Subjective Assessment of Text Quality on Smartphone Display with Super Resolution

Aya Kubota
Department of Information Science
Kogakuin University
Tokyo, Japan
em17008@ns.kogakuin.ac.jp

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

***Abstract*— It is only a decade ago that smartphones appeared on the market. However, the market has since grown rapidly, and people of all ages now use smartphones. In many cases, people read text on their smartphones, but depending on the design of a website, it may be difficult to read its text. By improving the resolution of the text, the readability of the text can be improved. One research area for increasing the resolution is Super Resolution (SR), which includes Non-Linear Signal Processing Super- Resolution SR (NLSP), a method that can be implemented on smartphones. However, NLSP has never been applied to text to improve readability. Text has many kinds of characters, such as Chinese characters, and alphabets of different languages. Features of these characters are different. We applied NLSP to Japanese text including Chinese characters, katakana and numbers, displayed on Liquid Crystal Display (LCD), and verified its effectiveness using a subjective assessment. In addition, we applied NLSP to English text and compared the difference between image quality text with and without NLSP. The subjective assessment results show that NLSP can increase the resolution of Japanese and English text. Thus, the assessment results for text on LCD are discussed in this paper.**

Keywords- Nonlinear signal processing; Super-Resolution; Subjective assessment; Smartphone.

I. INTRODUCTION

Smartphones have become daily necessities in modern society. In addition to processing communication functions, such as telephone and e-mail, it is possible to obtain information in real time via the Internet. When used for the above functions, text must often be read, in the form of operation buttons or explanatory text. Support functions to make text easier to read, such as changing the font size, are set in the application that is preinstalled in the operating system (such as mail, smartphone settings, etc.). However, there are websites that do not have a font size larger than a certain size even if the text is enlarged, and sites where the color of the background and the text is very similar. Therefore, problems, such as these can make it difficult to read the text.

Improving the resolution of the images can make it easier to read a text. One method to improve resolution is Super-Resolution (SR) technology [1]. Most 4K Televisions (TV) are equipped with SR. Non-Linear Signal Processing SR (NLSP) is an SR technology that can be embedded into smartphones [2]. The algorithm is simple and fast: hence, processing with software is possible, and smartphones with NLSP are already being sold on the market [3]. The effectiveness of NLSP is higher than that of other SR technologies [4][5], and NLSP is effective even in smartphone videos [6].

However, the effectiveness of NLSP for text on smartphone display has not been verified. In this study, we verify the effectiveness of using a smartphone with NLSP compared to one without NLSP.

Images processed with NLSP are introduced only to the display of the smartphone and there is no electric output of the processed image. Therefore, it is impossible to use an objective assessment because the objective assessment requires electric image signal with and without NLSP. Subjective assessment is the only way to assess the difference between the displays. However, subjective assessment is only a reflection of how we feel. It is difficult to ensure the reproducibility of subjective assessments. In addition the subjective assessments require observers and time to assess the image quality.

Although there are issues with the subjective assessment, the ITU-R has standardized subjective assessment methods. ITU-R BT.710 recommends experimental conditions to obtain reproducible results in subjective assessment experiments [7]. However, BT.710 does not mention practical quantitative scoring assessment, which is defined in BT.500. They are the Double Stimulus Continuous Quality Scale (DSCQS) and the Double Stimulus Impairment Scale (DSIS). In this experiment, we need to compare five smartphones and they are different manufactures products, and BT.500 and BT.710 do not meet our requirements. One of our authors developed a subjective assessment for multiple displays [7][8]. It applies best–worst method, and statistical analysis is introduced to analyze reproducibility. It shows good results if the images/videos are selected appropriately.

This paper is organized as follows. In Section II, the subjective assessment for multiple displays is explained. In Section III, NLSP is explained. In Section IV, the test images are presented and the experiments are explained. In Section V, the statistical analysis is adapted to the assessment results and in Section VI the analyzed result is discussed. Section VII is the conclusion of the paper.

II. QUANTITATIVE ASSESSMENT OF DIFFERENT SMARTPHONE DISPLAYS

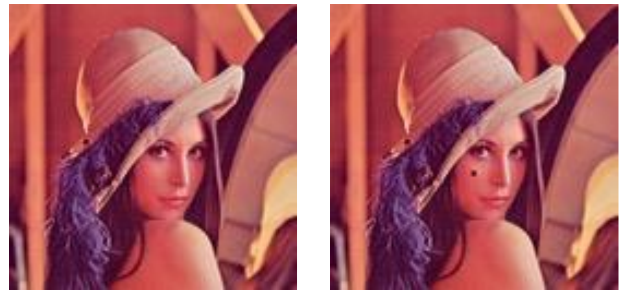
Smartphone displays show us images as optical signal. It is difficult to compare image quality between different types of smartphones. It is very difficult to conduct objective quantitative assessments between optical images. Until now, objective comparison of the image quality using different types of smartphone displays has not been reported. In this study, we introduced subjective assessment, and made it possible to quantitatively asses the image quality between different displays.

Objective assessment and subjective assessment are evaluation methods. Objective assessments analyze the signal and express high and low image quality by a numerical value. However, the results of an objective assessment do not always match with how we feel. For example, an original image is given in Figure 1(a), and the degraded image is given in Figure 1(b). The Peak Signal to Noise Ratio (PSNR) of the degraded image in Figure 1(b) is 40.1112dB. A PSNR 40dB is generally said to be a high image quality [9]; however, Figure 1(b) contains degradation in the form of a black square in the center of the image. When images include local degradation, the results of PSNR sometimes deviate from our feeling.

Thus, objective assessments cannot reflect image quality accurately. In addition, objective assessments require a comparison of the assessment image with the original image. As discussed in the previous section, signals processed inside the smartphone cannot be output anywhere outside the display. Therefore, assessment by signal analysis is impossible, and thus the experiment is conducted using subjective assessment.

The best–worst method was adopted as the assessment method using multiple displays. Normalized ranking method and paired comparison method are other assessment methods. Experimental stimuli are ranked at once in the normalized ranking method. The process of the method is simple; however, when differences between the stimuli are small, sometimes the differences cannot be detected because of large differences between stimuli influences. In the paired comparison method, stimuli are compared one on one and ranked. Two stimuli are selected, and the observers evaluate the stimuli based on the other. Thus, differences between stimuli can be obtained in detail. However, evaluation is performed for all the stimulus combinations, which places a heavy burden on the observers. In the best–worst method, observers select the best stimuli and the worst stimuli. After excluding the selected stimuli, the observers again select the best and the worst from the remaining stimuli.

Although normalized ranking method is common, the best-worst method can detect small differences more accurately than the normalized ranking method. Accuracy of the best-worst method is lower than the paired comparison. However, the time consumption of the best-worst method is shorter than that of the paired comparison. It means that observers' burden of the best-worst method is lower than that of the paired comparison. Therefore, in this paper, the best–worst method is adopted.



(a) Original image

(b) Degraded image (PSNR: 40.1112dB)

Figure 1. Objective assessment by PSNR

In this study, an assessment experiment was conducted using five smartphones. The test images are screenshots of a website containing text.

III. SUPER RESOLUTION

Super resolution technology is a method to improve image/video resolution, and mounted on most 4K TVs. Although smartphones that has 4K resolution display are for sale, images/videos that have 4K resolution are insufficient. Therefore, it is necessary to improve the image/video resolution. However, it is impossible to mount current mainstream SR technologies to smartphones due to the technical reason. In this section, problems of the conventional SR technologies if they are mounted to smartphones, and NLSP, which can solve the issue are explained.

A. *Super resolution for smartphones*

The purposes of TV and smartphone are different; therefore, performance difference, such as display size and processing speed, is great.

If conventional SR are mounted to smartphones, issues will occur. For example, image quality difference cannot be understood on small smartphone displays, and processing will be slow because processing works on software. Although designed hardware for implementation SR is mounted in TVs, smartphones have no space to mount new hardware.

Therefore, it is impossible to implement SR for TVs to smartphones. The size of the monitor becomes an important factor in seeing an SR processed image [10]. Much research on SR has been conducted. However, it does not discuss the difference in clarity of the image depending on the display size. Even if images are processed with SR, whether SR is effective or not on small smartphone displays has not been reported. SR studies freely select their processed image sizes to recognize the resolution improvement. Personal Computer (PC) monitors have been used to check image resolution. Although commercial HDTV sets with SR technology can be used (Tos, 2009 [11]), the screen sizes of HDTVs are 40 inches or larger. On the other hand, the display sizes of commercial smartphones are approximately 5 to 6 inches. It is difficult to recognize improvement with SR on a small display. Even if we can recognize resolution improvement on a large display, such as a PC monitor or HDTV, it is not always recognizable on smartphone displays. Therefore, if we are to implement SR technology, it is meaningless to implement the SR function unless resolution improvement is recognized. Smartphones are developed on the assumption that they are portable; therefore, the small devices are used to carry out many functions. Thus, it is impossible to add devices to a smartphone to use SR. There are two difficulties in implementing SR on a smartphone with limited resources. The first is the complexity of the SR algorithm. Many SR algorithms have been proposed (Farsiu et al. [10], Park et al. [12], Katsaggelos et al. [13], van Eekeren et al.[14], Panda et al. [15], Glasner et al. [16], Sun et al. [17], Dong et al. [18]). Super Resolution image Reconstruction (SRR) and Learning-Based Super Resolution (LBSR) are typical SR technologies, though many others have been proposed. However, all SR algorithms, including SRR and LBSR are difficult to use in real time for video because they require iteration to create a high-resolution image. Iteration is very time consuming and difficult to execute on the CPU/GPU of a smartphone. Although a non-iteration SRR algorithm for HDTV has been proposed (Matsumoto and Ida, 2010), the resolution is lower than that of a conventional HDTV and an additional device for implementation SRR is required.

The second difficulty is SR on smartphones must work on the CPU/GPU of a smartphone. Due to the space and power consumption, it is difficult to add a device for SR implementation to a smartphone. If we add a new device to a smartphone, the new parts will shorten battery duration owing to higher power consumption. Thus, to use SR on a smart-phone, it is necessary to work with the limited resources, such as the CPU/GPU, of a smartphone. The CPU/GPU executes many tasks, and resources, such as the memory bandwidth are limited. If sufficient CPU/GPU power and resources are not provided for the SR process, a video cannot be processed in real time, and frame drops can occur. In the worst case, the video will freeze. To overcome these difficulties, an SR algorithm for a smartphone must be simple and sufficiently light to work on CPU/GPU power and limited resources.

B. NLSP

NLSP is a simple and fast SR technique, which made it possible to implement SR to smartphones for the first time in the world.

The process is similar to enhancer that it increases resolution by emphasizing edges; however, NLSP emphasizes high-frequency components extracted from the input image using a nonlinear function [2]. Figure 2 shows the signal flow of NLSP. The input signal has two paths. The first path consists of a High-Pass Filter (HPF), Non-Linear Function (NLF), and a Limiter (LMT). This path generates high-frequency components that the original video does not have. High-frequency components include the edges and details of an image/video. HPF detects the edges of the input signal. Then, the detected edges are processed with the NLF. It can create high-frequency components not included the

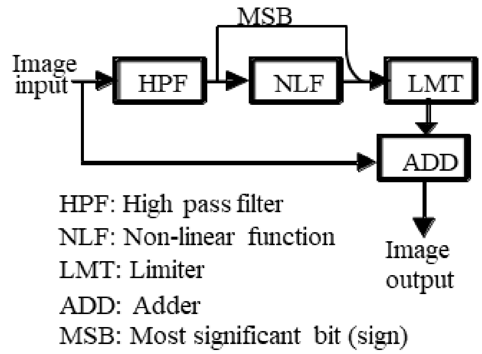


Figure 2. Block diagram of NLSP

input video. An example of an NLF is a cubic function ($f(x) = x^3$). The function can amplify the high-frequency components by as many as three times. We explain the NLF using the cubic function $f(x) = x^3$. It is well known that images and videos can be expressed by sine and cosine waves with Fourier series. If f(x) is assigned sinθ, it is changed to $(\sin\theta)^3$ using the cubic function. Similarly, if f(x) is assigned cosθ, it is changed to $(\cos\theta)^3$. $(\sin\theta)^3$ can be changed to $\sin(3\theta)$ and $(\cos\theta)^3$ can be changed to $\cos(3\theta)$. $\sin(3\theta)$ and $\cos(3\theta)$are harmonic waves, and the frequency is higher than the original video. The cubic function is just an example of a nonlinear function, and the NLF is used to create the high-frequency components by harmonic waves. The harmonic waves are generated only from the edges detected with the HPF. Flat areas do not have edges; therefore, there are no harmonic waves. The LMT saturates these large values to fit the harmonic waves to the video.

The second path is from the input, and it is directly connected to the Adder (ADD). The ADD adds the harmonic waves processed by the LMT to the original video. The process is conducted pixel by pixel.

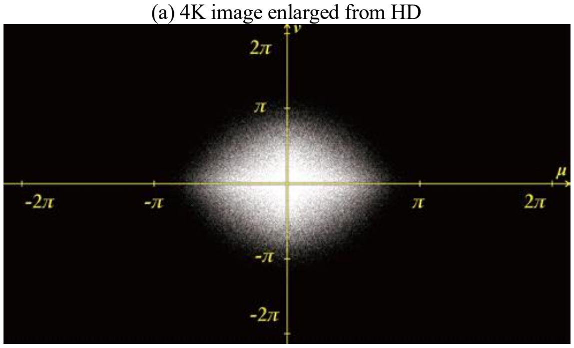
Therefore, the output of the ADD has high-frequency components not included the original video. This processing method can improve the resolution, and even generate high-frequency components that exceed the Nyquist frequency of the original video. This simple and fast algorithm has led to the development of real-time NLSP hardware.

Figure 3 shows an image processed with NLSP hardware. Figure 3(a) is an enlarged image from HDTV to 4K. Figure 3(b) is the NLSP processed result of Figure 3(a). Although Figure 3(a) is blurry, Figure 3(b) more clearly expresses the edge and details than Figure 3(a). Figures 3(c) and (d) are the two-Dimensional Fast Fourier Transform (2D-FFT) results of Figures 3(a) and (b) respectively.

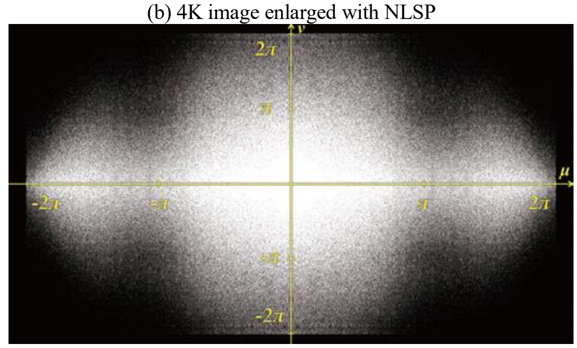
Figures 3(c) and (d) show the frequency characteristics in the frequency domain. The horizontal and vertical axis are the horizontal and vertical frequencies of the image. The center of the image shows low-frequency. The frequency is higher with distance from the center. Figure 3(d) has horizontal and vertical high-frequency components that are

(a) 4K image enlarged from HD

(b) 4K image enlarged with NLSP



(c) 2D-FFT result of Figure 3(a)

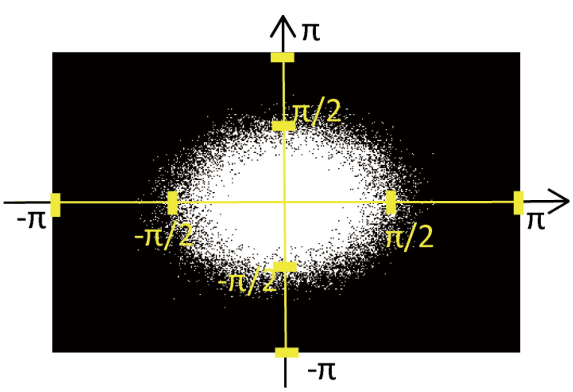


(d) 2D-FFT result of Figure 3(b)

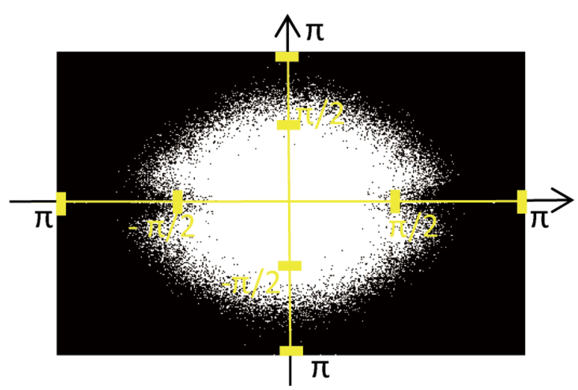
Figure 3. Image processed with real-time NLSP hardware

(a) Input image

(b) Image processed with smartphone emulator



(c) 2D-FFT result of Figure 4(a)



(d) 2D-FFT result of Figure 4(b)

Figure 4. Image processed with an NLSP smartphone

not present in Figure 3(c). This means that NLSP creates high-frequency components, and increases the resolution.

Although Super-Resolution image Reconstruction (SRR) and Learning-Based Super Resolution (LBSR) are the current mainstream SR technologies, they cannot be mounted to a smartphone. SRR is a technology that generates a high-resolution image from multiple degraded images [10]; however, the processing requires iteration. When the input image and output image have the same resolution, the technique is not very effective [19]. LBSR is a method that increases resolution using a database [20]. The effectiveness is affected by the database, and the processing requires both an expensive database and iteration. Thus, both of the above technologies require complex processing. In addition, their effectiveness is lower than that of NLSP [5][8].

Although the NLSP algorithm is very simple, whether NLSP can process videos in real-time on CPU/GPU of a smartphone has not been verified. One of the authors used a smartphone emulator to prove that NLSP can work normally on a smartphone [2]. Figure 4 shows the NLSP processed result with a smartphone emulator. Figure 4(a) shows a frame of an input video. Figure 4(b) shows the NLSP processed result of Figure 4(a) on a smartphone. Figure 4(a) more clearly expresses the edge and the details than Figure 4(a) does. Figures 4(c) and (d) show the 2D-FFT results of Figures 4(a) and (b) respectively. Figure 4(d) has horizontal and vertical high-frequency components not included Figure 4(c). The results show that NLSP can process in real-time and improve the resolution on a smartphone. A smartphone with NLSP has already been sold on the markets (Figure 5) [3].

Figure 5. Developed Smartphone with NLSP

IV. EXPERIMENT

The effect of image processing differs, depending on the image. We adjusted NLSP for text; hence, it was necessary to verify the effect of NLSP for text. When a new technology is developed, it is necessary to compare a processed image with an unprocessed image. Thus, in this experiment, a smartphone with NLSP and one without NLSP were compared. The result of the comparison indicates the effects of using NLSP. In addition, the experiment was conducted using smartphones from different manufacturers and verified the effect of NLSP in comparison with other technologies.

Text includes many types of characters, such as Chinese characters, hiragana, and alphabets. Each character has different features. Most Chinese characters consist of straight lines. Hiragana and alphabets consist of straight and curved lines. Therefore, even if NLSP is effective when applied to Chinese characters, we do not know whether NLSP is effective or not for characters that have different features. Thus, in this study, we conducted a subjective assessment to evaluate the image quality of Japanese text including Chinese characters, katakana, and hiragana. After, a similarly subjective assessment was conducted using English text.

A. *Experimental Condition*

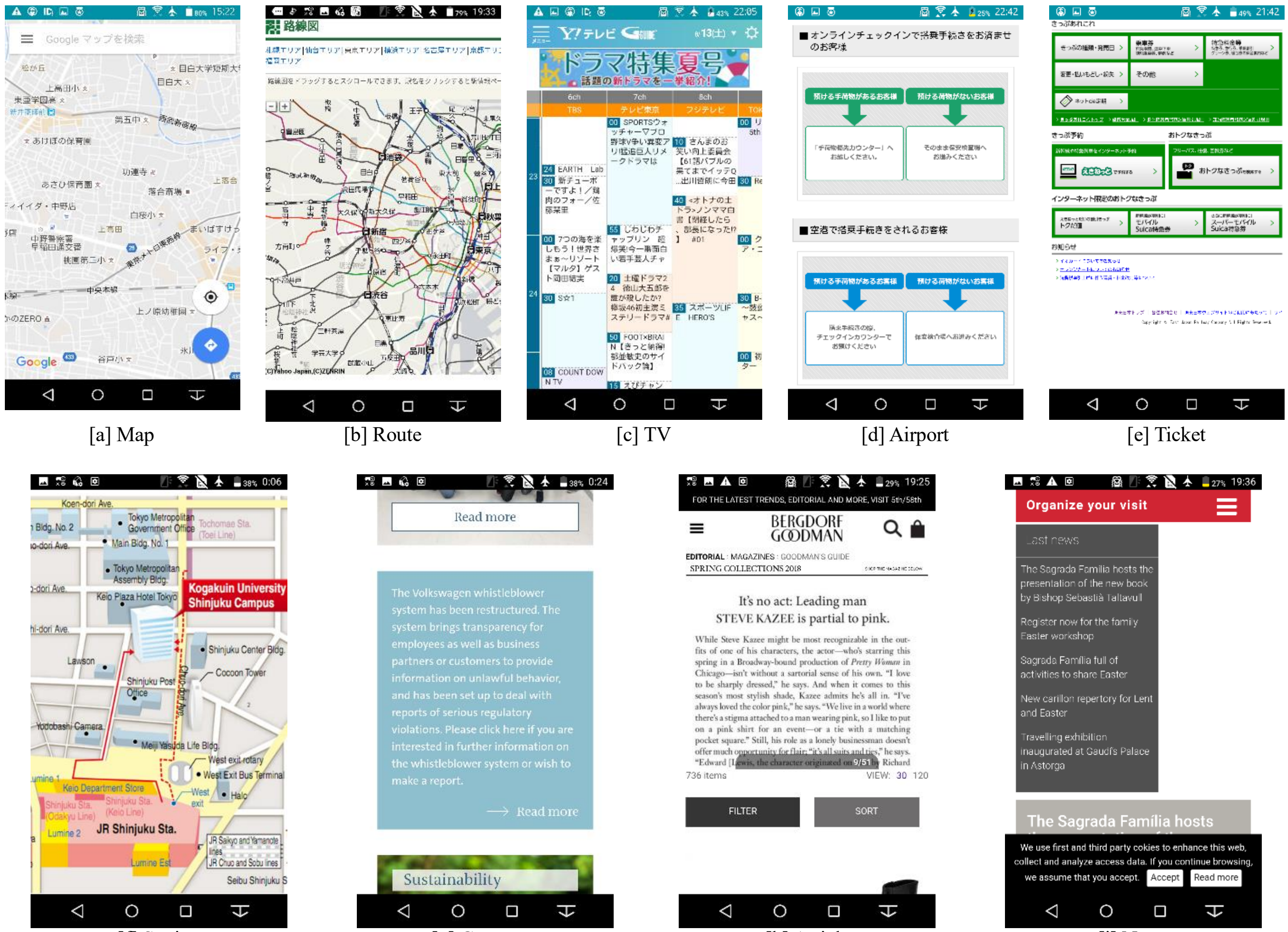
The observers were instructed on the experimental procedure, the meaning of resolution and the point of evaluation. Explanation of the resolution was conducted using training images to make the observers understand correctly. In addition, the observers were instructed not to consider the color, brightness, or noise of the image. When the observers purchase a smartphone, the viewing distance is different for each observer. Thus, observers could freely adjust the viewing distance. After evaluation, we investigated points where the observers gazed to judge whether the observers correctly evaluated differences in resolution.

B. *Test Images*

Nine screenshots of websites containing text were used as experimental images. Five images were Japanese text, and the others were English text. Japanese text images included hiragana, katakana, and Chinese characters. The images are shown in Figure 5. The resolution of all the images is WQHD. Figure 5 [a]-[e] shows the Japanese text images. Figure 5 [f]-[i] shows the English text images. The Japanese text images are of websites browsed by many people (a site for smartphones, a PC, a map). The site for smartphones is enlarged and viewed when the site has small text; therefore, an un-enlarged site image and two enlarged site images were used. One of the two enlarged images contained text with only small differences in color from the background color. Similarly, the three English text images are screenshots of websites for a smartphone and a map. When web articles written on PDF are browsed, the resolution is often low. Thus, one of the website screenshots for smartphones is a PDF article page.

C. *Observers*

At least 20 observers are required for adequate statistical analysis. In this experiment, 23 observers participated in the experiment and had normal visual acuity and color vision. Non-experts who do not work in the image industry cannot always distinguish image quality differences, even if experts can distinguish them. If there is a significant difference in the experiment using non-experts, the difference of image quality is great. Therefore, all the observers were non-experts.



[a] Map [b] Route [c] TV [d] Airport [e] Ticket

[f] Station [g] Company [h] Article [i] News

Figure 6. Test images

D. Experiment 1

In Experiment 1, NLSP was applied to Japanese text, and the subjective assessment to evaluate the image quality of NLSP for text was conducted. A smartphone with NLSP and a smartphone without NLSP, and different manufacturer's smartphones were used. This experiment shows that NLSP is more effective than the conventional SR technologies.

1) Experimental Equipment

Five smartphones were used in this experiment. To ensure that the results were not caused by display differences, two of the five smartphones featured the same terminal. One was a smartphone with NLSP (smartphone A), and the other was one without NLSP (smartphone B). The remaining three smartphones were smartphones from different manufacturers (smartphone C–E). The display resolution of smartphones A and B was WQHD (2560 × 1440), whereas that of the others was full HD (1920 × 1080). The brightness was adjusted to be close to the same brightness.

2) Experimental Method

The observers evaluated the image quality of the test image and ranked the five smartphones by resolution. The best-worst method was used in the experiment. First, the observers selected the best (1st rank) and the worst (5th rank) smartphones from the five smartphones. Second, the next best (2nd rank) and the next worst (4th rank) smartphones were selected in the same way from the remaining three smartphones. The remaining smartphone was ranked 3rd.

E. Experiment 2

In Experiment 2, English text images with and without NLSP were compared and evaluated.

1) Experimental equipment

Two smartphones were used to evaluate the image quality. These were the same terminal smartphones. One smartphone output images processed with NLSP to a display. The other output unprocessed images. When the image quality assessments were conducted using multiple displays, it was proven that an individual difference of displays did not affect the results [21].

The smartphones have 5.4 inch display, and the resolution is WQHD. The brightness was adjusted to be close to the same brightness on both devices.

The observers compared images displayed on smartphones, and chose the smartphone, which had the higher

TABLE I. Analysis result (Map)

l/k	R_l	f_{kl}					P_l	ε_l	$K_{\varepsilon l}$
		A	B	C	D	E			
1	5	22	1	0	0	0	90	0.1	1.28
2	4	1	2	3	17	0	70	0.3	0.52
3	3	0	9	9	2	3	50	-0.5	0.00
4	2	0	4	8	4	7	30	-0.3	-0.52
5	1	0	7	3	0	13	10	-0.1	-1.28
$\sum(f_{kl} \times K_{\varepsilon l})$		28.72	-8.74	-6.47	6.82	-20.33			
R_k		1.25	-0.38	-0.28	0.30	-0.88			
S_k^2		0.15	0.71	0.52	0.40	0.48			

resolution. To prevent prejudice from affecting the results, the state of NLSP (ON/OFF) was not revealed to the observers.

V. RESULTS

In this section, the results of the two experiments in the previous section are explained.

1) Experiment 1 results

The assessment results were analyzed, and the presence or absence of significant differences was identified. The assessment results were quantified, and the average scores representing the image quality of each stimulus were calculated [22]. The calculation requires a normalized score $K_{\varepsilon l}$, which can be calculated using P_l and ε_l. P_l is the average of each segment of the range from 0 to 100 separated into the number of stimuli. In this experiment, the number of stimuli, i.e., the number of smartphones (n), equals 5. The value ε_l is the median of each segment of the standard normal distribution separated into n segments. $K_{\varepsilon l}$ is the percentile of the standard normal distribution. Thus, $K_{\varepsilon l}$ is the distance from the average of the standard normal distribution. The values of $K_{\varepsilon l}$ were given as a normalized score according to rank. The average scores of the total score are the evaluation values for each stimulus.

The aggregate results of "Map" (Figure 3(a)) are shown in Table 1. The rows represent rank, and the columns represent stimuli (smartphones A–E). The values of intersection (f_{kl}) are the number of observers for stimulus k for rank l. Thus, f_{1A} indicates that 22 observers ranked the smartphone with NLSP (smartphone A) 1st.

First, rank is converted to a value. The higher the ranking, the higher the r_l value of the smartphone, where r_l is calculated as follows:

$$r_l = n - l + 1 \qquad (1)$$

The percentile values P_l are calculated using r_l as follows:

$$Pl = \frac{r_l - 0.5}{n} 100 \qquad (2)$$

The calculation results are shown in each row r_l, P_l of Table 1. Next, ε_l is calculated using (3) or (4). If the value of P_l is larger than 50, formula (3) is used. If the value of P_l is 50 or less, formula (4) is used. This is because the values of ε_l are calculated based on the point of the variance 0 of the standard normal distribution.

$$\varepsilon_l = \begin{cases} 1 - \dfrac{P_l}{100} & (P_l > 50) \quad (3) \\ \dfrac{P_l}{100} & (P_l \leq 50) \quad (4) \end{cases}$$

The calculation results are shown in row ε_l of Table I. $K_{\varepsilon l}$ is calculated using ε_l from the normal distribution table. The values of $K_{\varepsilon l}$ shown in Table I were given to each stimulus according to the ranking. The average scores (R_l) of the total scores ($\sum(f_{kl} \times K_{\varepsilon l})$) are the evaluation values of the stimulus. For example, the average score R_A is calculated as follows: $R_A = 28.72/23 \fallingdotseq 1.25$. The average scores and total scores are shown in Table 1. The average scores of "Map" (Figure 3(a)) are shown in the yardstick graph in Figure 4. The horizontal axis indicates the average score. The marks on the axis (oval, triangle, square, rhombus, and x) indicate the average scores of each stimulus (smartphone A, smartphone B, smartphone C, smartphone D, and smartphone E, respectively). The higher the average score, the higher the evaluation. In Table 1, the average score of smartphone A is the highest, indicating that smartphone A has the highest resolution.

A t-test was used to verify the significant difference between the stimuli. The variance of the average score (S_k^2) and the statistical quantity t_0 are calculated as follows:

$$S_k^2 = \frac{\Sigma\{fk_l \times (K\varepsilon_l)^2\}}{\sqrt{\Sigma(fk_l)}} - R_k^2 \qquad (5)$$

$$t_0 = \frac{R_x - R_y}{\sqrt{\sum(f_{kl})\,(S_x^2 + S_y^2)}} \sqrt{\sum(f_{kl}) \sum\{(f_{kl}) - 1\}} \quad (6)$$

The value $\sum(f_{kl})$ indicates the number of observers. x and y are stimuli. The calculation results are shown in Table 1. The values of t are calculated using the Degree of Freedom (DoF) from t distribution. In this experiment, the DoF is $\mathrm{DoF} = 2 * \sum(f_{kl}) - 2 = 46 - 2 = 44$. The t value of 1% significant level is $t_{1\%} = 2.414134$ and that corresponding to a 5% significant level is $t_{5\%} = 1.68023$. If the value of t_0 is larger than the value of $t_{5\%}$, there is a significant difference between stimuli.

Here, smartphone A is the highest, and smartphone D is the second highest. The t_0 value between smartphones A and D ($t_0(A, D)$) and the result of the t-test is as follows:

$$t_0(\mathrm{A, D}) = 10.33 > t_{1\%} \quad (7)$$

In (7), $t_0(A, D)$ is larger than $t_{1\%}$. This result indicates that smartphone A has a higher resolution than smartphone D and has a significance value of 1%. The results of the 3rd rank (smartphone C), 4th rank (smartphone B), and 5th rank (smartphone E) are as follows:

$$t_0(\mathrm{D, C}) = 4.13 > t_{1\%} \quad (8)$$

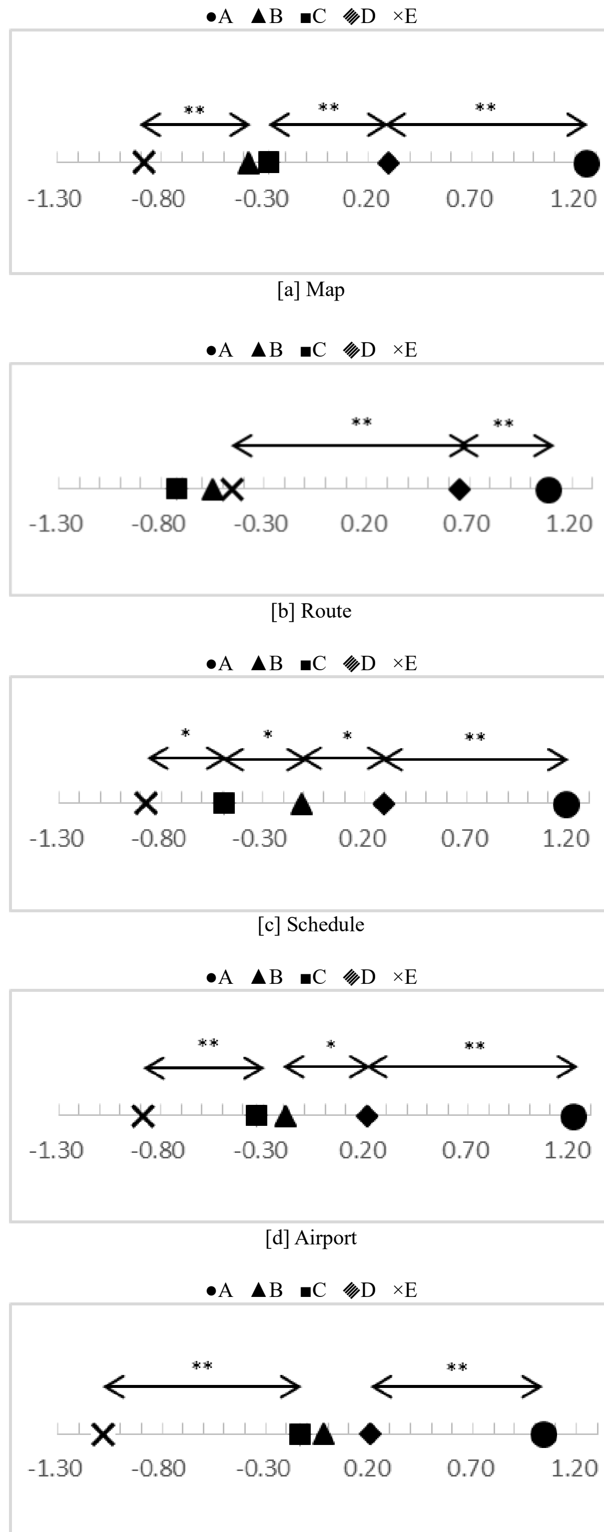
$$t_0(\mathrm{C, B}) = 0.53 > t_{1\%} \quad (9)$$

$$t_0(\mathrm{B, E}) = 2.77 < t_{5\%} \quad (10)$$

$t_0(D, C)$ and $t_0(C, B)$ are larger than $t_{1\%}$. Therefore, there are significant differences of 1% between smartphones D and C, and smartphones C and B. $t_0(B, E)$ is less than $t_{1\%}$ and $t_{5\%}$, indicating that there is no significant difference between smartphones B and E. The arrows indicate significant differences in the graph in Figure 4. The asterisks represent the level of significant difference between stimuli. "**" represents a significant difference of 1%, and "*" represents a significant difference of 5%. The analysis results of images [b–e] are shown in Figure 3 (b–e). Smartphone A has the highest resolution and significant differences of 1% between other smartphones in all the images. On the other hand, smartphone E has the worst resolution for all of the images and significant differences for four out of five images with the other smartphones.

2) Experiment 2 results

In Experiment 2, the observers compared the images processed with and without NLSP, and chose the smartphone, which had the higher resolution. We calculated ratio of each smartphones selected, and evaluated the statistical significant differences between the stimuli. In statistics, there are two important criteria about the significant difference. They are 95% and 99%. To obtain the 95% significant difference, at least 20 observers are required. If one out of twenty observers selects the smartphone with NLSP, the 95% significant difference between the stimuli is obtained. In contrast, if two observers select the smartphone with NLSP, the probability is 90%. In statistics, 90% does not indicate a significant difference. In this experiment, more than 20 observers participated. Thus, if 95% of the observers assess that the smartphone with NLSP has a higher resolution than the smartphone without NLSP, there is a significant difference of 95%.



[a] Map

[b] Route

[c] Schedule

[d] Airport

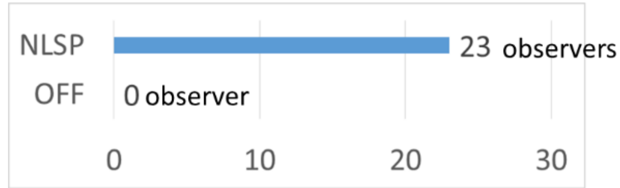
[e] Ticket

Figure 7. Assessment results (Experiment 1)

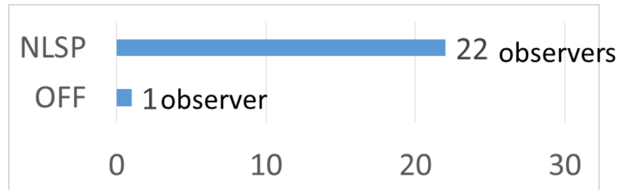
Figure 7 shows the results. The vertical axis represents the stimuli, the horizontal axis represents the number of observers. Here, the smartphone with NLSP is represented as

"NLSP," and the smartphone without NLSP is represented as "OFF."

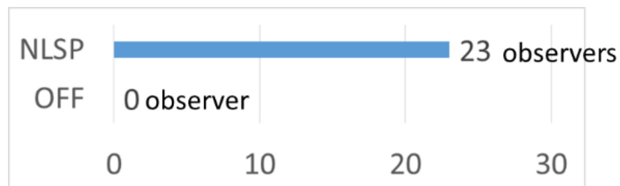
The results are explained using the result of "Station" (Figure 7[a]) and that of "Company" (Figure 7[b]). The graph of Figure 7[a] indicates that 23 observers selected NLSP as having high resolution, and no one selected OFF. The results show that all the observers, that is, 100% observers assessed that NLSP has a higher resolution than OFF. Therefore, NLSP has a higher resolution, and there is a significant difference of 99% between the stimuli. In Figure 7[b], 22 observers selected NLSP, and one observer selected OFF. More than 95% of observers selected NLSP, which indicates that the result has reproducibility of more than 95%. Figure 7 [c], [d] show the results of the other test images. The results show that NLSP has a higher resolution than OFF, and there are significant differences of more than 95% between the stimuli for all of the images.



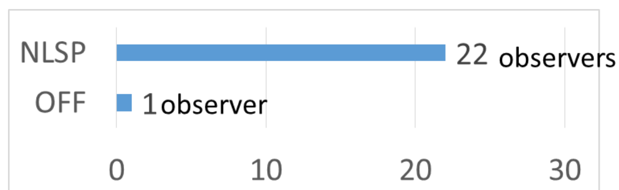
[a] Station



[b] Company



[c] Article



[d] News

Figure 8. Assessment results (Experiment 2)

VI. DISCUSSION

In Experiment 1, smartphone A (with NLSP) has the highest score and a significant difference of 1% between the other smartphones (which are either without NLSP or from different manufacturers) in all the images. The results indicate that NLSP is valid for text on smartphone displays. The same results were obtained for all the images. Thus, NLSP is valid for images other than the five images used in this paper. There are significant differences between smartphones without NLSP. It is assumed that the results were influenced by the internal processing differences.

In Experiment 2, the smartphone with NLSP has a higher resolution than the smartphone without NLSP, and there are statistical significant differences of 1% or 5% between the stimuli for all scenes. Significant differences were obtained in both Japanese and English texts containing characters with different features. Therefore, it is assumed that NLSP can improve the resolution of text on smartphone displays.

In this experiment, a gazing point was not specified for the observers. In addition, there were significant differences in all of the images when all the observers were non-experts. From the above, there are clear differences of image quality between the images with NLSP and those without NLSP.

The same results were obtained in Experiments 1 and 2. Therefore, different of effect according to language cannot be found. Although we cannot technically specify the font type of test images, bold letters may affect the subjective assessment results. However, it can be adjusted by parameter controls.

VII. CONCLUSIONS AND FUTURE WORK

Subjective assessments using smartphones with NLSP and smartphones without NLSP were conducted to verify the effectiveness of NLSP for texts. In Experiment 1, Japanese text including hiragana, katakana, and Chinese characters was used as test images. In Experiment 2, test images included English text images.

The results of Experiments 1 using five smartphones indicated that the image quality of a smartphone with NLSP is the highest, and there are significant differences between the other smartphones. In Experiment 2, the images with and without NLSP were compared using two smartphones. The results show that the smartphone with NLSP has a higher resolution than the smartphone without NLSP does, and there were significant differences for all test images.

From the above, it was proven that NLSP can improve the resolution of texts on smartphone displays.

The statistical analyses indicate that the experimental results are reproducible. The conclusion that a smartphone with NLSP has the highest image quality was obtained for all the images; therefore, both the assessment method and the analysis method in this experiment were valid as subjective assessment methods.

In future work, we will apply to more characters and type of fonts, such as bold letters, and verify the general performance. Although the NLSP has been implemented only to one model of smartphone, its processing can work regardless of the operating system. Therefore, our final target is to implement the NLSP on as many smartphone models as we can.

References

[1] A. Kubota and S. Gohshi, “Subjective Assessment for Text with Super Resolution on Smartphone Displays,” International Conference on Creative Content Technologies 2018 (CONTENT 2018), pp. 24-29, Feb. 2018.

[2] S. 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] http://www.fmworld.net/product/phone/f-02h/display.html?fmwfrom=f-02h_index. [retrieved: Nov. 2018].

[4] H. Shoji and S. Gohshi, “Subjective Assessment for Resolution Improvement on 4K TVs:Analysis of Learning-Based Super-Resolution and Non-Linear Signal Processing Techniques,” The Eleventh International Multi-Conference on Computing in the Global Information Technology (ICCGI2016), Vol. 2016, pp. 10-15, 2308-4529, Nov. 2016

[5] M. Sugie, S. Gohshi, H.Takeshita, and C. Mori,“Subjective Assessment of Super-Resolution 4K Video using Paired Comparison,”2014 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS-2014), pp. 42-47, Dec. 2014.

[6] C. Mori and S. Gohshi,“Image Quality of a Smartphone Display with Super-Resolution,”Proceedings of the ISCIE Inter-national Symposium on Stochastic System Theory and its Applications, Vol. 2016, pp. 340-344, 2016.

[7] ITU-R. BT.710-4, “Subjective Assessment Method for Image Quality in High-Definition Television,” Nov. 1998.

[8] H. Shoji and S. Gohshi, “Subjective Assessment for Learning-Based Super-Resolution and Non-Linear Signal Processing Techniques,” The 47th ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS’ 15), pp. 79-80, Dec. 2015.

[9] Y. Mukaigawa, K. Suzuki, and Y. Yagi, “Analysis of subsurface scattering under generic illumination,” 2008 19th International Conference on Pattern Recognition (ICPR), pp. 1-5, 1051-4651, Dec. 2008.

[10] S. Farsiu and M. Dirk Robinson, “Fast and Robust Multi-Frame Super-Resolution,” IEEE Trans Image Process 2004, Vol.13, no.10, pp. 1327-1344, 1941-0042, Oct. 2004.

[11] Toshiba (in Japanese), http://www.toshiba.co.jp/regza/detail/superresolution/resolution.html. [retrieved: Nov. 2018].

[12] S. H. Park, M. K. Park, and M. G. Kang, “Super-Resolution Image Reconstruction: A Technical Overview,” IEEE Signal Processing Magazine, 1053-5888/03, pp. 21-36, May. 2003.

[13] M. R. Banha, and A. K. Katsaggelos, “Digital Image Restoration,” IEEE Signal Processing Magazine, 1558-0792, Vol. 14 (2), pp. 24-41, Mar. 1997.

[14] A.W.M.Eekeren, K. Schutte, and L. J. van Vliet, “Multiframe Super-Resolution Reconstruction of Small Moving Objects,” IEEE Transactions on Image Processing, 1941-0042, pp. 2901-2912, Vol. 19, no. 11, Nov. 2010.

[15] S. Panda, M. S. R. S. Prasad, and G. Jena, “POCS Based Super-Resolution Image Reconstruction Using an Adaptive Regularization Parameter,” International Journal of Computer Science Issues, Vol. 8, issue 5, No. 2, 1694-0814, Sep. 2011.

[16] D. Glasner, S. Bagon, and M. Irani, “Super-Resolution from a Single Image,” International Conference on Computer Vision (ICCV), pp. 349-35, 2380-7504, Oct. 2009.

[17] J. Sun, Z. Xu, and H. Y. Shum, “Image Super-Resolution using Gradient Profile Prior,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 1063-6919, Jun. 2008.

[18] C. Dong, C. C. Loy, K. He, and X. Tanh, “Learning a Deep Convolutional Network for Image Super Resolution”, European Conference on Computer Vision (ECCV), Vol. 8692, pp. 184-199, 978-3-319-10593-2, Sep. 2014.

[19] S. Gohshi and I. Echizen, “Limitation of Super Resolution image Reconstruction,” The Journal of The Institute of Image Information and Television Engineers, Vol. 36, No. 48, ME2012-125, pp. 1-6, 978-9-8975-8129-8, Nov. 2012.

[20] W.T. Freeman, T.R. Jones, and E.C. Pasztorm, “Example-Based Super-Resolution,” IEEE Computer Graphics and Applications,” Vol. 22, No. 2, pp. 56-65, 1558-1756, Mar. 2002.

[21] C. Mori, M. Sugie, H. Takeshita, and S. Gohshi, “Method of Subjective Image Quality Assessment of Multiple Displays,” The Journal of the Institute of Image Electronics Engineers of Japan, Vol. 46, No. 2, pp. 315-323, Apr. 2017. (in Japanese)

[22] T. Fukuda and R. Fukuda, “Ergonomics handbook, ” Tokyo: Scientist press co.ltd, pp. 41-71, 2009. (in Japanese).