SUPER-RESOLUTION FOR 2K/8K TELEVISION USING WAVELET-BASED IMAGE REGISTRATION

Yasutaka Matsuo and Shinichi Sakaida

NHK (Japan Broadcasting Corporation)

ABSTRACT

We propose a super-resolution method to convert spatial resolution from 2K to 8K to utilize existing 2K HDTV video for 8K UHDTV broadcasting. The proposed method uses image registration for wavelet multi-scale bands of 2K video frames considering future hardware implementation to real-time processing. This image registration consists of alignment and assignment procedures. The wavelet multi-scale bands are extracted using wavelet decomposition of 2K video frames. The alignment is processed from 2K video frames to its wavelet low-frequency bands. By using the alignment result, the assignment is processed from its wavelet high-frequency bands to super-resolved high-frequency bands. After these alignment and assignment procedures, super-resolved 8K video frames are reconstructed using wavelet reconstruction with 2K video frames and their super-resolved high-frequency bands generated in the assignment procedure. Experiments showed that the method provides objectively better PSNR and SSIM measurements and subjectively better appearance than conventional super-resolution methods.

***Index Terms*—** Super-resolution, image registration, wavelet, 2K HDTV, 8K UHDTV

1. INTRODUCTION

The practical broadcasting of 8K ultra high-definition television (UHDTV) is scheduled to launch in Japan in December 2018 [1]. The 8K UHDTV, specified in ITU-R Recommendation BT.2020, has 7680 horizontal pixels × 4320 vertical lines [2]. If we utilize existing 2K high-definition television (HDTV) content in 8K UHDTV broadcasting, we need to expand that spatial resolution 16 times. In this paper, we therefore propose a super-resolution method of spatial resolution from 2K HDTV to 8K UHDTV considering future hardware implementation to real-time processing.

Prior work on super-resolution methods can be classified into two groups, as shown in Table I: learning-based and reconstruction-based. Two examples of learning-based super-resolution are example-based super-resolution [3], [4] and super-resolution convolutional neural networks (SRCNNs) [5], [6]. The learning-based super-resolution can correctly estimate unknown coefficients of super-resolved high-frequency bands over the Nyquist frequency of an original image if it has a suitable database set. However, it requires an extensive database set to ensure a high quality super-resolved image. In addition, hardware implementation is difficult because a huge number of repeated operations is required.

TABLE I. CLASSIFICATION OF PRIOR WORK ON SUPER-RESOLUTION.

- □ Learning-based super-resolution
  - ➢ Example-based super-resolution
  - ➢ Super-resolution convolutional neural networks
- □ Reconstruction-based super-resolution
  - ➢ Filtering-based super-resolution
    - ✧ Linear filtering
    - ✧ Non-linear filtering
  - ➢ Registration-based super-resolution
    - ✧ Multi-frame registration
    - ✧ Multi-scale registration ← <u>Proposed method</u>

In contrast, reconstruction-based super-resolution can be classified into two subgroups: filtering-based and registration-based. Two examples of filtering-based super-resolution are wavelet super-resolution with linear filtering [7], [8], [9] and total variation super-resolution with non-linear filtering [10], [11]. The filtering-based super-resolution can be implemented in hardware with comparative ease. However, unknown coefficients of super-resolved high-frequency bands cannot be correctly estimated because these unknown coefficients are generated in only the linear or non-linear filtering procedure. Two examples of registration-based super-resolution are multi-frame registration [12], [13] and multi-scale registration [14], [15]. The multi-frame registration can generate correct super-resolved components if the registration is precisely done with high accuracy. However, if an original image has large motion blur or occluded objects, the registration has low accuracy. In contrast, high-resolution images have high self-similarity because they have many similar objects in their high-resolution images. By utilizing this self-similarity, we created a super-resolution method to convert spatial resolution from 4K to 8K [14], [15]. In [15], registration can be precisely done with high accuracy using exhaust block

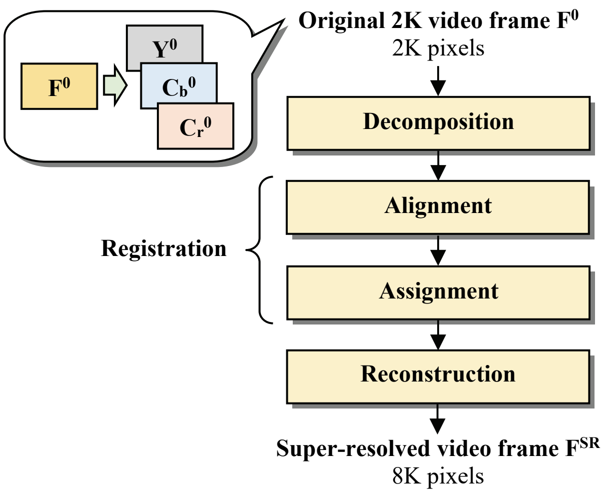

Figure 1. Overview of proposed method.

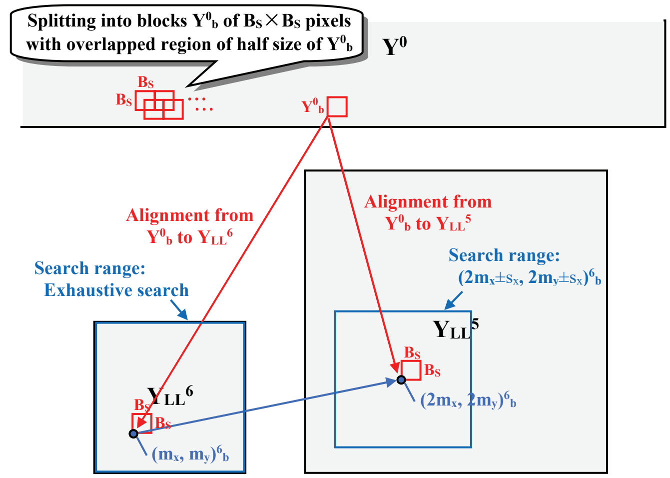

Figure 2. Details of alignment procedure (in case of n = 6, 5).

matching in wavelet multi-scale bands. In addition, registration is processed in each red (R), green (G), and blue (B) color plane considering the color sampling pattern (e.g., Bayer sampling pattern) of one-CMOS sensor in a 4K camera to increase the registration accuracy.

In this study, we developed a super-resolution method to convert the spatial resolution from 2K to 8K. In this conversion, the computational cost increases if [15] uses two steps and if the spatial resolution is enhanced from 2K to 4K and 4K to 8K. In addition, the color sampling pattern of a one-CMOS sensor does not need to be considered because the 2K HDTV video of TV broadcasting is taken by a three-CMOS sensor camera with a square-lattice sampling pattern. In consideration of these differences, in the proposed method, registration is processed in only one-step to enhance spatial resolution from 2K to 8K. To reduce the computational cost further, the registration is performed using step-search-like block matching in wavelet multi-scale bands. Section 2 shows the details of the method. Section 3 presents the results of experiments. Section 4 concludes this paper.

2. PROPOSED METHOD

Figure 1 shows an overview of the method. An original 2K video frame of 2K HDTV video is decomposed into multi-scale bands. This decomposition is performed using wavelet decomposition, and multi-scale bands consisting of low- and high-frequency bands are output. The registration is then processed in alignment and assignment procedures. The alignment procedure is performed from an original 2K video frame to its low-frequency bands. This alignment procedure is performed using step-search-like block matching. By using the alignment result, the assignment procedure was performed from high-frequency bands to super-resolved high-frequency bands. After these alignment and assignment procedures, each super-resolved 8K UHDTV video frame is reconstructed using wavelet reconstruction with an original 2K video frame and its super-resolved high-frequency bands with assignment coefficients. How the four units in Figure 1 are processed is explained in Subsection 2.1–2.4.

2.1. Decomposition

An original 2K video frame F^0 that has luma and chroma planes $\{F^0 \mid F = Y, C_b, C_r\}$ is decomposed using N-level wavelet decomposition in each plane. The wavelet filter uses the Cohen-Daubechies-Feauveau (CDF) 9/7-tap wavelet [16], which has a linear phase characteristic. The decomposition level N uses six levels decided in an exploratory experiment. Using this decomposition, multi-scale bands of six levels are extracted such as low-, horizontal high-, vertical high-, and diagonal high-frequency bands $\{F_{LL}^n, F_{LH}^n, F_{HL}^n, F_{HH}^n \mid n = 1, 2, 3, 4, 5, 6\}$.

2.2. Alignment

Y^0 is split into blocks Y^0_b of $B_S \times B_S$ pixels with an overlapping region of a half size of Y^0_b for horizontal and vertical directions. The alignment is performed from Y^0_b to Y_{LL}^6 using the exhaustive search block matching method, which utilizes an evaluation function of the sum of squared difference (SSD) value, shown in Figure 2 first. The result of this alignment, matching position $(m_x, m_y)^6_b$ is output. The alignment is then performed from Y^0_b to Y_{LL}^5, and the search range is $(2m_x \pm s_x, 2m_y \pm s_x)^6_b$, as shown in Figure 2. The alignments from Y^0_b to $\{Y_{LL}^n \mid n = 4, 3, 2\}$ are performed similarly. In these alignments, the same block size is utilized for each low-frequency band of Y_{LL}^n. As a result, similar objects of different sizes between Y^0 and Y_{LL}^n are matched.

2.3. Assignment

Coefficients of multi-scale bands of $\{Y_{LH}^n, Y_{HL}^n, Y_{HH}^n \mid n =$

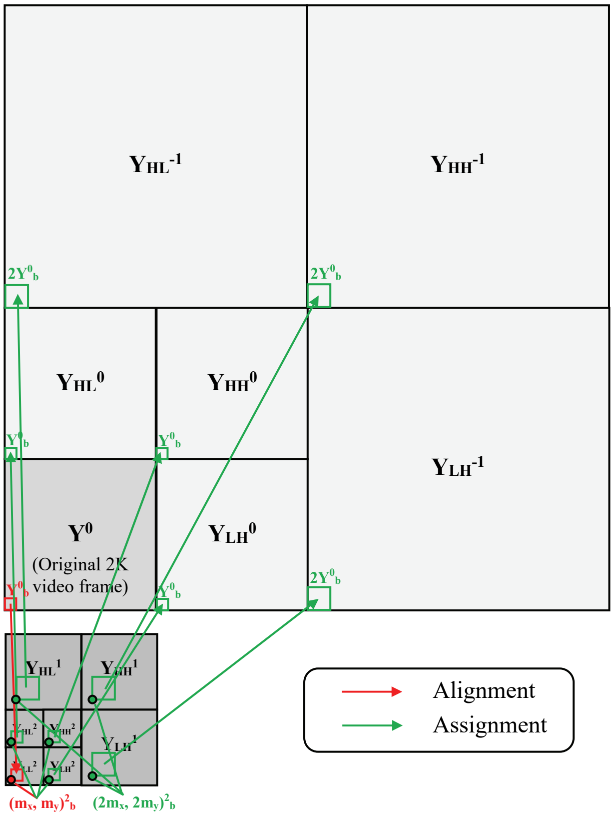

Figure 3. Details of assignment procedure (in case of n = 2).

6, 5, 4, 3, 2, 1} are assigned to super-resolved high-frequency bands $\{Y_{LH}^0, Y_{HL}^0, Y_{HH}^0, Y_{LH}^{-1}, Y_{HL}^{-1}, Y_{HH}^{-1}\}$ over the Nyquist frequency of Y^0 by using the alignment results, as shown in Figure 3. In this assignment procedure, if an arbitrary block $Y^0{}_b$ matches a position $(m_x, m_y)^n{}_b$ in the $\{Y_{LL}{}^n \mid n = 2, 3, 4, 5, 6\}$, coefficients of the same position $(m_x, m_y)^n{}_b$ in its high-frequency bands $\{Y_{LH}{}^n, Y_{HL}{}^n, Y_{HH}{}^n\}$ are assigned to the same block of $\{Y_{LH}^0, Y_{HL}^0, Y_{HH}^0\}$ with window function F_W. $Y^0{}_b$ and $(m_x, m_y)^n{}_b$ are then expanded twice for horizontal and vertical directions as $2Y^0{}_b$ and $(2m_x, 2m_y)^n{}_b$. If the difference between $2Y^0{}_b$ and the coefficients of position $(2m_x, 2m_y)^n{}_b$ in $Y_{LL}{}^{n-1}$ is less than the threshold value T_H, the coefficients of the same position $(2m_x, 2m_y)^n{}_b$ in high-frequency bands $\{Y_{LH}{}^{n-1}, Y_{HL}{}^{n-1}, Y_{HH}{}^{n-1}\}$ are assigned to the same block of $\{Y_{LH}^{-1}, Y_{HL}^{-1}, Y_{HH}^{-1}\}$ with F_W.

The assignment of C_b and C_r planes are processed similarly for the Y plane.

2.4. Reconstruction

A super-resolved 8K video frame F^{SR} of an original 2K video frame F^0 is generated using two-level wavelet reconstruction with using F^0, $\{F_{LH}^0, F_{HL}^0, F_{HH}^0\}$, $\{F_{LH}^{-1}, F_{HL}^{-1}, F_{HH}^{-1}\}$. As a result of this reconstruction, F^{SR} is output.

3. EXPERIMENT

The test video sequences used in this experiment are shown

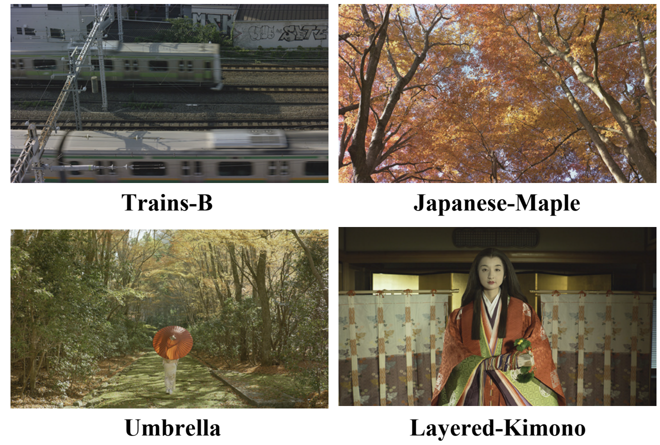

Figure 4. Test video sequences
(by using 60 frames per sequence).

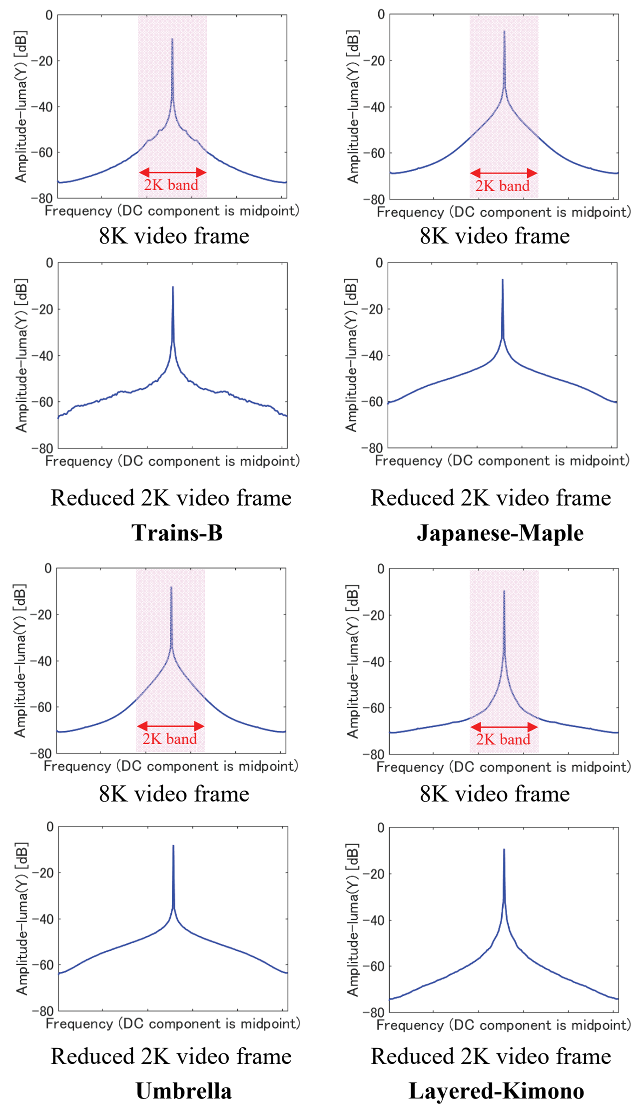

Figure 5. Frequency power spectrum of 8K and its reduced 2K video frames (Average power of 256-DFT point for horizontal and vertical directions).

in Figure 4. These sequences were distributed by the ITE [17]. They were shot using an 8K UHDTV camera [18] that had a square-lattice color sampling pattern of three-CMOS sensors. In this experiment, the spatial resolution of the 8K video sequence was reduced to 2K pixels to measure the PSNR and SSIM values after super-resolution. This reduction was performed using pixel interleaving after low-pass filtering with a Lanczos-3 filter. We confirmed that the shape of the frequency power spectrum of reduced 2K video frames is similar their original 8K video frames, as shown in Figure 5. Super-resolution methods were used in this experiment, as shown in Table II. In these super-resolution methods, the discrete and stationary wavelet decomposition (DSW) method [8] used the CDF 9/7-tap wavelet, the same as in the proposed method. The multi-frame registration (MFR) method used four video frames, two for the past direction and two for the future direction. The proposed method (PM) used parameters that were obtained from an exploratory experiment for which the B_S was {4, 6, 8}, S_x and S_y were 2 × B_S, F_W used the Blackman window, and the T_H was $B_S \times B_S \times 4$. The proposed method using 2 steps (PM2) expanded the reduced 2K video frame to 4K pixels using only assignment from $\{Y_{LH}^n, Y_{HL}^n, Y_{HH}^n\}$ to $\{Y_{LH}^0, Y_{HL}^0, Y_{HH}^0\}$, then this expanded 4K frame also expanded to 8K pixels in a similar way.

The averages of PSNR and SSIM of 60 frames per sequences are listed in Table III and Table IV. In these tables, the PSNR and SSIM obtained using the PM were 0.20–2.43 [dB] and 0.020–0.097 higher than the BC, MFR, SAR, and DSW methods. The PSNR and SSIM obtained using the PM were slightly lower than the PM2, such as 0.00–0.02 [dB] and 0.000–0.001. However, the computational cost of the PM was about 1/4 lower than that of the PM2 because it should calculate the alignment for expanding from 4K to 8K pixels. If the alignment of the PM was performed from $Y^0{}_b$ to $\{Y_{LL}^n \mid n = 6, 5, 4, 3, 2\}$ using the exhaustive-search block matching method instead of step-search-like block matching, the PSNR and SSIM increased about 0.14 [dB] and 0.008. However the computational cost of the PM increased about 256 times. In addition, the computational cost of the PM was about 2/3–1/256 lower than the DSW, SAR, and MFR methods. Cropped video frames obtained using the super-resolution methods are subjectively compared in Figure 6, where the PM obviously produced sharper and clearer image than that of the BC. These results demonstrate that the PM has better image quality than other super-resolution methods.

4. CONCLUSION

A super-resolution method for 2K/8K television using registration of wavelet multi-scale bands while considering the self-similarity of high-definition video frames was proposed. In this registration, an alignment procedure was performed using step-search-like block matching from an original 2K video frame to its low-frequency bands. By using these alignment results, the coefficients of its high-frequency bands were assigned to super-resolved high-frequency bands of the original 2K video frame to expand the spatial resolution from 2K to 8K pixels. The results of experiments showed that the proposed method provides objectively better PSNR and SSIM measurements and subjectively clearer super-resolved video frames than those provided using conventional super-resolution methods.

TABLE II. SUPER-RESOLUTION METHODS.

Super-resolution methods	Abbr.
Bi-cubic interpolation	**BC**
Multi-frame registration [13]	**MFR**
Synthetic aperture radar [11]	**SAR**
Discrete and stationary wavelet decomposition [8]	**DSW**
Proposed method using 2 steps (2K to 4K, then 4K to 8K)	**PM2**
Proposed method	**PM**

TABLE III. AVERAGES OF PSNR OF LUMA (Y) AND CHROMA (C_B, C_R) PLANES [DB].

			Super-resolution methods					
			BC	**MFR**	**SAR**	**DSW**	**PM2**	**PM**
Sequences	**Trains-B**	**Y**	31.18	31.48	32.13	32.32	32.60	32.60
		C_b	32.72	32.98	34.45	34.78	34.99	34.99
		C_r	40.99	41.45	42.89	43.10	43.33	43.31
	Japanese-Maple	**Y**	26.16	26.20	27.30	27.37	27.63	27.63
		C_b	32.91	33.11	34.94	34.97	35.20	35.20
		C_r	37.29	37.48	39.42	39.44	39.68	39.68
	Umbrella	**Y**	29.02	29.90	30.13	30.30	30.55	30.55
		C_b	29.40	30.58	31.30	31.47	31.68	31.68
		C_r	40.00	41.37	42.04	42.21	42.43	42.43
	Layered-Kimono	**Y**	34.97	35.38	35.99	36.04	36.29	36.27
		C_b	33.85	34.03	35.88	35.90	36.11	36.10
		C_r	41.66	41.90	43.64	43.70	43.91	43.90

TABLE IV. AVERAGES OF SSIM OF LUMA (Y) AND CHROMA (C_B, C_R) PLANES (STANDARD DEVIATION OF GAUSSIAN = 2).

			Super-resolution methods					
			BC	**MFR**	**SAR**	**DSW**	**PM2**	**PM**
Sequences	**Trains-B**	**Y**	0.821	0.829	0.838	0.843	0.865	0.864
		Cb	0.497	0.512	0.536	0.552	0.574	0.573
		Cr	0.612	0.631	0.652	0.666	0.687	0.686
	Japanese-Maple	**Y**	0.657	0.667	0.693	0.709	0.731	0.731
		C_b	0.635	0.648	0.687	0.695	0.716	0.716
		C_r	0.707	0.718	0.752	0.769	0.790	0.790
	Umbrella	**Y**	0.682	0.701	0.720	0.737	0.758	0.758
		C_b	0.419	0.440	0.469	0.482	0.503	0.503
		C_r	0.555	0.578	0.617	0.632	0.653	0.652
	Layered-Kimono	**Y**	0.340	0.357	0.368	0.374	0.396	0.395
		C_b	0.308	0.322	0.343	0.348	0.369	0.369
		C_r	0.361	0.375	0.392	0.401	0.421	0.421

BC **DSW** **PM**

Figure 6. Cropped super-resolution frames of Japanese-Maple.

5. REFERENCES

[1] M. Sugawara, S-.Y. Choi, and D. Wood, “Ultra-high-definition television (Rec. ITU-R BT.2020): A generational leap in the evolution of television,” IEEE Signal Processing Magazine, vol. 31, issue 3, pp. 170–174, May 2014.

[2] Y. Miki, T. Sakiyama, K. Ichikawa, M. Abe, S. Mitsuhashi, M. Miyazaki, A. Hanada, K. Takizawa, I. Masuhara, and K. Mitani, “Ready for 8K UHDTV broadcasting in Japan,” Proceedings of International Broadcasting Convention (IBC) 2015, September 2015.

[3] W.T. Freeman, T.R. Jones, and E.C. Pasztor, “Example-based super-resolution,” IEEE Computer Graphics and Applications Magazine, vol. 22, issue 2, pp. 56–65, March/April 2002.

[4] X. Li, K.M. Lam, G.Qiu, L. Shen, and S. Wang, “Example-based image super-resolution with class-specific predictions,” Journal of Visual Communication and Image Representation, vol. 20, issue 5, pp. 312–322, July 2009.

[5] C. Dong, C.C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, issue 2, pp. 295–307, February 2016.

[6] Waifu2x, http://waifu2x.udp.jp/index.html

[7] A. Temizel and T. Vlachos, “Image resolution upscaling in the wavelet domain using directional cycle spanning,” SPIE Journal of Electronics Imaging, vol. 14, issue 4, October 2005.

[8] D. Hasan and G. Anbarjafari, “Image resolution enhancement by using discrete and stationary wavelet decomposition,” IEEE Transactions on Image Processing, vol. 20, issue 5, pp. 1458–1460, May 2011.

[9] P.B. Ramdas, S.B. Keshav, and M.P. Pradeep, “Wavelet transform techniques for image resolution enhancement: a study,” International Journal of Emerging Technology and Advanced Engineering, vol. 2, issue 4, pp. 167–172, April 2012.

[10] S.D. Babacan, R. Molina, and A.K. Katsaggelos, “Total variation super resolution using a variational approach,” Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 641–644, October 2008.

[11] S.D. Babacan, R. Molina, and A.K. Katsaggelos, “Variational Bayesian super resolution,” IEEE Transactions on Image Processing, vol. 20, issue 4, pp. 984–999, April 2011.

[12] S.C. Park, M.K. Park, and M.G. Kang, “Super-resolution image reconstruction: a technical overview,” IEEE Signal Processing Magazine, vol. 20, issue 3, pp. 21–36, May 2003.

[13] D. Capel, Image mosaicing and super-resolution, Springer, U.S.A., 2004.

[14] Y. Matsuo, S. Iwasaki, Y. Yamamura, and J. Katto, “Wavelet domain image super-resolution from digital cinema to ultrahigh definition television dividing noise component,” Proceedings of IEEE Visual Communications and Image Processing (VCIP), HO5, no. 40, November 2012.

[15] Y. Matsuo and S. Sakaida, “Super-resolution method by registration of multi-scale components considering color sampling pattern and frequency spectrum power of UHDTV camera,” Proceedings of IEEE International Symposium on Multimedia (ISM), pp. 521–524, December 2016.

[16] A. Cohen, I. Daubechies, and J.-C. Feauveau, “Biorthogonal bases of compactly supported wavelets,” Communications on Pure and Applied Mathematics, vol. 45, issue 5, pp. 485–560, June 1992.

[17] Ultra-high definition / wide-color-gamut standard test sequences – series A, http://www.ite.or.jp/content/test-materials/uhdtv_a/

[18] T. Yamashita and K. Mitani, “8K extremely-high-resolution camera systems,” Proceedings of IEEE, vol. 101, issue 1, pp.74–88, January 2013.