SUPER-RESOLUTION IMAGE RESTORATION FROM IMAGE SEQUENCE

Cheng-Han Lee (李承翰), Chi-Fu Lin (林奇賦),Chiou-Shann Fuh (傅楸善)

Dept. of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
E-mail: r01922087@ntu.edu.tw, daky1983@gmail.com, fuh@csie.ntu.edu.tw

ABSTRACT

The field of super-resolution has a wide area of applications. In order to display relatively low-quality content on high-resolution displays, the need for super resolution algorithms has become an urgent market priority. A method of super-resolution based on project-onto-convex-sets (POCS) is proposed in this thesis. In the super-resolution process, a set of low-quality images is given, and a single improved-resolution image is desired. We adopt frequency-domain method to estimate motion and enhance the result of high-resolution image by logarithmic image processing model.

Keywords *Super Resolution; POCS; Multiple frames*

1. INTRODUCTION

Super-Resolution (SR) image reconstruction is a branch of image fusion for bandwidth extrapolation beyond the limits of traditional imaging systems [14]. It is a function procedure a high-resolution (HR) image from several low-resolution (LR) images which are sub-pixel shifted from each other covering similar region in the world [3].

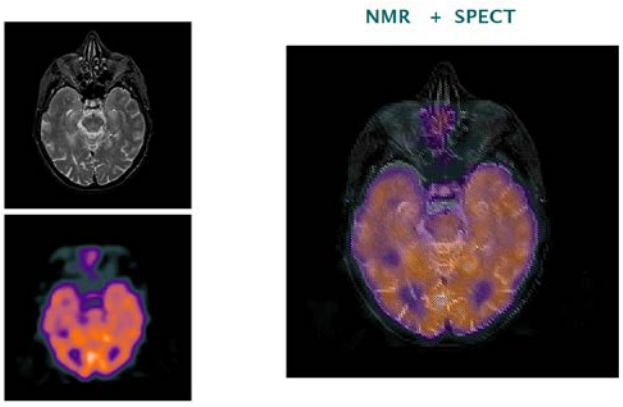
Problems of super-resolution in a number of imaging fields, such as satellite surveillance pictures and remote monitoring, where the size of the CCD (Charge-Coupled Device) array used for imaging may introduce physical limitations on the resolution of the image data. Medical diagnosis may be made more accurately if data from a number of scans can be combined into a single more detailed image. A clear, high-quality image of a region of interest in a video sequence may be useful for facial recognition algorithms, car number plate identification, or for producing a “print-quality” picture for the press.

In this paper, we propose a super resolution method that generates a high-resolution image four low-resolution images. This method uses four sub-pixel shift low-resolution images from image sequence and applies motion estimation and logarithmic image processing model to enhance high-resolution image.

The field of super-resolution has a wide area of application [3]. Although the concept of super-resolution is the same, the techniques of achieving HR imagery may not be the same for each application. For some applications such as real-time video surveillance or target detection, computational time is very important and requires a super-resolution technique with high accuracy and low computational cost. On the other hand, for certain applications such as astronomical imaging or text recognition, computational time is not a constraint and thus such applications can implement super-resolution techniques with high accuracy and a higher computational cost.

To name a few applications of super-resolution both in civilian and military domains: medical Imaging, remote sensing, target detection and recognition, radar imaging, forensic science, surveillance systems.

(a)



(b)

(c)

(d)

Fig. 1: (a) Satellite surveillance image [10]. (b) Image fusion with NMR (Nuclear Magnetic Resonance Spectroscopy) and SPECT (Single Photon Emission Computed Tomography) [6]. (c) Low-resolution image. (d)High-resolution image.

2. RELATED WORK

It is possible to have reconstruction-based super-resolution due to the reason that each low-resolution image contains sub-pixel different functions of the original scene. Sub-pixel registration differences or blur differences have contributed to this effect [14]. We can treat it as an observation model. Then, we can super-resolution as an inverse problem that need us to reverse the decimation and blur.

Each low-resolution pixel can be used as the integral of the high-resolution image if we employ a particular blur function. We can assume that the pixel locations in the high-resolution frame are not previously determined, along with the point-spread function that describes how the blur behaves [14]. Due to the reason that pixels are not continuous, this integral in the high-resolution frame is modelled as a weighted sum of high-resolution pixel values, and the Point-Spread Function (PSF) kernel will provide the weights.

Every low-resolution pixel contains some information of the world. The set of constraints will not be redundant if we have a set of low-resolution images with different sub-pixel registrations with respect to the high-resolution frame, or with different blurs. Each additional image will have contribute information to the estimate of the high-resolution image. We must account these factors for in a super-resolution model. Thus, the full picture of the process of generating a low-resolution image from high-resolution image generates a low-resolution image set [14].

2.1. Observation Model

The digital imaging system is not without problem due to the limitations of hardware, acquiring images with various kinds of degradations [12]. As always happened in sports photography or videos, the finite aperture time leads to motion blur. The finite sensor size will result in sensor blur, and aliasing effects due to the limited sensor density. These degradations are modeled fully or partially in different SR techniques.

Fig. 2 shows how a typical observation model relates HR image to LR image [12]. We input continuous real scenes to the imaging system, and they are well approximated as band-limited signals. We can generate the high-resolution digital image through sampling the continuous signal beyond the Nyquist rate (a) real scenes we desired. In super-resolution setting, there usually exists some kind of motion between the scene to capture and the camera. Multiple frames of the scene become the inputs to the camera are, connected by possibly local or global shifts, leading to image (b). Going through the camera, several kinds of blur will be incurred by these motion-related high resolution frames, such as optical blur and motion blur. These blurred images (c) are then down-sampled at the image sensors (e.g. CCD detectors) into pixels, by an integral of the image falling onto each sensor area. The sensor noise

and color filtering noise will affect these down-sampled images. In the end, blurred, decimated frames and noisy versions of the underlying true scene are captured by the low-resolution imaging system.

Fig. 2: The observation model of an imaging system relating a high-resolution image to the low-resolution observation frames [12].

2.2. Down-Sampling

The HR image is as stated ideally sampled at Nyquist frequency [13]. The sampling at high resolution can be said as an image covered by fine grid. One pixel at a certain gray level is represented by each cell in the grid. When down-sampling, a coarser grid is placed upon this fine grid. The cells in this new grid will each cover more than one cell in the fine grid. The result is the loss of the fidelity within the cell as it can only contain one gray level. The gray level is generated as being the level of one of the pixels $x = 1, \dots, L1$; $y = 1, \dots, L2$ denoted as phase. If the motion is regarded as global, i.e. no local changes in the picture the phase is set to the same when creating every LR pixel. We show the principle is shown in Figure 3, in which an HR image is down-sampled by a factor of 2. It is possible to create an ordered set of LR images that have all the information at phases suited for easy reconstruction in a simulated scenario.

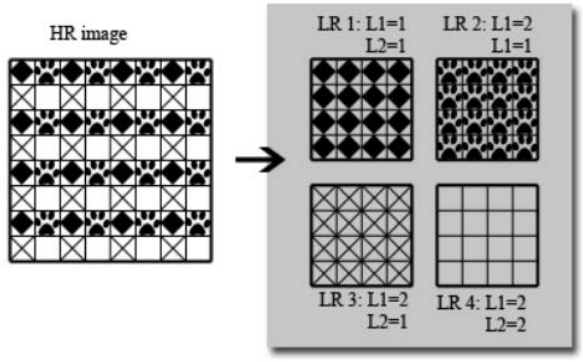


Fig. 3: An example of down-sampling [13].

2.2. Sub-Pixel Shift

In some circumstances, we get some images from the same scene. For example, one camera captures a scene many times or some cameras capture a scene one time. Sub-pixel shift are the multi-frames that have some different perspectives of the same scene. These low-resolution images with different sample-scene are important to restore high-frequency detail.

Sub-pixel shift estimation is identifying the shift in the x and y directions between two patches wherein, we can say that they have already been aligned in such a way that there are no integer shifts between them [7]. Traditionally, this problem has been solved by the resolution pyramids in which the sub-pixel shift problem is posed as an integer shift problem in higher resolution. However, the interpolation algorithm used for increasing the resolution limits such a technique.

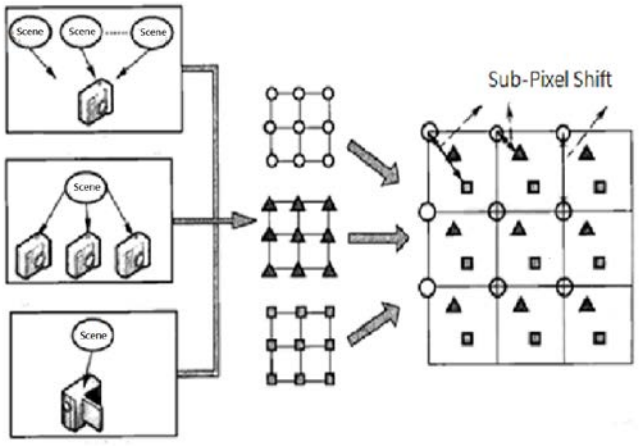


Fig. 4: Sub-pixel shift.

3. STATE-OF-THE-ART SR METHODS

3.1. Frequency Domain Methods

Frequency-domain SR methods typically rely on familiar Fourier transform properties, specifically the shifting and sampling theorems, for the removal of alias [15]. Tsai and Huang [17] were the first to explain the concept of multi-frame SR re-construction. They proposed a frequency domain observation model, by disregarding the image fusion: algorithms and applications effects of blurring and observation noise during image acquisition. The observed images were modeled as under-sampled images of a static, unknown scene with continuous spatial variables, which was subject to global translational motion. They assume ideal impulse sampling, but there is a sampling rate below the Nyquist rate. The shift and aliasing properties were used to formulate a system of equations that is related the aliased Discrete Fourier Transform (DFT) coefficients of the observed images to samples of the Continuous Fourier Transform (CFT) of the unknown scene. The observations were derived from the unknown scene.

Starting from the original continuous scene $z(x,y)$, global translational motion yields K shifted images:

$z^{(k)}(x,y) = f(x + \Delta_x^{(k)}, y + \Delta_y^{(k)})$ with $k \in \{1, 2, ..., K\}$. The CFT of the scene is given by $Z(u,v)$ and that of the translations by $Z^{(k)}(u,v)$. The shifted images are impulse sampled to yield K observed LR images: $y^{(k)}[m,n] = z(mT_x + \Delta_x^{(k)}, nT_y + \Delta_y^{(k)})$, where T_x and T_y denote the sampling periods in the x and y dimensions, respectively. The K corresponding 2-D DFTs are denoted by $Y^{(k)}[u, v]$. The CFT of the scene and the DFT of the shifted and sampled images are related via aliasing by

$$y^{(k)}[u,v] = \frac{1}{T_x T_y} \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} z^{(k)}\left(\frac{u}{M_1 T_x} + \frac{p_1}{T_x}, \frac{v}{M_2 T_y} + \frac{p_2}{T_y}\right) \quad (3.1.1)$$

Due to the shifting property of the Fourier transform, spatial shifting appears as phase shifting:

$$Z^{(k)}(u,v) = e^{j2\pi[\Delta_x^{(k)}u + \Delta_y^{(k)}v]} Z(u,v) \quad (3.1.2)$$

If $z(x,y)$ is band-limited, there exists $\Omega_x, \Omega_y \in N$ such that $Z(u,v) \to 0$ for $|u| \geqq \Omega_x/T_x$ and $|v| \geqq \Omega_y/T_y$ and the infinite summations in Eq. (3.1.1) are reduced to finite sums. Using the shifting property of Eq. (3.1.2) the relationship in Eq. (3.1.1) obtains the following matrix form:

$$Y = \phi Z \quad (3.1.3)$$

where Y is a $K \times 1$ vector; the k-th element of which contains the DFT coefficients $Y^{(k)}[u,v]$ of the observed frame $y^{(k)}[m,n]$; and ϕ is a matrix that relates the DFTs of the observed frames to the samples of the unknown CFT of $z(x,y)$ contained in vector Z.

3.2. Spatial Domain Methods

Many super-resolution spatial-domain methods are proposed in the past couple of years [12]. Since the HR image and LR images are related in a sparse linears system, spatial-domain methods overcome the difficulties of the frequency-domain methods. Prior knowledge might be used to constrain or regularize the super-resolution problem. We can apply many flexible estimators to the SR reconstruction.

The spatial-domain methods contain interpolation, deterministic regularized techniques, stochastic methods, iterative back projection, and projection onto convex sets among others.

3.2.1. Bicubic Interpolation

Given a sampled signal, its continuous counterpart can be approximated using some suitable interpolation kernel [2]. By applying successively 1D kernel interpolation on horizontal and vertical directions, we can 2D interpolation. For uniformly spaced data, we can the continuous-domain signal $Y(u,v)$ as,:

$$Y(u,v) = \sum_k \sum_l y(k,l) h\left(\frac{u - u_k}{\Delta u}\right) h\left(\frac{v - v_k}{\Delta v}\right) \quad (3.2.1)$$

where $(\Delta u, \Delta v)$ are sampling intervals, h() is the interpolation kernel and $\{y(k,l)\}$ represent the pixel array in the low-resolution grid. The SR signal is obtained by resampling (Eq. 3.2.1) on a finer grid. The cubic convolution kernel is given as,

$$h(s) = \begin{cases} 1.5|s|^3 - 2.5|s|^2 + 1 & 0 \le |s| < 1 \\ -0.5|s|^3 + 2.5|s|^2 - 4|s| + 2 & 1 \le |s| < 2 \\ 0 & 2 \le |s| \end{cases} \quad (3.2.2)$$

3.2.2. Iterative Back-Projection Techniques

Irani and Peleg considered the Super-resolution of monochrome and color low-resolution image sequences [8]. Based on computer-aided tomography [3], they derived an iterative back-projection algorithm. The algorithm starts with an initial guess (X^0) for the output high-resolution image and the imaging process (A) is simulated to generate low-resolution images (b^{sim}) based on the initial guess. These simulated low-resolution images are then compared with the observed ones (B) and the error generated between them is back-projected onto the initial guess via back-projection operator (A^{bp}), so to minimize the error iteratively.

$$X^t = X^0 + A^{bp}(b - b^{sim}) \quad (3.2.3)$$

Though the algorithm is about translational and rotational motion, the authors propose that other motions can also use the same concept. They look into multiple motion analysis including occlusion and transparency. The algorithm solves the issue of blur and noise successfully. However, the technique is unable to generate a unique solution because of the ill-posed nature of super-resolution.

3.2.1. Projection Onto Convex Sets

The constraint sets, in the POCS formulation, are used to define the feasible solution space for the super-resolution restoration [4]. Constraints are defined as convex sets in the vector space $\mathbb{R}^{N_1 \times N_2}$ which represents the space containing all super-resolution images. Sets that represent desirable characteristics of the solution are defined on this space. These sets encapsulate constraints such as fidelity to the observed data, positivity, bounded energy, smoothness, and so on. The solution space of the super-resolution restoration problem is, by design, the intersection of the convex constraint sets. POCS is an iterative procedure which, given any point in the vector space, locates a point which simultaneously satisfies all the constraint sets.

Given k convex constraint sets in $\mathbb{R}^{N_1 \times N_2}$ such that the intersection of the sets is nonempty, POCS projects a point in the vector space onto each constraint set in turn. Repeat the process until a point is reached which lies in the intersection of the k sets [3]. This iteration converges, provided that the constraint sets are convex.

POCS has attracted much attention in recent years in a multitude of image restoration and reconstruction applications. Three reasons for this stand out:

- Simplicity:
  POCS is very intuitive and generally simple to implement. The only potential source of difficulty is in determining the projection operators.

- Flexible spatial-domain observation model: Because the POCS method is typically formulated in the spatial domain, very general motion and observation models may be used. The complexity of the motion and observation model has little impact on the POCS solution procedure.

- Powerful inclusion of a-priori information: Perhaps the most useful aspect of the POCS formulation is the ease with which a-priori information may be included. It is generally simple to define convex constraint sets which incorporate desired solution characteristics. These sets may impose restrictions such as positivity or bounded energy which is difficult to represent in terms of cost function.

4. OUR METHOD

Our super-resolution method is based on the project-onto-convex-set method. We use a frequency domain motion estimation method to estimate the shift and rotate between those low-resolution frames [18].

Moreover, we use POCS methods to reconstruct high-resolution image from low-resolution images. The POCS method often causes aliasing at the edges of the result images. Therefore, we also adopt logarithmic image processing model method to improve the result of high-resolution image. The flow of our proposed algorithm is as follows:

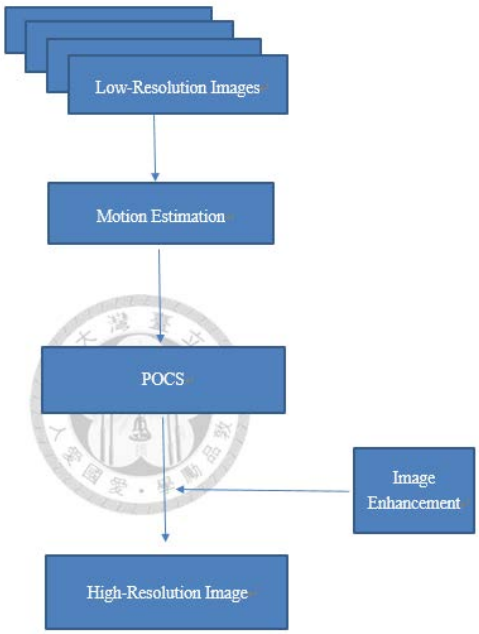


Fig. 5: The flowchart of our proposed algorithm.

4.1. Motion Estimation

We use a frequency domain method to estimate the motion between the reference image and each of the other images [18]. The motion can be described as a function of s three parameters: horizontal and vertical shifts, Δx_1 and Δx_2, and a planar rotation angle φ.

This frequency domain approach estimates the horizontal shifts, vertical shifts, and rotations separately. Assume the reference signal as $f_1(x)$ and $f_2(x)$ as shifted and rotated version.

This frequency domain approach estimates the horizontal shifts, vertical shifts, and rotations separately. Assume the reference signal as $f_1(x)$ and $f_2(x)$ as shifted and rotated version. $f_2(x) = f_1(R(x+\Delta x))$,

with $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\Delta x = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$, $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

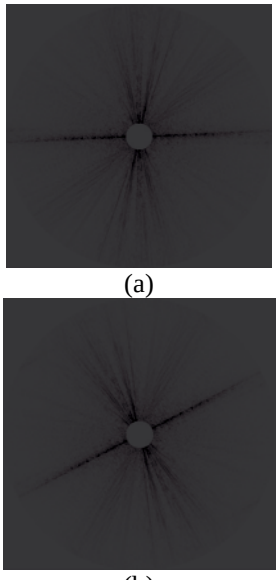
In Fourier domain, the above equation can be expressed as:

$$\begin{aligned} F_2(u) &= \iint_x f_2(x) e^{-j2\pi u^T x} dx \\ &= \iint_x f1(R(x+\Delta x)) e^{-j2\pi u^T x} dx \\ &= e^{-j2\pi u^T x} \iint_x f1(Rx') e^{-j2\pi u^T x'} dx', \end{aligned}$$

where $F_2(u)$ is the Fourier transform of $f_2(x)$, and $x' = x + \Delta x$ is the coordinate transformation. The relation between the amplitudes of the Fourier transforms can be computed as:

$$|F_2(u)| = |F_1(Ru)|,$$

where $|F_1(u)|$ is rotated version of $|F_2(u)|$ as the spatial domain (Fig. 6).



(a) (b)

Fig. 6: Rotation estimation [18]. (a) Frequency values of the reference image. (b) Frequency values of the rotated image (φ = 25 degrees).

Magnitudes $|F_1(u)|$ and $|F_2(u)|$do not depend on the shift value Δx because the spatial domain shift only affects the phase values of the Fourier transforms. Thus, we can estimate the rotation angle φ from the amplitudes of the Fourier transforms $|F_1(u)|$ and $|F_2(u)|$ first. Also, we can compute shift value Δx from the phase

difference between $F_1(u)$ and $F_2(u)$ after compensation for the rotation.

In our method, we use the center area (50 x 50 pixels) only of selected images to improve the efficiency of computation.

4.3. Project onto Convex Sets

The four main steps of our POCS method are listed as follows:

1. Up-sample the low-resolution images.
2. Align low-resolution images to reference images.
3. Combine up-sampled images.
4. Use linear filtering on combined image and iterate until it converges.

First, we up-sample by factor of 2. Based on the results of motion estimation, we can compute the shift of the up-sampled low-resolution images separately. Therefore, we are able to align those images with reference to low-resolution images.

Due to the sub-pixel shift, those images will have slight differences.

(a)

(b)

Fig. 7: Image aligned based on the motion estimation

After we align up-sampled images, we will obtain four up-sampled images with slight differences. We combine them together to fill some zero pixels with values. Then, we adopt linear filtering on all pixels to obtain high-resolution image. Meanwhile we determine if ΔI is greater than threshold, if yes it will go back to Step 3.

Otherwise, the high-resolution image is the result that we expect. The definition of ΔI is:

ΔI = norm ($Image_i$ – $Image_{i-3}$) / norm ($Image_i$), where $Image_i$ denotes the result image of the i-th iteration.

4.4. Image Enhancement

We enhance the result of high-resolution image by the logarithmic image processing model [5]. The algorithm can be implemented by:

$$\log(\overline{f}'(i,j)) = \alpha \log(\overline{a}(i,j)) + \beta[\log(\overline{f}(i,j) - \log(\overline{a}(i,j))]$$

where $\overline{f}$ is the transformation of a gray tone function f :

$$\overline{f} = 1 - \frac{f}{M},$$

where $\overline{f}(i,j)$ is the original gray tone function; $\overline{f}'(i,j)$ is the enhanced gray tone function; and $\overline{a}(i,j)$ is averaged image with window size (n x n) pixels.

The parameter α governs the contrast of the image, and β governs the sharpness of the image.

5. EXPERIMENTAL RESULTS

(a)

(b)

(c)

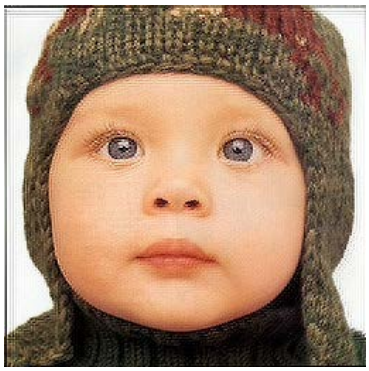
(d)

Fig. 8: Lena image. (a) Original. (b) IBP: PSNR: 19.1dB, 1 vote. (c) Bi-cubic interpolation: PSNR: 20.1dB, 8 votes. (d) Our method: PSNR: 22.1dB, 12 votes.

(a) (b) (c)

Fig. 9: Lena image details. (a) IBP. (b) Bi-cubic (c) Our method.

(a)



(b)

(c)

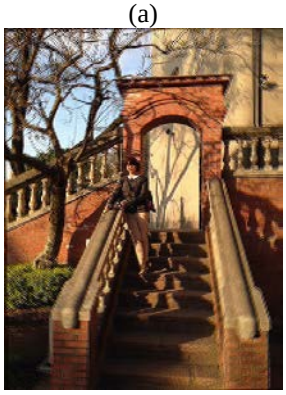
(d)

Fig. 10: Child image. (a) Original. (b) IBP: PSNR: 19.1dB, 2 votes. (c) Bi-cubic interpolation: PSNR: 20.1dB, 4 votes. (d) Our method: PSNR: 23.2dB, 15 votes.

(a) (b) (c)

Fig. 11: Child image details. (a) IBP. (b) Bi-cubic (c) Our method.

(a)



(b)

(c)

(d)

Fig. 12: image. (a) Original. (b) IBP: PSNR: 16.51dB, 7 votes. (c) Bi-cubic interpolation: PSNR: 19.18dB, 2 votes. (d) Our method: PSNR: 22.64dB, 12 votes.

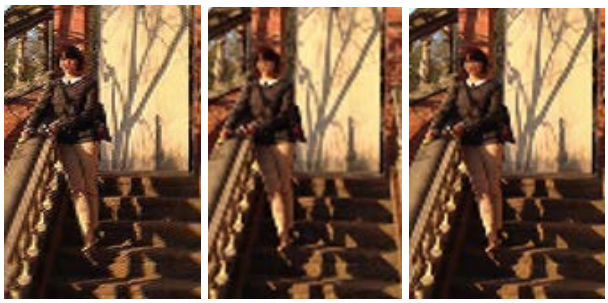


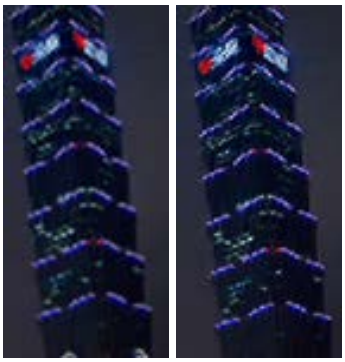
Fig. 13: Stair image details. (a) IBP. (b) Bi-cubic (c) Our method.

(a)

(b)

(c)

Fig. 14: Taipei 101 image. (a) Original. (b) Normalized convolution: PSNR: 27.75dB, 9 votes, execution time: 19s. (c) Our method: PSNR: 27.95dB, 12 votes, execution time: 1.1s.



(a) (b)

Fig. 16: Taipei 101 image details. (a) Normalized convolution. (b) Our method.

6. CONCLUSION AND FUTURE WORK

Super resolution is a challenging problem due to ill-pose. In our work, we propose a method on POCS method and combine interpolation to improve linear filtering. Moreover, we adopt logarithmic image processing model to enhance the super-resolution results. Our results show that the proposed method outperforms iterative-back-projection and interpolation methods in PSNR and subjective visual voting. Also, the execution time of our proposed method is better than structure-adaptive-normalized-convolution and have almost equal PSNR.

The result of our method in fence pattern and high-density texture sometimes may cause incorrect result which requires further improvement. Moreover image alignment is also imperfect. The boundary of result image has some black border.

Furthermore, we can try to use input low-resolution images which have different zoom. It may contain more information to cover real-scene, so that we can recover high-resolution image better.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Science and Technology of Taiwan, R.O.C., under Grants NSC 101-2221-E-002 -194 and NSC 102-2221-E-002 -177, and by Winstar Technology, Test Research, Qisda, Lumens Digital Optics, PolarLink, and Lite-on.

REFERENCES

[1] N. Alldrin, "Super-Resolution," UCSD Research Examination Paper, 2006.

[2] R. S. Babu, and K. S. Murthy, “A Survey on The Methods of Super-Resolution Image Reconstruction,” International Journal of Computer Applications, Vol. 15, No. 2, pp. 1-6, 2011.

[3] V. Bannore, Iterative-Interpolation Super-Resolution Image Reconstruction, Springer, Berlin, 2009.

[4] S. Borman, “Topics in Multi-frame Super-Resolution Restoration,” Doctoral Dissertation, University of Notre Dame, 2004.

[5] G. Deng, L. W. Cahill, and G. R. Tobin “The Study of Logarithmic Image Processing Model and Its Application to Image Enhancement,” IEEE Transactions on Image Processing, Vol. 4, No. 4, pp. 506-512, 1995.

[6] J. Flusser, F. Sroubek, and B. Zitova´, Image Fusion: Principles, Methods, and Applications, Tutorial EUSIPCO Lecture Notes, 2005.

[7] M. D. Gupta, S. Rajaram, T. S. Huang, and N. Petrovic, "Ordinal Regression Based Sub-Pixel Shift Estimation for Video Super-Resolution," EURASIP Journal on Advances in Signal Processing Vol. 2007, No. 85963, pp. 1-9, 2007.

[8] M. Irani and S. Peleg, “Improving Resolution by Image Registration,” CVGIP: Graphic Models and Image Processing, Vol. 53, No. 3, pp. 231–239, 1991.

[9] M. Irani and S. Peleg, “Motion Analysis for Image Enhancement: Resolution, Occlusion & Transparency,” Journal of Visual Communication and Image Representation, Vol. 4, pp. 324–335, 1993.

[10] MapMart, “Satellite Imagery,” http://www.mapmart.com/Products/satelliteImagery.aspx, 2014.

[11] K. Mathew and D. Shibu, “Super Resolution Image Reconstruction through Wavelet Based Technique,” International Journal of Engineering, Business and Enterprise Applications, Vol. 4, No. 2, pp. 161–165, 2013.

[12] P. Milanfar, Super-Resolution Imaging, CRC Press, Boca Raton, FL, 2010.

[13] A. Öhman, “Methods and Algorithms for Image Fusion and Super Resolution,” Master Thesis, Chalmers University of Technology, 2008.

[14] L. C. Pickup, “Machine Learning in Multi-frame Image Super-Resolution,” PhD Thesis, University of Oxford, 2009.

[15] T. Stathaki, Image Fusion: Algorithms and Applications, Academic Press, Waltham, MA, 2008.

[16] H. Takeda, P. Milanfar, M. Protter, and M. Elad, “Super-Resolution without Explicit Sub-pixel Motion Estimation,” IEEE Transactions on Image Processing, Vol. 18, No. 9, pp. 1958-1975, 2009.

[17] R. Tsai and T. Huang, “Multi-frame Image Restoration and Registration,” JAI Press, Vol. 1, pp. 317–339, 1984.

[18] P. Vandewalle, S. Susstrunk and M. Vetterli1, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-resolution," EURASIP Journal on Advances in Signal Processing, Vol. 2006, No. 71459, pp. 1-14, 2006.