Subjective Assessment for Text with Super Resolution on Smartphone Displays

Aya Kubota† Seiichi Gohshi‡
†‡Department of Information Science
Kogakuin University
Tokyo, Japan
e-mail: †em17008@ns.kogakuin.ac.jp, ‡gohshi@cc.kogakuin.ac.jp

***Abstract*— Smartphones appeared on the market only a decade ago. 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 text can be improved. One research area for increasing the resolution is super resolution (SR), which includes nonlinear signal processing super- resolution SR (NLSP), a method that can be implemented on smartphones. However, NLSP has never been applied to text in order to improve readability. We applied NLSP for text displayed on liquid crystal display (LCD), and verified its effectiveness. Thus, in this paper, the assessment results for text on LCD are discussed.**

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

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, which could be on 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 texts is not very different. Problems, such as these can therefore make it difficult to read text.

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

However, the effectiveness of NLSP for text on smartphone display has not been verified. In this study, we verify the effectiveness of using smartphone with NLSP compared 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 the subjective assessments. The subjective assessments also requires observers and time to assess the image quality.

Although there are issues about the subjective assessment, ITU-R standardized subjective assessment methods. ITU-R BT.710 recommends experimental conditions to obtain reproducible results in subjective assessment experiments [6]. 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 our case we need to compare five smartphones and they are different manufactures products, BT.500 and BT.710 do not meet our requirements. One of our authors developed an subjective assessment for multiple displays [6][14]. 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 test images are presented and 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. ASSESSMENT METHOD

Objective assessment and subjective assessment are evaluation methods. Objective assessments analyze the signal and expresses high and low of image quality by a numerical value. However, results of 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 [7], but Figure 1(b) contains degradation in the form of a black square in the center of the image. When images

 ISBN: 978-1-61208-611-8

(a) Original image (b) Degraded image (PSNR: 40.1112dB)

Figure 1. Objective assessment by PSNR

include local degradation, the results of PSNR sometimes deviate from our feeling.

Thus, objective assessments cannot accurately reflect image quality. In addition, objective assessments require comparing the assessment image with the original image. As discussion 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, but 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 observers evaluate the stimuli based on the other. Thus, differences between stimuli can be obtained in detail. However, evaluation is performed for all 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, observers again select the best and the worst from the remaining stimuli. The best–worst method can detect differences more accurately than the normalized ranking method, and the best–worst method is a smaller burden for observers than the paired comparison method. Therefore, the best–worst method is adopted in this paper.

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

III. NLSP

NLSP is a simple and fast SR technique. 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 [8]. The nonlinear function can generate high-frequency components that are not included in the original image. These high-frequency components express edges and details of the image. An example nonlinear function is the cubic function ($f(x) = x^3$). The function can amplify the high-frequency components by as much as three times. Figure 2 shows an example of NLSP processed image. Figure 2(a) is an original image. Figure 2(b) is a NLSP processed image. Figure 2(b) has more details, such as edges of mountain and the surface of it than the original image.

Super-resolution image reconstruction (SRR) and learning-based super resolution (LBSR) are the current mainstream SR technologies. SRR is a technology that generates a high-resolution image from multiple degraded images [9], but the processing requires iteration. When the input image and output image have the same resolution, the technique is not very effective [10]. LBSR is a method that increases resolution using a database [11]. The effectiveness is affected by the database, and the processing requires both an expensive database and iteration. Thus, both the above technologies require complex processing. In addition, their effectiveness is lower than that of NLSP [12][13].

IV. Experiment

The effect of image processing differs, depending on the images. We adjusted NLSP for text; hence, it was necessary to verify the effect of NLSP for text. 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 verifies the effect of NLSP in comparison with other technologies.

A. Experimental equipment

Five smartphones were used in this experiment. To ensure that the results are 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 smartphone 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.

B. Test images

Five screenshots of websites containing text were used as experimental images. The images are websites browsed by many people (a site for smartphones, a PC, a map). The site for smartphones are enlarged and viewed when the site has small texts, so an unenlarged 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. The images are shown in Figure 3. The resolution of all the images is WQHD.

C. Observers

At least 20 observers are required for adequate statistical analysis. In this experiment, 23 observers participating in the experiment had normal visual acuity and color vision. Non-experts who do not work in the image industry cannot


(a) Original image


(b) NLSP processed image

Figure 2. Example of NLSP processed image


(a) Map


(b) Route


(c) TV


(d) Airport


(e) Ticket

Figure 3. Test images

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 large. Therefore, all observers were non-experts.

D. *Experimental method*

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, observers select 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 are selected in the same way from the remaining three smartphones. The remaining smartphone was ranked 3rd.

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 observers understand correctly. In addition, the observers were instructed not to consider the color, the brightness or noise of the image. When the observers purchase a smartphone, the viewing distance is different for each observer. Thus, the observers could freely adjust the viewing distance. After evaluation, we investigated points where the observers gazed to judge whether observers correctly evaluated differences in resolution.

V. 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 [14]. 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:

TABLE I. ASSESSMENT RESULTS (Figure 3(a) 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			

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

The percentile values P_l are calculated using r_l as follows:

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

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.

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

The calculation results are shown in row ε_l of Table 1.

$K_{\varepsilon l}$ is calculated using ε_l from the normal distribution table. The values of $K_{\varepsilon l}$ shown in Table 1 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, which indicates 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:

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

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

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 $\text{DoF} = 2 * \sum(f_{kl}) - 2 = 46 - 2 = 44$. 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:

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

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:

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

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

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

$t_0(D, C)$ 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 $t_{5\%}$, which indicates 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 in all images and significant differences for four out of five images with the other smartphones.

VI. DISCUSSION

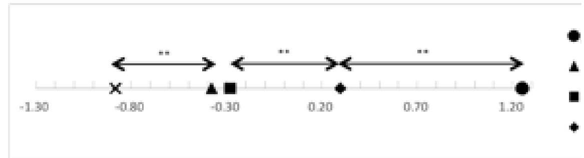
Smartphone A (with NLSP) has the highest score and a significant difference of 1% between the other smartphones (which are either without NLSP or from different manufacturers) in all the images. The results indicate that NLSP is valid for text on smartphone displays. The same results were obtained for 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 this experiment, a gazing point was not specified for the observers. In addition, there are significant differences in all the images when all the observers are non-experts. From the above, there are clear differences of image quality between images with NLSP and those without NLSP.

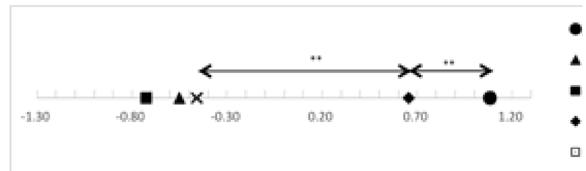
VII. CONCLUSIONS

Subjective assessments using smartphones with NLSP and those without NLSP were conducted to verify the effectiveness of NLSP for texts. The results of experiments 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.

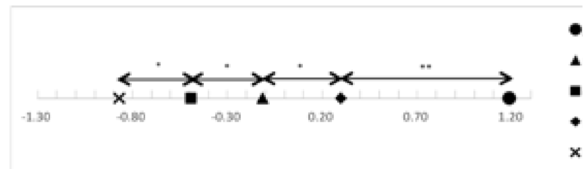
Statistical analyses indicate that the experimental results are reproducible. The conclusion that a smartphone with NLSP has the highest image quality was obtained for all images, therefore, both the assessment using the best-worst method and the analysis method in this experiment were valid as subjective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



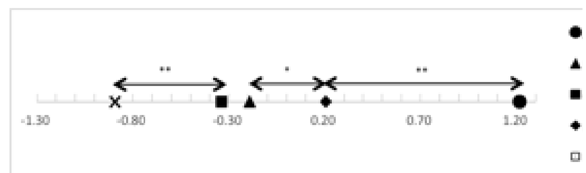
(a) Analysis result "Map"



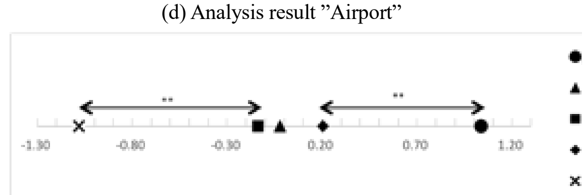
(b) Analysis result "Route"



(c) Analysis result "TV"



(d) Analysis result "Airport"



(e) Analysis result "Ticket"

Figure 4. Analysis results

REFERENCES

[1] 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.

[2] http://www.fmworld.net/product/phone/f-02h/display.html?fmwfrom=f-02h_index. [retrieved: Mar. 2017]

[3] 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), pp.10-15, Nov. 2016.

[4] M. Sugie, S. Gohshi, H. Takehisa, and C. Mori, "Subjective Assessment of Super-Resolution 4K Video using Paired Comparison," , 2014 International

[5] 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 International Symposium on Stochastic System Theory and its Applications, Vol.2016, pp.340-344, 2016.

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

[8] Manisha and P. Ahlawat, "Enhanced DNA Based Cryptography," International Journal of Advance Research in Computer Science and Management Studies, vol.3, Issue 6, pp.452-455, Jun. 2015.

[9] S. Gohshi, "A new signal processing method fo video: reproduce the frequency spectrum exceeding the Nyquist frequency," MMSys '12 Proceedings of the 3rd Multimedia Systems Conference, pp.47-52, Sep. 2012.

[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, Oct. 2004.

[11] 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, Nov. 2012.

[12] 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, Mar. 2002.

[13] 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.17-22, Dec. 2014.

[14] H. Shoji and S. Gohshi, “Subjective Assessment for Learning-Based Super-Resolution and Non-Linear Signal Processing Techniques,” The Eleventh International Multi-Conference on Computing in the Global Information Technology (ICCGI 2016), pp.10-15, Nov. 2016.

[15] T. Fukuda and R. Fukuda, “Ergonomics handbook,” ISBN978-4-86079-036-3, Scientist press co.ltd, pp.41-71, Tokyo, 2009. (in Japanese)