Image Quality of a Smartphone Display with Super-Resolution

Chinatsu Mori† Seiichi Gohshi‡
†‡ Dept. of Informatics, Kogakuin University
1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo, 163-8677, Japan
E-mail: † ed15002@ns.kogakuin.ac.jp, ‡ gohshi@cc.kogakuin.ac.jp

Abstract

The smartphone market has been growing rapidly and various smartphone content services are provided. However, the lack of bandwidth for the mobile network limits the smartphone video services to real-time. Most of the videos for smartphones on the internet are low-bitrate and blurring. In our previous works, we proposed a new SR technology that uses nonlinear signal processing (NLSP) and we demonstrated its theoretical capability and image quality for 4K TV sets. We applied NLSP to a smartphone. However, the capability of the NLSP on the smartphone display has not been tested. A subjective assessment using best-worst method was performed to verify the performance of NLSP on the smartphone display. The assessment was conducted by comparing the image qualities of the smartphones with and without NLSP and different manufacturers’ smartphones. Assessment data were statistically analyzed, and the assessment results proved that the image quality of the smartphone with NLSP is most superior in resolution.

1 Introduction

Smartphones are currently available on the market, and there are various content services provided for the smartphones. The mobile network is not sufficient for real-time video services due to the lack of bandwidth, and most of the videos for smartphones on the internet are low-bitrate. While a video coding process, the video data are compressed by a reduction of the high frequency components in the videos. This process causes blurring in the videos and the low-bitrate videos for the smartphones are blurring.

Image quality is an important factor for consumers to use the video devices. Display resolution of the video devices typically represents an index value of its image quality, and manufacturers sold the smartphones with the high-resolution display, such as full-HD and 4K. However, improving the image quality of the content is also required.

Super-resolution (SR) [1][2] is one of the technologies to improve the image quality. Unlike the classical image enhancement, SR can create the high frequency components not existing in the original signal. SR was introduced in TV sets [1][2].

In our previous works, an SR using nonlinear signal processing (NLSP) was proposed, and its image quality for 4K TV sets was also been proven [3][4][5][6]. We applied NLSP to a smartphone. However, the capability of SR on smartphone displays, which are smaller than TV sets, has not been proven. Display sizes of the smartphones sold in 2014 are about 4.5-5.5 inches and it means a tenth of conventional TV sets. There is no evidence that SR can improve the image quality on smartphone displays. Thus, the evaluation of NLSP on smartphone displays is necessary and we compare the image qualities of the smartphones with and without NLSP, and existing manufacturers’ smartphones.

While the evaluation of the capability of SR is required, there is no standard method for the evaluation. The signal analysis in the frequency domain is a way to analyze the image resolution. However, the SR processed signal is only sent to the display of the devices, and there is no way to take the video signals from the smartphone devices. Since the signal analysis cannot be used, a subjective assessment is the only way to evaluate the performance of SR embedded in the video devices.

Subjective assessments are human psychophysical experiments. Numerous subjective assessment methods were reported in the areas of psychology and psychophysics, and they were applied to various purposes [7][8][9]. Our previous works, we applied Scheffe’s paired comparison to assess an image quality of SR, and proved the usefulness of the assessment [7][5][6]. However, the assessment using paired comparison has an issue that becomes extremely time-consuming and fatiguing to the assessors in proportion to the number of the stimuli. In other method, a ranking method is the quicker assessment method [8] and this method was applied to an image quality assessment of CR images [10]. Nevertheless, the ranking method is inclined to cause dispersion of the assessors’ opinions.

To overcome these problems, we anew introduce another assessment method, best-worst method [9]. This method can contribute a simple, quick procedure and high accuracy data collection. In this paper, the assessment method is described and we also verify the
usefulness of this method, including the image qualities of the smartphone displays.

2 NLSP

In this section, our SR algorithm is explained.

Various SR algorithms have been proposed and some of them are equipped on TV sets [1][2]. Most are very complex and require special hardware devices to work in real time. Application of them to smartphones is impossible because SR has to work over software.

Most SR methods are based on linear processing theories and pattern matching methods. Our SR method uses non-linear signal processing [3][4]. Figure 1 shows the signal flow of NLSP. The process is separated from the input image. First process including the high pass filter, the cubic function, and the limiter, creates the high frequency components. The high pass filter detects edges in the input image. The cubic function is one of the non-linear functions, and outputs \( y = x^3 \). This function generates the third harmonic waves of the edges detected by the high pass filter. The waves are generated only from the detected edges, and have high frequency components the input image does not have. The output of the cubic function becomes very large. If the input image has 8 bit depth, the range of the output of the high pass filter is from -255 to 255, and thus that of the cubic function is from \((-255)^3\) to \((255)^3\). The limiter fits the waves to the image. In the second process, the output of the limiter and the input image are added by the adder. The output image has high frequency components not existing in the original input image. The basic algorithm of NLSP is the same as Enhancer except the cubic function. This algorithm can improve the resolution Enhancer cannot, and can work in real time.

3 Subjective Assessment Method

Our subjective assessment using best-worst method is explained in this section.

The five smartphones (A, B, C, D, and E) shown in figure 2 were used as stimuli for the assessment. The four smartphone models sold by different manufacturers (a, b, c, and d) were selected, and NLSP was applied to the manufacturer a’s model. The manufacturers and the state of NLSP are described in table 1. The smartphones A and B are the same models of the manufacturer a. The NLSP processed video can be output onto the display by the setting of SR processing for NLSP to be enabled (ON). If the setting of it is disabled (OFF), the unprocessed video signal is output. The smartphones C, D, and E are respectively different manufacturers’ models. Display resolution of the five smartphones is full-HD (1920 \( \times \) 1080). The assessment was conducted by comparing the image qualities when full-HD videos are displayed. The smartphones were respectively ranked according to the image quality in resolution by using best-worst method.

3.1 Best-Worst Method

The procedure of best-worst method is a repetition of a choice of a pair of the best and worst objects from a set. The assessment process of best-worst is as below.

1. An observer selects the best and worst (highest and lowest resolution) smartphones from the five smartphones. The best smartphone is of the 1st rank, and the worst smartphone is of the 5th rank.
2. The observer selects the best and worst smartphones in the same way from the remaining three smartphones. The best smartphone is of the 2nd rank, and the worst smartphone is of the 4th rank. Then the last remaining smartphone is of the 3rd rank.

4 Experiment

In this section, the experimental conditions are given.

4.1 Apparatus

The smartphones for the assessment were flagship models that each manufacturer sold in 2014. They have 5.1-5.5 inch displays that the resolution is full-HD (1920 \( \times \) 1080). The same video sequence was repeated between the smartphones, and the observers compare the image qualities.

Table 1: Stimuli

<table>
<thead>
<tr>
<th>Smartphones</th>
<th>Manufacturers</th>
<th>NLSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>ON</td>
</tr>
<tr>
<td>B</td>
<td>a</td>
<td>OFF</td>
</tr>
<tr>
<td>C</td>
<td>b</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>E</td>
<td>d</td>
<td>-</td>
</tr>
</tbody>
</table>
4.2 Test Sequences

We selected five test sequences for the assessment [11]. The length of each test sequence is 10-15 second. The test sequences were taken by a consumer digital video camera with MPEG-4 format and HDTV (1080p) resolution, and they have very high frequency elements without camera works, such as tilting and panning. The test sequences are shown in figure 3 (1)-(5). The ovals on figure 3 (1)-(5) indicate high-resolution areas that are easy to assess the resolution. Test sequence “Basilica” (figure 3 (1)) has details in a roof, a surface of a pillar, and characters of a wall. Test sequence “Market” (figure 3 (2)) has details in a stone pavement, window frames, and a surface of a house wall. Test sequence “Square” (figure 3 (3)) has details in a surface of a building, a fence, and a mark of the building. Test sequence “Castel 1” (figure 3 (4)) has details in a surface of a brick building, and a stone pavement. Test sequence “Castel 2” (figure 3 (5)) has details in two statues and a bilateral surface of a building. These areas were presented to the observers.

4.3 Observers

23 observers participated in the experiment. They are non-video experts who do not work in the video industry and they have normal visual acuity and color vision [11].

4.4 Experimental Condition

The experiment was conducted in a room with general lighting conditions. Prior to each assessment, a training session was held for the observers to understand the meaning of resolution. The experimental procedure and the evaluation points were properly provided to the observers using a dummy video sequence.

Observers take smartphones in their hands and compare the image resolution. The assessment areas in the test sequences were not specified. The two-four high-resolution areas of the ovals on figure 3 (1)-(5) were presented to the observers as the examples of the assessment areas. The viewing distance during the assessment was not specified to reproduce the viewing environment because it depends on each user. The observers were asked to assess only resolution, including saturated-highlights. Saturated-highlights mean the loss of details that occurs in the white areas in the image with excessive luminance level. Other image quality factors, such as noise, luminance, and color, were not considered in the assessment. Figure 4 shows a photograph of the experiment.

5 Results

In this section, the analysis process and the assessment results are discussed.

Using best-worst method in 3.1, the ranks (1st-5th) were given to each stimulus (smartphones A, B, C, D, and E) as the evaluation value. (E.g. the smartphone A ranks 1st, the smartphone B ranks 3rd, the smartphone C ranks 2nd, the smartphone D ranks 4th, and the smartphone E ranks 5th.) The rank data were normalized by the distance scale since the average values of the rank do not have mathematical meanings. The normalized scores corresponding to each rank were given to the stimuli and the average scores for each stimulus were calculated [12]. The statistical differences between the average scores were evaluated by t-test.

The assessment result for “Basilica” sequence is shown as table 2. The row \( l \) indicates the ranks (1-5), and the column \( k \) indicates the stimuli (A-E). The values of \( f_{kl} \) are the numbers of the observers for the stimulus \( k \) as the rank \( l \). (E.g. \( f_{C2} \) (15) means that 15 observers ranked the smartphone C as the 2nd.)

The normalized distances for each stimulus were calculated as follow. The column \( r_l \) in table 2 represents the score corresponding to the rank \( l \). The number of the stimuli (5) is represented as \( n \). The column \( P_l \) indicates the mean value of the \( l \)th range, which is divided from the 0-100 range into \( n \) ranges. The column \( K_{cl} \) is the normalized quantile of \( P_l \) on the Gaussian distribution.
and it was used as the normalized score for the rank \( l \). The normalized scores were given to each stimulus. The row \( \Sigma(f_{kl} \times K_{\epsilon l}) \) is the total of the normalized scores for the stimuli \( k \). The row \( R_k \) means the average scores and the row \( S_k^2 \) means the variances for the stimuli \( k \). Each average score represents the scale value of the image quality in resolution.

Figure 5 shows the average scores for “Basilica” sequence. The horizontal axis indicates the average scores of the stimuli respectively (smartphone A, smartphone B, smartphone C, smartphone D, and smartphone E). The higher the averages indicate the higher the evaluation of resolution. The significance test is necessary because the differences in the average scores are not guaranteed.

Here, t-test was used to evaluate the statistical differences between the average scores of the stimuli. A statistics quantity \( t_0 \) for t-test is as follows:

\[
t_0(x, y) = \frac{(R_x - R_y)\sqrt{N - 1}}{\sqrt{S_x^2 + S_y^2}}
\]

where \( x \) and \( y \) are stimuli, where \( N \) is the number of observers (23). The degree of freedom of the distribution for the \( t_0 \) is \( \phi = 2N - 2 = 44 \). A critical \( t \) value for the 0.01 significance level with the degree of freedom \( \phi \) is \( t_{1\%} = 2.41413 \), and that for the 0.05 significance level is \( t_{5\%} = 1.68023 \) [12]. If \( t_0(x, y) \) is greater than the critical \( t \) values, the null hypothesis was rejected and there is the statistical difference between the average scores of the stimuli \( x \) and \( y \).

In the result of “Basilica” sequence as shown figure 5, the highest smartphone is A, and the second highest smartphone is C. The \( t_0 \) value for the smartphone A and the smartphone C \( (t_0(A, C)) \) is as follows:

\[
t_0(A, C) = 4.35831 > t_{1\%}
\]

Since equation 2 satisfies \( t_0 > t_{1\%} \), the statistical difference with the 0.01 significance level between the smartphones A and C is obtained. In addition, the 3rd highest smartphone is B, the 4th highest smartphone is D, and the lowest smartphone is E.

### 6 Discussion

The average score of the smartphone A with NLS is higher than the smartphone B, without NLS, and the other manufacturers’ smartphones. It implies NLS can improve the image quality in resolution, and the image quality of the smartphone with NLS is superior to the other manufacturers’ smartphones. The assessment results prove the usefulness of the image quality of NLS on smartphone displays. There are statistical significance differences between all the stimuli. The asterisks (\(*\)/\(*\)) in figure 5 represent the statistical difference between the stimuli indicated by the arrow. The 0.01 significance is shown as “\(*\)”, and the 0.05 significance is shown as “*”. The results for the other test sequences are shown as “\(*\)”. The statistical differences are not shown as “\(*\)”.

Equation 3 does not satisfy \( t_0 \), but satisfies \( t_0 > t_{5\%} \). Thus, there is the statistical difference between the smartphones C and B with the 0.05 significance level. Equations 4 and 5 satisfy \( t_0 > t_{5\%} \). Thus, the statistical differences between the smartphones B and D, and the smartphones D and E, are obtained with the 0.01 significance level. There are the statistical differences between all the stimuli. The smartphones (\(*\)/\(*\)) in figure 5 represent the statistical difference between the stimuli indicated by the arrow. The 0.01 significance is shown as “\(*\)”, and the 0.05 significance is shown as “*”. The results for the other test sequences are shown in figures 6-9. All results have the similar tendencies. The averages are high in the order of A, C, B, D, and E. The average score of the smartphone A equipped with NLS is highest of the five smartphones for all cases.

### Table 2: Analysis result (Basilica)

<table>
<thead>
<tr>
<th>( l \setminus k )</th>
<th>( f_{kl} )</th>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
<th>( D )</th>
<th>( E )</th>
<th>( r_l )</th>
<th>( P_l )</th>
<th>( \epsilon_l )</th>
<th>( K_{\epsilon l} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>90</td>
<td>0.1</td>
<td>1.28155</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>6</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>70</td>
<td>0.3</td>
<td>0.52440</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td>12</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>50</td>
<td>0.5</td>
<td>0.00000</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>30</td>
<td>0.3</td>
<td>-0.52440</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>10</td>
<td>0.1</td>
<td>-1.28155</td>
</tr>
<tr>
<td>( \Sigma(f_{kl} \times K_{\epsilon l}) )</td>
<td>24.874</td>
<td>2.330</td>
<td>9.905</td>
<td>-8.915</td>
<td>-28.194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R_k )</td>
<td></td>
<td>1.081</td>
<td>0.101</td>
<td>0.431</td>
<td>-0.388</td>
<td>-1.226</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( S_k^2 )</td>
<td></td>
<td>0.199</td>
<td>0.181</td>
<td>0.291</td>
<td>0.077</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( n \): Number of stimuli
\( N \): Number of observers
\( r_l = n - l + 1 \)
\( P_l = \frac{(r_l - 0.5)}{100} \times 100 \)
\( \epsilon_l = \begin{cases} n - P_l & (P_l > 0.05) \\ 100 - P_l & (P_l \leq 0.05) \end{cases} \)
\( K_{\epsilon l} = \begin{cases} \Phi^{-1}(\epsilon_l) & (P_l > 0.05) \\ -\Phi^{-1}(\epsilon_l) & (P_l \leq 0.05) \end{cases} \)
\( \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \)
\( R_k = \frac{\Sigma(f_{kl} \times K_{\epsilon l})}{N} \)
\( S_k^2 = \frac{\Sigma(f_{kl} \times K_{\epsilon l}^2)}{N} - R_k^2 \)
7 Conclusion

In this paper, NLSP was applied to a smartphone, and the subjective assessment of smartphone displays was conducted. Best-worst method was used for the assessment. The results were statistically proven that the image quality of the smartphone with NLSP is superior in resolution to the other smartphones without NLSP.

The assessment method described in this paper is applicable to other assessments. It is necessary to consider other factors for NLSP image quality, such as noise and color.

References