Performance Verification of Super-Resolution Image Reconstruction

Masaki Sugie
Department of Information Science,
Kogakuin University
Tokyo, Japan
Email: em13010@ns.kogakuin.ac.jp

Seiichi Gohshi
Department of Information Science,
Kogakuin University
Tokyo, Japan
Email: gohshi@cc.kogakuin.ac.jp

Abstract—In this paper, the specific limitation of super-resolution image reconstruction (SRR) is reported. In many SRR studies, specially made content was used, rather than general television and cinema content, but the amount of aliasing differs between the two. The presence of aliasing distortion is assumed during the original degradation of SRR, but a low-resolution image that we want to transform into a higher resolution may not have aliasing. It is not known whether it is possible to create a high-resolution image with SRR in the absence of aliasing, and the real power of SRR can be discussed only if general content, i.e., content that does not have aliasing, is used. The validity of SRR can be determined by examining whether new frequency components above the Nyquist frequency of the low-resolution image are created by showing a two-dimensional Fourier power spectrum of the output. This has not been performed in previous studies. In this study, we perform frequency analysis for SRR assuming that general content does not have aliasing. From the results of our experiments, in conjunction with the frequency component, we prove that SRR does not work with general content.

I. INTRODUCTION

Studies of “super-resolution (SR)” have been performed [1][2][3] and these include “super-resolution image reconstruction (SRR)” as the current mainstream method of generating a high-resolution image from a plurality of low-resolution images, which is then reproduced on a television. SRR assumes that a high resolution original exists. By applying a pre-defined camera model degradation process to the original image, we can create low-resolution images that are assumed from the input image. SRR is a technique of calculating a high-resolution image by iterations through these low-resolution images [2][3] by assuming that the original image is a special case that suffers component aliasing distortion during degradation of the SRR [1][2][3][4]. There is a phase shift of aliasing between the low-resolution image groups caused by a shift of sampling position. By referring to these aliasing components, the high-frequency components of the original image can be restored. Iteratively correcting the differences by referring sub-pixel aliasing components between multiple images makes it possible to produce a high-resolution image.

The low-resolution image that we intend to transform into a higher resolution is a still image from every single frame in a video broadcast on a television or cinema screen. However, because the anti-aliasing process is performed before transmission by the charge-coupled device (CCD) of a camera, there is no aliasing in these images [5]. The aliasing used for SRR is caused by the re-sampling of the high-resolution image. If such a high-resolution image does not exist, there is no aliasing. If input video has a motion of the object very slow, aliasing may be obtained by an interlace process. Progressive scanning is used for video display on television, while a video that has been transmitted in the interlace mode must be converted to progressive format. There is a possibility that aliasing may still occur with interfield interpolation, even if the previous field cannot be referred very well; however, this is a very rare case. The interfield interpolation process to prevent aliasing is used in all televisions, indicating that there is no merit in the implementation of complex SRR processing.

Still images and videos have different properties. Figs. 1 and 2 are examples of the same object taken by a still camera and a video camera, respectively. Fig. 1 taken by a still camera is not too much blurred, whereas motion blur is created in each...

Fig. 1. Fast moving object (still camera)

Fig. 2. Fast moving object (video camera)
frame of the video (Fig. 2). Because the shutter open time is approximately 1/50—1/60s, a greater motion blur occurs when it is 1/24 s in a frame of a movie. Similarly, motion blur is affected by the shutter speed, and frames per second becomes a problem in video. Video can be considered as a continuous still image. Degradation of resolution due to accumulation of blur at the time of imaging is present in the content of movie and high-definition television broadcasts in general. Therefore, the idea does not hold. To refer the aliasing components due to the motion effect of the object with sub-pixel levels between multiple images, accurate motion estimation at this level is required. Motion effect denotes the change in the position of the subject due to movement. However, aliasing does not naturally exist because of motion blur. In addition, the abovementioned estimation of motion becomes very difficult. Motion blur is coupled with the change in object size within the movie. There is no algorithm to select the appropriate pixels from the object that has changed its size, and hence, practical application is difficult. Motion blur is dominant in videos.

Therefore, a model that assumes aliasing is unlikely to effectively function if it is assumed to be implemented in the product. For SRR to be valid, it must be band limited with a low-pass filter (LPF) to prevent aliasing in the image, and it must be verified by frequency analysis to determine whether the effect can be predicted. If it is valid, this can be confirmed by checking whether new frequency components above the Nyquist frequency of the low-resolution image are created. A two-dimensional Fourier power spectrum of the output result will suffice. So far, this has not been achieved; thus, it remains unknown whether high-definition resolution enhancement is possible in the absence of aliasing images by SRR. SRR requires enormous processing time because it is necessary to iterate several times. However, to achieve real-time processing in television, there is a precedent for one-time iterations [6]. There has been no verification of the relationship between the output and the number of iterations by SRR.

In this paper, we perform frequency analysis for SRR by assuming that typical content, such as images and individual frames of video, does not contain aliasing. With regard to the discussion on the frequency component, we test whether it is possible to create high-quality images, even in the absence of aliasing. Moreover, we consider the relationship between the output and the number of iterations in SRR.

II. SRR

In this paper, we use bilateral total variation (BTV), which is a regularization technique devised by Farsiu [2], to verify SRR. X is the original high-resolution image. Multiple low-resolution images \( \hat{Y} \) are created by the down-sampling degradation \( D \) and motion effect \( F \) of \( X \) during image capturing. By defining the matrix process of degradation, it is possible to restore \( X \) to \( \hat{Y} \) by multiplying the inverse of the matrix. Fig. 3 shows the degradation process in SRR. Simple decimation is assumed to be the deterioration process \( D \), with aliasing included in the low-resolution image that is created. However, it is necessary to assume the bandwidth limitation by LPF in the process itself. Defocus of the point-spread function is also considered a deterioration process in [2], but it is not considered as an LPF by the CCD in the camera.

SRR is an inverse or minimization problem. The inverse problem for images can be solved by using constraints called regularization. If noise degradation factors other than what is assumed are included, there is no primary solution. BTV is a regularization constraint "to remove the noise while holding the edges of the image [2]." In this study, we use the steepest descent method to solve the minimization problem for SRR. Correct the error between the low-resolution image group \( Y_k \) and the images deteriorate again by the deterioration process estimation \( D_k \), while \( F_k \) restores image \( X_n \), and the new estimation image \( \hat{X}_{n+1} \) is created using a regularization term as a constraint. \( X_n+1 \) is an answer to SRR if the error between \( X_n \) and \( X_{n+1} \) has converged to a value approaching zero by the iteration process. Equation (1) is a model equation if SRR is solved by the steepest descent method with BTV. In this paper, we verified this by using the equation algorithm (1).

\[
\hat{X}_{n+1} = \hat{X}_n - \beta \left\{ \sum_{k=1}^{N} F_k^T D_k^T \text{sign}(D_k F_k \hat{X}_n - Y_k) \\
+ \lambda \sum_{l=1}^{P} \sum_{m=0}^{P} \alpha \left[ I - S^{l-m}_{y} S^{l}_{x} \right] \text{sign}(\hat{X}_n - S^{l}_{x} S^{m}_{y} \hat{X}_n) \right\} \\
(l + m \geq 0)
\]

\( \beta \) is a scalar defining the step size in the direction of the gradient. \( S^{l}_{x} \) and \( S^{m}_{y} \) have a shifting effect to horizontal and vertical directions respectively. \( S^{-1}_{x} \) and \( S^{-1}_{y} \) define the transposition of matrices \( S^{l}_{x} \) and \( S^{m}_{y} \), respectively, and have a shifting effect in the directions opposite to that of \( S^{l}_{x} \) and \( S^{m}_{y} \). Variables \( \alpha, \beta, \lambda, N, \) and \( P \) are values greatly affected by processing results. When assuming the implementation of products, it would be difficult to change each irregular value.

![Fig. 3. Deterioration process in SRR](image-url)
for each image. On the basis of the pre-experiment and the set value in the literature [2], in this study, we aimed to unify the values for all images. In addition, the $F$ value for validation and deterioration of the $D$ matrix was assumed to be known. This way, it was possible to refer the aliasing due to the phase shift. SRR is an algorithm that solves an ill-posed problem by regularization constraints, which have been proposed. Equation (2) has been proposed in [2], and equation (3) has been proposed in [7].

$$\hat{X} = ||Y - H\hat{X}||^2_2 + \lambda\gamma(\hat{X})$$  \hspace{1cm} (2)$$

$$J(X) = ||Y - HY||^2_2 + \lambda||X||^2$$ \hspace{1cm} (3)

However, regularization constraints merely regulate provisional pre-information of the image. A constraint condition that is suitable for all images and videos is not known. It is unlikely that this condition can exist. Therefore, squeezing into one solution is difficult to achieve, even by regularization.

III. LPF

In this study, we adopted an LPF filtering process directly multiplied by filter coefficients in the kernel for each pixel value of the image. This approach has the advantage of a short processing time and is easy to design. Furthermore, ringing does not occur because there are no rapidly changing pixel values. The coefficient of the LPF was [1/4, 1/2, 1/4]. By performing the smoothing process using the kernel in both parallel and vertical directions, all pixels were smoothed evenly. For each image, smoothing was performed multiple times to the extent that components above the Nyquist frequency were band limited.

IV. Experiments

The down-sampling reduction process $D$ that simulates camera shooting created a reduced image in two cases: with and without an LPF. The down-sampling ratio by $D$ performed on two patterns, which were 1/16 (vertically and horizontally 1/8) and 1/64 (vertically and horizontally 1/4). We examined to observe if they could be restored to high definition. In this experiment, validation of SRR used the algorithm of formula (1) that used the above BTV. Each variable in the expression was unified with the following values:”$\alpha = 0.6, \beta = 0.5, \lambda = 0.02, N = 16$, and $P = 3$.”

In SRR, the peak signal-to-noise ratio (PSNR) is related to the number of iterations, and it tends to increase with the number of iterations. The purpose of previous papers was to improve the PSNR by changing the constraint conditions of SRR, and therefore, they terminated the iteration process if the number of iterations exceeded 100. If the purpose is to improve the PSNR, there should not be a limit set on the number of iterations. In this paper, the iteration process was continued, even if the number of iterations exceeded 100. The termination condition that is affected only by change in the PSNR was set, using the following equation (4).

$$||PSNR(\hat{X}_{n+1}) - PSNR(\hat{X}_n)|| < 1.0E - 5$$ \hspace{1cm} (4)

We used 11 images (512×512 resolution) from the standard image database of “The USC-SIPI Image Database” [8]. We targeted brightness in all images. Performances metrics were PSNR and power spectrum (PS), and we calculated the number.
V. RESULTS AND DISCUSSION

The experiments were performed using 11 different images, but here we concentrate on the results using the image shown in Fig. 4 as "Lenna." Fig. 5 is a PS of "Lenna."

A. Resolution and aliasing

Figs. 6-9 are the PS and SRR results in an image that has been subjected to 1/16 reduction process for "Lenna." The border in PS is the position of the Nyquist frequency of the low-resolution image. The results of SRR without the LPF are (1-a)-(1-d), and the results of SRR with an LPF are (2-a)-(2-d).

In the reduced image without an LPF (Fig. 6 (1-a)), the high-frequency components of the original image are folded as aliasing. As a result, they are present in a greater portion than the Nyquist frequency (the position of the frame in the figure) in Fig. 7 (1-b). In addition, owing to the folding effect, jaggies are introduced in the edge component of the image (which especially occur in the vicinity of the hat brim in Fig. 6 (1-a)). The PSNR was calculated at a value higher than the reduced image (Fig. 6 (2-a)) using the LPF. In the reduced image using an LPF (Fig. 6 (2-a)), the edge is smooth compared to the pictures without LPF, because the LPF provided appropriate band limitation. Observing at the PS in Fig. 7 (2-b), the high-frequency component is not present outside the Nyquist frequency. The PSNR is less in the image without the LPF; however, because the components of the original image were not included, it is the image closest to the original input image.

Fig. 8 and 9 are the images of Figs. 6 and 7, respectively, restored using SRR. The image without LPF came closest to the original image. The PSNR was 37.36 dB. Observing at the PS in Fig. 9 (1-d), the high-frequency components that exceed the Nyquist frequency of the reduced image is restored. However, it is apparent that this result is a consequence of using a high-frequency component of the original image. Furthermore, part of the high-frequency component in Fig. 9 (1-d) was confirmed in an area (hereafter called NULL Area) where there was no frequency component. This NULL area caused by the folded position of aliasing of the low-resolution image was affected, indicative of the frequency components that cannot be restored there. Using an LPF produced a low-resolution image that failed to restore high-frequency components compared to the pictures without LPF.

The PSNR was 27.40 dB, approximately 10 dB lower than in the case of no LPF. Observing at the PS in Fig. 9 (2-d), components that exceed the Nyquist frequency were not restored. We presume this is because it could not refer to the high-frequency components of the original image owing to the band-limitation effect by the LPF.

B. Number of iterations and convergence result

Fig. 10 is a graph of the SRR result at the time of the 1/16 reduction exhibiting the progress of PSNR by iteration. Without the LPF, the number of iterations rapidly increased to approximately 300, but the rate of climb then markedly decreased, reaching saturation when PSNR was beyond approximately 37dB after approximately 600 iterations, eventually settling at 1219. When using the LPF in the same process, the number of iterations converged at 100, and consequently, the reconstruction effect was not observed at all. When using an LPF, there are few reference ingredients for the reconstruction and fewer estimates of the high-frequency component than in the case without LPF. It is thought that, consequently, processing settled comparatively early. Where SRR is intended to be implemented for television, it is necessary to limit the number of iterations for real-time processing [6]. However, several iterations are required to increase the resolution. If high definition is a priority, it is unlikely that SRR can be realized for real-time processing in television broadcasts.

C. Difference between the reduction ratios

Fig. 11 and 12 are the result of the images restored by the reduced images that had been subjected to 1/64 reduction for "Lenna" by SRR. Similar to the case of 1/16 reduction, the components were restored above the Nyquist frequency without an LPF (Fig. 11 (1-a), Fig. 12 (1-b)). The use of an LPF shows a low-definition image (Fig. 11 (2-a), Fig. 12 (2-b)) with unrestored components above the Nyquist frequency. Further, the NULL area is further increased, compared to 1/16 collapsed without the LPF (Fig. 12 (1-b)). It was evident that the impact of the position of Nyquist frequency being increased further than in the case of 1/16 reduction. The number of iterations for convergence with 1/64 reduction at the time was 4110 without the LPF, and 4108 times with it. The total...
From the above results, trends with and without an LPF were also observed if the reduction ratio was changed. High definition was not possible with the LPF. Furthermore, if the reduction ratio was different, the required number of iterations and the PSNR results varied.

D. Difference between the experimental image

Changes due to the presence or absence of an LPF followed the same trend as in the other experimental images. However, the iterations and the PSNR were different. For example, we show the SRR result of 1/16 reductions with an LPF in "Mandrill" (Fig. 13, Fig. 14) and "Fighter" (Fig. 17, Fig. 18). Figs. 15 and 16 are "Mandrill" results, Figs. 19 and 20 are "Fighter" results. As in the case of "Lenna," SRR could not restore the high-frequency components beyond the Nyquist frequency, and the output was a low-resolution image. If an image frequency component was different, the restoration effect was reduced owing to its improper aliasing references by the LPF. Where a relatively large number of high-frequency components were included, such as in "Mandrill" (Fig. 13), the PSNR was low. Because the high-frequency component of the reduced image is limited by the bandwidth limitation in the LPF, faithful reconstruction of the high-frequency components due to SRR is difficult. The number of iterations for convergence with 1/16 reduction in "Mandrill" was 2013 without the LPF, and 113 with it. The number of iterations for convergence with 1/16 reduction in "Fighter" was 1437 without the LPF, and 120 with it.

From the above results, the same lack of higher definition is shown using the LPF, regardless of the experimental images. Similar to the results of each reduction factor, the number of iterations and the resulting PSNR changed. This indicates that
SRR has no versatility with respect to the differences in the type of image and resolution.

E. Discussion on the application to general content

In this subsection, we again consider the differences between still images and videos, as described in Section I, and discuss the performance of SRR for general content. As mentioned in Section I, we want to make higher-resolution images from input images of each frame of a video broadcast on television and movies. In this paper, low-resolution images in the SRR process were assumed as the input images. The assumption in this validation was quite advantageous for SRR.

The degradation element was small compared with the actual video frames. Motion blur, as shown in Fig. 2, and expansion deterioration were not assumed. The motion estimation that significantly affects recovery results in SRR was obtained by moving the still images in the existing experiment. Motion effect $F$ was merely caused by the range of a few pixels and was assumed to be a simple parallel movement. In addition, the deviation matrix was known. In motion estimation, because it is necessary that object is referenced to the position of the same pixel in another frame as a matrix, a high-definition object must be reflected in the frame. Accurate motion estimation at sub-pixel level is difficult and is expensive to implement.

Video content has changes in the size of the object, motion effect, and motion blur, thereby creating poor conditions and making accurate estimation of the deviation matrix impossible. In addition, if applied in real-time video, iterations are limited. The results of sub-sections B and D required at least 100 iterations for the generation of a single image, and this number varied from image to image. Implementation of each frame of the video in real time is impossible. Further, it is questionable whether processing results become unified in terms of resolution quality.

When implementing video content broadcast on television and in cinemas, all the above factors must be estimated and considered. SRR for video is degraded further than the experiment images, with worse results. Above all, even if everything works ideally, the components above the Nyquist frequency of the input video cannot be created, which means that the SRR cannot be a resolution enhancement for television or movies. Higher-resolution images cannot be created with the SRR unless aliasing is included in low-resolution images.

The experiment image was assumed to be a general scene from a movie or video. Accordingly, because these images are not high definition, SRR indicates it to be impractical for general content. It was shown in section V that higher-resolution images cannot be created with SRR unless aliasing is included in low-resolution images.

In recent years, "Super-Resolution by nonlinear processing" [9] that can create new high-frequency components above the Nyquist frequency of the image by referring to the frequency component of an image have been studied. This technique can perform high-speed processing and is relatively simple. Our future study is to compare the difference in performance between SR using nonlinear processing and SRR using frequency analysis.

VI. CONCLUSION

In this study, it was assumed that aliasing is not present in commonly available content. Furthermore, by performing frequency analysis on SRR of a low-resolution image with no aliasing, we examined whether SRR can create a higher-resolution image. From the results of our experiment, it became clear that generating components above the Nyquist frequency of low resolution images by the SRR is impossible. Even if we try to create a higher-resolution image by using SRR, it cannot produce high-definition images. In addition, the required number of iterations and the PSNR result in the SRR calculate a different value for each image. SRR can be applied to the frame of a moving image, but there is no guarantee that from the above results, the effect obtained in the same processing count is unified in all frames.

REFERENCES