Subjective Assessment about Super Resolution Image Resolution

Seiichi Gohshi, Hiroyuki Sekiguchi, Yoshiyasu Shimizu, Takeshi Ikenaga

Abstract—Super resolution (SR) technologies are now being applied to video to improve resolution. Some TV sets are now equipped with SR functions. However, it is not known if super resolution image reconstruction (SRR) for TV really works or not. Super resolution with non-linear signal processing (SRNL) has recently been proposed. SRR and SRNL are the only methods for processing video signals in real time. The results from subjective assessments of SSR and SRNL are described in this paper. SRR video was produced in simulations with quarter precision motion vectors and 100 iterations. These are ideal conditions for SRR. We found that the image quality of SRNL is better than that of SRR even though SRR was processed under ideal conditions.

Keywords—Super Resolution Image Reconstruction, Super Resolution with Non-Linear Signal Processing, Subjective Assessment, Image Quality

I. INTRODUCTION

CRT displays have been replaced with LCD displays and the resolution of LCD has been constantly improving. HDTV resolution (1920 x 1080) has been very common in recent years through the use of LCD technologies. Recently, 4K (4000 x 2000 pixels) LCDs have been available on the market. Although we have inherited various kinds of content, most of it comes in low resolution. Format conversion and enlargement are necessary to fit low resolution videos and images onto high resolution displays. Enlarging inevitably causes blurring. Improving resolution is a crucial requirement and many methods are being proposed. There are very many possibilities of using super resolution (SR) technology in broadcasting, the medical field, security, and other industries. SR has been researched for many years to resolve the issues with it [1][2][3][4]. The application of SR is expected to become widespread since high resolution displays are becoming more common in proportion to their reasonable cost. Super resolution image reconstruction (SRR) is a very common SR technology and it has been applied to TV sets and BD players. However, the limits of SRR are increasing by a factor of two [5]. The factor of two is a very low limit and larger factors are necessary for many applications. The HDTV broadcasting service has adopted the interlaced video system. Although CRT accepts interlaced video, LCD cannot. The broadcasting content has to be converted to progressive video format since LCD only accepts this. Although commercial HDTV LCDs are called full HDTV, they cannot display the broadcasting interlaced HDTV format directly. They have to convert the interlace video to the progressive video.

Seiichi Gohshi is with Kogakuin University, 1-24-2, Nishi-Shinjuku, Shinjuku-ku, Tokyo, Japan 1638677 (e-mail: gohshi@lcc.kogakuin.ac.jp).

Hiroyuki Sekiguchi was with Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, Japan 169-8655

Yoshiyasu Shimizu is with Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, Japan 169-8655

Takeshi Ikenaga is with Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo, Japan 169-8655

This means that full HDTV has double the information of digital HDTV broadcasting. 4K displays have four times the resolution of full HDTV. The information on 4K displays is eight times that of HDTV broadcasting. Blurring is increased when HDTV broadcasting content is converted and displayed on 4K TV. The enlarging factor of eight is far larger than two, which is over the limit of SRR. One of the definitions of SR is "reproducing the high frequency elements that exceed the Nyquist frequency" [6]. Unfortunately, this is not possible as long as SRR is being applied. SRR creates high resolution images from the finest pixels of the original image, which means the reconstructed highest resolution is the same as that in the original image. SRR has become a euphemism for SR since SRR is the only SR that can be applied to real-time video systems. However, there is no evidence that SRR actually works for real video. Video sequences that were used in the previous research were different from those in regular TV content. The camera stops without panning or zooming and only a few objects move at the same size in the frame. This is very rare and it has no applications in real life since the moving objects change in size even on security camera video. Although SRR based SR is used for TV, we do not actually know if SRR really works for TV content. It may just be the case of "The Emperor's New Clothes" if it is applied to TV. The evaluations of SRR have therefore not been sufficient.

Super resolution non-linear signal processing (SRNL) is a method of SR [7] that has recently been proposed. SRNL does not need frame memories or iterations. It can produce higher frequency elements than the Nyquist frequency. Although both SRR and SRNL have aimed at improving the resolution of real-time video, no comparisons of SRR and SRNL have yet been done. Image quality is usually measured with subjective assessments. Video that is processed with SRR in ideal conditions is compared to SRNL in this paper. The complexity and necessary resources for SR are also important. A well-known issue with SRR is iterations. Although SRR can be done with a small number of iterations, this does not disclose the true capabilities of SRR. The number of iterations is 100 in this paper. Although this is large enough to make it converge to SR images, it is not a practical number for video systems due to the delay and cost of 100 frame memories. The flicker in SRR has not previously been mentioned but will be discussed in the following section. SR with parabolic interpolation has been proposed [8]. Although parabolic interpolation produces clearer and sharper images, there is no proof that it can create high frequency elements exceeding the Nyquist frequency. The images that are interpolated with parabolic interpolation are subjectively the same as the original images in terms of resolution [10]. Since SR with parabolic interpolation does not produce satisfying effects when HDTV is converted to 4K TV, it is not discussed in this paper.

II. ISSUES WITH SRR

SRR has two main problems, which are the limits of resolution and flicker.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:9, 2012

SRR uses many frame memories and is a method that chooses the finest pixels in these frame memories. Video has a stronger relation with the temporal axis than that with the vertical and horizontal axis. SRR makes it possible to improve resolution since the probability of obtaining pixels with the highest resolution increases if many frames can be used. However, it is not practical to use many frame memories on TV. Although one frame SR was developed for this reason, it does not work up to the limits of SRR. One frame SRR chooses the finest pixels from homothetic objects in the same frame [5][9]. The idea is based on a similar object existing in the same frame. This is almost the same idea as that in fractal based video coding [11]. Unfortunately, fractal video coding only works for limited images such as sunflowers and ferns. Intra-frame SRR for real time video does not theoretically work although it might work for still images [5][9]. Since our evaluations of the ultimate image quality of SRR is discussed in this paper, intra-frame SR is not.

III. LIMITATIONS OF RESOLUTION WITH SRR

Fig. 1 shows the idea of the SRR for video to improve resolution from a full enlargement. There are four sequential frames on the right. There is an enlarged frame on the left and its size is double horizontally and vertically. The ratio is 1: 4 and this is the ratio of full HDTV to 4K TV. Fig. 1 only outlines the idea behind SRR and it only uses four low resolution frames to reconstruct the high resolution frame. More than 16 frames are necessary for SRR to reconstruct a high resolution image for 1:4 enlargement in an actual system [2]. The white circles and rectangles in the four frames at right are assumed to have the highest resolution in the four frames. These highest resolution areas are used to create the 4K TV frame on the left. Other pixels in the 4K TV frame have been chosen from the original four low frames and the 4K TV frame has been reconstructed by exploiting these pixels.
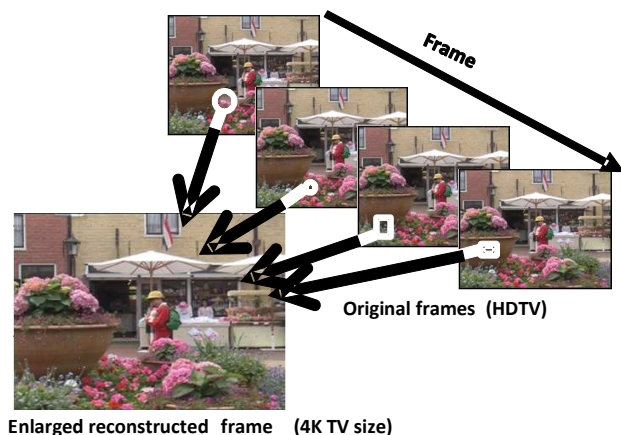

Fig. 1 SSR for enlargement

The highest resolution of the reconstructed 4K TV frame is the same as that of the original four frames with the highest resolution. Although there have been many options and proposals regarding SRR, this problem are critical in that the reconstructed image cannot exceed the limits of resolution of the original image.

The ratio is 1: 8 if we convert the interlaced HDTV to 8KTV. The ratio of enlargement from conventional analogue TV to HDTV is almost 1: 10. It is impossible to create HDTV resolution from analogue TV content with SRR because of the large scaling factor. We would not need digital HDTV broadcasting services if this were possible. Broadcasting companies could continue analogue services and they would not have any need for digital broadcasting equipment. TV sets with SRR cannot create HDTV quality with conventional SR technologies.

IV. FLICKER CAUSED BY SRR

Another issue that has not been mentioned before is flicker. TV and cinema content is created using various kinds of camera operations including panning, tilting, and zooming. The average luminance levels of frames are slightly different with these camera operations even though our eyes cannot perceive this. An SRR algorithm chooses the finest pixels in the original frames that have slightly different luminance levels. There is no guarantee pixels with the same luminance levels will be chosen. Pixels in the same position in the reconstructed frames are not selected from the same luminance levels. This means that pixels in the reconstructed frames are chosen from different frames that have different luminance levels. The DC levels of pixels in reconstructed frames change frame by frame for this reason. This creates flicker. Flicker does not occur if pixels are chosen in the same frame.
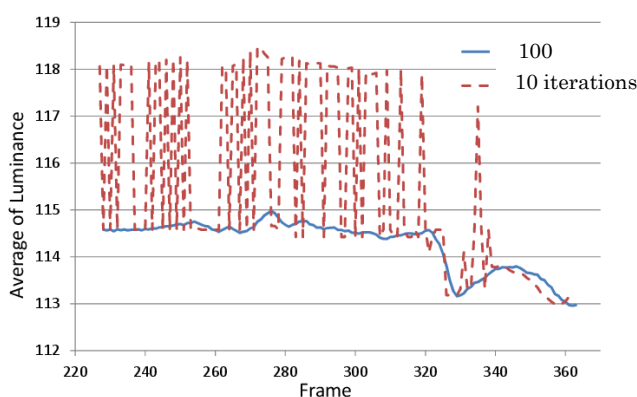

Fig. 2 Flicker artifact

Fig. 3 Flicker caused by SRR

However, the reconstructed frame cannot provide high resolution since high resolution pixels such as those of self-similar objects for SRR do not exist in regular content. Flicker does not happen if the camera is not operated by panning, tilting, and zooming such as in stationary security cameras.

The content used in previous research with SRR did not include camera operations and luminance levels did not change.The content used very slow camera operationsto reduce the luminance level change.

Flicker is visible in flat areas in the image. The white rectangle in Fig. 2 is the surface of a pod. It is a flat area and it is an appropriate area to evaluate flicker. We conducted computer simulations with the video shown in Fig. 2 for 450 frames and created doubly enlarged frames with the SRR algorithm.

Fig. 3 shows the average changes in the luminance level in the white rectangular area in Fig. 2 from frame numbers 220 to 370. The horizontal axis plots the frame number and the vertical axis means the average luminance level. The solid line represents the average luminance level of 100 iterations and the dotted line represents the average luminance level of 10 iterations. The solid line reconstructed with 100 iterations provides a smooth change in the luminance level while the dotted line for 10 iterations yields an active change in luminance, i.e., flicker. Fig. 3 suggests that much iteration reduce flicker, which has not been mentioned before. One hundred iterations use 100 frames, which is duration of 1.67 s (16.7 ms x 100 frames) for video, which is not an acceptable delay for live broadcasting.

The iterations under the constraint decrease flicker since it averages the luminance level of frames. The condition at the end of the iterations is that the total sum of error defined by L_n norm becomes smaller than the preset value. L_n norm usually means L_1 or L_2 norm. The L_n norm value just means minimizing error in the whole image and there is no guarantee that it will provide the best image quality for all images since the regions of interest differ by person. However, minimizing the error defined by L_n norm averages the luminance level and reduces its vibration in the reconstructed frames.

V. SRR VIDEO FOR SUBJECTIVE ASSESSMENTS

We created SRR video with 17low resolution (LR) frames including the previous eight and the following eight framesand 100 iterations because we needed to know the SRR limits under the best conditions. The best image quality of SRR was obtained under the best simulation conditions. According to our simulations, 17 LR frames and 100 iterations were necessary to produce sufficient image quality for subjective assessment. Although we conducted simulations with more iterations and lower resolution frames, there were no differences in subjective image quality. Seventeen frame memories of full HDTV and 100 iterations are not a practical method of signal processing for commercial products such as TVs and BD players due to cost and delay. However, these conditions are essential at least to evaluate SRR. Using these best conditions for SRR provides the best quality and this means that it compromises image quality and cost. However, there are no advantages to apply SRR to video if it does not provide superior video quality from the results of subjective assessment of SRR since video is created under ideal conditions.
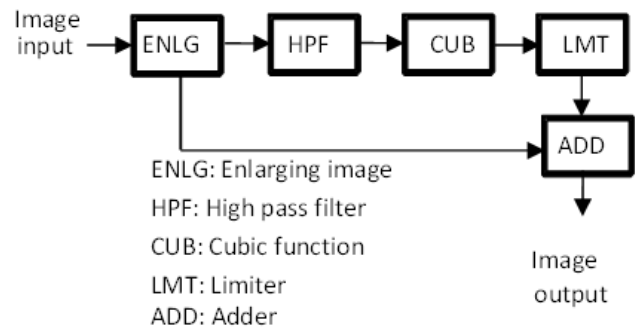

Fig. 4 The proposed method

VI. SRNL

SRNL was proposed fairly recently and it can create higher frequency elements that the original image does not have without iterations or frame memories. Although SRR cannot create high resolution images (HRIs) with just one still image, SRNL can. Fig. 4 shows the flow of signal processing with SRNL. The image is input at the top left and is enlarged with a linear digital filter such as a Lanczos or Bicubic filter. The enlarged image is distributed to the high pass filter (HPF) and the adder (ADD). The HPF detects edges in the image and the edges are cubed. The cubed edges can create high frequency elements that the input image does not have. It is well-known that the image is expanded by a Fourier series that consists of a constant value $a_0 \cos(0 \cdot \omega)$, $a_n \cos(n\omega)$ and $b_n \sin(n\omega)$ functions. Here $n = \pm 1, \pm 2, \cdots$ .The constant value is the luminance level and the edges are represented by $a_n \cos(\cdot)$ and $b_n \sin(\cdot)$ functions. The CUB generates $(a_n \cos(n\omega))^3$ and $(b_n \sin(n\omega))^3$ from $a_n \cos(n\omega)$ and $b_n \sin(n\omega)$, $a_n^3 \cos^3(n\omega)$ and $b_n^3 \sin^3(n\omega)$ can be modified to $a_n^3 \cos(3\omega)$ and $b_n^3 \sin(3\omega)$. This means that third harmonics elements are generated with CUB.

An example processed with SRNL is discussed here. An input image is shown in Fig. 5, which is 256x256 pixels. It was processed with the SRNL algorithm presented in Fig. 4. First, Fig. 5 is doubled horizontally and vertically and it becomes the 512x512 pixel image shown in Fig. 6. The doubled image is processed with HPF and CUB. CUB generates the high frequency edges that the original image does not have.
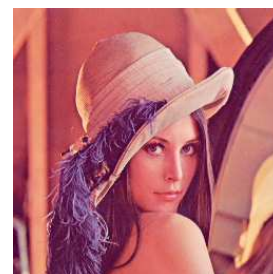

Fig. 5 256x256 pixel image

World Academy of Science, Engineering and Technology
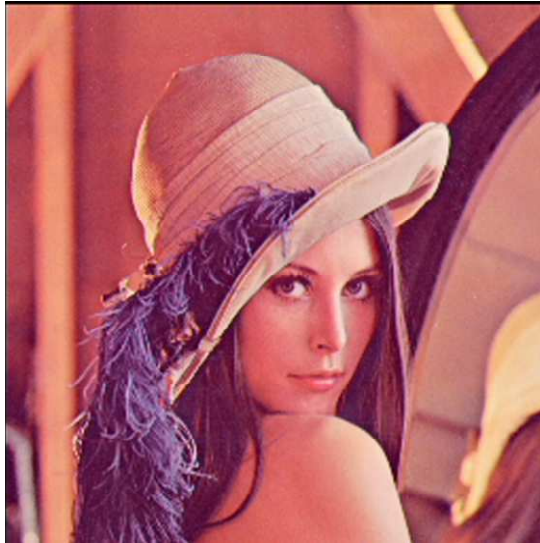International Journal of Computer and Information Engineering
Vol:6, No:9, 2012

Fig. 6 Doubled image in Fig.5 (512x512)

Fig. 7 SRNL processed image with Fig.6

Large and small edges are deleted with LMT. The output of LMT is added to the doubled image and the SRNL processed image shown in Fig. 7 is created. Fig. 8 has the two dimensional FFT results for Fig. 5 and Fig. 9 has the two dimensional FFT results for Fig. 7. Since Fig. 5 is a quarter the size of Fig. 7, the Nyquist frequency of Fig. 5 is half that of Fig. 7. Comparing Fig. 8 with Fig. 9, Fig.8 has a repeated spectrum and Fig. 9 has higher frequency elements that Fig. 8 does not have. This means that SRNL can create higher frequency elements that the original image does not have. The higher frequency elements in Fig. 9 exceed the Nyquist frequency of Fig. 8. High resolution frames can be created with one LR frame by using SRNL.

512x512 SRR HRI may be required here to compare with Fig. 7. However, SRR cannot create HRI with a single 256x256 image such as that in Fig. 5. One more advantage of SRNL is that iterations are unnecessary, which means it is a very light algorithm. The load of SRNL and SRR will be compared in Section Ⅷ.

VII. SUBJECTIVE ASSESSMENT

SRR and SRNL were subjectively assessed to compare their image quality. The ITE sequences for the assessment were selected from the six ITE test sequences in Figs. 10, 11, 12, 13, 14, 15[11]. "Sequence 3" means the No. 3 test sequence in the ITE test sequence and 320x240 or 360x240 mean the resolutions. Each sequence has 450 frames. All of these sequences were enlarged and processed with SRNL and SRR.

Fig. 16 shows the processed results for the first frame of Sequence 2 with SRNL. The original Sequence 2 has 360x240 resolution and the enlarged processed sequences has 720x480 resolution.

Fig. 17 shows the SRR processed results for the first frame of Sequence 2, which also has the same resolution as that in Fig. 16, i.e., 720x480. The SRR simulation conditions are summarized in Table I. SRR uses 17 LR frames including the previous eight and the following eight frames. The motion vector precision and iteration are in sufficient numbers. None of the objects in Sequence 2 moved separately and only the camera was panned horizontally. The camera shook a little vertically at the beginning of Sequence 2. This yielded different sampling points. These were the best conditions for SRR. The resolution in Fig. 17 is not better than that in Fig. 16 even though the best conditions for SRR were adopted.
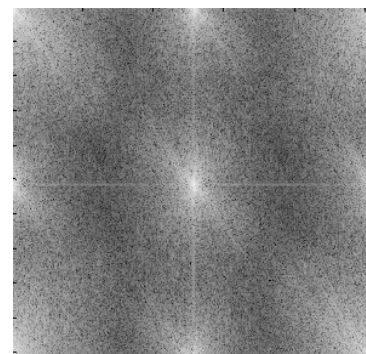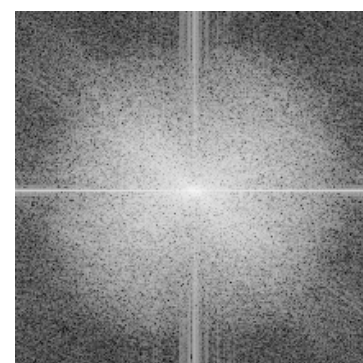

Fig. 8 2D FFT Result of Fig.5

Fig. 9 2D FFT Result of Fig.7

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:9, 2012

TABLE I
SIMULATION CONDITIONS FOR SRR

Items	Parameters
Number of LR frames	17 frames (previous 8 and following 8)
Motion vector precision	1/4 pixel
Number of iterations	100

TABLE II
GRADES OF SUBJECTIVE ASSESSMENT

Meanings	Score
SRNL is excellent	2
SNRL is good	1
Even	0
SRR is good	-1
SRR is excellent	-2

Fig. 10 ITE test sequences 2
(360x240)

Fig. 11 ITE test sequences 3
(320x240)

Fig. 12 ITE test sequences 7
(360x240)

Fig. 13 ITE test sequences 8
(320x240)

Fig. 14 ITE test sequences 16
(360x240)

Fig. 15 ITE test sequences 30
(320x240)

Nineteen observers were asked to only assess the resolution of the video sequences processed with SRNL and SRR using a scale with five grades. The grades are categorized in TableIIand are scaled from -2 to 2. Two monitors were prepared for SRNL and SRR and each sequence was repeated until an observer determined the score.

Zero in this assessment meant the images processed with SRNL and SRR were assessed as being the same. Fig.18 plots the results of subjective assessment by 19 observers. The vertical axis represents the score from -2 to 2 in the assessment. The dots are the average for observers for all sequences and the lines with the dots are the standard deviation. SRR and SRNL may have resolution of similar quality if the deviation crosses the horizontal axis with the zero score. However, none of the results crossed the zero line and all of them indicated that SRNL was better than SRR. The differences were larger than the three standard deviations, except for Sequence 30, which had a significant difference in subjective assessments.

The ratio between the original sequences and the enlarged sequences is 1: 4, which is the same ratio as that between full HDTV to 4K TV. As previously mentioned 4K TV is available on the market. However, not much content has been created for 4K displays since there is no 4K broadcasting or 4KBlu-ray. Even movies are created in full HDTV resolution. Important content for 4K TV is that converted from HDTV. Although SRR is a common SR technology, the limit of enlarging SRR is by a factor of two and the enlargement factor in this case would be four. This means that SRR cannot provide sufficient resolution for enlargement from HDTV to 4K TV. According to our subjective assessment, SRNL provides better results than SRR at the same ratio of enlargement from HDTV to 4K TV.

Fig. 16 Enlarged and processed with SRNL

Fig. 17 Enlarged and processed with SRR

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:9, 2012

VIII. COMPLEXITY AND LOAD

SRNL is a very simple algorithm and it does not consume resources excessively. The loads of SRNL and SRR are compared in this section. We used the CPU running time in the simulations with SRNL and SRR. The simulation for SRR included enlargement, motion vector estimation with quarter pixel precision, and the iteration process of SRR. The simulation for SRNL included enlargement and the algorithm given in Fig. 4. The specifications of the machine used for the simulations are listed in Table III. Table IV summarizes the simulation times for Sequence 2 in Fig.10. SRR requires 69890.5 s for 450 frames and SRNL requires 88 s. The ratio between SRR to SRNL is almost 1000: 1. This proves that SRNL has a lighter load than SRR.

The CPU load in software simulation is in proportion to the load on hardware including cost. Although further research is necessary, 4K TV has the possibility of being equipped with SRNL as real-time hardware. SRNL also has the possibility of being used for software image and video applications because of its cost efficiency.

TABLE III
MACHINE SPECIFICATIONS

CPU	Intel Core I7
Memory	8 GB
Software	Visual Studio with Open CV

TABLE IV
CPU LOAD

Algorithm	Processing time (s)
SRR(100 iterations)	69890.5
SRNL	88.0

IX. CONCLUSION

SRNL was compared with SRR as typical conventional SR. It was proved that the resolution of SRNL was better than that of SRR through subjective assessments. Six video sequences were used in the assessments and SRNL had better resolution with a difference of three times the standard deviation in five of six sequences. The loads of SRNL and SRR were compared with the CPU running time. The ratio between them was almost 1000: 1. Although SRR has been researched for many years as typical SR, it does not work and cannot improve the resolution of all images and video. The load of SRR is not light enough to be installed in real-time hardware. SRNL improves resolution for various kinds of video content and is a light algorithm. SRNL has the possibility of being used in real-time hardware and used in the security, medical, and broadcasting industries. Further research on SRNL is necessary to use it in practical applications.
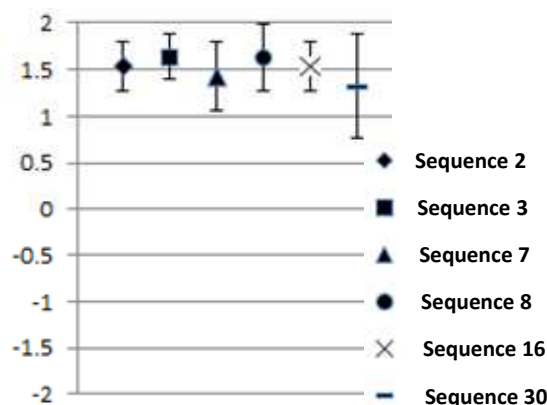

Fig. 18 Results of the subjective assessment

REFERENCES

[1] SubhasisChaudhuri,"Super-Resolution Imaging", Kluwer Academic Publishers, ISBN978-0-7923-7471-8. Boston/Dordrecht/LondonFarusi
[2] A. Katsaggelos, R. Molina, and J. Mateos, "SuperResolution of Images and Video"}, Synthesis Lectures on Image, VideoMultimedia Processing, Morgan Claypool Publishers, May 2010. LaVergne. TN USA.
[3] S. Cheol, M. K. Park, and M. G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview",IEEE Signal Processing Magazine, 1053-5888/03, May, 2003.
[4] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multi-frame Super-resolution", IEEE Transactions on Image Processing , vol. 13, no. 10, pp. 1327-1344, October 2004
[5] D. Glasner, S. Bagon, and M. Irani, "Super-Resolution from a Single Image", International Conference on Computer Vision (ICCV), October 2009.
[6] http://www.toshiba.co.jp/regza/detail/superresolution/resolution.html(in Japanese).
[7] S. Gohshi, "A new signal processing for video" , MMSys'12, ACM, Feb. 2012, ISBN: 978-1-4503-1131-1.
[8] T. Ida et al., "Improving resolution of video", Patent ID:JP.4116649, http:// www.ipdl.inpit.go.jp/homepg_e.ipdl (in Japanese).
[9] N. Matsumoto and T. Ida, "A Study on One Frame Reconstruction-based Super-resolution Using Image Segmentation", IEICE Technical Report \relax, SIP2008-6,IE2008-6 (2008-04) (in Japanese).
[10] G.Yang, and F. Yu, "Parabolic Interpolation with adaptive error compensation", Proceedings of the Second International Symposium on Networking and Network Security (ISNNS '10), Jinggangshan, P. R. China, 2?4, April, 2010, pp. 267-270 :ISBN978-952-5786-09-1.
[11] M.F.Barnsley,"Fractals Everywhere", Morgan Kaufmann, ISBN 0-12-079069-6. Harcourt Place, 32 Jamestown Road, London NW17BY. UK.
[12] ITE Test Sequence ver.1, http://www.nes.or.jp/gaiyo/pdf/ite_ hyoujundouga_sample.pdf