A REAL-TIME EXAMPLE-BASED SINGLE-IMAGE SUPER-RESOLUTION ALGORITHM VIA CROSS-SCALE HIGH-FREQUENCY COMPONENTS SELF-LEARNING

Chang Su and Li Tao

DMS-Lab, Samsung Research America, Irvine, CA, USA
Email: { changsu.01, l.tao }@samsung.com

ABSTRACT

In this paper, we propose a fast and dictionary-free example-based super-resolution (EBSR) algorithm to solve the contradiction in EBSR methods of their high performance in achieving high visual quality and their low efficiency and high costs. With a novel cross-scale high-frequency components (HFC) self-learning strategy, the missed HFC of a high-resolution (HR) image are approximated from its low-resolution counterparts. A high-quality estimation of the HR image is thus obtained by compensating the HFC to its initial guess. Simulations show that the proposed algorithm gets comparable results to the state-of-the-art EBSR but with much higher efficiency and lower costs.

Index Terms— Image super-resolution, image upsampling, image quality, self-learning, self-similarity

1. INTRODUCTION

Image super-resolution (SR) techniques have been extensively researched in recent years. To a given low-resolution (LR) image, the key to a successful SR processing is to recover the high-frequency components (HFC) that are not effectively recovered by upsampling in its high-resolution (HR) counterpart. Interpolation methods cannot recover the HFC and perform poorly in practice [1]. Modern SR techniques recover the lost HFC with either reconstruction-based (RB) or example-based (EB) strategies. The RBSR [2, 3, 4] reconstructs a HR image from multiple LR images [1, 5]. But they suffer from sub-pixel level registration errors, and the high costs of iteration-based post-processing such as iterative back-projection (IBP). The EBSR [6, 7, 8, 9] recovers the lost HFC by representing the HFC with the priors of the correspondence between the LR and HR HFC, e.g., neighbor embedding [10] or sparse representation [6]. Although it performs the best among all image upsampling methods, it is iterative and computationally expensive, and not suitable for commercial applications.

Efforts had been made recently to accelerate the EBSR. J. C. Yang *et. al.* [11] adopt in-place example regression for fast EBSR. But the method still needs off-line training and extra storages for building its first-order regression model. R. Timofte *et. al.* [12, 13] relax the l_0-norm constraint of the sparse representation to l_2-norm constraint and obtain a closed-form solution of sparse representation via Ridge regression [14]. The HR HFC thus can be directly computed from the pre-computed projection matrices. Their researches show that the algorithm outperforms many popular EBSR algorithms in both output visual quality and efficiency. However, it requires huge resource to save the projection matrices and not efficient enough for big-size image upsampling applications such as full-HD (FHD) to UHD conversion. In this paper, we propose an efficient and dictionary-free EBSR algorithm that obtains comparable high-quality HR images as the conventional EBSR methods do but greatly improve the efficiency and the feasibility for commercial broadcasting applications. Section 2 presents an overview of the proposed algorithm. Section 3 describes the details of the proposed algorithm. Experimental results and evaluations are given in Section 4. Finally, Section 5 concludes this paper.

2. OVERVIEW

Let $\mathbf{I}_k$ be an image in scale k with the scaling factor s, where $s = 2^k$. Knowing that real-world images have great content redundancy in different scales [15], the basic idea of the proposed algorithm is to recover the lost HFC of $\mathbf{I}_k$ from its lower scales. In this paper, we focus our discussions on the market urgently needed FHD to UHD conversion problem, i.e., estimating a HR image $\mathbf{I}_1$ from the LR image $\mathbf{I}_0$ with $s = 2$. Figure 1 shows the flow diagram of the proposed algorithm.

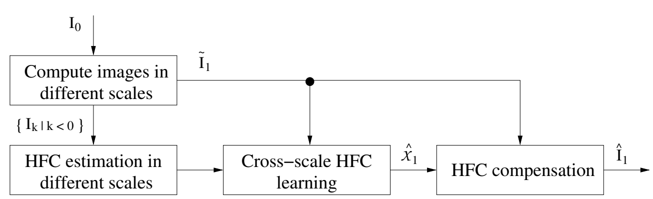


Fig. 1. The flow diagram of the proposed EBSR algorithm.

As shown in Fig.1, the initial guess of $\mathbf{I}_1$, denoted $\tilde{\mathbf{I}}_1$, and the lower-scale images $\{\mathbf{I}_k | k < 0\}$ are firstly computed from

$\mathbf{I}_0$. Then the HFC that may not be recovered in each scale by upsampling, as well as their initial guesses, are computed. The third, an efficient cross-scale HFC learning is applied to $\tilde{\mathbf{I}}_1$, $\{\mathbf{I}_k|k \leq 0\}$, and their corresponding HFC maps to form the LR and the HR HFC correspondence. An estimation of the HR HFC, denoted $\hat{\mathcal{X}}_1$, is thus computed. Finally, an estimation of $\mathbf{I}_1$ is generated by compensating $\hat{\mathcal{X}}_1$ to $\tilde{\mathbf{I}}_1$. Section 3 presents the details of the proposed algorithm.

3. THE PROPOSED EB-SR ALGORITHM

In this section, we present the details of the proposed algorithm. Sec. 3.1 models the proposed algorithm. Sec.3.2 gives the details of the HFC computations as well as the proposed cross-scale HFC self-learning strategy.

3.1. Solution Modeling

Let $\mathcal{X}_k$ be the HFC map in the k-th scale that are not recovered by image upsampling from scale $(k-1)$, $\mathbf{I}_1$ can be represented as a base layer that is lack of HFC and a detail layer that is abundant in HFC [16]. The initial guess of $\mathbf{I}_1$, i.e., $\tilde{\mathbf{I}}_1$, is a good candidate of the base layer of $\mathbf{I}_1$, namely,

$$\mathbf{I}_1 = \tilde{\mathbf{I}}_1 + \mathcal{X}_1. \tag{1}$$

As shown in (1), a well estimated $\mathcal{X}_1$ will lead to a high-quality estimation of $\mathbf{I}_1$. An initial guess of $\mathcal{X}_1$, denoted $\tilde{\mathcal{X}}_1$ can be conveniently obtained by interpolating the lost HFC map in scale 0, i.e., $\tilde{\mathcal{X}}_1 = \uparrow_2 [\mathcal{X}_0]$, where $\uparrow_2 [\cdot]$ is the upsampling operator with factor 2. Noting that interpolation does not recover any new HFC, $\tilde{\mathcal{X}}_1$ is not suitable for estimating the HR image $\mathbf{I}_1$. Inspired by Yang *et. al.*'s work [11], we assume a projection function $f(\cdot)$ that projects a HFC patch in the initial guess $\tilde{\mathcal{X}}_k$, denoted $\tilde{p}_k$, to the real one p_k in $\mathcal{X}_k$. With Taylor expansion, $f(\tilde{p}_k)$ can be expanded at its lower scale counterpart $\tilde{p}_{k-1}$ thus a HFC patch of $\mathcal{X}_1$, denoted p_1, can be represented as

$$p_1 = f(\tilde{p}_1) = f(\tilde{p}_0) + f'(\tilde{p}_0)(\tilde{p}_1 - \tilde{p}_0) + O(\cdot), \tag{2}$$

where $O(\cdot)$ is the high-order residual. Note that the proposed algorithm is different from Yang *et. al.*'s work [11]: 1) Yang *et. al.* expand the projection function at image contents, the proposed method uses Taylor expansion to represent the HFC not recovered in upsampling; and 2) Yang *et. al.* build the regression model by off-line training and save the priors in dictionaries, the proposed algorithm, as shown in Sec. 3.2, approximates $f'(\cdot)$ based on the self-similarity property of image contents, and it is dictionary-free. Considering $f(\tilde{p}_0) = p_0$, and omitting $O(\cdot)$, we get an estimation of p_1, denoted $\hat{p}_1$, as

$$p_1 \approx \hat{p}_1 = p_0 + f'(\tilde{p}_0)(\tilde{p}_1 - \tilde{p}_0). \tag{3}$$

In practice, it is very difficult to determine $f(\cdot)$. Based on the content-redundancy between different scales of real-world images, we propose to directly approximate $f'(\tilde{p}_0)$ from scale -1. Let B_0 and B_{-1} be the most similar patches between $\mathbf{I}_0$ and $\mathbf{I}_{-1}$. It is highly probable that the HFC of B_0 and B_{-1} are also similar. Let p_{-1} and $\tilde{p}_{-1}$ be the corresponding HFC patches in $\mathcal{X}_{-1}$ and its initial guess $\tilde{\mathcal{X}}_{-1}$, respectively, p_0 can be approximated with $\tilde{p}_0$, p_{-1}, and $\tilde{p}_{-1}$ by (3) but between scales 0 and -1. From (3), we have

$$f'(\tilde{p}_{-1}) \approx \frac{p_0 - p_{-1}}{\tilde{p}_0 - \tilde{p}_{-1}}. \tag{4}$$

According to the self-similarity property of real-world contents, it is reasonable that we estimate $f'(\tilde{p}_0)$ from $f'(\tilde{p}_{-1})$. Note that $f'(\tilde{p}_{-1})$ is approximated from scales 0 and -1, it may have outliers that are caused by delicate detail loss due to downsampling. Especially, in the case of strong but thin structures such as fine grass or trees, (4) may lead to high outliers. To further improve the robustness of the proposed method to different contents, we restrict the value of $f'(\cdot)$ to be within an experimental determined range $[T_l, T_h]$, where $T_l < T_h$. Thus, we estimate $f'(\tilde{p}_0)$ as

$$f'(\tilde{p}_0) = \begin{cases} T_l & : \ f'(\tilde{p}_{-1}) < T_l \\ f'(\tilde{p}_{-1}) & : \ f'(\tilde{p}_{-1}) \in [T_l, T_h] \\ T_h & : \ \text{otherwise.} \end{cases} \tag{5}$$

With (5), the proposed algorithm directly obtains $f'(\tilde{p}_0)$ from the given image thus need not any training or dictionaries. This greatly increases the efficiency and decreases the costs.

3.2. HFC computation and HR image approximation

In this section, we compute the HFC maps that are necessary to estimate $\mathbf{I}_1$ in Sec.3.1. Theoretically, $\mathcal{X}_0$ can be obtained by subtracting $\uparrow_2 [\mathbf{I}_{-1}]$ from $\mathbf{I}_0$. But in such a way, $\tilde{\mathcal{X}}_0$ has to be computed independently, and may not keep the high correspondence to $\mathcal{X}_0$ due to the content loss in downsampling. To solve the problem, we assume that $\mathbf{I}_0$ and $\tilde{\mathbf{I}}_1$ share the same base image $\mathbf{I}_0^b$ but with different HFC, thus blur is included in $\tilde{\mathbf{I}}_0$. In this paper, we compute $\mathbf{I}_0^b$ by further removing HFC from $\tilde{\mathbf{I}}_0$ with a low-pass filter $LP_h(\cdot)$. From (1), we have,

$$\begin{cases} \mathcal{X}_0 &= \mathbf{I}_0 - \mathbf{I}_0^b \\ \tilde{\mathcal{X}}_0 &= \tilde{\mathbf{I}}_0 - \mathbf{I}_0^b, \end{cases} \tag{6}$$

where $\mathbf{I}_0^b = LP_h(\tilde{\mathbf{I}}_0)$.

Similarly, we can get $\mathcal{X}_{-1}$ and $\tilde{\mathcal{X}}_{-1}$ with $\mathbf{I}_{-1}^b = LP_h(\tilde{\mathbf{I}}_{-1})$, where $\tilde{\mathbf{I}}_{-1} = \uparrow_2 [\mathbf{I}_{-2}]$. However, loss in delicate structures may be serious in scale -2 thus the obtained HFC may contain considerable content differences instead of HFC changes. To maintain the robustness of later HR image approximation, we adopt double-filtering to compute $\mathbf{I}_{-1}^b$ to avoid applying downsampling in scale -1, as shown in (7),

$$\mathbf{I}_{-1}^b = LP_t(LP_h(\mathbf{I}_{-1})), \tag{7}$$

where $LP_t(\cdot)$ and $LP_h(\cdot)$ are only different from their standard deviations. With (7), $\mathcal{X}_{-1}$ and $\tilde{\mathcal{X}}_{-1}$ are obtained similarly as shown in (6) but with $\mathbf{I}_{-1}$, $\tilde{\mathbf{I}}_{-1}$, and $\mathbf{I}^b_{-1}$.

With (6) and (7), all items in (3) — (5) are known. Let coordinates $\mathbf{x}_k$ determines an image patch $B(\mathbf{x}_k)$ in $\mathbf{I}_k$. To each $\tilde{B}_1(\mathbf{x}_1)$ of $\tilde{\mathbf{I}}_1$, we locate its first N best matched patches $\{B_0(\mathbf{x}_0^n)\}$ in $\mathbf{I}_0$, where N is an integer, and $n = 1, 2, \cdots, N$. Thus, the N most similar HFC patches of $\tilde{p}_1(\mathbf{x}_1)$ in $\mathcal{X}_0$, as well as their initial guess in $\tilde{\mathcal{X}}_0$, are located as $\{p_0(\mathbf{x}_0^n)\}$ and $\{\tilde{p}_0(\mathbf{x}_0^n)\}$, respectively. Similarly, to each $p_0(\mathbf{x}_0^n)$, we can locate its most similar HFC patch in $\mathcal{X}_{-1}$, namely $p_{-1}(\mathbf{x}_{-1}^n)$, as well as its initial guess $\tilde{p}_{-1}(\mathbf{x}_{-1}^n)$ in $\tilde{\mathcal{X}}_{-1}$. From (3), we estimate $\hat{p}_1(\mathbf{x}_1)$ as

$$\hat{p}_1(\mathbf{x}_1) = \frac{1}{N}\sum_{n=1}^{N}[p_0(\mathbf{x}_0^n) + \beta_n(\tilde{p}_1(\mathbf{x}_1) - \tilde{p}_0(\mathbf{x}_0^n))], \quad (8)$$

where β_n is the $f'(\cdot)$ estimated from the n-th corresponding HFC patch pair between scale 0 and -1. From (4), we have

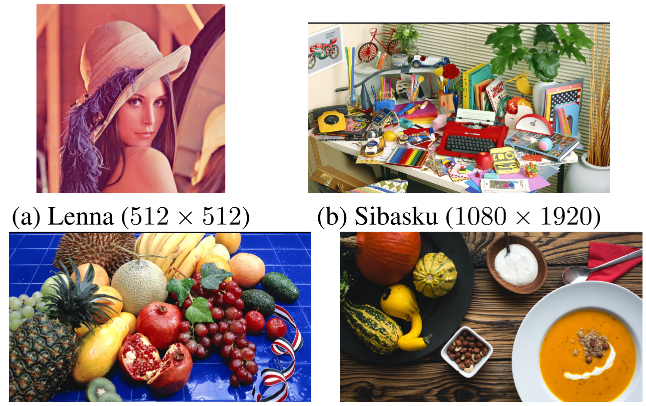
$$f'(\tilde{p}_{-1}(\mathbf{x}_{-1}^n)) = \frac{p_0(\mathbf{x}_0^n) - p_{-1}(\mathbf{x}_{-1}^n)}{\tilde{p}_0(\mathbf{x}_0^n) - \tilde{p}_{-1}(\mathbf{x}_{-1}^n)}, \quad (9)$$

and β_n is then determined by (5) and (9).

4. EXPERIMENTAL RESULTS AND EVALUATIONS

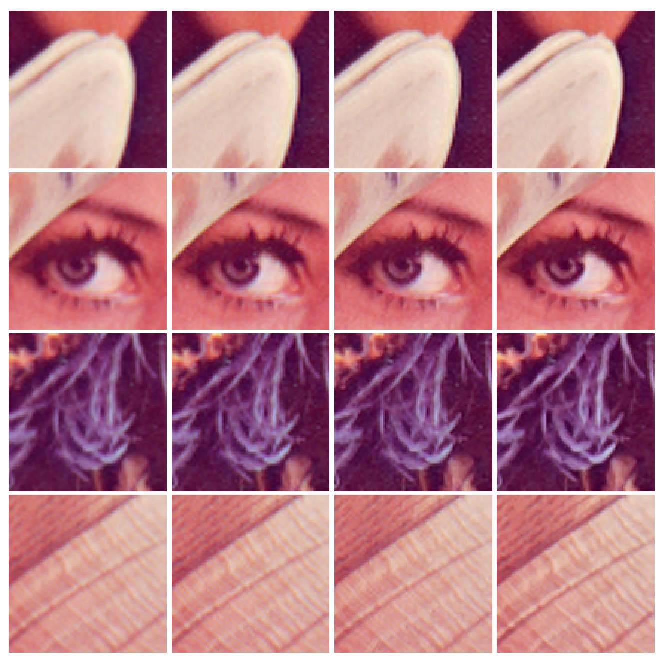
Performance evaluations are applied to the proposed algorithm and 1) the bicubic interpolation [17] , which is still popular in commercial applications, 2) the fractal SR (F-SR) [18], which is so far the fastest iteration-based SR referencing the methods evaluated in [12, 13, 19], and 3) the Timofte's EBSR (T-EBSR) [13], which is one of the most efficient and state-of-the art EBSR algorithms [13]. Note that we set the number of iterations in F-SR to 6 to obtain comparable results, and adopt the 1024-atom dictionaries in the T-EBSR provided by the authors. The parameters of the proposed algorithm are experimentally set as: the size of all patches as well as the low-pass filter are 3×3; the search range of the cross-scale patch matching is 5×5; T_l and T_h in (5) are set to -2.4 and 2.4, respectively; the standard deviations of $LP_h(\cdot)$ and $LP_t\cdot$ used in (6) and (7) are set to 1.25 and 1.0, respectively; and N is set to 2 in (8). All test algorithms are applied to the Samsung static test image database, which contains over 90 images with different image contents. Note that the test images are independent to the images we used to determine the parameters of the proposed algorithm. Figure 2 shows parts of the test images that are discussed in this section.

Figure 3 shows the examples of the experimental results of "Lenna". As shown in Fig.3, the bicubic interpolation performs poorly and introduces serious blur in the results. The F-SR performs better but its results are still blurry. The T-EBSR and the proposed algorithm perform well to different contents. The proposed algorithm outperforms the T-EBSR in edges and complex structures (row 1 - 2). It successfully recovers the HFC that are not recovered by upsampling and obtain natural and sharp structures. The T-EBSR obtains more details than the proposed algorithm in delicate textures (row 3 and 4). However, the method tends to generate jaggies along edges. This is because that the HFC that the T-EBSR recovered are in fact synthesized by the dictionary atoms, and the atoms obtained by training may not always match complex real-world contents well.



(a) Lenna (512×512) (b) Sibasku (1080×1920) (c) Fruits (1080×1920) (d) Soup (1080×1920)

Fig. 2. The given low-resolution test images.

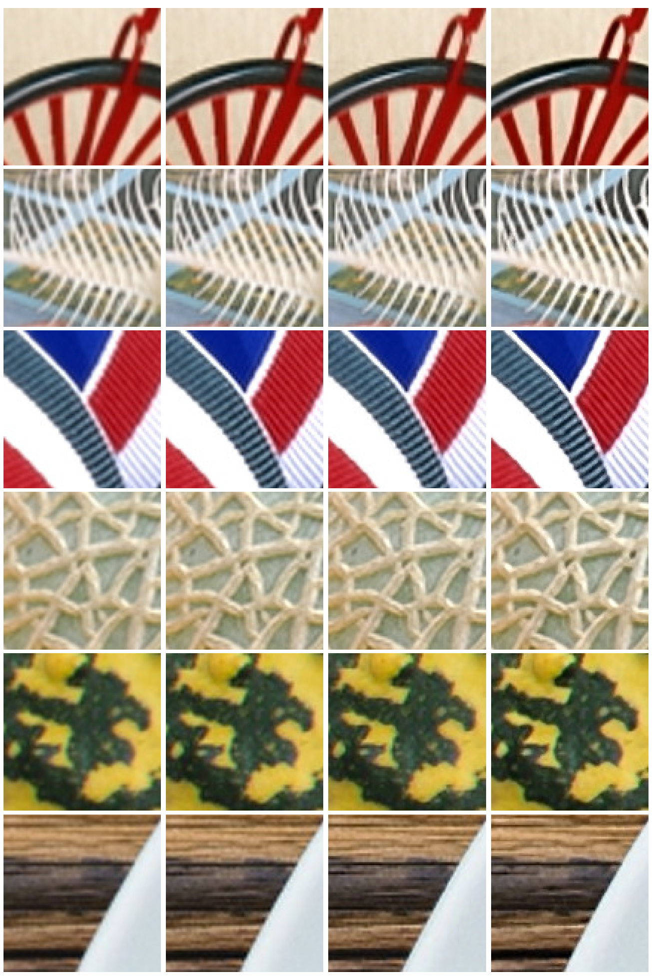


(a) bicubic (b) F-SR (c) T-EBSR (d) prop.

Fig. 3. Comparison with "Lenna".

Figure 4 shows the examples of the experimental results with the FHD test images, and the results confirms our observations in Fig.3. The bicubic interpolation cannot recover any

HFC thus its results are all blurry. The F-SR clearly outperforms the bicubic interpolation. However, it does not recover enough HFC thus its results still have perceivable blur. Also, small distortions may be introduced along strong edges in the F-SR results. Both the T-EBSR and the proposed algorithm performs well in all tests, but they performs differently to different contents. As shown in row 1 and 2 ("Sibasku") of Fig.4, the proposed algorithm significantly outperforms the T-EBSR by obtaining sharp and natural edges and complex structures. The T-EBSR tends to introduce small distortions along strong but thin structures. Row 3 and 4 of Fig.4 ("Fruits") show another example of the superior of the proposed algorithm in HFC recovering around edges and delicate structures. Row 5 and 6 of Fig.4 ("Soup") show the mixture of edges and fine details. As can be seen, the proposed algorithm gets higher quality edges than T-EBSR does, and T-EBSR recovers more details than the proposed algorithm. But small distortions along delicate structures can be seen in the T-EBSR's results.



(a) bicubic (b) F-SR (c) T-EBSR (d) prop.

Fig. 4. Comparison with FHD test images (row 1 and 2: "Sibasku"; row 3 and 4: "Fruit"; and row 5 and 6: "Soup").

Implemented in C++ and under Linux OS, we apply the proposed algorithm as well as the reference algorithms to 20 randomly selected FHD images (1080×1920) to obtain UHD images (2160×3840). Table 1 shows the average time consuming of each of the algorithms. As shown in Table 1, the Bicubic interpolation is the most efficient, and the F-SR is the next most efficient, followed by the proposed, and the T-EBSR methods. However, the bicubic and the F-SR methods do not effectively recover the HFC that are not recovered by upsampling. The proposed algorithm generates the comparable results to the T-EBSR algorithm, and even the T-EBSR is one of the most efficient EBSR methods, the proposed method is about 7.7 times faster than it. More important, the proposed algorithm is not iterative, and need not any training or dictionary. It is more feasible to commercial applications.

Table 1. The efficiency of different SR algorithms (in sec.).

	Bicubic Intp.	F-SR	T-EBSR	Prop.
Time	0.4353	0.8663	58.9523	7.6998

5. CONCLUSION

An efficient and dictionary-free example-based super-resolution (EBSR) algorithm is proposed for UHD broadcasting applications. With a novel cross-scale self-learning strategy, the lost high-frequency components (HFC) of a high-resolution (HR) image are estimated from its lower scales. A visually pleased HR image estimation is thus obtained by compensate the HFC to its initial guess. Due to the cross-scale HFC self-learning, offline training and dictionaries, which are necessary in many high-performance EBSR, are not required in the proposed algorithm. Simulations show that the proposed algorithm effectively recovers the HFC that are not effectively recovered by upsampling thus obtains sharp and natural HR estimations. Although it may slightly lose few details comparing to the state-of-the-art EBSR methods, the proposed algorithm outperforms the EBSR in edges and complex structures, and does not introduce perceivable structure distortions in delicate contents. Due to its non-iterative framework, the proposed algorithm is hardware friendly, efficient, and suitable for commercial applications.

This study shows that the computationally expensive content representation in EBSR can be replaced by much more efficient cross-scale self-learning to obtain the correspondence of the HFC between the LR and the HR images. This finding presents a potential feasible solution to apply high-performance EBSR methods in commercial broadcasting applications, such as the 4K UHD-TV.

6. REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstuction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] B. Wan, L. Meng, D. Ming, H. Qi, Y. Hu, and K. Luk, "Video image super-resolution restoration based on iterative back-projection algorithm," in *Int. Conf. On Computational Intelligence For Measurement Systems And Applications*, Hong Kong, China, May 2009, IEEE, pp. 46–49.

[3] Y. H. Cho, K. Y. Hwang, H. Y. Lee, and D. S. Park, "Generation of high resolution image based on accumulated feature trajectory," in *Int. Conf. on Image Processing*, Hong Kong, China, Sept. 2010, IEEE, pp. 1997–2000.

[4] H. Su, L. Tang, Y. Wu, D. Tretter, and J. Zhou, "Spatially adaptive block-based super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1031–1045, Mar. 2012.

[5] S. Borman and R. L. Stevenson, "Super-resoltuion from image sequences — a review," in *Midwest Symp. On Circuits And Systems*, Nore Dame, IN, Aug. 1998, IEEE, pp. 374–378.

[6] J. C. Yang, J. Wright, T. S. Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[7] R. Zeyde, M. Protter, and M. Elad, "On single image scale-up using sparse-presentation," in *Int. conf. On Curves and Surfaces*, Avignon, France, Jun. 2010, pp. 711–730, Springer.

[8] C. Y. Kim, K. Choi, and J. B. Ra, "Example-based super-resolution via structure analysis of patches," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 407–410, Apr. 2013.

[9] J. F. Yu, X. B. Gao, D. Tao, X. L. Li, and K. B. Zhang, "A unified learning framework for single image super-resolution," *IEEE Trans. On Neural Networks And Learning Syst.*, vol. 25, no. 4, pp. 780–792, Apr. 2014.

[10] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Int. Conf. On Computer Vision and Pattern Recognition*, Washington, DC, USA, Jun. 2004, vol. 1, pp. 4321–4326, IEEE.

[11] J. C. Yang, S. Jose, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Int. Conf. On Computer Vision And Pattern Recognition*, Portland, OR, Jun. 2013, IEEE, pp. 1059–1066.

[12] R. Timofte, V. De Smet, L. Van Gool, KU Leuven, and ETH Zurich, "Anchored neighborhood regression for fast example-based super-resolution," in *Int. Con. On Computer Vision*, Sydney, NSW, Dec. 2013, pp. 1920–1927, IEEE.

[13] R. Timofte, V. De Smet, and L. Van Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian Conf. On Computer Vision*, Singapore, Nov. 2014, pp. 1–15.

[14] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-posed Problems*, Winston & Sons, 1977.

[15] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Int. Conf. on Computer Vision*, Kyoto, Japan, Sep. 2009, pp. 349–356, IEEE.

[16] K. M. He, J. Sun, and X. O. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[17] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[18] Y. C. Wee and H. J. Shin, "A novel fast fratal super resolution technique," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1537–1541, Aug. 2010.

[19] C. Y. Yang, C. Ma, and M. H. Yang, "Single-image super resolution: a benchmark," in *European Conf. On Computer Vision*, Zurich, Switzerland, Sept. 2014, vol. 8692, pp. 372–386.