

between point  $(x, y)$  and point  $(u, v)$  on the high-resolution image.

$$LAD_i(x, y; u, v) = \sum_{(m,n) \in w} |I_i(x+m, y+n) - I_o(u+m, v+n)|$$

$$LR_i(x, y) = \arg \min_{(u,v)} LAD_i(x, y; u, v)$$

$$LT_i(x, y) = LR_i(x, y) - (x, y) * \text{MagnificationFactor}$$

In order to get more accurate image registration and then reconstruct the high-resolution image of a moving object, we choose interesting points of corresponding pairs under the following constraints.

- The gradient at an interesting point should be larger than a threshold.

For each interesting point on a low-resolution image, we look for the corresponding point on the simulated high-resolution image where higher local-complexity around the point is required.

- The translation between each corresponding pair should not be zero.

Our goal is to reconstruct a moving object on a stationary background so we consider the zero-translated points as background. These points should not be chosen as interesting points.

Under the constraints, we can find a set of corresponding pairs. We use the mode translation of the set to represent the translation of image  $i$ . Set  $P_i$  is the set of interesting points of image  $i$ .

$$T_i = M(\{LT_i(x, y) | (x, y) \in P_i\})$$

where  $M(A)$  is the mode of list  $A$ .

### 2.3.2 Global Matching Technique

Global matching function  $GR(i)$  searches the corresponding position  $(u, v)$  of low-resolution image  $i$ . Function  $GR(i)$  minimizes the absolute difference within the whole image  $GAD(u, v)$ .

$$GAD_i(u, v) = \sum_{(x,y) \in i} |I_i(x, y) - I_o(u+x, v+y)|$$

$$GR(i) = \arg \min_{(u,v)} GAD_i(u, v)$$

Then, the translation  $T_i$  of the image  $i$  is  $GR(i)$ .

## 2.4 User-defined Boundary

To improve the speed and the accuracy of image registration, we only look for corresponding pairs of the moving object. Thus we choose interesting points inside a user-defined boundary. For global matching function, the user-defined boundary should be bounded in the object, i.e. each pixel on the area should belong to the object as well, so that the interesting points will not lie on the background and mis-registration caused by occlusion can be eliminated. For local matching function, the user-defined area could be larger than the object. The point belonging to background can be ignored since the relative translation is zero as described in Section 2.3.1. For objects that we cannot use a rectangular area to bind, we suggest applying local matching function to calculate the translations.

## 3. Automatic Selection from Image Sequences

With a large number of image sequences, it not only costs much time to reconstruct a high-resolution image but also reduces the quality if some images are mis-registered. We propose a novel way to select a minimal number of useful images. To reconstruct a high-resolution image of magnification factor of  $n$ , we only need one image to get sufficient information for each mod-translation (modulus of translation). Mod-translation for image  $i$  is defined as  $T_i \bmod \text{MagnificationFactor}$ . Our algorithm can select the best image for each mod-translation. Thus, we exploit the most useful and minimal number of images to reconstruct high-resolution images.

### 3.1 Automatic Selection with Global Matching Technique

We propose two criteria to select the better image from two images with the same mod-translation.

For two images  $i$ ,  $j$  having the same mod-translation and  $(u_i, v_i) = T_i$ ,  $(u_j, v_j) = T_j$ , we select image  $i$  if

$$GAD_i(u_i, v_i) < GAD_j(u_j, v_j)$$

Most registration has a  $\min GAD_i(u, v)$  of nonzero because the intensities of simulated high-resolution are produced by interpolation. If the initial guess is reasonably correct, the real translation of image  $i$  having smaller  $GAD_i(u_i, v_i)$  will be closer to an integral grid so the error would be minimized after the real translation is rounded to  $T_i$ .

### 3.2 Automatic Selection with Local Matching Technique

Mis-registered images would reduce the quality of high-resolution images. Therefore, we discard these mis-registered images and select the most useful and minimal number of low-resolution images by comparing the remaining images with the same mod-translation.

Image  $i$  that should not be discarded has the following criteria.

- The number of interesting points,  $\#P_i$ , under the constraints described in Section 2.3.1 should be larger than a threshold.
- The ratio of the mode of the translation,  $\#\{(x, y) | (x, y) \in P_i, LT_i(x, y) = T_i\} / \#P_i$ , should be larger than a threshold.
- The ratio of the second mode of the translation,  $\#\{(x, y) | LT_i(x, y) = M(LT_i(p, q) - T_i), (x, y)(p, q) \in P_i\} / \#P_i^2$  should be smaller than a threshold.

For two images  $i$  and  $j$  having the same mod-translation, and  $(u_i, v_i) = T_i$  and  $(u_j, v_j) = T_j$ , we select image  $i$  if

$$a. \quad s_i^2 < s_j^2$$

The variance of  $\{LT_i(x, y) | (x, y) \in I_i\}$  is defined

as  $\mathbf{s}_i^2 = \mathbf{s}_{xi}^2 + \mathbf{s}_{yi}^2$ . Symbols  $\mathbf{s}_{xi}^2$  and  $\mathbf{s}_{yi}^2$  are

the variances of the translation values along  $x$  and  $y$  axes respectively. When we calculate variances, the noises should not be taken into consideration. A noise is labeled if the number of the translation is one. If the variance is smaller, the registration is more satisfactory for each interesting point and is closer to the real answer.

Table 1 indicates the performance of our system with and without automatic selection.

Table 1. The performance of our system with and without automatic selection. We use five sets of 62x62 low-resolution images and the magnification factor is 3. (Measured with Intel Pentium III and 128MB RAM)

|                           |                   | Run Time<br>(seconds) | PSNR<br>(db) |
|---------------------------|-------------------|-----------------------|--------------|
| Local Matching Technique  | With Selection    | 496.2                 | 26.78        |
|                           | Without Selection | 582.4                 | 26.66        |
| Global Matching Technique | With Selection    | 75.8                  | 26.78        |
|                           | Without Selection | 155.4                 | 26.66        |

#### 4. Image Enhancement Post-processing

In order to make the super-resolution images much clearer and more recognizable, we add a post-processing that applies some basic image enhancement techniques [5].

Edge sharpening method improves the resolvability of the image. In our system we apply Laplacian

mask  $\begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$  for convolution. After high-pass

filtering, the image becomes sharp-edged and the reconstructed image is more easily recognized (as shown in Figure 2.).

Besides, local histogram equalization is used to make the image more adaptive to human eyes and median filter is applied so as to remove impulse noises. Both of those image enhancement techniques are helpful for human recognition in our system.

## 5. Results

### 5.1 Reconstructing High-resolution Images with Moving Simulated Camera

We simulate a camera by taking an image as original scene and down-sampling the original scene into several pictures. Using the simulated camera, we take pictures beginning at different points, i.e. the simulated camera moves when taking pictures. Then, our algorithm takes these pictures as inputs and reconstructs a high-resolution image iteratively and magnification factor of length is 4. The aim is to reconstruct high-resolution images of the whole scene so the user-defined area in registration should be the same with the area of low-resolution images. The performance is good after sufficient iterations, as shown in Figures 3 and 4.

### 5.2 Reconstructing High-resolution Objects from Image Sequences of Moving Object

In section 4.1, we simulate a camera taking pictures when moving on a static scene. In this section, we take 27 pictures of a moving object with a real camera. On each picture, only the object moves slightly and the background stays immobile. Our aim is to reconstruct the high-resolution image of that object and magnification factor of length is 2. To improve the speed and accuracy of registration, we specify an area within the object. As the number of iteration increases, on the high-resolution image, the object becomes clearer while the background becomes blurry and words are more discernible on the edge sharpened high-resolution image as Figure 5 shows.

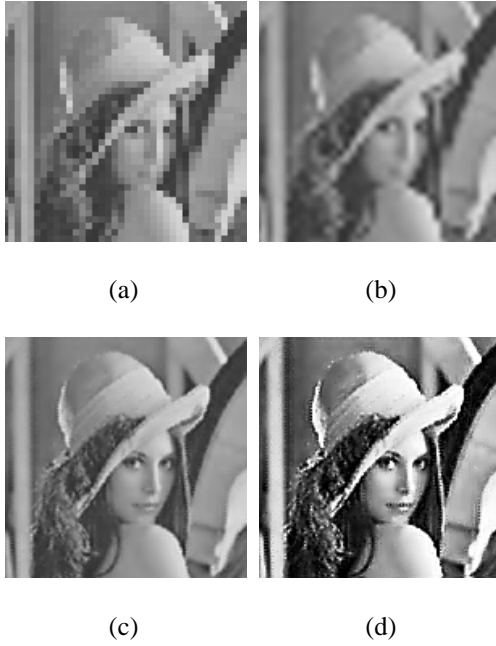


Figure 2. (a) One of low-resolution images. (b) Initial guess. (c) Reconstructed image after 100 iterations. (d) Enhanced final output image.

## 6. Conclusions

We have developed an image reconstruction system that constitutes improved super resolution iterative method, intelligent selection from image sequences, and final image enhancement process.

First, we suggest a complex initial guess using 3<sup>rd</sup> order interpolation in order to reduce the number of iterations required and improve the performance of image registration. Second we propose a better image registration method, including using gradient constraint, user-defined boundary, and translation thresholding, which tends to capture only the information of the moving object instead of the stationary background and allows the reconstruction of image sequences of a moving object in a scene. Then we introduce a novel idea of intelligent image selection. By filtering out redundant and useless images, the system runs

dramatically faster. Besides, because we discard poor-quality images, final image quality will be better. Finally we add a post-processing of image enhancement that contains edge crispening and local histogram equalization to make the target objects in image sequences more recognizable.

## Reference

- [1] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP: Graphical Models and Image Proc.*, Vol. 53, pp. 231-239, 1991.
- [2] R. Y. Tsai and T. S. Huang, "Multiframe Image Restoration and Registration," in *Advances in Computer Vision and Image Processing*, Vol. 1 (T. S. Huang, ed.), pp. 317-339, Greenwich, CT: Jai Press, 1984.
- [3] P. Cheeseman, B. Kanefsky, R. Kruft, J. Stutz, and R. Hanson, "Super-Resolved Surface Reconstruction from Multiple Images," *NASA Technical Report FIA-94-12*, 1994.
- [4] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-Resolution Image Reconstruction for Lower-Resolution Image Sequences and Space-Varying Image Restoration," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. III, pp. 169-172, San Francisco, CA, 1992.
- [5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.
- [6] W. K. Pratt, *Digital Image Processing*, 2nd Ed., Wiley, New York, 2001.



Figure 3. Results of our proposed method with a fixed scene and a simulated moving camera. (a) One of low-resolution images. (b) Initial guess. (c) Reconstructed image after 100 iterations. (d) Enhanced final output image.

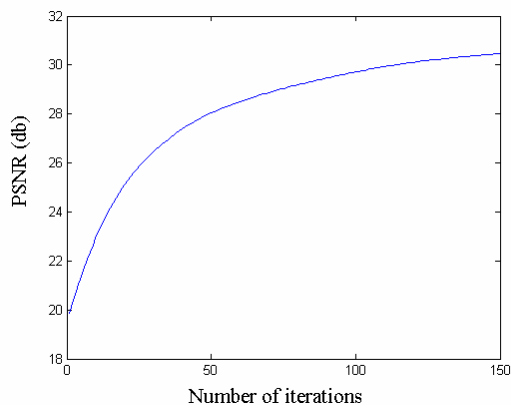
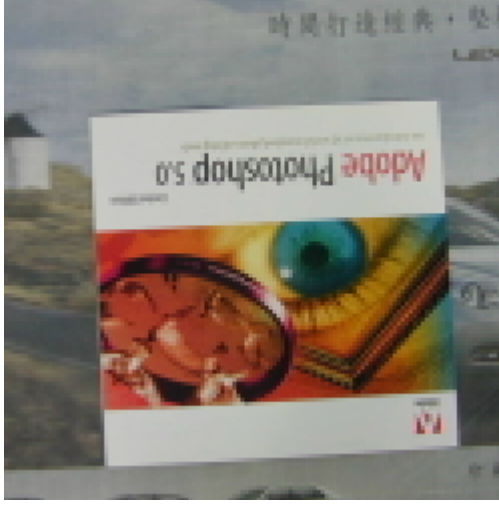


Figure 4. PSNR of iteratively output images. As the number of iterations grows, the performance, evaluated by PSNR, converges.

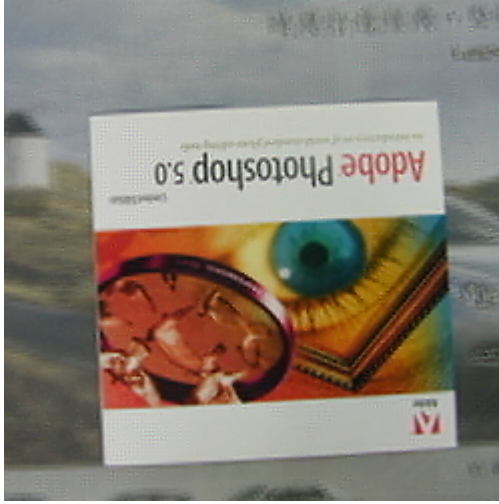




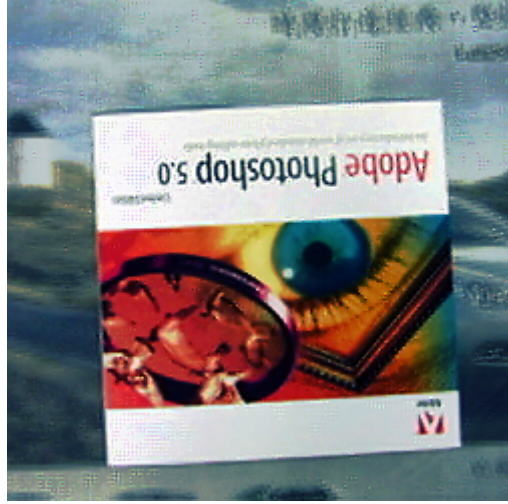
(a)



(b)



(c)



(d)

Figure 5. Results of our proposed method with a moving scene and a fixed real camera. (a) One of low-resolution images. (b) Initial guess. (c) Reconstructed image after 100 iterations. (d) Enhanced final output image.

---

1.

$$\text{MSE} = \sum \frac{[f(i, j) - F(i, j)]^2}{N^2}$$

$$\text{RMSE} = \sqrt{\text{MSE}}$$

$$\text{PSNR} = 20 \log_{10} \left( \frac{255}{\text{RMSE}} \right)$$

2. Define –

$$A - x = \{a \mid a \in A, a \neq x\}$$

where  $A \subset R^n \times R^n$  and  $a, x \subset R \times R$ .

For example,  $\{(x_1, y_1)(x_1, y_1)(x_2, y_2)(x_3, y_3)\} - (x_1, y_1) = \{(x_2, y_2)(x_3, y_3)\}$