Reducible Dictionaries for Single Image Super-Resolution based on Patch Matching and Mean Shifting

Rasti, Pejman; Nasrollahi, Kamal; Orlova, Olga; Tamberg, Gert; Moeslund, Thomas B.; Anbarjafari, Gholamreza

Published in:
Journal of Electronic Imaging

DOI (link to publication from Publisher):
10.1117/1.JEI.26.2.023024

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Journal of Electronic Imaging

JElectronicImaging.org

Reducible dictionaries for single image super-resolution based on patch matching and mean shifting

Pejman Rasti
Kamal Nasrollahi
Olga Orlova
Gert Tamberg
Thomas B. Moeslund
Gholamreza Anbarjafari

Pejman Rasti, Kamal Nasrollahi, Olga Orlova, Gert Tamberg, Thomas B. Moeslund, Gholamreza Anbarjafari, "Reducible dictionaries for single image super-resolution based on patch matching and mean shifting," *J. Electron. Imaging* **26**(2), 023024 (2017), doi: 10.1117/1.JEI.26.2.023024.

Reducible dictionaries for single image super-resolution based on patch matching and mean shifting

Pejman Rasti,[a] Kamal Nasrollahi,[b] Olga Orlova,[c] Gert Tamberg,[c] Thomas B. Moeslund,[b] and Gholamreza Anbarjafari[a,d,*]
[a]University of Tartu, Institute of Technology, iCV Research Group, Tartu, Estonia
[b]Aalborg University, Visual Analysis of People Laboratory, Aalborg, Denmark
[c]Tallinn University of Technology, School of Science, Division of Mathematics, Department of Cybernetics, Tallinn, Estonia
[d]Hasan Kalyoncu University, Department of Electrical and Electronic Engineering, Gaziantep, Turkey

Abstract. A single-image super-resolution (SR) method is proposed. The proposed method uses a generated dictionary from pairs of high resolution (HR) images and their corresponding low resolution (LR) representations. First, HR images and the corresponding LR ones are divided into patches of HR and LR, respectively, and then they are collected into separate dictionaries. Afterward, when performing SR, the distance between every patch of the input LR image and those of available LR patches in the LR dictionary is calculated. The minimum distance between the input LR patch and those in the LR dictionary is taken, and its counterpart from the HR dictionary is passed through an illumination enhancement process. By this technique, the noticeable change of illumination between neighbor patches in the super-resolved image is significantly reduced. The enhanced HR patch represents the HR patch of the super-resolved image. Finally, to remove the blocking effect caused by merging the patches, an average of the obtained HR image and the interpolated image obtained using bicubic interpolation is calculated. The quantitative and qualitative analyses show the superiority of the proposed technique over the conventional and state-of-art methods. *© 2017 SPIE and IS&T* [DOI: 10.1117/1.JEI.26.2.023024]

Keywords: super-resolution; dictionary reduction; interpolation kernel; image processing.
Paper 161064 received Dec. 24, 2016; accepted for publication Apr. 7, 2017; published online Apr. 22, 2017.

1 Introduction

The resolution of an image is one of its important properties. This is, however, limited to the capabilities of the image acquisition device, mainly the imaging sensor.[1] The size of this sensor defines the available spatial resolution of produced images. Thus, for increasing the spatial resolution of an image, one could increase the sensor density by reducing the spaces allocated to each pixel on the sensor or simply by increasing the size of the sensor. The latter solution is expensive and slows down the imaging process, while the former solution decreases the amount of light incident on each sensor, which in turn increases the shot noise.[1]

Applying various signal processing techniques is the other approach for enhancing the resolution of an image. One of the famous techniques for this purpose is super resolution (SR).[2–5] The basic idea behind SR methods is to recover/estimate one or more high resolution (HR) image(s) from one or more low resolution (LR) images.[6,7] Huang and Tsai[8] as pioneers of SR proposed a method to improve the spatial resolution of satellite images of earth, where a large set of translated images of the same scene are available. They showed that SR, using multiple offset images of the same scene and a proper registration, can produce better HR images compared with spline interpolation. Since then, SR methods have become common practice for many applications in different fields, such as remote sensing,[9,10] surveillance video,[11–14] medical imaging such as ultrasound, magnetic resonance imaging, and computerized tomography scan,[15–18] optical character recognition problems, and face recognition.[19–23]

Different techniques have been developed for performing SR.[2,24,25] One such technique is based on sparse representation. In Ref. 26, a downsample version of HR images is considered as LR images, whose patches were assumed to have a sparse representation with respect to an overcomplete dictionary of prototype signal-atoms where signals are described by sparse linear combinations of these atoms. They showed that the HR image can be correctly recovered from the downsampled signal using the compressed sensing principle. Their method illustrated the effectiveness of sparsity as a prerequisite for regularizing the otherwise ill-posed SR problem.

Yang et al.[27] presented an approach to SR, based on sparse signal representation. They showed that image patches can be represented as a sparse linear combination of elements from a properly chosen complete dictionary. They sought a sparse representation for each patch of the LR input image, and then used the coefficients of this representation to generate the HR output. Their proposed method enforced the similarity of sparse representations between the LR and HR image patch pairs with respect to their own dictionaries. The HR image patch was generated by applying the HR image patch dictionary and the sparse representation of an LR image patch. They used their algorithm for general image SR and the special case of face hallucination.

The sparse representation is used to generate a scaled-up image in Ref. 28, proposed by Zeyde et al. They started their work using the concept of the method of Refs. 26 and 27 but simplified the overall process both in terms of the computational complexity and the algorithm architecture, using a

*Address all correspondence to: Gholamreza Anbarjafari, E-mail: shb@ut.ee

different training approach for the dictionary-pair, and introduced the ability to operate without a training-set by bootstrapping the scale-up task from the given LR image.

Peleg and Elad[29] used a statistical prediction model based on sparse representations of LR and HR image patches to generate an HR image. Their proposed method allowed them to avoid any invariance assumption, which was a common practice in sparsity-based approaches treating this task. Minimum mean square error estimation is used to predict the HR patches, and the resulting scheme had the useful interpretation of a feedforward neural network. They also recommended data clustering and cascading several levels of the basic algorithm.

Zhu et al.[30] introduced an idea for quick single image SR, which was tailored to the specifications of sparse representation, developed on the basis of self-example learning. The desirable efficiency of the foregoing method arises from the replacement of the regular singular value decomposition (SVD) by the K-SVD, which was aimed at accelerating the computation upon utilizing more simplistic approximation, along with the orthogonal matching pursuit algorithm, being able to satisfy the conditions of the aforementioned approach more appropriately via entailing considerably fewer signals.

In this work, we contribute to the field of image SR by introducing the following new steps:

- Compared with the work in Ref. 27, we do not employ a sparse representation but utilize separate dictionaries for LR and HR patches.
- The minimum distance between input LR patches and LR patches in the LR dictionary is used for choosing a corresponding HR patch in the HR dictionary.
- A mean shift method is used for illumination enhancement to avoid blocking effects in the super-resolved image.

Furthermore, we show that the dictionaries used in the proposed system can be reduced in size quite a lot, without affecting the performance of the system. The quantitative and qualitative experimental results show the superiority of the proposed method over the state-of-the-art of Ref. 29, which uses sparse representation, and the more recent technique of Ref. 31, which uses deep learning techniques based on convolutional neural networks (CNN).

The rest of this paper is organized as follows. A detailed overview of the proposed method is presented in Sec. 2. Section 3 reports and discusses the results of the experiments carried out. Finally, Sec. 4 concludes the paper.

2 Proposed System

In this section, first we present some notation for our work. Then, the way we have built the LR and HR dictionaries is discussed. Then, we continue with the details of the proposed system.

The LR and HR images are represented as matrix $\Psi_l \in \mathbb{R}^{N_l \times M_l}$ and $\Psi_h \in \mathbb{R}^{N_h \times M_h}$, where $N_h = \alpha N_l$, $M_h = \alpha M_l$, and $\alpha > 1$ is some integer scale-up factor. The blur operator is denoted by $H: \mathbb{R}^{N_h \times M_h} \rightarrow \mathbb{R}^{N_h \times M_h}$, and the decimation operator for a factor α in each axis is denoted by $Q: \mathbb{R}^{N_h \times M_h} \rightarrow \mathbb{R}^{N_l \times M_l}$, which discards rows and columns from the input image (nearest neighbor interpolation function is used as a decimation function in this work). Two acquisition models are commonly used in the literature[29] to describe how an LR image is generated from an HR image, and each of them has a different rationale. The first assumes that prior to decimation, a known low pass filter is applied on the image

$$\Psi_l = Q[H(\Psi_h)] + v, \tag{1}$$

where v is an additive noise in the acquisition process. The corresponding problem of reconstructing Ψ_h from Ψ_l is also referred to in the literature as zooming deblurring.[32] The second acquisition model[27,29,31] assumes that there is no blur prior to decimation, namely $\Psi_l = Q\{\Psi_h\} + v$, so image reconstruction is cast as a pure interpolation (zooming) problem. In other words, the problem is only filling out the missing pixels between the original pixels in the input LR image, which remain unaltered in the recovered HR image. In this work, the second model is considered, and the images are assumed to be noise free, i.e., $v = 0$.

Let $P^k = R_n^k \Psi$ be an image patch of size $n \times n$ centered at location k and extracted from the image Ψ by the linear operator R. Hence, the LR and HR patches are extracted as

$$P_l^k = R_n^k \Psi_l \quad P_h^k = R_{\alpha n}^k \Psi_h, \tag{2}$$

where α is the scale-up factor, h refers to HR, and l refers to LR. Extracted patches of HR images and their corresponding LR patches are saved in two dictionaries of HR patches D_h and LR patches D_l, respectively. Hence, every $P_l^k \in D_l$ has a correspondence in D_h. Th mapping between these two pairs is expressed by f as

$$P_h^k = f(P_l^k). \tag{3}$$

The main motivation for the proposed model is the desire to predict a missing HR detail for each LR patch via a pair in the created D_l and D_h dictionaries. Following the block diagram of the proposed system, shown in Fig. 1, we first find all the patches of the input LR image, $\tilde{\Psi}_l$, using R_n^q centered at location q

$$\tilde{P}_l^q = R_l^q \tilde{\Psi}_l. \tag{4}$$

Then the minimum distance between each patch, $\tilde{P}_l^q$, and all patches in D_l is calculated by

$$d_\kappa = \min_k [d(P_l^k, \tilde{P}_l^q)] \quad \text{where } d(P_l^k, \tilde{P}_l^q) = \sqrt{\sum_i (\tilde{P}_{l_i}^q - P_{l_i}^k)^2}, \tag{5}$$

where κ refers to the index of the patch in D_l, which has the minimum distance from the $\tilde{P}_l^q$ of the input LR image.

Having found a patch in the LR dictionary with a minimum distance to P_l^q, its HR corresponding patch is found in D_h and replaced in the HR image by

$$\tilde{P}_h^q = f(P_l^\kappa). \tag{6}$$

To avoid a sudden change of illumination,[33] a simple illumination enhancement (mean shift) is applied to $\tilde{P}_h^q$ by moving its mean, $\mu_{\tilde{P}_h^q}$, toward the mean of $\tilde{P}_l^q$, $\mu_{\tilde{P}_l^q}$, using

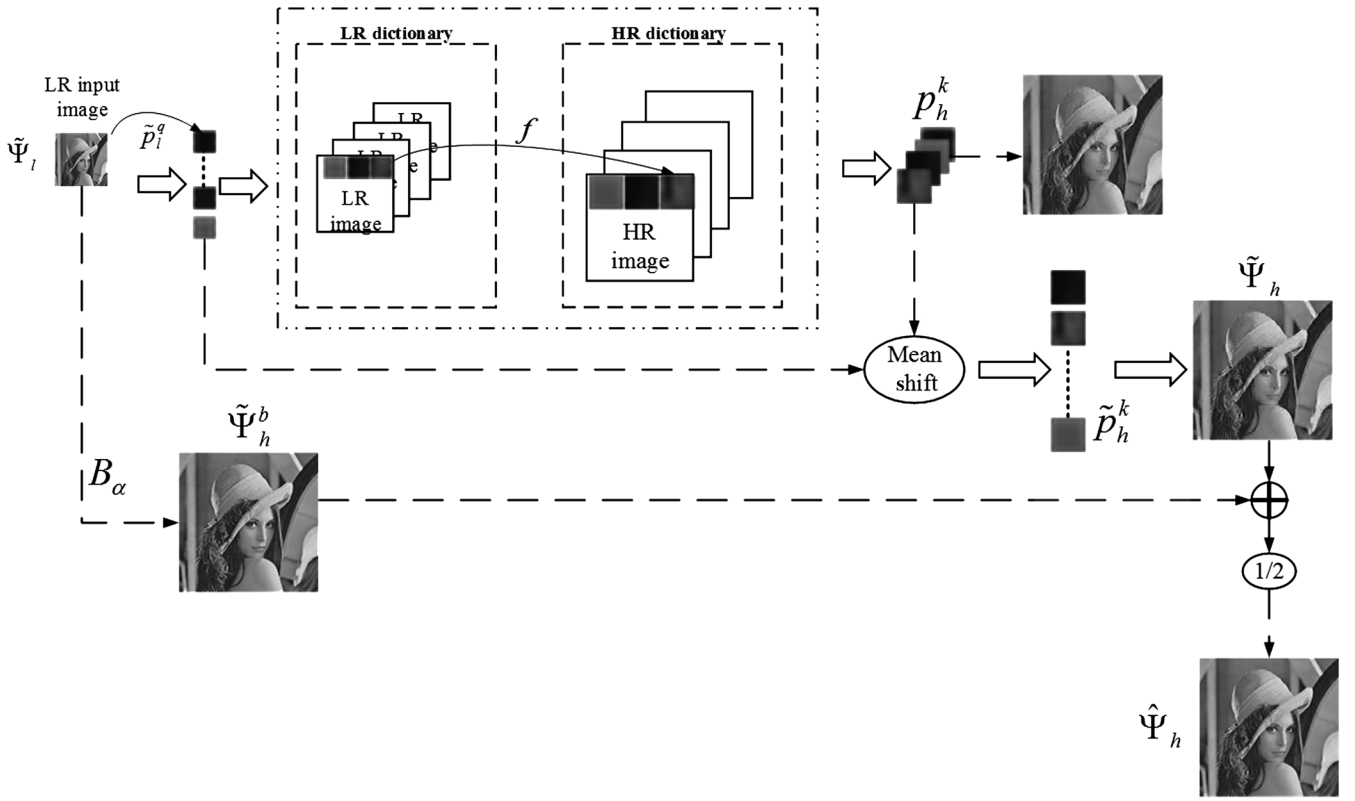

Fig. 1 The block diagram of the proposed system.

$$\tilde{P}_h^q = \tilde{P}_l^q \times \frac{\mu_{\tilde{P}_l^q}}{\mu_{\tilde{P}_h^q}}. \tag{7}$$

This process is repeated until the last patch of the input LR image. Finally, all founded HR patches are merged together according to their location to get the HR image, $\tilde{\Psi}_h$. The generated image has some blocking effect, which is not desired. To remove this effect, the LR input image, $\tilde{\Psi}_l$, is also interpolated using bicubic interpolation with the same scaling factor, using

$$\tilde{\Psi}_h^b = B_\alpha\{\tilde{\Psi}_l\}, \tag{8}$$

where B is the bicubic interpolation operator with a scaling-up factor of α. Finally, the HR image, $\hat{\Psi}_h$, is calculated by averaging the HR image obtained by merging patches, $\tilde{\Psi}_h$, and the HR image obtained by bicubic interpolation, $\tilde{\Psi}_h^b$, as shown in

$$\hat{\Psi}_h = \frac{\tilde{\Psi}_h + \tilde{\Psi}_h^b}{2}. \tag{9}$$

The general steps of the proposed single image SR method are summarized and shown in Algorithm 1 and are illustrated in Fig. 1.

One of the main constraints in dictionary-based SR algorithms is the huge size of the dictionaries. There exist many patches within the dictionaries that are very similar to one another. As one of the contribution of this work, to reduce the number of patches in the dictionary, we tried to find the set of patches that spans the vector space, which includes all the patches as well as the structural similarity between patches. For the first approach, the selected patches are independent or almost independent from each other. To find the independent and almost independent patches, we used SVD. Then, the Euclidean distance between the vector of

Algorithm 1 Single image SR schema.

Input: LR image and scale-up factor.

Output: HR image.

1 **Image interpolation** using bicubic interpolation with the scale-up factor to generate a scaled-up image from $\tilde{\Psi}_l$.

2 **Extract LR patches** centered at locations q from the LR image.

3 for q do

4 **Compute the minimum distance between the LR patch** and the LR patches in D_l.

5 **Find the corresponding HR patch** with the LR patches that has minimum distance.

6 **Enhance the illumination** of HR patch.

7 **Replace the HR patch** in $\tilde{\Psi}_h$.

8 end **for**

9 **Find the average between** $\tilde{\Psi}_h$ and $\tilde{\Psi}_h^b$.

10 **Generate HR** image.

singular values of patches have been calculated. The patches who are discriminant enough from the other ones, i.e., the calculated cross distance is bigger than threshold value, τ, are kept in the dictionary. The threshold, τ, is selected based on the application. In this work, we reduced the size of dictionary by 30%, and the visual quality dropped only 0.07 dB (discussed more in the next section).

Another approach of dictionary reduction proposed in this work is based on a structural similarity index (SSIM). To remove patches that are structurally similar to other patches, a similarity between all patches is calculated using SSIM; then, patches with a similarity index more than threshold τ are removed from the dictionaries. Similar to the previous approach, the threshold, τ, is selected based on the application. In this approach, we reduced the size of dictionary by 45%, and the visual quality dropped only 0.08 dB (discussed more in the next section).

3 Experimental Results

In this work, 581 HR images from different databases, namely, the LFW face database[34] and some standard test image databases,[35] are used to make the HR dictionary, D_h, and its corresponding LR dictionary, D_l. These images are selected from different categories, such as faces images, natural images, and texture images. Some of these images are shown in Fig. 2.

For making the LR dictionary, the HR images are downsampled by factor of 2 and 4, i.e., in our experimental results, we conduct SR with scaling factors of $\alpha = 2$ and $\alpha = 4$. Then patch sizes of 8×8, 16×16 for making HR dictionaries and patch sizes of 4×4 are chosen for making the LR dictionary. All HR images used for this work are set on the size of 256×256. Also, the input LR test images of all SR techniques used for the experimental results were obtained by downsampling their HR original counterparts using nearest neighbor kernel. It should be mentioned that the test images are not used in constructing the dictionaries.

For testing, many well-known benchmark images (such as Butterfly, Comic, Flowers, Foreman, Girl, Lena, Man, Pepper, Starfish, and Zebra) were used.

The peak signal-to-noise ratio (PSNR) and SSIM are used to evaluate the imperceptibility characteristics quality measurement. Tables 1–4 show the PSNR values in dB and SSIM values for sparse coding based network (SCN),[36] the CNN based SR method of Ref. 31, Peleg and Elad's system of Ref. 29, and the proposed SR system (with and without mean shift) for the aforementioned images, respectively. It can be seen from these tables that the proposed system, even without the mean shift step, outperforms the other methods, and, involving the mean shift, improves the results even further.

Table 1 PSNR results of the proposed method with a scaling factor of 2 compared with Refs. 29, 31, 36, where bold numbers show the best performance.

	SCN[36]	SRCNN[31]	Peleg and Elad[29]	Proposed method without mean shift	Proposed method with mean shift
Butterfly	21.53	22.45	22.95	23.63	**23.67**
Comic	20.08	21.03	21.68	22.63	**22.67**
Flowers	22.56	23.58	24.13	24.95	**25.03**
Foreman	30.58	30.33	30.55	32.71	**32.83**
Girl	29.64	30.95	31.69	32.23	**32.40**
Lena	25.00	24.93	25.42	28.12	**28.37**
Man	22.71	23.68	24.32	25.32	**25.43**
Pepper	25.30	27.21	27.63	30.69	**31.63**
Starfish	23.50	24.52	25.12	26.05	**26.16**
Zebra	19.90	20.75	21.67	22.40	**22.40**
Average of 1000 face images	30.99	31.95	32.33	32.14	**32.72**
Average	24.71	25.58	26.14	27.35	**27.57**

For the Lena image, the PSNR of the proposed method is 3.37, 3.50, and 3.01 dB higher than those of the SCN,[36] CNN-based technique of Ref. 31, and SR method of Peleg and Elad's,[29] respectively.

Table 1 shows that, for instance, for the pepper image, the PSNR of our proposed super-resolved image is about 4.4 and 4 dB higher than those of the SRCNN[31] and the Peleg and Elad's method in Ref. 29, respectively. The average of the

Fig. 2 Some images used in constructing the dictionaries.

Table 2 SSIM results of the proposed method with a scaling factor of 2 compared with Refs. 29, 31, 36.

	SCN[36]	SRCNN[31]	Peleg and Elad[29]	Proposed method without mean shift	Proposed method with mean shift
Butterfly	0.8217	0.8269	0.8415	0.8549	0.8665
Comic	0.7321	0.7360	0.7468	0.7777	0.7879
Flowers	0.7585	0.7740	0.7857	0.8151	0.8253
Foreman	0.8859	0.8852	0.9011	0.9123	0.9232
Girl	0.7646	0.7997	0.8291	0.8403	0.8407
Lena	0.8318	0.8347	0.8416	0.8805	0.8826
Man	0.7251	0.7431	0.7638	0.7901	0.8013
Pepper	0.8526	0.8811	0.8941	0.9539	0.9598
Starfish	0.7954	0.8071	0.8205	0.8404	0.8520
Zebra	0.7041	0.7259	0.7502	0.7633	0.7802
Average of 1000 face images	0.9535	0.9453	0.9516	0.9412	0.9562
Average	0.8536	0.8145	0.8296	0.8518	0.8614

Table 3 PSNR results of the proposed method with a scaling factor of 4 compared with Refs. 29, 31, 36, where bold numbers show the best performance.

	Proposed algorithm	SRCNN[31]	Peleg and Elad[29]	SCN[36]
Butterfly	**18.97**	18.67	14.90	17.32
Comic	**18.31**	17.71	14.47	16.56
Flowers	**20.69**	19.83	19.16	18.65
Foreman	**27.73**	26.80	25.52	26.22
Girl	**28.61**	28.26	26.08	26.69
Lena	**23.68**	22.00	21.84	21.95
Man	**21.15**	20.64	18.97	19.51
Pepper	**23.94**	23.56	21.56	21.89
Starfish	**21.35**	21.11	18.80	19.63
Zebra	**17.50**	16.11	15.84	15.58
Average of 1000 face images	**27.31**	27.19	26.79	27.01

Table 4 SSIM results of the proposed method with a scaling factor of 4 compared with Refs. 29, 31, 36.

	Proposed algorithm	SRCNN[31]	Peleg and Elad[29]	SCN[36]
Butterfly	0.6639	0.6633	0.4845	0.6120
Comic	0.5575	0.5160	0.3631	0.4887
Flowers	0.5900	0.5894	0.5611	0.5497
Foreman	0.8794	0.8156	0.8190	0.7897
Girl	0.6756	0.7178	0.6448	0.6525
Lena	0.8895	0.7128	0.8111	0.6996
Man	0.6189	0.5654	0.6708	0.5172
Pepper	0.8877	0.7754	0.8305	0.7188
Starfish	0.6917	0.6672	0.5623	0.6123
Zebra	0.5480	0.4796	0.5025	0.4457
Average of 1000 face images	0.9002	0.8516	0.7843	0.8189

PSNR results of our proposed method is ∼1.4 and 0.7 dB more than the SRCNN method in Ref. 31 and Peleg and Elad's method in Ref. 29, respectively.

To have a better understanding of the performance of the proposed algorithm, 1000 facial images from Ref. 34 have been super-resolved using the conventional and state-of-the-art techniques as well as the proposed SR algorithm. The quantitative results show that, on average for these 1000 facial images, the proposed algorithm improves the PSNR values by 1.73, 0.77, and 0.39 dB for SCN,[36] SRCNN,[31] and Peleg and Elad's method,[29] respectively.

Figures 3(b)–3(e) show the visual comparison of the images produced by the SCN,[36] the CNN-based SR method of Ref. 31, Peleg and Elad's system of Ref. 29 and the proposed SR system (with and without mean shift) for the Girl and Peppers images. The experimental results show that the proposed method performs better than the conventional and the state-of-the-art methods.

3.1 *Observations on Dictionary Size*

In the previous section, we have shown that the proposed SR algorithm outperforms state-of-the-art SR algorithms in terms of known measurements criteria of PSNR and SSIM. In this section, we show that the dictionaries that are used in the proposed system are reducible to a great degree, with a negligible reduction in the performance of the system. Tables 5 and 6 report performance reduction of the system in terms of PSNR, when dictionary size reduction is done using the SSIM technique mentioned in the previous section. The size of LR patches is 4×4 and 8×8, in Tables 5 and 6, respectively. These tables as well as Fig. 4 show that average PSNR results of tested images do not drop significantly when ∼50% of dictionary size is reduced.

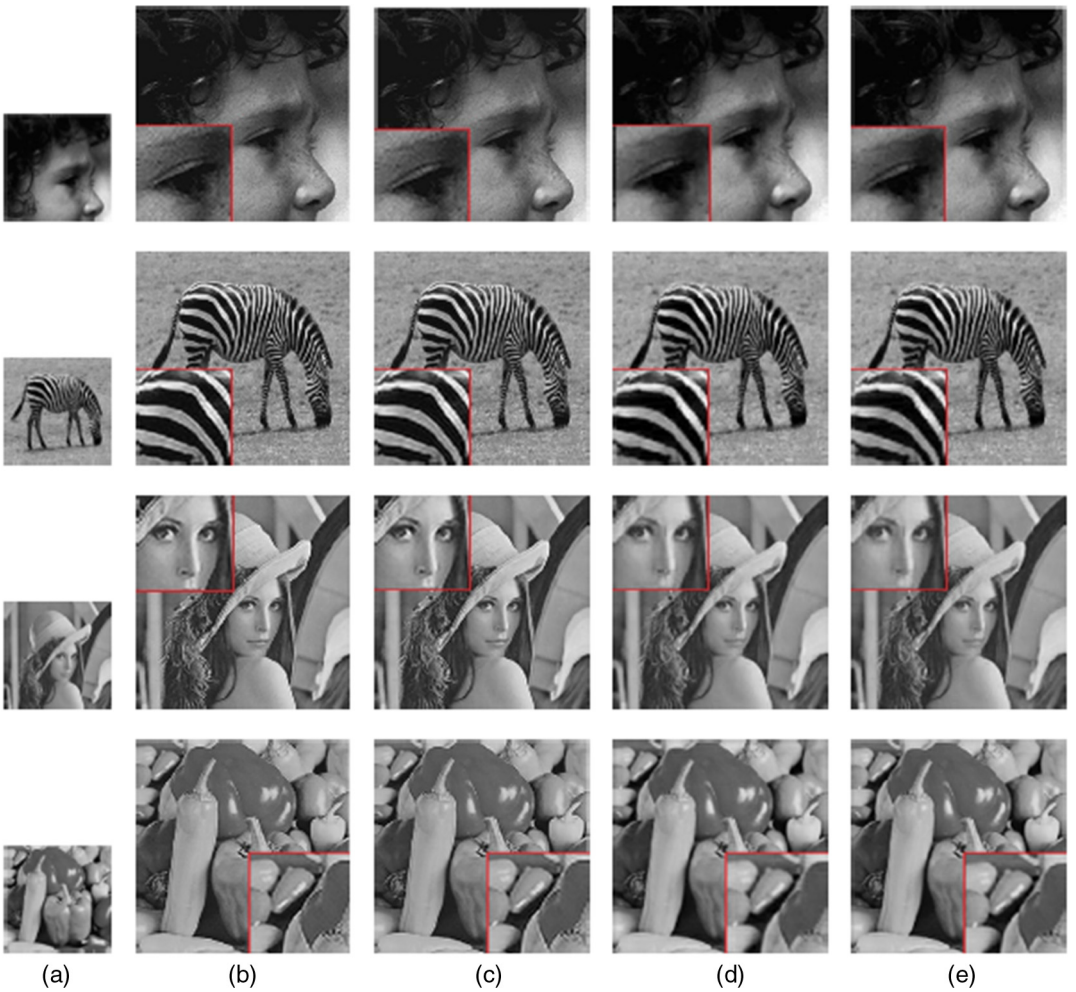

Fig. 3 Visual comparison of the proposed method against those of Ref. 31 and 29 with $\alpha = 2$: (a) the original LR image, (b)–(e) the results of SRCNN,[31] Peleg and Elad's method,[29] the proposed method without mean shift and the proposed method with mean shift, respectively.

Table 5 Average PSNR results of applying SSIM approach of dictionary reduction on LR patches size of 4×4 and HR patches size of 8×8.

Reduction (%)	Average PSNR	No. of patches	PSNR difference to original dictionary
0	27.53	8,13,056	0
13	27.51	5,42,469	0.02
37	27.48	5,12,472	0.05
40	27.475	4,87,010	0.055
46	27.45	4,42,799	0.08
51	27.4	4,02,128	0.13
56	27.35	3,60,958	0.18
67	27.24	2,72,202	0.29
73	27.18	2,22,830	0.35

Table 6 Average PSNR results of applying SSIM approach of dictionary reduction on LR patches size of 8×8 and HR patches size of 16×16.

Reduction (%)	Average PSNR	No. of patches	PSNR difference to original dictionary
0	26.77	1,37,984	0
13	26.77	1,20,736	0
15	26.76	1,16,997	0.01
18	26.76	1,13,528	0.01
22	26.76	1,07,136	0.01
27	26.72	1,00,201	0.05
33	26.69	92,435	0.08

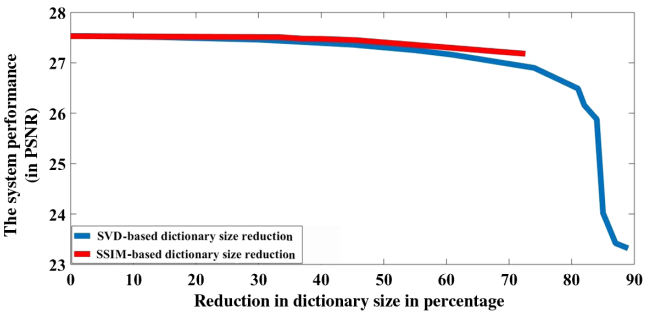

Fig. 4 Effect of dictionary reduction on performance of the algorithm.

A similar observation is seen in Table 7 and Fig. 4 when dictionary reduction is based on the SVD approach. It shows that a significant drop in performance of the system happens only when ∼80% of the size of the dictionary is reduced. Figure 5 shows a visual comparison between using the original size of the dictionary and a reduced version of that using two approaches of SSIM-based dictionary size reduction and SVD-based dictionary size reduction.

Since the dictionary size is reduced by 73%, the speed of the system compared with the original size of dictionary is improved by 62.5%

Table 7 Average PSNR results of applying SVD approach of dictionary reduction on LR patches size of 4×4 and HR patches size of 8×8.

Performance reduction (in %)	Average PSNR	No. of patches	PSNR difference to original dictionary
0	27.53	8,13,056	0
3	27.53	7,87,623	0
14	27.51	6,96,376	0.02
30	27.46	5,67,158	0.07
45	27.36	4,45,066	0.17
55	27.25	3,66,260	0.28
61	27.16	3,13,891	0.37
71	26.96	2,37,120	0.57
74	26.9	2,11,181	0.63
81	26.49	1,57,728	1.04
82	26.16	1,42,412	1.37
84	25.88	1,31,880	1.65
85	24.02	1,22,298	3.51
87	23.42	1,04,083	4.11
89	23.32	92,646	4.21

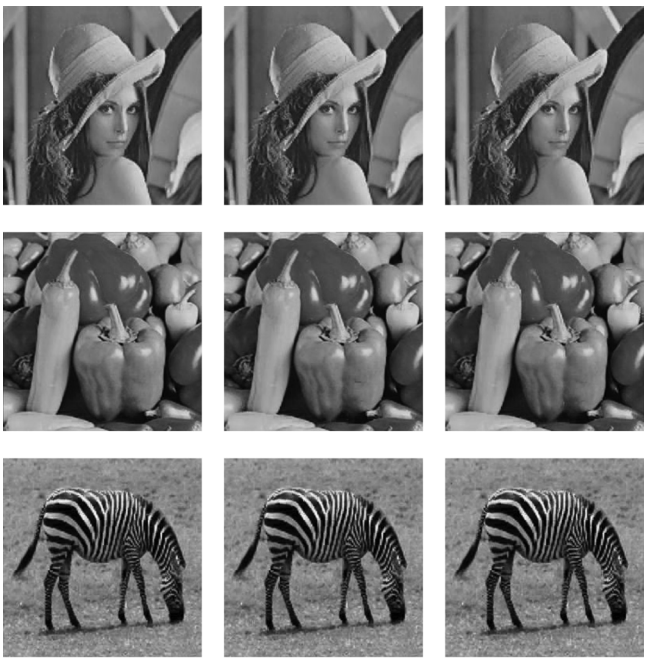

Fig. 5 Visual comparison of using the original size of the dictionary against those of SVD-based dictionary size reduction and SSIM-based dictionary size reduction algorithms where ∼56% of dictionary size is reduced. (a) The super-resolved image using the original size dictionary, (b) and (c) the results of SSIM-based dictionary size reduction and SVD-based dictionary size reduction.

4 Conclusion

In this work, a single-image SR method based on generating a dictionary from pairs of HR and their corresponding LR images was proposed. First, HR and LR pairs were divided into patches to make HR and LR dictionaries, respectively. The initial HR representation of an input LR image was calculated by merging the HR patches from the HR dictionary with those LR patches that have the closest distance to the patches of the input LR image. Each selected HR patch was processed further by passing through illumination enhancement processing to reduce the noticeable change of illumination between neighbor patches in the super-resolved image. To reduce the blocking effect, an average of the obtained SR image and the bicubic interpolated image was calculated. The quantitative and qualitative analyses of the experimental results showed the superiority of the proposed technique over conventional and state-of-the-art techniques.

Acknowledgments

This work has been partially supported by Estonian Research Grant (PUT638), the Estonian Centre of Excellence in IT (EXCITE) funded by the European Regional Development Fund.

References

1. P. Milanfar, *Super-Resolution Imaging*, CRC Press, New Year (2010).
2. K. Nasrollahi and T. B. Moeslund, "Super-resolution: a comprehensive survey," *Mach. Vision Appl.* **25**(6), 1423–1468 (2014).
3. H. Demirel and G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition," *IEEE Trans. Image Process.* **20**(5), 1458–1460 (2011).
4. K. Nasrollahi et al., "Deep learning based super-resolution for improved action recognition," in *Int. Conf. on Image Processing Theory, Tools and Applications (IPTA '15)*, pp. 67–72, IEEE (2015).

5. P. Rasti et al., "Improved interpolation kernels for super resolution algorithms," in *Int. Conf. on Image Processing Theory, Tools and Applications (IPTA '16)*, IEEE (2016).
6. J. Yang and T. Huang, "Image super-resolution: historical overview and future challenges," in *Super-Resolution Imaging*, pp. 20–34 (2010).
7. P. Rasti, H. Demirel, and G. Anbarjafari, "Image resolution enhancement by using interpolation followed by iterative back projection," in *Signal Processing and Communications Applications Conf. (SIU)*, pp. 1–4, IEEE (2013).
8. T. S. Huang and R. Y. Tsay, "Multiple frame image restoration and registration," in *Advances in Computer Vision and Image Processing*, pp. 317–339 (1984).
9. F. Li, X. Jia, and D. Fraser, "Universal HMT based super resolution for remote sensing images," in *15th IEEE Int. Conf. on Image Processing (ICIP '08)*, pp. 333–336, IEEE (2008).
10. P. Rasti et al., "Wavelet transform based new interpolation technique for satellite image resolution enhancement," in *IEEE Int. Conf. on Aerospace Electronics and Remote Sensing Technology (ICARES '14)*, pp. 185–188, IEEE (2014).
11. M. Cristani et al., "Distilling information with super-resolution for video surveillance," in *Proc. of the ACM 2nd Int. Workshop on Video Surveillance and Sensor Networks*, pp. 2–11, ACM (2004).
12. F. C. Lin et al., "Investigation into optical flow super-resolution for surveillance applications," in *APRS Workshop on Digital Image Computing: Pattern Recognition and Imaging for Medical Applications*, B. C. Lovell and A. J. Maeder, Eds., pp. 73–78, The University of Queensland, Brisbane (2005).
13. P. Rasti et al., "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *Int. Conf. on Articulated Motion and Deformable Objects*, pp. 175–184, Springer (2016).
14. T. Uiboupin et al., "Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring," in *Signal Processing and Communication Application Conf. (SIU '16)*, pp. 437–440, IEEE (2016).
15. J. A. Kennedy et al., "Super-resolution in pet imaging," *IEEE Trans. Med. Imaging* **25**(2), 137–147 (2006).
16. J. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.* **2**(1), 1–36 (1998).
17. K. Malczewski and R. Stasinski, "Toeplitz-based iterative image fusion scheme for MRI," in *15th IEEE Int. Conf. on Image Processing (ICIP '08)*, pp. 341–344, IEEE (2008).
18. S. Peled and Y. Yeshurun, "Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging," *Mag. Reson. Med.* **45**(1), 29–35 (2001).
19. D. Capel and A. Zisserman, "Super-resolution enhancement of text image sequences," in *Proc. 15th Int. Conf. on Pattern Recognition*, Vol. **1**, pp. 600–605, IEEE (2000).
20. P. H. Hennings-Yeomans, S. Baker, and B. V. Kumar, "Recognition of low-resolution faces using multiple still images and multiple cameras," in *2nd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS '08)*, pp. 1–6, IEEE (2008).
21. O. G. Sezer, Y. Altunbasak, and A. Ercil, "Face recognition with independent component-based super-resolution," *Proc. SPIE* **6077**, 607705 (2006).
22. B. K. Gunturk et al., "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Image Process.* **12**(5), 597–606 (2003).
23. K. Nasrollahi and T. B. Moeslund, "Finding and improving the key-frames of long video sequences for face recognition," in *Fourth IEEE Int. Conf. on Biometrics: Theory Applications and Systems (BTAS '10)*, pp. 1–6, IEEE (2010).
24. G. Anbarjafari and H. Demirel, "Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image," *ETRI J.* **32**(3), 390–394 (2010).
25. H. Demirel and G. Anbarjafari, "Discrete wavelet transform-based satellite image resolution enhancement," *IEEE Trans. Geosci. Remote Sens.* **49**(6), 1997–2004 (2011).
26. J. Yang et al., "Image super-resolution as sparse representation of raw image patches," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, IEEE (2008).
27. J. Yang et al., "Image super-resolution via sparse representation," *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010).
28. R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Int. Conf. on Curves and Surfaces*, pp. 711–730, Springer, Berlin, Heidelberg (2010).
29. T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.* **23**(6), 2569–2582 (2014).
30. Z. Zhu et al., "Fast single image super-resolution via self-example learning and sparse representation," *IEEE Trans. Multimedia* **16**(8), 2178–2190 (2014).
31. C. Dong et al., "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
32. G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.* **21**(5), 2481–2499 (2012).
33. G. Anbarjafari, "An objective no-reference measure of illumination assessment," *Meas. Sci. Rev.* **15**(6), 319–322 (2015).
34. G. B. Huang et al., "Learning to align from scratch," in *Conf. on Neural Information Processing Systems (NIPS)* (2012).
35. "Standard test image," http://www.imageprocessingplace.com/root_files_V3/image_databases.htm (04 May 2016).
36. Z. Wang et al., "Deep networks for image super-resolution with sparse prior," in *Proc. of the IEEE Int. Conf. on Computer Vision*, pp. 370–378 (2015).

Pejman Rasti received his MSc degree from the Electrical and Electronic Engineering Department of Eastern Mediterranean University in 2014. He has been working in the field of image processing, and he is currently involved in research projects related to static image and video sharpening and super-resolution. He is currently studying as a PhD student in iCV Group at University of Tartu.

Kamal Nasrollahi received his MSc in computer engineering and his PhD in electrical engineering from Tehran Polytechnic, Iran, 2007 and Aalborg University, Denmark, 2010, respectively. He is currently employed as an associate professor in the Visual Analysis of People Lab at Aalborg University. He has been involved in four national and international research projects. His research interests include facial analysis systems, biometrics recognition, and inverse problems.

Olga Orlova received her MSc degree in applied mathematics from Tallinn University of Technology in 2014. Her scientific interests concern approximation theory, in particular sampling operators. She is currently studying as a PhD student at Tallinn University of Technology.

Gert Tamberg received the degree in technical physics (applied mathematics) from Tallinn University of Technology 2004. From 2005 to 2009 he was a lecturer, from 2009 senior researcher at Tallinn University of Technology. His scientific interests concern approximation theory, in particular sampling operators. He is a member (from 2009 board member) of the Estonian Mathematical Society, AMS, EMS, IEEE, and SIAM.

Thomas B. Moeslund received his PhD and MSc degrees from Aalborg University in 2003 and 1996, respectively. He is currently the head of the Visual Analysis of People Lab and the head of the Media Technology Section both at Aalborg University, Denmark. His research interest is focused on all aspects of automatic analysis of images and video data especially imagery involving people. He has published 5 books and more than 120 journal and conference papers.

Gholamreza Anbarjafari heads the intelligent computer vision (iCV) research group in the Institute of Technology, University of Tartu, Estonia. He is an IEEE senior member and the vice chair of the Signal Processing/Circuits and Systems/Solid-State Circuits Joint Societies Chapter of the IEEE Estonian section. He has published over 100 scientific works. He has been the TCP of ICOSST, ICGIP, SampTA, and SIU.