

Subjective Assessment of Super-Resolution -High-Resolution Effect of Nonlinear Signal Processing-

Chinatsu Mori, Masaki Sugie, Hirohisa Takeshita, Seiichi Gohshi
Kogakuin University

Department of Informatics

1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-8677 Japan

em13018@ns.kogakuin.ac.jp, em13010@ns.kogakuin.ac.jp, em13012@ns.kogakuin.ac.jp, gohshi@cc.kogakuin.ac.jp

Abstract—Results of a subjective assessment of super-resolution (SR) technology for 4K TV sets are reported. 4K TV sets are currently available in the market, and some are equipped with SR functionality. Recently, SR technology that uses nonlinear signal processing (NLSP) and can work in real time has been proposed. The proposed method can enhance images without enlargement. The subjective assessment was performed by comparing 4K videos with and without NLSP on 4K TV sets. The assessment method was a combination of Scheffe’s paired comparison and ITU-R BT.500. Assessment data were statistically analyzed, and the results proved that NLSP is superior in resolution.

I. INTRODUCTION

4K TV sets are currently available on the market. They provide high quality images with quadruple the resolution of HD. However, although 4K TV sets are widely available, 4K video content is not; up-converted HD video content is still used with 4K TV sets.

Image enlargement is required to display low-resolution images on high-resolution displays. Note that enlarging of an image causes blurring. Thus, image resolution should be improved, and almost all consumer TV sets are equipped with a sharpness function (i.e., an enhancer or an unsharp mask). The sharpness function algorithm is simple and can work in real time; thus, it is widely used for video devices. However, the sharpness function can enhance edges but cannot actually improve resolution.

Super-resolution (SR) technology is a method for improving the resolution of images. Unlike the sharpness function, SR can reproduce high frequency spectra that the sharpness function cannot create.

While the capabilities of SR technologies for TV sets need to be improved, the evaluation of SR technology image quality is also required. A TV manufacturer defines SR as ‘technology that can create the high frequency spectra higher than the Nyquist frequency’ [1]. It is possible to theoretically analyze resolution by comparing SR and non-SR processed video signals. However, there is no way to take the video signals after the SR signal process from the TV sets. Thus, a subjective assessment is the only way to evaluate the image qualities of SR technologies embedded in video devices.

Our team has proposed a new SR technology using nonlinear signal processing (NLSP), and its theoretical capability has been proven [2]. In addition, we have demonstrated improved subjective image quality in 4K videos up-converted from HD [3].

The proposed method can enhance images without enlargement. However, the high-resolution effect has not been tested.

In this paper, our purpose is to verify the high-resolution effect of NLSP. We perform a subjective assessment wherein candidates are 4K videos with and without NLSP.

II. SUBJECTIVE ASSESSMENT METHOD

An experiment for the assessment was conducted by comparing 4K videos with and without NLSP. Two consumer grade 4K TV sets of the same model were used. The 4K TV set used in the experiment is shown in Figure 1. Observers watched the TVs and compared the image resolution qualities. This was intended to reproduce the conditions by which shoppers compare similar items at a store.

ITU-R BT.500 is a common subjective assessment of the quality of television pictures [4]. BT.500 defines certain experimental conditions, such as viewing and lighting conditions, test sequences, and observers. However, the BT.500 standard is typically used to assess the quality of several test sequences on only one display device. Thus, we determined experimental conditions, such as the number of observers and the length of test video sequences, on the basis of BT.500, with the exception of unadaptable conditions for several displays.

III. EXPERIMENT

A. Scheffe’s Paired Comparison

Scheffe’s paired comparison method was used for the assessment. Using a pair of 4K videos with and without NLSP, observers scored the videos on a 5-grade scale (-2 to +2). The rating scale for the assessment is shown in Table I. The results of the assessment were statistically analyzed using analysis of variance. Prior to the experiment, we conducted a training session to explain resolution and the experimental method to each observer. The observers were asked to assess only resolution and ignore other factors, such as noise and color.

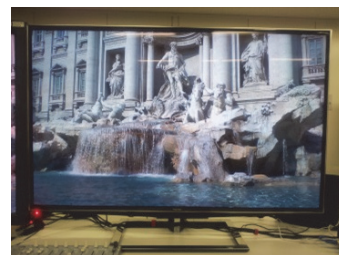


Fig. 1. Consumer grade 4K TV

TABLE I. RATING SCALES

Score	Description	Detailed Information.
2	Excellent	The resolution is better than other one.
1	Good	The resolution is little better than other one.
0	Even	The resolution is the same as other one.
-1	Poor	The resolution is little poorer than other one.
-2	Bad	The resolution is poorer than other one.

B. Apparatus

A system diagram of the experiment is shown in Figure 2. A 4K video player was used to show video sequences to the observers, and a video signal was simultaneously distributed to each 4K TV set. In the process for non-NLSP, the original video signal was output to the 4K TV set without any processing. In the process for NLSP, external NLSP hardware was connected between the video player and one of the 4K TV sets. Then, the output signal after NLSP was displayed on the 4K TV. The NLSP hardware is shown in Figure 3.

C. Test Sequences

Test video sequences for the experiment were taken by a consumer video camera. The resolution is 4K (3840×2160), and the format is MPEG-4. Five test video sequences that are appropriate for the assessment of resolution were selected. Most did not include pan and tilt scenes. The length of each test sequence is between 10 s and 15 s in reference to BT.500. Figure 4 shows the test video sequences for the experiment. The ovals on the figures indicate high-resolution areas. The observers were asked to watch these areas and determine assessment scores.

D. Observers

Thirty observers participated in the experiment. The observers are non-experts, and they have normal visual acuity and color vision in reference to BT.500.

E. Experimental Conditions

Normal lighting conditions for a room were selected for the experiment to reproduce a viewing environment in which consumers choose a TV set at a store. The observers could freely change their viewing distance during the assessment. Figure 5 shows a photo of the experimental environment.

IV. RESULTS AND DISCUSSION

The assessment results for the ‘‘Cherry Tree’’ sequence are shown in Figure 6. Figure 6 shows a graph of the average scores and the standard deviations of each stimulus. Here, ‘‘NLSP’’ indicates stimulus of the process for NLSP, and ‘‘Non-NLSP’’ indicates stimulus of the process without NLSP. The horizontal axis is the average scores. The average scores were obtained by dividing the total assessment scores by the number of observers. The bars extending from the marks are standard deviations. The graph shows that the average score for NLSP is 1.73 and that of non-NLSP is -1.27. Note that the average score for NLSP is higher than that of non-NLSP. Other assessment results are shown in Figures 7, 8, 9, and 10. All results demonstrate similar tendencies.

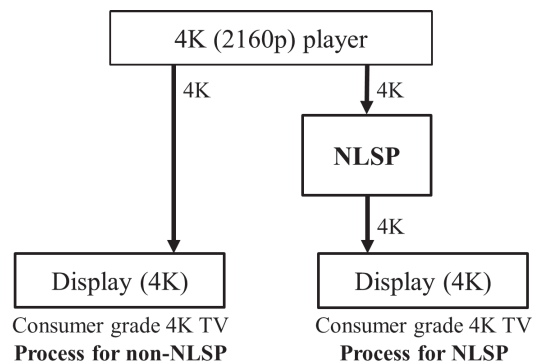


Fig. 2. System diagram



Fig. 3. NLSP hardware

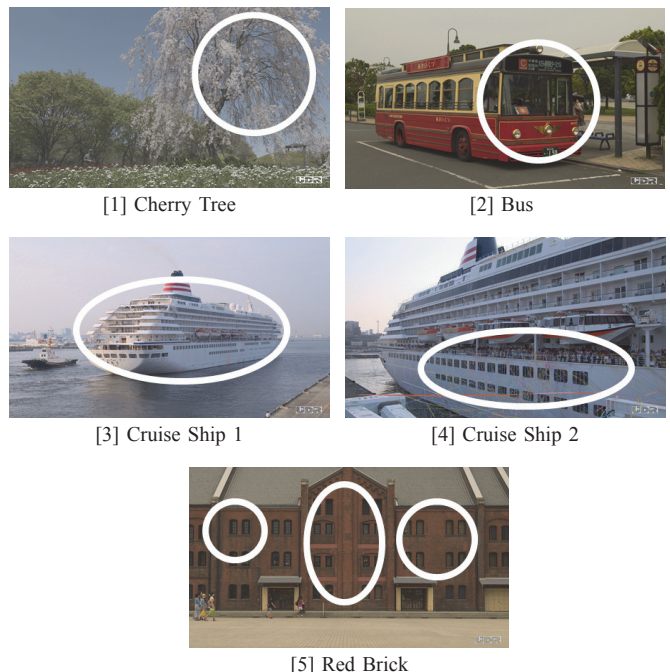


Fig. 4. Test sequences



Fig. 5. Experimental environment

Analysis of variance was used to assess significant differences. The results of the analysis of variance for the “Cherry Tree” sequence are shown in Table II. Table II shows the sum of squares, degrees of freedom, and mean square values [5]. F_0 denotes a value for an F-test. F_0 is obtained by the quotient of the mean square of the stimuli and that of the residual. Here, a critical F value for the 0.01 significance level is $F_{1\%} = 7.093$. If F_0 is greater than $F_{1\%}$, then the null hypothesis was rejected and there is significant difference between the stimuli. F_0 for the “Cherry Tree” sequence is $F_0 = 667.330 > F_{1\%}$. Similarly, other analysis of variance results for other sequences are shown in Tables III, IV, V, and VI. All results satisfy $F_0 > F_{1\%}$, and significant differences between the stimuli are observed.

As a result, NLSP stimuli have higher average scores than non-NLSP stimuli. In addition, the 0.01 significant differences were detected in all test sequences. Thus, it is statistically proven that video signals with NLSP obviously differ from the original video signals, and the image resolution quality of NLSP is superior. It is assumed that similar results would be obtained for other video sequences because the experimental method is reproducible. These results prove that NLSP can effectively enhance images.

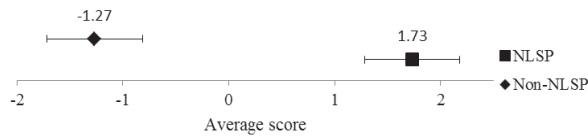


Fig. 6. Average score (Cherry Tree)



Fig. 7. Average score (Bus)

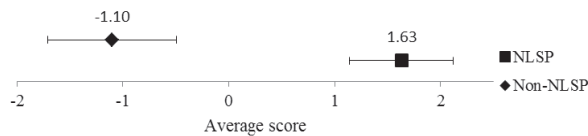


Fig. 8. Average score (Cruise Ship 1)

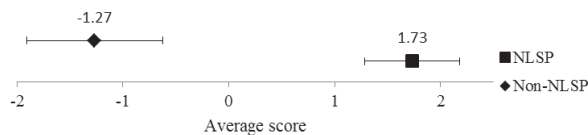


Fig. 9. Average score (Cruise Ship 2)

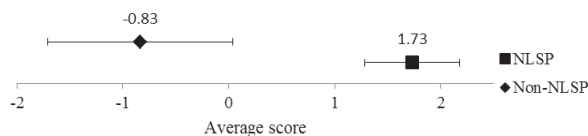


Fig. 10. Average score (Red Brick)

V. CONCLUSION

A subjective assessment was performed to test the high-resolution effect of NLSP. The assessment was conducted by comparing NLSP and non-NLSP processed 4K videos on 4K TV sets. The results statistically proved that NLSP can effectively enhance images. A conventional SR function is performed for only enlarged images, whereas the proposed method is useful for enhancing images without enlargement. In this study, we have only assessed the resolution. In the future, we intend to assess other factors, such as noise and color.

REFERENCES

- [1] <http://www.toshiba.co.jp/regza/detail/superresolution/resolution.html> (in Japanese)
- [2] Seiichi Gohshi, “A new signal processing method for video: Reproduce the frequency spectrum exceeding the Nyquist frequency,” MMSys '12 Proceedings of the 3rd Multimedia Systems Conference, pp.47-52, Sep.2012.
- [3] Masaki Sugie, Seiichi Gohshi, Hirohisa Takeshita, and Chinatsu Mori, “Subjective assessment of Super-resolution: Paired comparison with 4K TV,” IEICE Technical Report, vol. 114, no. 233, IE2014-48, pp. 51-56, Dec.2014. (in Japanese)
- [4] Rec. ITU-R BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” ITU-R, 2002.
- [5] Tadahiko Fukuda, Ryoko Fukuda, “Ergonomics handbook,” ISBN978-4-86079-036-3, Scientist press co.ltd, Tokyo, 2009. (in Japanese)

TABLE II. ANALYSIS OF VARIANCE (CHERRY TREE)

Factor	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Stimuli	135.000	1	135.000	667.330**
Residual	11.733	58	0.202	-
Total	146.733	59	-	-

TABLE III. ANALYSIS OF VARIANCE (BUS)

Factor	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Stimuli	117.600	1	117.600	475.870**
Residual	14.333	58	0.247	-
Total	131.933	59	-	-

TABLE IV. ANALYSIS OF VARIANCE (CRUISE SHIP 1)

Factor	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Stimuli	112.067	1	112.067	367.917**
Residual	17.667	58	0.305	-
Total	129.733	59	-	-

TABLE V. ANALYSIS OF VARIANCE (CRUISE SHIP 2)

Factor	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Stimuli	135.000	1	135.000	441.541**
Residual	17.733	58	0.306	-
Total	152.733	59	-	-

TABLE VI. ANALYSIS OF VARIANCE (RED BRICK)

Factor	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Stimuli	98.817	1	98.817	204.448**
Residual	28.033	58	0.483	-
Total	126.850	59	-	-

** : 1% Significance difference ($F_0 > F_{1\%}$), $F_{1\%} = 7.093$