

IMAGE RECONSTRUCTION WITH IMPROVED SUPER-RESOLUTION ALGORITHM

CHIEN-YU CHEN, YU-CHUAN KUO and CHIOU-SHANN FUH*

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

**fuh@csie.ntu.edu.tw*

In this paper we propose a technique that reconstructs high-resolution images with improved super-resolution algorithms, based on Irani and Peleg iterative method, and employs our suggested initial interpolation, robust image registration, automatic image selection and image enhancement post-processing. 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 restoration; image enhancement; super resolution; image registration; interpolation; image selection.

1. Introduction

Due to environmental constraints and resolution of image sensors, 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,³ frequency domain methods,⁶ and Bayesian statistical methods.¹

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

2. Improved Irani and Peleg Iterative Method

2.1. Brief description of traditional Irani and Peleg method

Irani³ 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 defined by point-spread-function (PSF) of the image sensor; 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 results 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 iteration. Repeatedly apply the above process until the reference frame h converges to a satisfactory result after several iterations.

2.2. Improved initial guess

When the magnification factor and reconstruction image sizes become larger, the computation time gets 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 the initial guess is done merely once at the beginning of the process, the complexity of the whole Irani and Peleg method does not depend on the complexity of the initial guess, which is based on interpolation techniques. Here we introduce only first order (bilinear), third order (cubic), and fifth order interpolation that take different numbers of neighboring pixels into account and then evaluate the performances of super resolution algorithms with first to fifth orders initial guess

by measuring Peak Signal-to-Noise Ratio (PSNR)^a between the original image and reconstructed images from simulated low-resolution image sequences.

First order, or bilinear interpolation considers two unknown variables. Assume the interpolation function is $y = f_1(x) = ax + b$, and known neighboring pixels include $(0, A)$ and $(1, B)$; then

$$\begin{aligned} \begin{pmatrix} A \\ B \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \\ \Rightarrow \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} A \\ B \end{pmatrix} \\ &= \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} A \\ B \end{pmatrix}. \end{aligned}$$

Third order, or cubic, interpolation considers four unknown variables. Assume 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

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

Fifth order interpolation considers six unknown variables. Assume the interpolation function is $y = f_5(x) = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$, and known neighboring pixels include $(-2, A)$, $(-1, B)$, $(0, C)$, $(1, D)$, $(2, E)$, and $(3, F)$; then

^aMSE = $\sum \frac{[f(i,j) - F(i,j)]^2}{N^2}$
 RMSE = $\sqrt{\text{MSE}}$
 PSNR = $20 \log_{10}(\frac{255}{\text{RMSE}})$

$$\begin{pmatrix} A \\ B \\ C \\ D \\ E \\ F \end{pmatrix} = \begin{pmatrix} -32 & 16 & -8 & 4 & -2 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 32 & 16 & 8 & 4 & 2 & 1 \\ 243 & 81 & 27 & 9 & 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = \begin{pmatrix} -32 & 16 & -8 & 4 & -2 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 32 & 16 & 8 & 4 & 2 & 1 \\ 243 & 81 & 27 & 9 & 3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} A \\ B \\ C \\ D \\ E \\ F \end{pmatrix}$$

$$= \begin{pmatrix} -0.0083 & 0.0417 & -0.0833 & 0.0833 & -0.0417 & 0.0083 \\ 0.0417 & -0.1667 & 0.2500 & -0.1667 & 0.0417 & 0 \\ -0.0417 & -0.0417 & 0.4167 & -0.5833 & 0.2917 & -0.0417 \\ -0.0417 & 0.6667 & -1.25 & 0.6667 & -0.0417 & 0 \\ 0.05 & -0.5 & -0.3333 & 1 & -0.25 & 0.0333 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix}.$$

Similarly, other orders of interpolation are also solved for coefficients of $f_n(x)$. Applying $f_n(x)$ in two-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 as demonstrated in Fig. 1. In Fig. 1(a), to interpolate the value of pixel P in a two-dimensional image, A' , B' , C' and D' are computed first. Known values of A_1 , B_1 , C_1 and D_1 are used to determine the coefficients of the interpolation function $f_3(x)$. Then A' is interpolated using one-dimensional interpolation. Similarly, B' , C' and D' are determined according to A_i , B_i , C_i and D_i , for $i = 2, 3$ and 4 . Finally the value of pixel P is computed with one-dimensional interpolation in vertical direction. In Fig. 1(b), one-dimensional third order interpolation is used to determine the value of pixel P with known values of A' , B' , C' and D' .

We observe that different orders of interpolation result 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 and 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, third order interpolation ranks the best choice of initial guess if both complexity and reconstructed image quality

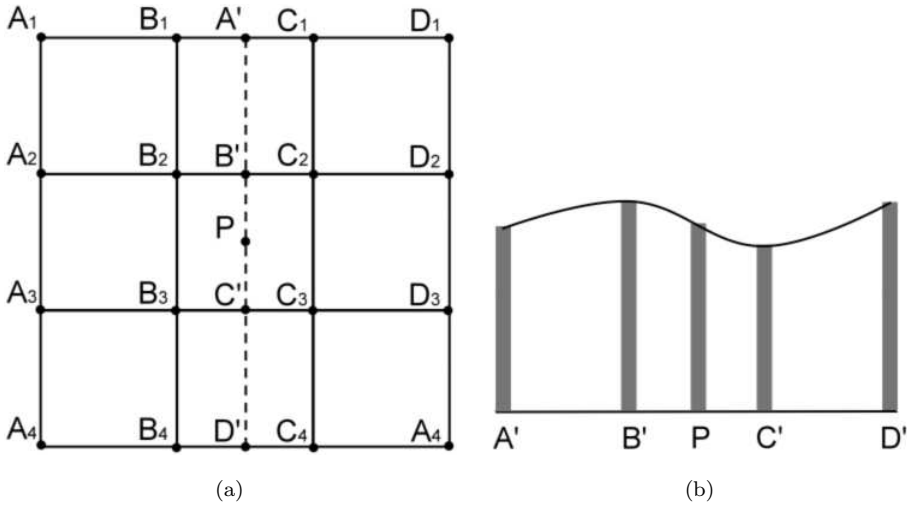


Fig. 1. Third order interpolation in two-dimensional image.

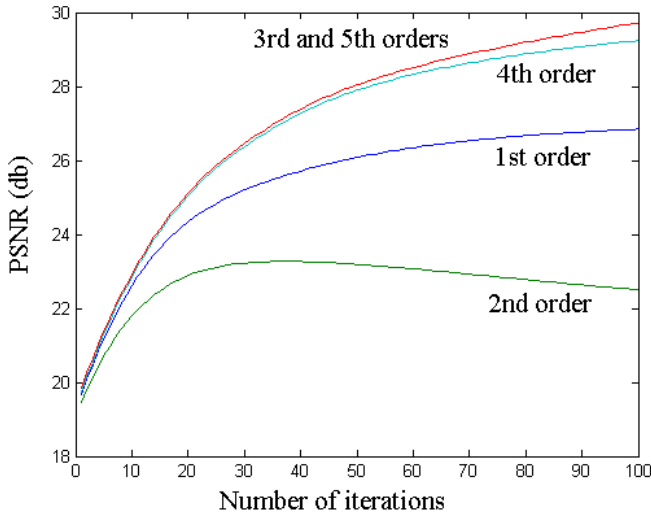


Fig. 2. Performance with first to fifth order of interpolation applied for initial guess. The curves of neutral pictures will not differ because neutral pictures can be considered color pictures in which every pixel has the same intensities in red, green and blue channels. When we use interpolation of second order, the image quality degraded because the initial guess is not precise enough.

are concerned. We evaluate the performance of different orders of interpolation by PSNR between the original and reconstructed images. Figure 2 shows that using initial guess with different orders of interpolation has different PSNR convergence rates. Blue, green and cyan curves represent first, second, and fourth orders, respectively. Performance with third and fifth orders of interpolation achieves similar results as the red curve shows. If the initial guess is not precise enough, it will lead

to misregistration and then result in degradation under a growing of number of iterations.

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 high-resolution 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.

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

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

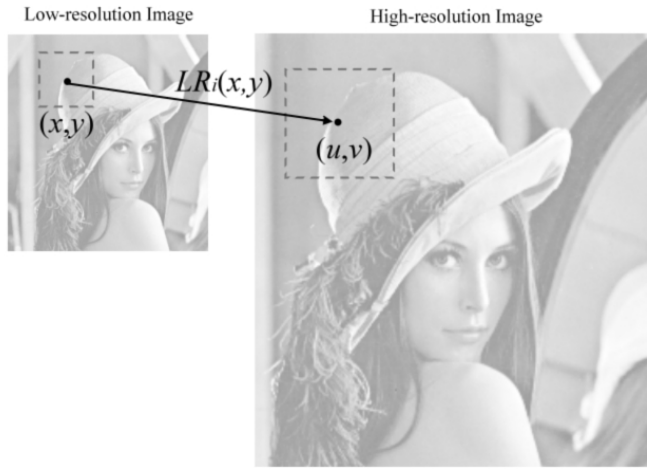
$$LT_i(x, y) = LR_i(x, y) - (x, y) * \text{Magnification Factor}.$$

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.

- (a) 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.
- (b) 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 as the set of interesting points of image i .

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



(a)



(b)

Fig. 3. (a) Image registration using local matching technique, (b) Image registration using global matching technique.

If A is a list of symbols, $M(A)$ represents the mode of the list; that is, the symbol which shows the most times.

The local matching technique used in Improved Image Reconstruction System, or IIRS, is shown in Fig. 3(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 $GR(i)$.

The results of global matching technique used in IIRS are shown in Fig. 3(b).

2.4. User-defined boundary

To improve the speed and 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 misregistration 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 Sec. 2.3.1. For objects within which we cannot define a sufficiently large rectangular area, 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 misregistered. 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, as illustrated by the example in Fig. 4. Figure 4(a) shows the translations and mod-translations of 12 images and in Fig. 4(b), the mod-translations are fitted into nine grids. There are redundant images when (mod-translation.x, mod-translation.y) equals to (2,0), (0,2), or (0,1).

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

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

i	T_i	Mod-translation
1	(0, 0)	(0, 0)
2	(5, -2)	(2, 1)
3	(4, 5)	(1, 2)
4	(-3, -1)	(0, 2)
5	(10, -8)	(1, 1)
6	(4, -3)	(1, 0)
7	(-2, -7)	(1, 2)
8	(9, 8)	(0, 2)
9	(0, 4)	(0, 1)
10	(4, -3)	(2, 0)
11	(-4, -7)	(2, 2)
12	(8, -3)	(2, 0)

(a)

		Mod-translation.x		
		0	1	2
Mod-translation.y	i			
	0	1	6	10, 12
	1	9	5	2
	2	8, 4	3, 7	11

(b)

Fig. 4. An example of automatic selection when the magnification factor of length is 3.

Most registration has a $\min GAD(u, v)$ 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

Misregistered images would reduce the quality of high-resolution images. Therefore, we discard these misregistered 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.

- (a) The number of interesting points, $\#P_i$, under the constraints described in Sec. 2.3.1 should be larger than a threshold.
- (b) 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.
- (c) The ratio of the second mode of the translation,

$$\#\{(x, y) | LT_i(x, y) = M(LT_i(p, q) - T_i), (x, y)(p, q) \in P_i\} / \#P_i^b$$

should be smaller than a threshold.

^bDefine —

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

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

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

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) $\sigma_i^2 < \sigma_j^2$.

The variance of $\{LT_i(x, y) | (x, y) \in I_i\}$ is defined as $\sigma_i^2 = \sigma_{x_i}^2 + \sigma_{y_i}^2$. Symbols $\sigma_{x_i}^2$ and $\sigma_{y_i}^2$ 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 closer to the real answer.

Table 1 indicates the performance of system with and without automatic selection. Local Matching (LMT) and Global matching techniques (GMT) are both considered.

Table 1. The performance of IIRS with and without automatic selection. We use five sets of 62×62 low-resolution images and the magnification factor is 3. (The results are measured on machines with Intel Pentium III and 128MB RAM.)

		Computation	
		Time	PSNR
		(seconds)	(db)
LMT	With Selection	155.4	26.78
	Without Selection	582.4	26.66
GMT	With Selection	75.8	26.78
	Without Selection	496.2	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.²

Edge sharpening method improves the resolvability of the image. In IIRS 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 Fig. 5.

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 IIRS.



Fig. 5. Results of our proposed method with a fixed neutral 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.

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 as the area of low-resolution images. The performance is good after sufficient iterations, as shown in Figs. 6 and 7. From Fig. 7, we discovered that as the number of iterations grows, the performance, evaluated by PSNR, converges.

5.2. *Reconstructing high-resolution objects from image sequences of moving object*

In Sec. 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

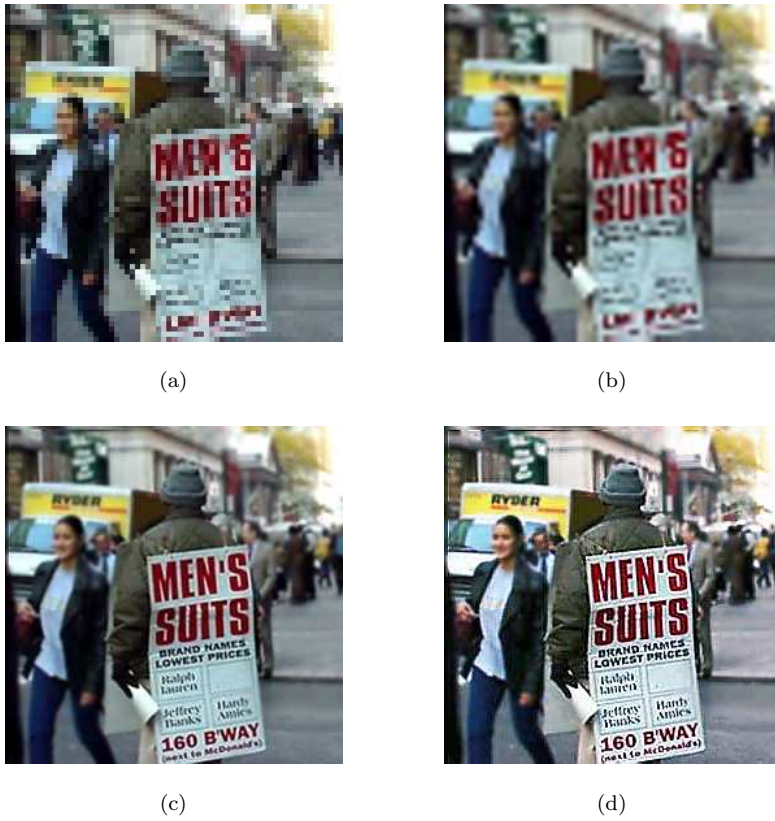


Fig. 6. Results of our proposed method with a fixed scene and a simulated moving camera. (a) One of low-resolution images. (b) Initial guess. (c) Reconstructed image after 100 iterations. (d) Enhanced final output image.

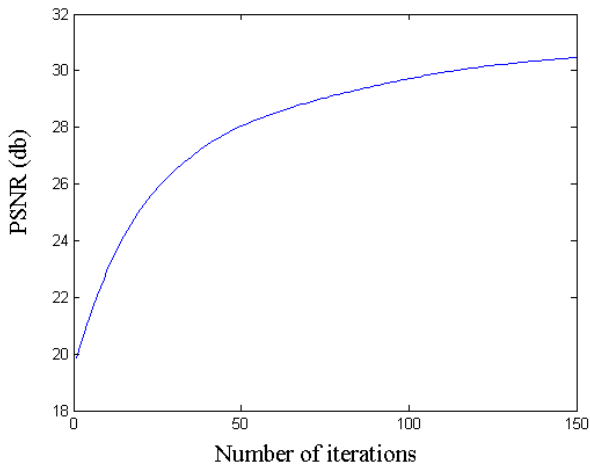


Fig. 7. PSNR of iteratively output images.



Fig. 8. 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.

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 Fig. 8 shows.

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 third 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

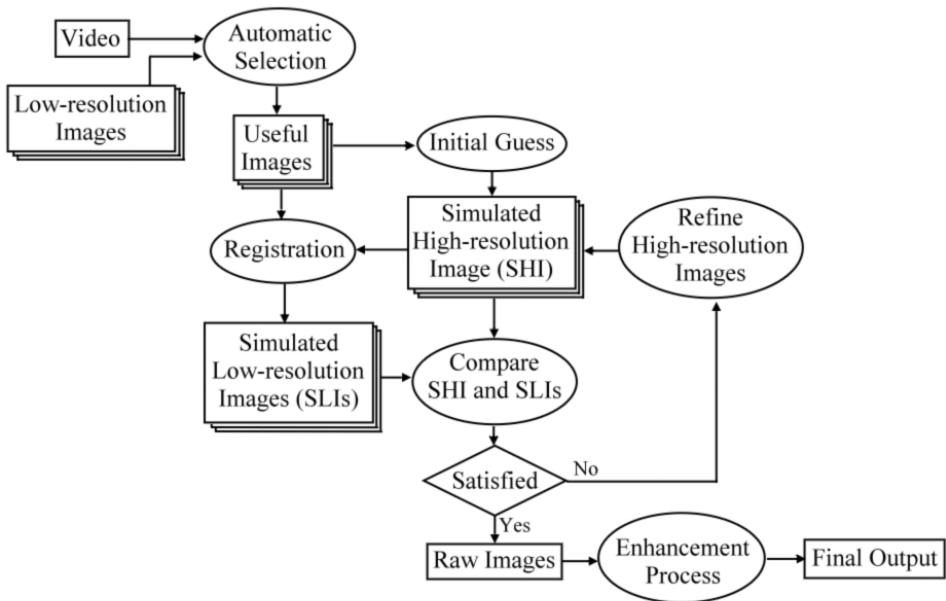


Fig. 9. The workflow diagram of IIRS.

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. Figure 9 depicts the workflow of our proposed system.

References

1. P. Cheeseman, B. Kanefsky, R. Krufft, J. Stutz and R. Hanson, Super-resolved surface reconstruction from multiple images, NASA Technical Report FIA-94-12, 1994.
2. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Addison-Wesley, Reading, MA, 1992).
3. M. Irani and S. Peleg, Improving resolution by image registration, *CVGIP: Graphical Models and Image Proc.*, 1991, Vol. 53, pp. 231–239.
4. W. K. Pratt, *Digital Image Processing*, 2nd edition (Wiley, NY, 2001).
5. 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 Int. Conf. Acoustics, Speech, and Signal Processing*, 1992, San Francisco, CA, Vol. III, pp. 169–172.
6. R. Y. Tsai and T. S. Huang, Multiframe image restoration and registration, in *Advances in Computer Vision and Image Processing*, ed. T. S. Huang, Vol. 1 (Jai Press, Greenwich, CT, 1984), pp. 317–339.



University in 2002.

Chien-Yu Chen is currently working on a M.S. degree in the Department of Computer Science at Stanford University. He received his B.S. degree in computer science and information engineering from National Taiwan



University of New York at Stony Brook.

Yu-Chuan Kuo received his bachelor's degree from the Department of Computer Science and Information Engineering, National Taiwan University in 2002. He is currently a graduate student in State University

of New York at Stony Brook. His research interests include computer vision, computer graphics and digital image processing.



Chiou-Shann Fuh received the B.S. degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1983, the M.S. degree in computer science from the Pennsylvania State University, University Park, PA, in 1987, and the Ph.D. degree in computer science from Harvard University, Cambridge, MA, in 1992. He was with AT&T Bell Laboratories and engaged in performance monitoring of switching networks from 1992 to 1993. He was an Associate Professor in the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan from 1993 to 2000 and then promoted to a Full Professor.

His current research interests include digital image processing, computer vision, pattern recognition, mathematical morphology, and their applications to defect inspection, industrial automation, digital still camera, and digital video camcorder such as color interpolation, auto exposure, auto focus, auto white balance, color calibration and color management.