JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 27, 1107-1122 (2011)

Video Super-Resolution by Motion Compensated Iterative Back-Projection Approach*

CHEN-CHIUNG HSIEH[1], YO-PING HUANG[2], YU-YI CHEN[1] AND CHIOU-SHANN FUH[3]
[1]*Department of Computer Science and Engineering*
Tatung University
Taipei, 104 Taiwan
E-mail: cchsieh@ttu.edu.tw
[2]*Department of Electrical Engineering*
National Taipei University of Technology
Taipei, 106 Taiwan
E-mail: yphuang@ntut.edu.tw
[3]*Department of Computer Science and Information Engineering*
National Taiwan University
Taipei, 106 Taiwan
E-mail: fuh@csie.ntu.edu.tw

Traditionally, uniform interpolation based approach is adopted to enhance the image resolution from a single image. Due to the one and only one image, the quality of the reconstructed image is thus constrained. Multiple frames as additional information are utilized to do super-resolution for higher-resolution image. If we have enough low-resolution images with observed sub-pixels, the high-resolution image can be reconstructed. To deal with general cases, we adopted non-uniform interpolation by iterative back-projection to estimate the high resolution image. Motion compensation is used to accurately back-project the kernel and make the process converge efficiently. Motion masks are produced for useful images/regions selection and sub-pixel blocks matching are used to do motion estimation. Objects are assumed to move slightly between two consecutive images. Thus, erroneous motion vectors could be corrected by the center of motion vector clusters. From experimental results, the PSNRs of proposed method were higher than the others, ranging from 0.5 to 1.6 dB. The difference values of the high frequency parts were also greater from 0.63% to 4.86%. It demonstrated the feasibility of the proposed method.

Keywords: super-resolution, image enlargement, motion compensation, iterative back projection, *k*-means clustering

1. INTRODUCTION

A video camera has limited spatial resolution which depends on the number of charge-coupled device (CCD) sensors. Due to the higher cost with hardware technology, software approach is usually adopted to reconstruct high resolution image. Super-resolution is the process of generating a raster image with a higher resolution than its source. The source consists of one or more images. Van Ouwerkerk [1] gave a survey on super-resolution in which single-frame super-resolution is also known as image scaling, interpolation, zooming, or enlargement.

Received July 28, 2009; revised November 11, 2009; accepted January 19, 2010.
Communicated by Tyng-Luh Liu.
* This paper was supported by the Institute for Information Industry, R.O.C., under project "High quality display adaptation technique for mobile video device," 2006.

The classical way of obtaining super-resolution image is by using kernels such as 3 × 3 or 5 × 5 windows. These are the basic techniques in image enlargement and they involve calculation of data values from the known pixels such as the nearest-neighbor interpolation, bilinear interpolation, and bi-cubic interpolation. Performing this re-sampling by using kernels is easy and effective, but not optimal in terms of quality. Specific algorithms such as smart interpolation by anisotropic diffusion (SIAD) [2], new edge-directed interpolation (NEDI) [3], and locally-adaptive zooming algorithm (LAZA) [4] have been designed for scaling up images that deliver higher quality results. However, these methods have a constraint on the number of available pixels in a single image.

For the reason of additional information, multiple images are considered to resolve such problems. The basic idea is illustrated as in Fig. 1. Multiple frames based approach [5] is to reconstruct a single high-resolution image from a sequence of low resolution images, and is also referred to as super-resolution (SR). It works only if the frames are shifted by fractions of a pixel from each other. The super-resolution algorithm is to aggregate those registered sub-pixels contained in the smaller original frames to produce a larger image. However, some restrictions itemized as followings are made with regard to different applications.

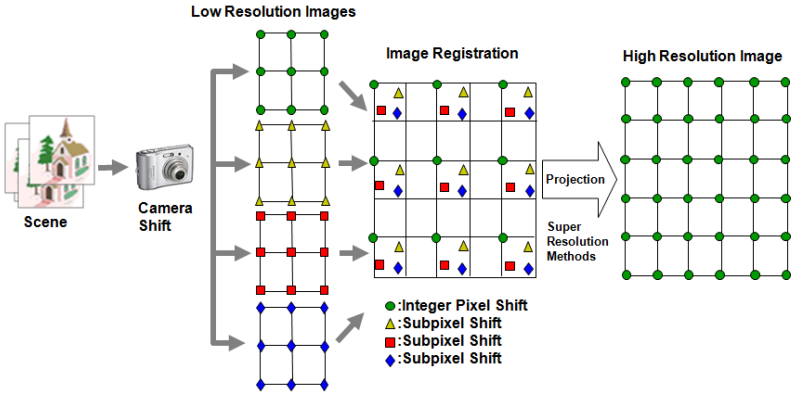


Fig. 1. Basic idea of super-resolution.

- Movement: If the objects in image are completely still or only move a little, the super-resolution image could not be well reconstructed for the lack of exposed sub-pixels. Images expansion for objects doing slow motion and translational motion are suggested in the literature [6, 7].
- Sampling rate: On the other hand, the captured images may be blurred and not useful to super-resolution when the objects move too fast. Therefore, the temporal sampling rate could not be too slow.
- Illumination: If the original images were captured with much noise, overexposure, or underexposure, we could not reconstruct the good high-resolution images for the error-prone image registration.

High-resolution images are useful and important in many applications such as medical imaging, astronomical imaging, and video surveillance. In recent years, videos from cheap

low quality cameras increase more and more, and the demand of higher resolution from these low resolution images attracted a lot of attentions. An initial version of this paper appeared in [8]. Here, we formulate the back projection function [9] as a linear translation within a short time period and the convergence time is greatly reduced. In the following section, we will briefly review the multiple frames approaches used to do super resolution from the viewpoint of learning or not. In section 3, we propose our method which adopted motion compensation in iterative back projection to deal with diversity of real world situations. In section 4, we design experiments to cover scenes with multiple moving objects and moving camera. Both visual and numerical verification of the results are given. Finally, discussions and conclusions are made in the last section.

2. RELATED WORKS

2.1 Learning Based Non-uniform Interpolation Approaches

Since there are multiple frames available, learning could be used both supervised to train interpolators and unsupervised to perform clustering. Neural network image scaling using spatial errors (NNSE) [10] has to be trained by "showing" example images. The consistency among neighboring nodes can be checked by the estimated or guessed higher resolution patches. However, learning makes the process more complicated and they could be applied only in the learned domain.

The main idea of network model based approach is to use some high-resolution images as references for the aim of learning how to sharpen an image. These references are high-resolution images, containing low-, mid-, and high-frequency information as training set. In general, the image is scaled up (with a factor of 2) by interpolating the missing pixels with bi-cubic method firstly. Then it results a larger image, but the high frequency information is missing. To estimate the missing high-frequency, the training set provides the various approximate patches to compensate the zoomed image.

There are several different models to learn the neighborhood relationships in those network model based algorithms. Shen *et al.* [11] introduced the Bayesian maximum a *posteriori* probability (MAP) method for image enlargement. The reconstruction of the super-resolution image is placed into a statistical framework by using the MAP estimation. Since super-resolution is an ill-posed problem, this method incurring *a priori* constraints to transform the problem into a well-posed problem. Another variant is to use Markov network [12] to model the relationships between high- and low-resolution patches, and between neighboring high-resolution patches. It is applied in an iterative algorithm, which usually converges in finite turns.

2.2 Motion Based Non-uniform Interpolation Approaches

Motion based systems for enhancing image resolution of video is mainly composed of three components [5]. The first and the most important step is image registration or motion estimation which takes a set of low-resolution frames as input and produces motion information from one to another frame. The second is the super-resolution algorithm which combines low-resolution frames and motion information to reconstruct a high resolution

image. Finally, image restoration is applied to the up-sampled image to remove blurring and noises. The differences among existing works are subject to what observation model is adopted, in which particular domain (spatial or frequency) the algorithm is applied, what type of reconstruction method is employed, and so on. The three stages corresponding to the three components are stated as below:

1. Motion estimation: *i.e*., registration (if the motion information is not known).
2. Non-uniform interpolation: adopted to produce an improved resolution image.
3. De-blurring process: dependent on the observation model.

With the image registration information, the HR image on non-uniformly spaced sampling points could then be obtained. The direct or iterative reconstruction procedure is followed to produce uniformly spaced sampling points [13]. Vandewalle *et al*. [14] proposed a frequency domain technique to precisely register a set of aliased images, based on their aliasing-free low-frequency part. A high resolution image is reconstructed using bi-cubic interpolation. It is the initial guessed SR image which greatly influenced the quality of reconstructed images. Wang *et al.* [15] proposed a scheme to enhance this step. The outlier registration with the high definition pixel precision is obtained by comparing warped high definition frames with the reconstructed SR image, and the adverse influence can be eliminated in calculating low definition difference to accelerate the convergence rate of the SR reconstruction and improve the quality of reconstructed images. Rochefort *et al.* [16] proposed an affine motion observation model. This model is based on a decomposition of affine transforms into successive shear transforms. They demonstrated the observation model leads to better results in the case of variable scale motions.

Irani and Peleg [9] formulated the non-uniform interpolation of SR as an iterative back-projection (IBP) process. The high-resolution image reconstruction is accomplished by iteratively minimizing the difference between the given low-resolution images and the simulated low-resolution images generated from down-sampling the current guessed high-resolution image. The advantage of IBP is that it is intuitively and easily understood. However, it has no unique solution due to the ill-posed nature of the inverse problem, and it has some difficulty in choosing the back-projection kernel.

Chen *et al*. [17] employed improved initial guess by bi-cubic interpolation, robust image registration, automatic image selection, and image enhancement to do super resolution. When the target of reconstruction is a moving object with respect to a stationary camera, high-resolution images can still be reconstructed. However, corresponding pairs of the moving object must be chosen carefully and manually. In addition, image registration is critical in the performance since it re-defines each pixel on the high-resolution image per iteration using the information of the corresponding pixel on the low-resolution images.

In this paper, we utilize sum of absolute difference (SAD) and normalized cross-correlation (NCC) as the matching criterion to do motion estimation. Objects are assumed to move slightly between two consecutive images. Thus, erroneous motion vectors could be corrected by the center of motion vector clusters. By motion vector clustering, general applications with multiple moving objects could be processed easier. Motion compensation is then used to accurately model the back-projection kernel and make the process converge efficiently. The details are described in the following sections.

3. SYSTEM ARCHITECTURE

To save the learning for model based approaches, overcome the intervening of human operations in [17], and increase the accuracy of super resolution, we proposed the motion compensated iterative back projection method. Fig. 2 shows the system architecture which is mainly composed of the proposed motion vector estimation and clustering, and the modified back projection based reconstruction of high-resolution image. In the following, we will discuss how these functional blocks work. Before processing, we have to remove noise because super-resolution will also enhance noise. Median filter or Wiener filter are usually adopted to smooth out noises. As to useful image selection, images with little motion are duplicate and discarded.

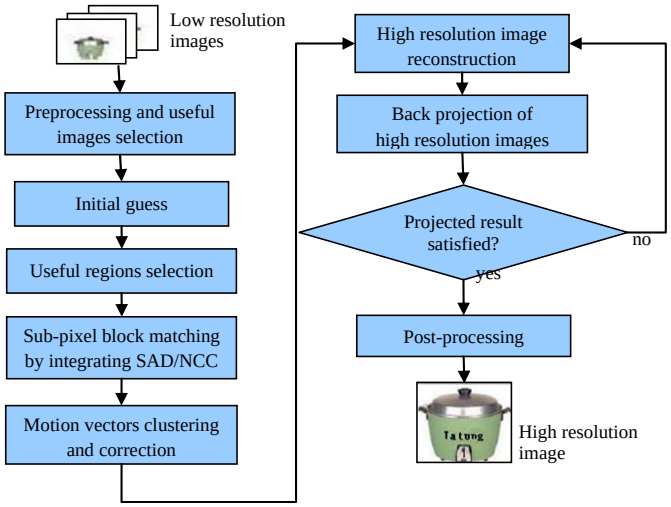


Fig. 2. Flowchart of the proposed system.

The frame used to do super resolution is called the reference frame. Firstly, bi-cubic interpolation is adopted to give the initial guess. Each successive frame of the same scene is processed with the super-resolution algorithm until a scene change is detected. The initial guess will affect the performance of the system. If we have good initial guess, the process of super-resolution will converge faster. On the contrary, the result of super-resolution will not converge or be worse.

3.1 Motion Estimation by Integrating SAD and NCC

Usually, some images are duplicate or even without motions when they were taken. Those images will not improve the quality of reference image but incur the burden of system. For this reason, we can remove them by measuring the quantity of movement. The difference between the current image and the reference image, and the difference between the current image and the previous image are computed. Then, the two difference images are intersected to produce moving regions. Non-moving regions and the boundary parts are excluded to reduce the execution time.

Motion estimation plays a very important role in super-resolution. According to the magnification factor m, we need to match current image with reference image to 2^{-m} scale of sub-pixel precision. Each matched image represents a different sub-pixel moving distance. These matched images of the different sub-pixel movements are then used to do super resolution. In this paper, images are segmented into 8×8 blocks and motion estimation is conducted by block matching.

The matching criterions are the sum of absolute difference (SAD) and normalized cross-correlation (NCC). The SAD, between a block in the current frame f_k and that block after it displaced by a motion vector $d = (u, v)$ in the reference frame f_l, is formulated as in Eq. (1).

$$SAD(u, v) = \sum_{x, y \in B} |f_k(x, y) - f_l(x+u, y+v)|, \tag{1}$$

where $f(x, y)$ denotes the pixel value at (x, y) within the matching block B. The estimated motion vector is then given as $d = (u, v) \mid \min SAD(u, v)$ for all possible (u, v) within the search window.

The other criterion normalized cross-correlation (NCC) complement with SAD could resist changing brightness better. The correlation between two signals is a standard approach for feature detection. The NCC stresses the features of image and is formulated as in Eq. (2).

$$NCC(u, v) = \frac{\sum_{x, y \in B} [f_k(x, y) - \bar{f}_k(x, y)][f_l(x+u, y+v) - \bar{f}_l(x+u, y+v)]}{\left\{ \sum_{x, y \in B} [f_k(x, y) - \bar{f}_k(x, y)]^2 \sum_{(x+u, y+v) \in B'} [f_l(x+u, y+v) - \bar{f}_l(x+u, y+v)]^2 \right\}^{0.5}}, \tag{2}$$

$\bar{f}_k(x, y)$ and $\bar{f}_l(x+u, y+v)$ are the means of the current block B and the corresponding block B' in the reference image, and $NCC(u, v)$ is the correlation coefficient between block B and B'. The estimated motion vector d is given as $d = (u, v) \mid \max NCC(u, v)$. In our system, blocks matched if $w_s SAD(u, v) + w_n NCC(u, v)$ exceeds a given threshold, where w_s and w_n are the weights.

In real world applications, some mismatched motion vectors still exist. Thus, motion vectors correction must be done to ensure the quality of super resolution. The motion vectors could be clustered by k-means algorithm [18] for the extraction of k individually moving objects. The algorithm is of non-supervised learning and k is automatically determined. By assuming rigid moving objects undergoing translational motion between successive frames, each cluster corresponds to a moving object. Within each cluster, the motion vectors should be consistent and the erroneous motion vectors within a cluster could therefore be corrected. Continuous motion could be assumed in a sequence of frames if video is captured at the speed of 24-30 frames per second.

3.2 Motion Compensated Back Projection Method

We can use the motion vectors between reference image and the low-resolution im-

age to do super-resolution. By comparing the real low-resolution images with the simulated low-resolution image that is back projected from the reconstructed super-resolution image in the nth iteration, the differences are used to improve the $(n + 1)$th iteration of super-resolution image. Repeatedly apply this process as shown in Fig. 3 until the error ε converges to a satisfactory result.

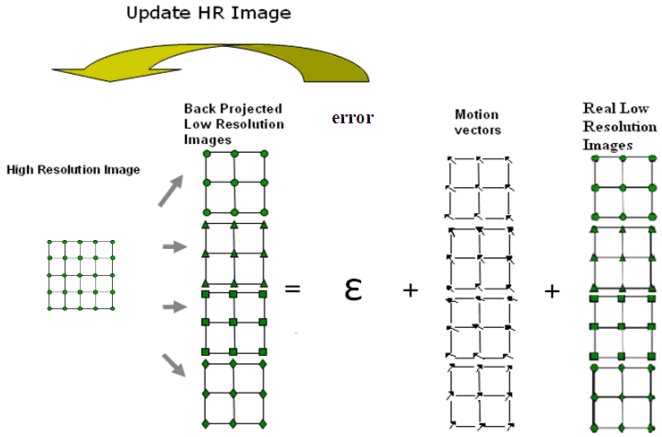


Fig. 3. Iterative back-projection (IBP) by motion compensation.

The imaging process of g_k at nth iteration is simulated by Eq. (3), where g_k is the kth observed image; f is the super-resolution image; h is the blurring operator, something like a Gaussian blurring kernel, defined by point-spread-function (PSF) δ of the sensor of the digital camera; T_k is some transformation operator that may happen to the high-resolution image; and $\downarrow s$ is the down-sampling operator. The iterative scheme of the super-resolution is updated by Eq. (4).

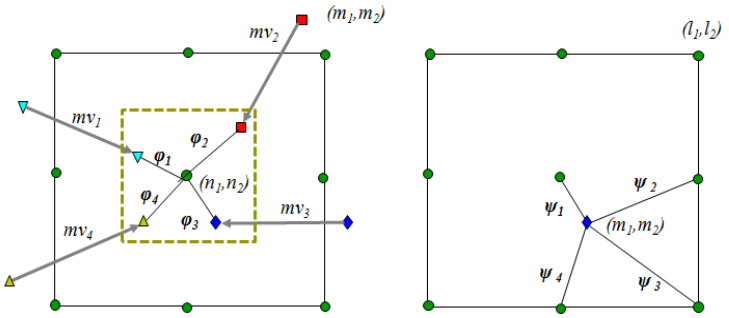
$$g_k^{(n)} = (T_k(f^{(n)}) * h) \downarrow s, \tag{3}$$

$$f^{(n+1)} = f^{(n)} + 1/K \sum_{k=1}^{K} T_k^{-1}(((g_k - g_k^{(n)}) \uparrow s) * p), \tag{4}$$

where K is the total number of low-resolution images that are used; p is the de-blurring operator; $\uparrow s$ is the up-sampling operator; and $f^{(n)}$ is the reconstructed result after nth iteration. The IBP scheme to estimate the HR image is then expressed by

$$f^{(n+1)}(n_1, n_2) = f^{(n)}(n_1, n_2) + 1/K \sum_{(m_1, m_2) \in \gamma_k^{(n_1, n_2)}} |g_k(m_1, m_2) - g_k^n(m_1, m_2)| \times p[m_1, m_2; n_1, n_2], \tag{5}$$

where g_k^n $(= f^{(n)}h)$ is the simulated LR image from the approximation of f after n iteration, $\gamma_k^{(n_1,n_2)}$ denotes the set $\{(m_1, m_2) \in g_k | (n_1, n_2) \in f$ is influenced by $(m_1, m_2)\}$ as shown in Fig. 4 (a), $p[m_1, m_2; n_1, n_2]$ is a back-projection kernel that determines the contribution of the error to $f^{(n)}(n_1, n_2)$ properly, and h is a blurring operator determined by the PSF of the



(a) (b)

Fig. 4. (a) The back projection kernel p; (b) The forward projection kernel h.

sensor projects from $f^{(n)}$ to g_k^n as shown in Fig. 4 (b). In other words, $\gamma_k^{(n_1,n_2)}$ represents the points that are within the block centered at (n_1, n_2) than the other points in $f^{(n)}$.

In this paper, $p[m_1, m_2; n_1, n_2]$ are the weights between (m_1, m_2) and (n_1, n_2), and Eq. (5) can be reformulated as Eq. (6).

$$f^{(n+1)}(n_1, n_2) = f^{(n)}(n_1, n_2) + 1/K \sum_{k=1}^{K} |g_k(m_1, m_2)\varphi_k| - 1/4 \sum_{i=1}^{4} f^{(n)}(l_1, l_2)\phi_i, \tag{6}$$

where $(m_1, m_2) + mv_k \in$ block of (n_1, n_2), $(l_1, l_2) - mv_k' \in$ block of (m_1, m_2), φ_k and ϕ_i are usually set to inverse proportional to distance, and g_k^n is the simulated LR image approximated from $f^{(n)}$ as shown in Fig. 4 (b). Note that mv_k' is the motion vector in the reconstructed HR image and is given in Eq. (7).

$$mv_k'(m_1, m_2) = mv_k(m_1, m_2) \times \frac{n_1}{m_1} \times \frac{n_2}{m_2} \tag{7}$$

The IBP super resolution algorithm is then given as follows,

Algorithm: IBP super resolution.
Input: Initial guess of HR image f. A set of LR images g_k, $k = 1 \sim K$, along with their motion masks and mv_k to f.
Output: Reconstructed HR image f.

Step 1: Back project each non-stationary point (m_1, m_2) in LR image by $f^{(n+1)'}(n_1, n_2) = 1/K \sum_{k=1}^{K} |g_k(m_1, m_2)\varphi_k|$ to reconstruct the HR image, where φ_k is the normalized inverse of the Euclidean distance between $2(m_1, m_2) + mv_k'(m_1, m_2)$ and (n_1, n_2), and the displaced (m_1, m_2) is within the kernel of (n_1, n_2).

Step 2: Project the non-stationary point (l_1, l_2) in HR image by $g_k^n = 1/4 \sum_{i=1}^{4} f^{(n)}(l_1, l_2)\phi_i$ to simulate the LR image, where ϕ_i is the normalized inverse of the Euclidean distance between the simulated point (m_1, m_2) and (l_1, l_2), and $(l_1, l_2) - mv_k'(m_1, m_2)$ is within the kernel of (m_1, m_2). To simplify the computation, $g_k^n = f^{(n)}(l_1, l_2)h$ was used to calculate $g_k^n p$ which results $f^{(n)}h * p$.

Step 3: Repeat steps 1 and 2 if $\varepsilon = \sum_{n_1} \sum_{n_2} (f^{(n+1)'}(n_1, n_2) - f^{(n)}(n_1, n_2))$ is greater than a given threshold.

In this algorithm, *h* and *p*, functional multiplicative inverse each other, play very important roles for convergence. Since the coefficients φ and ψ in *h* and *p* are both set to be normalized inverse of distance, *h* * *p* would produce the original image theoretically. It was pointed out in [11] that the choice of *p* affects the characteristics of the solution. Therefore, *p* is utilized as an additional constraint which represents the desired property of the solution. Irani and Peleg [12] formulated the error minimization by translation, affine, and moving planar surface motion models. However, their algorithm could only detect and track one dominant object at a time. In this paper, the minimization is restricted to be translational motion and *h* * *p* could be treated as a kind of PSF δ. By the remark in [12], $\|\delta - h * p\|$ would be less than one when the 2-D image motions of the tracked object consist of only translations and rotations. Therefore, our algorithm would converge at an exponential rate as proved by Theorem 3.2 in [12]. Still, we could deal with multiple moving objects concurrently.

After super-resolution process, the reconstructed high-resolution image may have noises. In order to make it clearer and more recognizable, post-processing is recommended. In this paper, adaptive histogram equalization is used to refine the high-resolution image.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

Experimental results of the proposed system were demonstrated to show its feasibility in real world applications. The experimental environments were described as follows,

- Personal computer (PC): Intel Pentium 4 3.0 GHz CPU and 1024 MB DDR RAM.
- Capture device: digital video camera with resolution ranging from 100 × 100 ~ 300 × 300.
- Processing: programs were written in MATLAB and the magnification factor is basically 2 × 2 (4 × 4 and 8 × 8 are optional). The number of images used to do super resolution within a second is 30 frames.

4.1 Super Resolution of Multiple Moving Objects

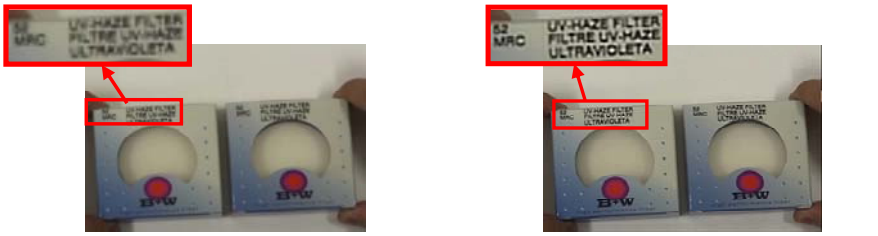
We recorded several minutes of the real world scene with two moving cars in the same direction on road. A blue sedan is in front and a bigger van behind is occluded in the lower parts. The magnification factor is 2 by 2 and the result is shown in Fig. 5. Similarly, Fig. 5 (b) looks vague due to limited information. Though the car license plate in Fig. 5 (c) looks clear, Chen *et al.*'s result still not so good at edges. Especially, there were lots of noises in the upper left parts which belong to the rear window of the big van. It is worse that human may not be able to recognize the van. However, by using motion vectors clustering, the reflections from window together with the complex background would not be over amplified to produce noises by our algorithm as shown in Fig. 5 (d).

Another example is given in Fig. 6 to demonstrate the system capability to recover the high frequency regions as indicated in the upper rectangles. Most of the characters

(a) Low resolution image. (b) Bi-cubic interpolation method.

(c) Chen *et al.*'s method. (d) Proposed super-resolution method.

Fig. 5. Results from different approaches to the SR image of real world LR images with multiple moving objects.



(a) The HR images by bi-cubic method. (b) The super resolution image generated by proposed method.

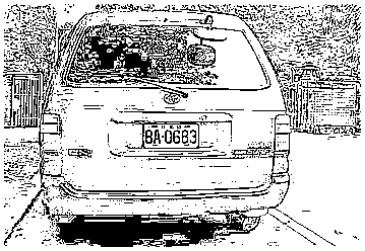
Fig. 6. Another example demonstrated the feasibility of proposed method. Two boxes move in opposite directions.

enlarged by bi-cubic could not be recognized. On the contrary, most characters are recognizable as shown in the super resolution image by our method.

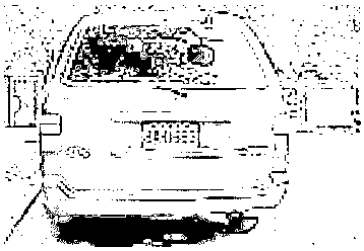
4.2 Super Resolution of Camera Motion

We panned the camera to record a scene with a red van parked by road side as shown in Fig. 7. The image sequence was down-sampled for the demonstration of proposed motion compensated super-resolution. Then we could compare the resulting super resolution image with the original high resolution image. High pass filter was used to get the finer detail parts of the original image, Chen *et al.*'s HR image, and our motion compensated HR image as shown in Figs. 7 (a), (b), and (d), respectively. By visual inspection, we could find that there were more details recovered in Fig. 7 (d).

The peak signal to noise ratio (PSNR) based on mean squared error (MSE) was used to measure image quality. The MSE was simply the mean of the squared differences for

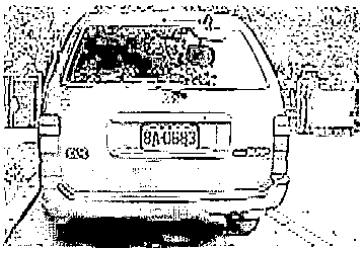


(a) The high frequency parts of the original high-resolution image.



(b) The high frequency parts generated by the Chen *et al.*' method.

(c) The super resolution image generated by the proposed method.



(d) The high frequency parts of (c).

Fig. 7. Ground truth comparisons of the high frequency regions between Chen *et al.*'s and our method.

every pixel. Let $f(i, j)$ denote the value of pixel of the original image, $F(i, j)$ the value of pixel of the compared image, r the number of rows and c the number of columns, the MSE can be obtained by using Eq. (8). The PSNR can be obtained from the MSE and the maximum signal value by using Eq. (10). The PSNR is expressed in decibels (dB) and a higher value corresponds to a lower error and thus a higher quality.

$$MSE = \frac{1}{r \times c} \sum_{i=1}^{r} \sum_{j=1}^{c} [\, f(i, j) - F(i, j)]^2, \tag{8}$$

$$RMSE = \sqrt{MSE}, \tag{9}$$

$$PSNR = 20 \log_{10} (\frac{255}{RMSE}). \tag{10}$$

Table 1 summarizes both the PSNRs and the differences among high frequency parts of the super resolution image produced by different techniques. The PSNR of proposed

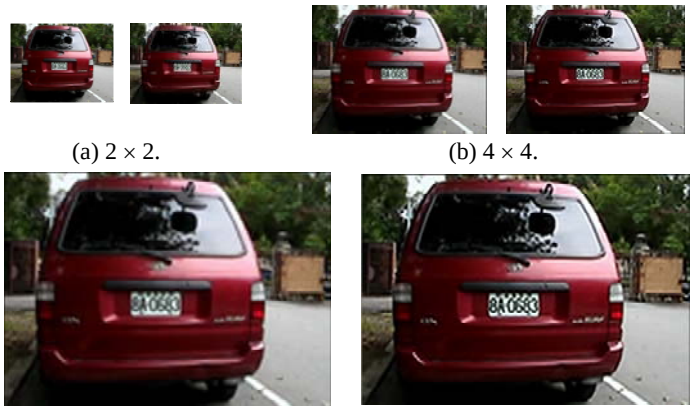
Table 1. The PSNR and difference in percent of the high frequency parts.

Technique	PSNR	Difference
The zero-order interpolation	16.8937	22.77%
The bilinear interpolation	17.4835	19.54%
The bi-cubic interpolation	17.8292	19.09%
Chen *et al.*'s method	18.0232	18.54%
The proposed method	18.5311	17.91%

method was higher than the others, ranging from 0.5 to 1.6 dB. The differences of the high frequency parts were also greater than the others from 0.63% to 4.86%. From experimental results, the quality of the resulting image was improved by the proposed method. Theoretically, the quality of reconstructed SR image would be better if more LR images are used as shown in Eq. (5).

4.3 Super Resolution of Scale 4 × 4/8 × 8

In the previous experiments, the super resolution scale was 2 × 2. Here, a previous video sequence was used to do super resolution of scale 4 × 4 (from 2 × 2) and 8 × 8 (from 4 × 4) as shown in Fig. 8. Different scaling factor images by the same super resolution algorithm were listed in the same row for visual inspection. Figs. 8 (a)-(c) show the LR van image were enlarged to 240 × 192 – 960 × 768. We could find that there were jigsaw phenomena in the left 960 × 768 (8 × 8) image by bi-cubic for the insufficient information within only an image. As to the right images reconstructed by the proposed method, they were sharpener than bi-cubic's. The reason was that the high frequency parts were kept and used to generate finer detail for the next scaling factor.



(a) 2 × 2. (b) 4 × 4.

(c) 8 × 8.

Fig. 8. Another example for different scaling factors 2 × 2, 4 × 4, and 8 × 8. The left was by bi-cubic and the right was by the proposed method.

Though super resolution could enlarge image, the higher scaling factor, the higher possibility that vague phenomenon occurred. It is due to lesser sub-pixels from the lower resolution images. These halo artifacts around the strong edges could be reduced by introducing more LR images. Theoretically, we need 4, 16, and 64 sub-pixels for the 2 × 2, 4 × 4, and 8 × 8 sub-pixel shifts for the reconstruction of super resolution images.

4.4 Execution Time with Image/Region Selection

The cost of computation largely depends on the motion estimation which in turn depends on the dimensions of the observed low resolution images. Table 2 shows the execution time for the four typical experiments without selection and with selection of useful

Table 2. Execution time of the proposed method with and w/o selection.

Example	Size	Without Selection	With Selection
Book	120 × 120	270s	254s
Green Car (Fig. 5)	141 × 96	263s	181s
Boxes (Fig. 6)	216 × 144	790s	503s
Red Car (Fig. 7)	140 × 96	250s	148s

images from the 30 LR images. Frames with little motion were discarded. Furthermore, we use motion masks or moving regions to speed up the motion estimation and the execution time was reduced when these selections are done. Still, the time spent depends on the quality of captured video and the size of moving regions. In the experiments, all the moving objects occupied large regions and moved in most of the frames. Therefore, the time saved was not so apparent. The processing speed can also be improved by using efficient programming language like C with sophisticated software/hardware motion estimation methods.

In practice, we are interested in some specific regions like license plates and the characters, thus we can choose interesting regions to do super resolution instead of the whole image. The processing time would be further reduced. The average execution time for region of size under 6,000 pixels is within one minute in our experiments.

5. CONCLUSIONS

To deal with general applications, we proposed a feasible and faster motion compensated iterative back projection based super-resolution algorithm. Firstly, the initial high-resolution image is guessed by the bi-cubic interpolation. Secondly, the current image is segmented into blocks for doing accurate sub-pixel motion estimation. By integrating both SAD and NCC features which are complemented in nature, the block matching is quite good in the dimensions of brightness and color. In addition, the non-moving regions (background) are ignored by motion masks to reduce block matching time. Then, *k*-means motion vector clustering is used to segment different moving objects and filter out the erroneous ones.

The high-resolution images are reconstructed by adding the sub-pixel motion vectors to the low resolution images. By projecting the HR image back to LR image for comparison with other real LR images, we could measure the error or quality of resulting HR image. The back projection kernel weights are set linearly proportional to the motion vector. Repeatedly apply this process until the reference frame converged to a satisfactory result or all useful low resolution images are consumed. Note that the reference frame is denoised and frames with little motion are considered as duplicated. Finally, image equalization is adopted to improve the quality of the image.

In the experiments, multiple moving objects and moving cameras as shown in Figs. 5-7 were tested. The PSNR of the proposed method was higher than the others, ranging from 0.5 to 1.6 dB. From the viewpoint of high frequency parts, our method could recover more detail parts. The amount of recovered high frequency parts were also higher than the other methods ranging from 0.63% to 4.86%. The proposed method is better than the tra-

ditional ones and could deal with image sequence of multiple moving objects and moving camera. As to the executing time, it consumes lots of computations to improve the whole scene. In real applications, interesting moving regions can be selected to do super resolution to save time. For regions with size under 6,000 pixels, we could produce results less than a minute. As to future works, objects undergoing unrestricted quick motion are too complex and unpredictable. Motion vectors clustering needs to be refined to take the factor of rotation and deformation into consideration. Accordingly, the back projection function could also be modified.

REFERENCES

1. J. D. van Ouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, Vol. 24, 2006, pp. 1039-1052.
2. S. Battiato, G. Gallo, and F. Stanco, "Smart interpolation by anisotropic diffusion," in *Proceedings of the 12th International Conference on Image Analysis and Processing*, 2003, pp. 572-577.
3. X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, Vol. 10, 2001, pp. 1521-1527.
4. S. Battiato, G. Gallo, and F. Stanco, "A locally-adaptive zooming algorithm for digital images," *Image Vision and Computing Journal*, Vol. 20, 2002, pp. 805-812.
5. S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, Vol. 20, 2003, pp. 21-36.
6. S. Chaudhuri and D. R. Taur, "High-resolution slow-motion sequencing: How to generate a slow motion sequence from a bit stream," *IEEE Signal Processing Magazine*, Vol. 22, 2005, pp. 16-24.
7. M. Ben-Ezra, A. Zomet, and S. K. Nayar, "Video super-resolution using controlled subpixel detector shifts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, pp. 977-987.
8. C. C. Hsieh, Y. P. Huang, Y. Y. Chen, C. S. Fuh, and W. J. Ho, "Video super-resolution by integrated SAD and NCC matching criterion for multiple moving objects," in *Proceedings of the 10th IASTED International Conference on Computer Graphics and Imaging*, 2008, http://www.csie.ntu.edu.tw/~fuh/personal/VideoSuperResolutionbyIntegratingSADandNCC.pdf.
9. M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *Journal of Visual Communications and Image Representation*, Vol. 4, 1993, pp. 324-335.
10. C. Staelin, D. Greig, M. Fischer, and R. Maurer, "Neural network image scaling using spatial errors," HP Laboratories, Israel, October 2003.
11. H. Shen, P. Li, L. Zhang, and Y. Zhao, "A MAP algorithm to super-resolution image reconstruction," in *Proceedings of the 3rd International Conference on Image and Graphics*, 2004, pp. 544-547.
12. W. T. Freeman and E. C. Pasztor, "Markov networks for super-resolution," in *Proceedings of the 34th Annual Conference on Information Sciences and Systems*, 2000, http://www.merl.com/papers/TR2000-08/.
13. S. P. Kim and N. K. Bose, "Reconstruction of 2-D band limited discrete signals from

non-uniform samples," in *Proceedings of IEEE Conference on Radar Signal Processing*, 1990, pp. 197-204.
14. P. Vandewalle, S. Süsstrunk, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on Applied Signal Processing*, 2006, pp. 1-14.
15. C. Wang, P. Xue, and W. Lin, "Improved super-resolution reconstruction from video, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, 2006, pp. 1411-1422.
16. G. Rochefort, F. Champagnat, G. L. Besnerais, and J. F. Giovannelli, "An improved observation model for super-resolution under affine motion," *IEEE Transactions on Image Processing*, Vol. 15, 2006, pp. 3325-3337.
17. C. Y. Chen, Y. C. Kuo, and C. S. Fuh, "Image reconstruction with improved super-resolution algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 18, 2004, pp. 1-15.
18. T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, "An efficient *k*-means clustering algorithm: analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 881-892

Chen-Chiung Hsieh (謝禎冏) received his B.S., M.S., and Ph.D. degrees in the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan, in 1986, 1988, and 1992, respectively. During Dec. 1992 to Jan. 2004, he was with the Institute for Information Industry (III) as a vice director. From Dec. 2004 to Jan. 2006, he joined Acer Inc. as a senior director. He is presently an Associate Professor in the Department of Computer Science and Engineering at Tatung University, Taipei, Taiwan. His research area is mainly focused in image and multimedia processing.

Yo-Ping Huang (黃有評) received his Ph.D. in Electrical Engineering from Texas Tech University, TX, U.S.A. He is currently a Professor in the Department of Electrical Engineering and Secretary General at National Taipei University of Technology, Taiwan. His research interests include data mining, artificial intelligence, multi-touch display system, RFID and QR code applications, and application systems design for handheld devices. Prof. Huang is a senior member of the IEEE and a fellow of the IET.

Yu-Yi Chen (陳佑易) received his B.S. and M. S. degrees in the Dept. of Computer Science and Engineering, Tatung University, in 2005 and 2007, respectively. His research interests include image processing and video surveillance.

Chiou-Shann Fuh (傅楸善) 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 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, digital video camcorder, and camera module such as color interpolation, auto exposure, auto focus, auto white balance, color calibration, and color management.