Dictionary-Based interpolation technique for text quality enhancement

Sung In Cho and Suk-Ju Kang

Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea

ABSTRACT
This paper presents a new dictionary-based interpolation technique to improve text quality when the resolution of an image is increased, so that the text can be displayed on a high-resolution display device. The proposed algorithm analyzes an image and extracts the shape of the text from the image. Further, it encodes and decodes the pattern of the text, and enhances the legibility of the text using a pre-trained code dictionary. Therefore, the proposed method improves text quality in terms of sharpness. In the experiments, the proposed algorithm outperformed benchmark methods for all test images. Specifically, the proposed method reduced the blur index by up to 0.112.

ARTICLE HISTORY
Received 18 May 2020
Accepted 23 September 2020

KEYWORDS
Index terms: interpolation; image up-scaling; text enhancement; resolution enhancement

1. Introduction
The resolution of displays, such as the liquid crystal display (LCD) and the organic light emitting diode (OLED) display, has been increasing in response to demands for clearer and more vivid images. High resolution is now possible because of the improvement of electronic devices and display manufacturing processes. For example, the improvement of television systems has led to an increase in their resolution from high definition (HD) to full high definition (FHD), and recently, to ultra-high definition (UHD). However, even if a display device has a UHD resolution, the increase in the resolution of the image contents requires the replacement of broadcasting equipment and an increase in the transmission bandwidth. To solve this problem, image up-scaling, a technique that increases the resolution of an image using two-dimensional interpolation, is used. Figure 1 shows an example of two-dimensional interpolation in a case wherein the resolution of the display device is twice higher than that of the image content. In this case, new pixels are generated using an interpolation technique based on either the current pixel or the current and neighboring pixels. The performance of the up-scaling system depends greatly on the interpolation performance.

The basic interpolation techniques are bilinear and bicubic interpolations [1–4]. These methods use linear interpolation, and hence, require little computation. However, they do not exactly produce the interpolated pixel in the edge area, where the luminance changes non-linearly, because a new pixel is generated by linearly connected pixels. To solve this problem, several interpolation techniques have been proposed. The methods used in [5] and [6] are based on fitting local fifth-order polynomials. They consider a small number of neighboring pixels, and hence, have low computational complexity. The method in [7] uses the raised cosine pulse for interpolation. It has several advantages, such as fast implementation using the fast Fourier transform (FFT) and simple hardware architecture using a finite impulse response (FIR) filter. The methods in [8] and [9] use polynomials matched by the values of the function at various points and those of the first-order derivatives, thus achieving a higher degree of continuity. Conventional interpolation methods focus on improving the smoothness among neighboring pixels for generating an interpolated pixel. However, if these methods are used to increase the resolution of a text image, a blurring artifact at the boundary of the text occurs. When the width of a text string is thin and the background color is uniform, such as in web pages, the legibility of text can be seriously degraded.

In this paper, a dictionary-based interpolation technique is proposed to improve the text quality when the width of a text string is thin and the background color is uniform, such as in web pages. The proposed method separates the text from the background of the image. Specifically, it analyzes the shape of the text and changes such shape to enhance the legibility of the text using the
2. Proposed method

Figure 2 shows the entire architecture of the proposed up-scaling system. First, the RGB input image is converted to a YCbCr image. Next, Y signals are encoded into serial bits by blocks using the majority recognition module and the block encoding module. Based on the pre-processed code, the input code is classified and the most suitable match in the dictionary is selected. Finally, interpolation is performed after the selected bits are decoded.

2.1. Encoding stage

In this stage, the shape of the text is encoded after the text image is extracted from an input image. Specifically, using the Y signal as an input, a luminance histogram is generated for pixels in the current and neighboring blocks (as shown in Figure 3), and the highest majority, i.e. the background luminance, which is defined in (1), is extracted.

\[ l_{maj} = \arg \max_{0 \leq l \leq 2^N} \{ \text{Hist}(l_i) \}, \]  

where \( l_i \) denotes the luminance of the \( i \)-th pixel and \( l_{maj} \) denotes the luminance of the majority of the pixels, i.e. the background luminance. \( N \) stands for the number of bits used to represent the pixel luminance. In this case, eight bits were used.

Based on the luminance of the background, the shape of the text for a current block is encoded as shown in
Figure 3 according to (2).

\[
C = \begin{cases} 
1, & \text{if } |l_{maj} - l_{in}| < M, \\
0, & \text{otherwise},
\end{cases} \quad (2)
\]

where \( l_{in} \) denotes the luminance of an input pixel and \( M \) denotes the acceptable margin between which the luminance may fluctuate due to noise. \( M \) is set experimentally to eight. \( C \) stands for the one-bit code.

Using (2), the nine-bit code (\( \text{Code}_{in} \)) is extracted for the \( 3 \times 3 \) block, as shown in Figure 4(a). For example, in Figure 4, the luminance of the background and the luminance of the text are white and gray, respectively. Hence, \( l_{maj} \) and \( l_{text} \) are 255 and 100, respectively. Therefore, in the case of the text, the difference between its luminance and the background luminance is higher than \( M \), and the one-bit code is 1. In the case of the background, the difference between its luminance and the luminance of the text is lower than \( M \), and the one-bit code is 0. For the \( 3 \times 3 \) block, encoding is performed from the top-left to the bottom-right, and 100 010 001 is extracted, as shown in Figure 4. Subsequently, this code and the luminance of the text enter the decoding stage.

2.2. Decoding stage

In the decoding stage, the input code is analyzed and the legibility of the text is improved by changing its shape. For this, a code dictionary that contains pre-trained codes is used. Specifically, in the code dictionary, an image with a low resolution (LR) is paired with an image with a high resolution (HR), as shown in Figure 4(b). The pairing of the LR and HR images in the dictionary is defined by a user to improve the text legibility. If the code dictionary receives the input code (\( \text{Code}_{in} \)), as shown in Figure 4, the same code is searched for in the dictionary and the code (\( \text{Code}_{out} \)) for the HR image is selected.

Next, the interpolation module converts the code for the HR image into the luminance of the output image, as shown in (3).

\[
l_{out} = \begin{cases} 
l_{text}, & \text{if } C = 1, \\
l_{maj}, & \text{otherwise},
\end{cases}
\quad (3)
\]

where \( l_{text} \) and \( l_{out} \) denote the luminance of the pixel of the text and the final luminance of the output image, respectively. For example, in Figure 4, when \( \text{code}_{in} \) is 100 010 001, the same code is searched for and \( \text{code}_{out} \), which is 110000 111000 011100 001110 000111 000011, is outputted. Next, the interpolation module decodes this code into the luminance, and accordingly, the boundary of a diagonal line is improved, as shown in Figure 4(c).

3. Experiment results

The performance of the proposed method was evaluated subjectively and objectively. First, the text qualities of sample text images constructed using the proposed and benchmark methods were visually compared. Second, the text qualities of the proposed and benchmark methods were assessed in terms of their blur, using a perceptual blur metric [10]. Third, the computation times of the proposed and benchmark methods were calculated. The interpolation algorithms in [5] (method 1), [7] (method 2), [9] (method 3), and [4] (method 4) were used as the benchmark methods. For the test images, sample text images in several languages, i.e., English, Japanese, Chinese, German, and Korean, were used. Each language
Figure 4. Encoding and decoding in the proposed method for the current block in Figure 3: (a) the input code and the image with a low resolution, (b) the code dictionary, and (c) the output code and the image with a high resolution.

category consisted of 20 images. Images with an HD resolution (1280 × 720) were used, and their resolutions were converted to WQHD (2560 × 1440).

For the subjective evaluation, Figure 5 shows samples of the English and Japanese text images. The benchmark methods (methods 1, 2, 3, and 4) used neighboring pixels to enhance smoothness in the boundaries, and hence, blurred the boundaries of the text. Methods 3 and 4 used a large number of neighboring pixels, and the blur is more marked than that in methods 1 and 2, as shown in Figure 5. On the other hand, the proposed dictionary-based interpolation method considered not only the smoothness but also the sharpness of the boundaries. Therefore, the blur was significantly reduced, unlike in the benchmark methods. Additionally, Figure 5 shows that the staircase effect of the text was reduced, unlike in the benchmark methods. This is because the proposed method, using the pre-trained code dictionary, performed interpolation only when the boundaries had to be improved.

For the objective evaluation, the blur indexes of the proposed and benchmark methods were assessed over the blur index of all the test images. The blur indexes ranged from 0 to 1, where a higher index indicates poorer image quality with respect to blur. Figure 6 shows graphs of the blur indexes of all the test images. The blur index of the proposed dictionary-based interpolation is lower than the blur indexes of the conventional methods for all the test images. In the case of the languages with many strokes, i.e., Chinese, Japanese, and Korean, the blur reduction effect was enhanced in the proposed method. Table 1 lists the average blur indexes of the proposed and benchmark methods. The average blur indexes of methods 1, 2, 3, and 4 were 0.072, 0.069, 0.114, and 0.162, respectively. On the other hand, the blur index of the proposed method was 0.058. Specifically, the average blur indexes of the proposed method were up to 0.021, 0.017, 0.067, and 0.112, which are lower than those of methods 1, 2, 3, and 4, respectively. This is because the proposed method could maintain the sharpness of the
Figure 5. Comparison of the subjective qualities of the sample test images of English (top) and Japanese (bottom) text using (a) method 1, (b) method 2, (c) method 3, (d) method 4, and (e) the proposed method.

Figure 6. Blur indexes of the proposed and benchmark methods for the sample test images of: (a) English text, (b) Chinese text, (c) Japanese text, and (d) Korean text.
Table 1. Blur indexes of the proposed and benchmark methods.

| Test image | Method 1 | Method 2 | Method 3 | Method 4 | Proposed |
|------------|----------|----------|----------|----------|----------|
| Image Sequence(# of frames) | German | 0.055 | 0.00 | 0.053 | 0.00 | 0.099 | −0.05 | 0.151 | −0.10 | 0.050 |
| | Chinese | 0.059 | −0.01 | 0.057 | 0.00 | 104 | −0.05 | 0.154 | −0.10 | 0.053 |
| | English | 0.10 | 0.050 | 0.151 | −0.07 | 0.175 | −0.11 | 0.063 |
| | Japanese | 0.079 | −0.02 | 0.076 | −0.02 | 0.120 | −0.06 | 0.165 | −0.11 | 0.060 |
| | Korean | 0.081 | −0.02 | 0.079 | −0.02 | 0.120 | −0.06 | 0.163 | −0.10 | 0.063 |

Table 2. Computation times of the proposed and benchmark methods.

| Computation Time (ms) |
|-----------------------|
| Method 1 | Method 2 | Method 3 | Method 4 | Proposed |
| 19.50 | 8.60 | 42.50 | 3.94 | 4.73 |

4. Conclusion

In this paper, a proposal was presented for a dictionary-based interpolation for text quality enhancement when low-resolution image content is up-scaled to a high-resolution display device. The proposed algorithm extracts the shape of the text and the background luminance from an image, and then adaptively applies interpolation to improve the text legibility using a pre-trained code dictionary. The proposed method improves the text quality in terms of sharpness, which was experimentally validated using various test images. In these experiments, the proposed method reduced the blur index by up to 0.112, higher than did the benchmark methods.

Acknowledgment

This research was supported by the Ministry of Science and ICT (MSIT) of Korea under the Information Technology Research Center support program (IITP-2020-2018-0-01421) and supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP) and the National Research Foundation of Korea grant funded by the Korean government (MSIT) (No. 2020M3H4A1A02084899). Suk-Ju Kang received a B.S. degree in Electronic Engineering from Sogang University, Republic of Korea, in 2006, and a Ph.D. degree in Electrical and Computer Engineering from Pohang University of Science and Technology in 2011. From 2011 to 2012, he was a Senior Researcher at LG Display, where he was also a project leader for resolution enhancement and multi-view 3D system projects. Then, from 2012 to 2015, he was an Assistant Professor of Electrical Engineering at the Dong-A University, Busan. He is currently an Associate Professor of Electronic Engineering at Sogang University. His current research interests include image analysis and enhancement, video processing, multimedia signal processing, circuit design for display systems, and deep learning systems.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Ministry of Science and ICT (MSIT) of Korea under the Information Technology Research Center support program (IITP-2020-2018-0-01421) and supervised by the Institute of Information & Communications Technology Planning & Evaluation (IITP) and the National Research Foundation of Korea grant funded by the Korean government (MSIT) (No. 2020M3H4A1A02084899).

Notes on contributors

Sung In Cho (member, IEEE) received a B.S. degree in Electronic Engineering from Sogang University, South Korea in 2010, and a Ph.D. degree in Electrical and Computer Engineering from Pohang University of Science and Technology in 2015. From 2015 to 2017, he was a Senior Researcher at LG Display; and from 2017 to 2019, an Assistant Professor of Electronic Engineering at Daegu University. He is currently an Assistant Professor of Multimedia Engineering at Dongguk University, Seoul. His research interests include image analysis and enhancement, video processing, multimedia signal processing, and circuit design for LCD and OLED systems.
Suk-Ju Kang (member, IEEE) received a B.S. degree in Electronic Engineering from Sogang University, South Korea in 2006, and a Ph.D. degree in Electrical and Computer Engineering from Pohang University of Science and Technology in 2011. From 2011 to 2012, he was a Senior Researcher and a Project Leader for resolution enhancement and multi-view 3D system projects at LG Display; and from 2012 to 2015, an Assistant Professor of Electrical Engineering at Dong-A University, Busan. He is currently an Associate Professor of Electronic Engineering at Sogang University. In 2019, he received the Young IT Engineer of the Year Joint Award from IEIE/IEEE. His research interests include image analysis and enhancement, video processing, multimedia signal processing, circuit design for display systems, and deep learning systems.

References

[1] J. Allebach, and P.W. Wong, Edge-Directed Interpolation, Proc. IEEE Int. Conf. Image Process. 3, 707–710 (1996).
[2] J.W. Hwang, and H.S. Lee, Adaptive Image Interpolation Based on Local Gradient Features, IEEE Signal Process. Lett. 11 (3), 359–362 (March 2004).
[3] T.M. Lehmann, C. Gonner, and K. Spitzer, Survey: Interpolation Methods in Medical Image Processing, IEEE Trans. Med. Imaging 18 (11), 1049–1075 (Nov. 1999).
[4] M.A. Nuno-Maganda, and M.O. Arias-Estrada, Real-Time FPGA-Based Architecture for Bicubic Interpolation: An Application for Digital Image Scaling, Proc. IEEE Int. Conf. Reconfig. Comput. FPGAs, 8 (2005).
[5] M. Hahn, L. Kruger, and C. Wohler, 3D Action Recognition and Long-Term Prediction of Human Motion, Computer Vision Syst. 5008, 23–32 (2008).
[6] D. Unat, T. Hromadka, and S.B. Baden, An Adaptive Sub-Sampling Method for In-Memory Compression of Scientific Data, Data Compression Conf., 262–271 (March 2009).
[7] A. Roy, and J.F. Doherty, Raised Cosine Filter-Based Empirical Mode Decomposition, IET Signal Process. 5 (2), 121–129 (Apr. 2011).
[8] G. Petrizzuoli, M. Cagnazzo, F. Dufaux, and B. Pesquet-Popescu, Wyner-Ziv Coding for Depth Maps in Multi-view Video-Plus-Depth, IEEE Inter. Conf. Image Process., 1817–1820 (Sept. 2011).
[9] C.C. Tseng, and S.L. Lee, Design of Fractional Delay Filter Using the Hermite Interpolation Method, IEEE Trans. Circuits. Syst. I 59 (7), 1458–1471 (July 2012).
[10] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric, Proc. SPIE Conf. Hum. Vision Electron. Imaging 6492 (Feb. 2007).