This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

Multi-level Modified Finite Radon Transform Network for Image Upsampling

Yang Zhao, Ronggang Wang, *Member, IEEE,* Wenmin Wang, and Wen Gao, *Fellow, IEEE*

***Abstract*—Local line-like feature is the most important discriminate information in image upsampling scenario. In recent example-based upsampling methods, grayscale and gradient features are often adopted to describe the local patches, but these simple features cannot accurately characterize complex patches. In this paper, we present a feature representation of local edges by means of a multi-level filtering network, namely multi-level modified finite Radon transform network (MMFRTN). In the proposed MMFRTN, the modified finite Radon transform (MFRT) is utilized in the filtering layer to extract the local line-like feature; the nonlinear layer is set to be a simple local binary process; for the feature pooling layer, we concatenate the mapped patches as the feature of local patch. Then we propose a new example-based upsampling method by means of the MMFRTN feature. Experimental results demonstrate the effectiveness of the proposed method over some state-of-the-art methods.**

***Index Terms*—image upsampling, super-resolution, PCANet**

I. Introduction

IMAGE upsampling, also known as single-image super-resolution or image upscaling, is a fundamental task for various computer vision and image processing applications, such as image editing, UHDTV/HDTV, medical image processing, and object recognition. In many scenarios, only low-resolution image (LRI) is available. Hence, recovering a high-resolution image (HRI) from LRI is immensely demanded.

Traditional interpolation-based upsampling methods [1], [2] often produce some unnatural artifacts, e.g., blurring, jaggy, and ringing effects. To suppress these artifacts, many methods have been proposed in recent years. In general, these methods can be summarized into two categories: edge-directed methods and example-based methods.

It is known that edges are more obvious than other areas for image perception [3], edge-directed methods aim to recover sharp and natural edges by enforcing some edge knowledge or extra constraints, such as smooth edge [4], geometric regularity [5], gradient prior [6]-[8], [46], local structure constraint [32], and different interpolated grids [9]-[12]. There were also some edge-enhancing algorithms that sharpen edges via different treatments, e.g., contrast enhancement [13], de-blurring and de-convolution [14], [33], and edge energy term constraint [15].

This work was partly supported by the grant of National Science Foundation of China 61370115, 61402018, 61305006, 61175022, China 863 project of 2015AA015905, the grant of China Postdoctoral Science Foundation, No. 2014M550016, Shenzhen Fundamental Research Project, and Shenzhen Peacock Plan.

The authors are with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, 2199 Lishui Road, Shenzhen 518055, China (e-mail: zyknight@mail.ustc.edu.cn; rgwang@pkusz.edu.cn; wangwm@pkusz.edu.cn; wgao@pku.edu.cn).

Example-based methods often utilize high-resolution (HR) samples to get the co-occurrence prior between HR and low-resolution (LR) image patches. This type of method was first proposed in [16] and was further developed in [17]-[21], and [36]-[45]. Chang *et al.* [36] proposed a super-resolution method based on neighbor embedding and the manifold learning method locally linear embedding (LLE). In [17], Yang *et al.* proposed an effective learning method based on sparse representation and achieved impressive results. Many sparse representation based methods were then proposed, such as bilevel dictionary learning [18], beta process joint dictionary learning [40], statistical prediction model based method [41], non-locally centralized sparse representation [42], and sub-dictionaries method [19]. In [20] and [21], single-image local self-example-based methods were introduced by considering the similarities of local patches in different scales of the same image. Recently, Turkan *et al.* [38], [39] presented super-resolution methods by means of neighbor embedding and single-image local-self-exemplar.

In example-based upsampling scenario, the feature of LR patch is utilized as "index" to search corresponding exemplars from the dictionary. Extracting accurate local feature is thus very important for image upsampling. However, most recent example-based methods merely utilize simple features, such as mean-value-removed grayscale, contrast-normalized grayscale [16], first- and second-order derivative [36], [17]-[19], [40]-[42], and applying PCA dimensionality reduction to the first- and second-order derivative [43], [45]. These simple features cannot well characterize complex local edges due to their limited discriminative capability. Therefore example-based methods with these simple features may produce unnatural artifacts around some complex edges such as zigzag edges and coastline-like edges. Is there better feature representation of local patch for image upsampling? This is the main issue to be addressed in this paper.

Numerous algorithms have been proposed to extract local features, such as LBP [22], SIFT [23], and HOG [24]. Recently, deep networks model drew significant attention, e.g., deep neural networks (DNNs) [25], convolutional deep neural network (ConvNet) [26], [27], and wavelet scattering networks (ScatNet) [28]. Later, Chan *et al.* [29] proposed a simple two-level PCA net (PCANet) and achieved state-of-the-art performance in several challenging vision tasks. These deep networks tried to extract multi-level representation with the idea that higher-level features could represent more abstract semantics of the data. However, these networks were not designed to extract specified feature for image upsampling scenario, and we thus proposed a specific line-filter based

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

network to characterize the local edges for image upsampling.

In example-based upsampling methods, similar known exemplars are utilized to reconstruct the HR edges. Hence an accurate representation of line-like feature is very important. Motivated by effective ConvNet and PCANet, we propose a multi-level modified finite Radon transform network (MMFRTN) to represent the line-like feature. In the proposed MMFRTN, the modified finite Radon transform (MFRT) [30] is utilized instead of convolution filter bank or PCA filters in filter layer; the nonlinear layer is set to be a simple local binary process; for the feature pooling layer, we use the concatenated mapped-patches as the local patch feature. Then we propose a new example-based upsampling method by means of the MMFRTN feature and locally linear embedding. Experimental results demonstrate the effectiveness of the proposed method.

Overall, The main technical contributions of this work are summarized as follows.

1) We demonstrate that better feature representation of local edge can improve image upsampling result. Recent example-based upsampling methods have paid less attention to explore better local features. In this paper, a filter network based feature is proposed for image upsampling, and this feature is proven to be more effective than traditional local features.
2) We propose an effective MFRT-filter-based network to extract local line-like feature. In image upsampling scenario, the extracted feature is supposed to accurately characterize the local edge, such as the location, direction, shape, and acutance. The MFRT can be regarded as line filter bank with bandwidth of 1, and it is very sensitive to the tiny edges and borderlines. As a result, the proposed MMFRTN can reconstruct sharp and clear HR edges.

The following paragraphs of this paper are organized as follows. Section II briefly reviews the MFRT, and Section III presents the proposed MMFRTN. Section IV describes the proposed MMFRTN based image upsampling method. Section V shows the experimental results and this paper is concluded in Section VI.

II. Brief review of MFRT

The Radon transform was first proposed by Radon in 1917. Radon transform accentuates linear features by integrating image intensity along all possible lines, thus it can be used to detect linear trends in the image. Afterward, Matus and Flusser proposed the finite Radon transform (FRT) for finite length signals [31]. In order to eliminate the wrap around effect caused by the modulo operations in FRT, Huang *et al.* [30] proposed a modified finite Radon transform (MFRT) by removing the modulo operations. The MFRT was defined as follows.

Denoting $Z_p = \{0, 1, ..., p-1\}$, where p is a positive integer, the MFRT of real function $f[x, y]$ on the finite grid Z_p^2 is defined by,

$$r[L_k] = MFRT_f(k) = \sum_{i,j \in L_k} f[i, j] \tag{1}$$

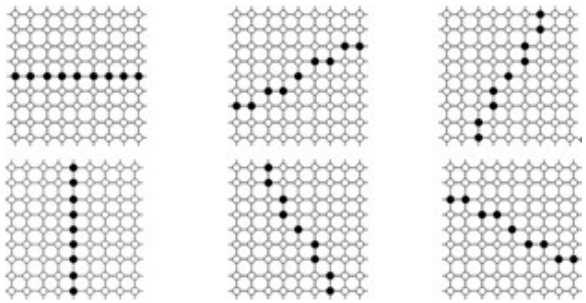

Fig. 1. 9×9 MFRT in six directions.

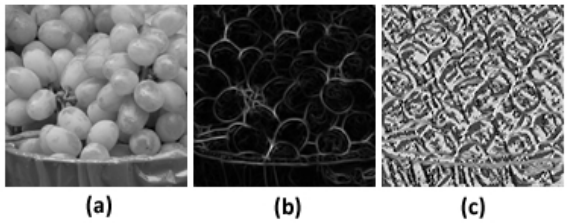

Fig. 2. An example of MFRT filtering results, (a) input image, (b) MFRT magnitude map, (c)MFRT phase map.

where L_k denotes the set of points that make up a line within the lattice Z_p^2, which means:

$$L_k = \{(i, j) : j = S_k(i - i_0) + j_0, i \in Z_p\} \tag{2}$$

where (i_0, j_0) denotes the center point coordinate of the lattice Z_p^2, and k means the index value corresponding to the slope of S_k. That is to say, different k denotes different slopes of L_k. For any given k, the summation of the line that passes through the center point (i_0, j_0) of Z_p^2, is calculated. It should be pointed out that all lines in different directions have the same number of pixels.

In MFRT, the direction θ_k and the energy e of the center point $f(i_0, j_0)$ within the lattice Z_p^2 are calculated by the following formulas,

$$\theta_{k(i_0,j_0)} = arg(min_k(r[L_k])), k = 1, 2, ..., N \tag{3}$$

$$e_{(i_0,j_0)} = |min_k(r[L_k])|, k = 1, 2, ..., N \tag{4}$$

In this paper, the MFRT is treated as a set of linear filters along different directions, as illustrated in Fig. 1. After the MFRT filtering, we can get the following two maps: the MFRT magnitude map consisting of the energy e at each pixel, and the MFRT phase map consisting of the direction θ_k at each pixel. Fig. 2 shows an example of MFRT magnitude map and MFRT phase map, from which we can see that the MFRT can well characterize edges and local line-like features.

III. The proposed multi-level MFRT network

As illustrated in Fig.3, the proposed MMFRTN consists of three stages, i.e., the multi-level filtering stage, the local binary quantization and mapping stage, and the final feature output stage. We describe each component of the MMFRTN in follows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

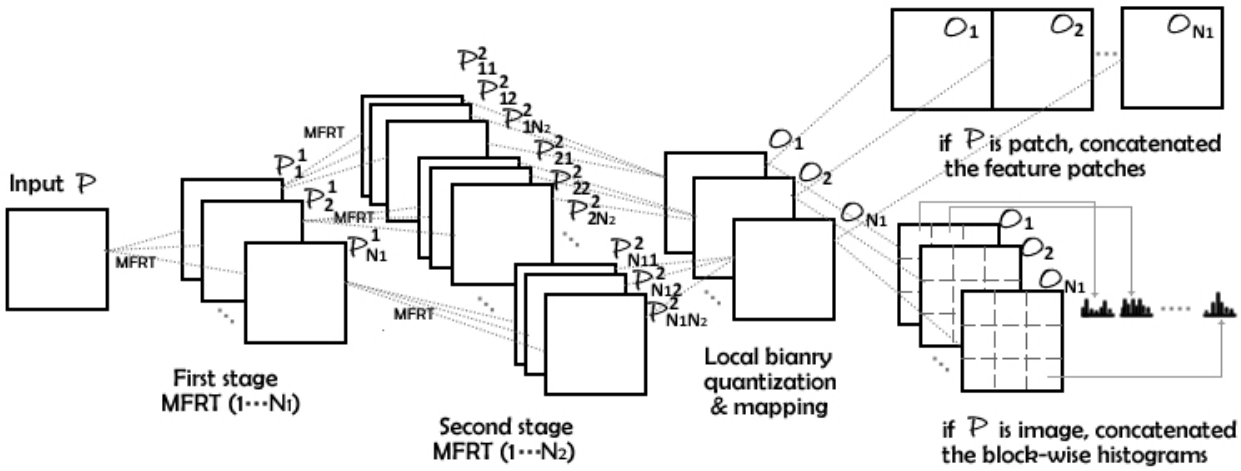

Fig. 3. The illustration of multi-level MFRT network.

A. The Multi-level MFRT filtering stages

Supposing that P is an input with the size of $m \times n$. The first stage of MMFRTN is MFRT filtering with different sizes of filter banks,

$$P_i^1 = MFRT_i(P), i = 1, 2, ..., N_1 \tag{5}$$

where $MFRT_i$ denotes i-th MFRT filtering process, N_1 is the total number of MFRT filter banks at the first stage, and P_i^1 is the filter response with corresponding i-th MFRT. For different index i, MFRT filter banks with different sizes are used. The size of MFRT depends to the size of input P. If P is a small image patch, the MFRT is set to a small size, e.g., 3×3, or 5×5, and ensure that the size of the filter is smaller than the size of input patch. Otherwise, the size of MFRT filter can be set much larger to contain more discriminative information, e.g., 11×11, or 27×27. The P is zero-padded before MFRT filtering so that the output P_i^1 has the same size with input P. Note that the MFRT filter has two kinds of responses, i.e., the magnitude map and the phase map. Here we don't distinguish these two responses since they can share the same MMFRTN framework.

The outputs of the first stage are,

$$P^1 = [P_1^1, P_2^1, ..., P_{N_1}^1] \in \mathscr{R}^{mnN_1}$$

The outputs P^1 are further inputted to the second stage. The second stage repeats the same filtering process as the first stage. For each input P_i^1 of the second stage, the outputs are calculated by,

$$P_{ij}^2 = MFRT_j(P_i^1), j = 1, 2, ..., N_2 \tag{6}$$

where N_2 is the total number of MFRT filter banks at the second stage. As a result, the outputs of the second stage are:

$$P^2 = [P_{11}^2, P_{12}^2, ..., P_{1N_2}^2, P_{21}^2, P_{22}^2, ..., P_{2N_2}^2, ..., P_{N_11}^2, P_{N_12}^2, ..., P_{N_1N_2}^2] \in \mathscr{R}^{mnN_1N_2}$$

The number of outputs of the second stage is $N_1 \times N_2$. Similarly, the MFRT filtering process can be simply repeated to build more stages if a deeper network is required.

B. Local binary quantization and mapping stage

For each input P_i^1 of the second stage, the N_2 outputs $\{P_{ij}^2\}_{j=1}^{N_2}$ are produced after the MFRT filtering process. We first convert these real-valued outputs to N_2 binary maps, and then encode these N_2 binary maps to an integer-valued map.

Here we use a local binary quantizing and mapping method to convert the N_2 outputs $\{P_{ij}^2\}_{j=1}^{N_2}$ to an integer-valued map O_i as follows,

$$O_i = \sum_{j=1}^{N_2} 2^{j-1} LB(P_{ij}^2) \tag{7}$$

where $LB(\cdot)$ is a local binary process for the output P_{ij}^2,

$$LB(x) = \begin{cases} 1 & \text{if } x \geq \frac{1}{N_P} \sum_{p=1}^{N_P} x_p \\ 0 & \text{if } x < \frac{1}{N_P} \sum_{p=1}^{N_P} x_p \end{cases} \tag{8}$$

where x_p denotes the p-th $(p = 1, ..., N_P)$ neighbor pixel of the central pixel x, and N_P is the total number of the neighbor pixels. In this paper, the eight pixels around x are selected as the neighbor pixels. Comparing with the heaviside operator, local binary process has two benefits: first, the local quantization is more suitable for extracting the local feature; second, the local binary step is robust to global monotonous transform caused by illumination variance.

After the binary quantizing and mapping, the outputs of this stage are:

$$O = [O_1, O_2, ..., O_{N_1}] \in \mathscr{R}^{mnN_1}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

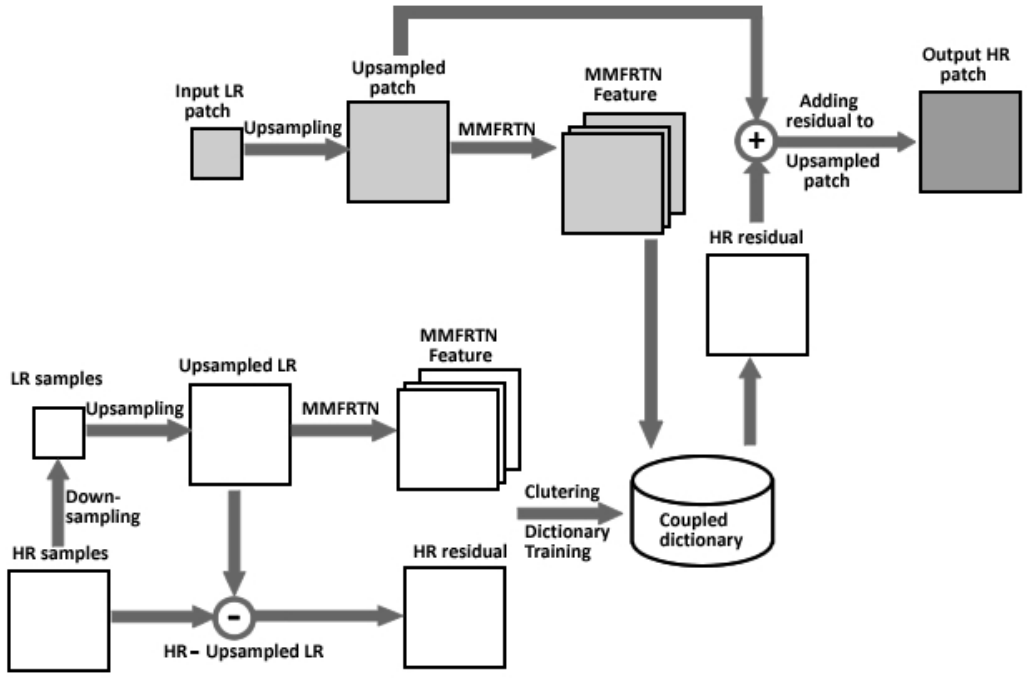

Fig. 4. The MMFRTN based image upsampling framework.

C. Output features

For image upsampling scenario, the given input P is a small patch. We thus extract the feature of local patch by directly concatenating the $O_i(i = 1, 2, ..., N_1)$ to form a new feature patch,

$$f_{patch} = [O_1, Q_2, ..., O_{N_1}] \in \mathscr{R}^{mnN_1}$$

Note that the proposed MMFRTN can also be applied in image classification scenarios by utilizing an entire image as input P. In this case, we partition each O_i into B blocks, and then concatenate the histograms of all blocks into one vector $Bhist(O_i)$ (with $2^{N_2}B$ bins). At last, all the histograms are connected to represent the global feature of an image [1],

$$f_{global} = [Bhist(O_1), Bhist(O_2), ..., Bhist(O_{N_1})]^T \in \mathscr{R}^{(2^{N_2})N_1 B}$$

IV. The proposed MMFRTN based image upsampling method

A. Upsampling Framework

As illustrated in Fig.4, the proposed upsampling framework contains two stages: a coupled dictionary learning stage in which MMFRTN feature/HR residual pairs are learned from training samples, and a reconstruction stage in which the input LR patches are reconstructed by means of the trained coupled dictionary.

[1]The proposed MMFRTN feature is also verified on a palm-print recognition database. More details and demo codes for image classification scenario are available at
http://zycv.890m.com/zyProjectPages/MMFRTNsupplementary.html

In the learning stage, various LR patches and corresponding HR patches are selected as training samples. The MMFRTN features of upsampled LR patches are used as the features of LR patches. The residuals between HR patches and upsampeld LR patches represent the difference between upsampled patches and real HR patches. Hence, we utilize the MMFRTN feature/HR residual pairs as the coupled examples. In the proposed method, we calculate over hundred thousands of MMFRTN feature/HR residual pairs from various training images, and the traditional bicubic interpolation is used to upsample these LR patches. Then these coupled examples are clustered to N_c centers by means of k-means clustering to form the coupled dictionary as shown in Fig.4. Note that the coupled example used in the clustering process is the concatenation of MMFRTN feature vector and HR residual vector. Hence, each clustered dictionary atom contains two parts, i.e., the MMFRTN feature and its corresponding HR residual.

In the reconstruction stage, the HR residual of input LR patch is reconstructed by means of corresponding MMFRTN feature and the coupled dictionary. The proposed method focuses on the feature extraction, and thus the MMFRTN feature can be applied to many dictionary-based reconstruction methods, such as simplest nearest neighbor (NN) matching, neighbor embedding [36], sparse representation [17], and sub-dictionaries [19].

In this paper, we utilize the reconstruction method based on locally linear embedding as in [36]. In the LLE based reconstruction methods [36],[38], and [39], the manifolds in LR and HR feature space are assumed to have similar local geometry, and thus the patches in the HR feature domain can be estimated as a weighted average of local neighbors using the same weights as in the LR feature domain.

Given a LR patch $\boldsymbol{y}$, the MMFRTN feature $F\boldsymbol{y}$ is firstly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

computed, and then its K-NN MMFRTN features in the dictionary are searched. The $F\boldsymbol{y}$ can be reconstructed as the weight average of its K-neighbors, and the optimal weights are achieved by minimizing the reconstruction errors, as follows,

$$\min_{\alpha} ||F\boldsymbol{y} - \boldsymbol{N}_l\alpha||^2 \quad s.t. \sum_{k=1}^{K} \alpha_k = 1 \tag{9}$$

where α denotes the K-weights, and $\boldsymbol{N}_l$ consists of K nearest neighbors of MMFRTN feature $F\boldsymbol{y}$. The optimal weights α can be calculated as in [36], and then the same weights are utilized to reconstruct the HR residual $\boldsymbol{x}$ in the HR space,

$$\boldsymbol{x} = \boldsymbol{N}_h\alpha \tag{10}$$

where $\boldsymbol{N}_h$ consists of K HR residuals corresponding to K-NN MFRTN features in the coupled dictionary. At last, the reconstructed HR residual $\boldsymbol{x}$ is added to the bicubic interpolated LR patch to obtain the final HR patch.

Note that the MMFRTN feature is encoded by a set of binary maps. Therefore the Hamming distance is used to measure the distance between two MMFRTN features, which can be calculated as,

$$d = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} P_a(i,j) \otimes P_b(i,j)}{mn} \tag{11}$$

where P_a and P_b denote two image patches with $m \times n$ size, and $\otimes$ is a XOR operation of corresponding binary codes.

B. Implementation Details

For image upsampling, the input of MMFRTN P is a local patch, and the output of MMFRTN is the feature patch f_{patch} accordingly. In MFRT filtering stages, we use the filter banks with two sizes of 3×3 and 5×5, and for each size, the magnitude maps and phase maps are all used. When the size of MFRT is 3×3, the filter bank consists of linear filters along the 4 different directions of $0°, 45°, 90°$, and $135°$. When MFRT size is 5×5, the filter bank consists of filters along the 8 different directions of $0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°$, and $157.5°$. Hence, suppose an upsampled LR patch $P \in \mathscr{R}^{mn}$ is given, the outputs of the first and second MFRT filtering stages are $P^1 \in \mathscr{R}^{4mn}$ and $P^2 \in \mathscr{R}^{16mn}$ respectively. The final output of MMFRTN is $f_{patch} \in \mathscr{R}^{4mn}$.

In order to extract the local feature in a small scale, the filter size is set as 3 or 5. Thus the size of input P should be equal or greater than 5×5. Fig.5 shows the average PSNR on ten test images with different patch sizes. For 2X magnification, it can be found that the PSNR decreases along with the patch size increases, and the highest PSNR is obtained with the smallest patch size of 5×5. When the upsampling factor increases to 3 or 4, the patch sizes need to be set larger to contain enough information of local patch. In this paper, the sizes of upsampled LR patches and HR patches are set as 5×5, 8×8, 12×12, and 24×24 for 2X, 3X, 4X, and 8X magnifications, respectively.

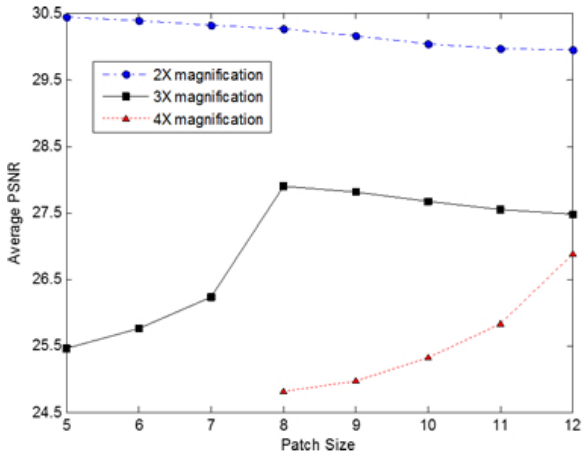

Fig. 5. Average PSNR values with different patch sizes.

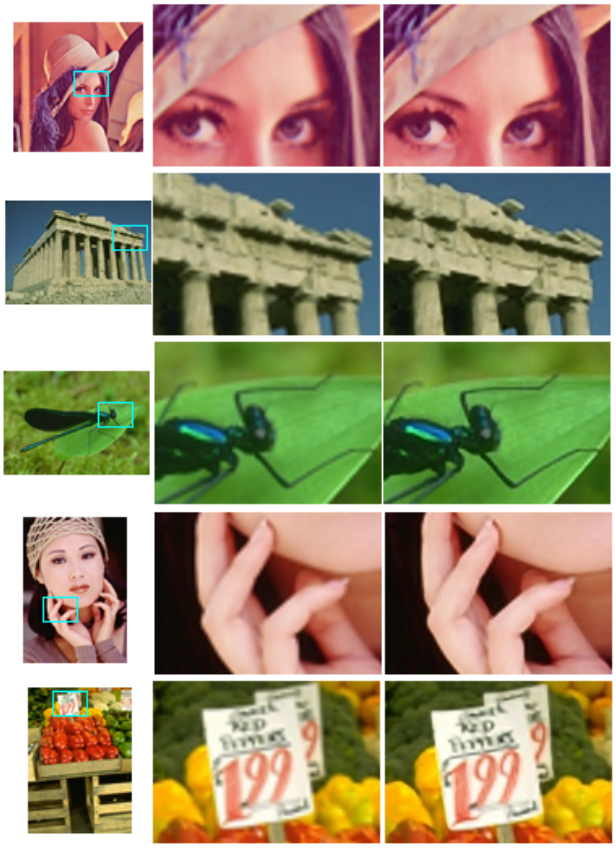

Fig. 6. Illustration of upsampled results with our method. The left column shows the selected close-up areas; the middle column shows the close-ups upsampled with the bicubic interpolation; the right column shows the close-ups upsampled with the proposed method.

V. Experiments

A. Training and testing image sets

In our experiment, we use the same training set of images proposed by Yang *et al*. [17] for the proposed method and other example-based methods. The training set contains 90 natural images downloaded from the Web, and each image has been downsampled to obtain its LR image. In this paper,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

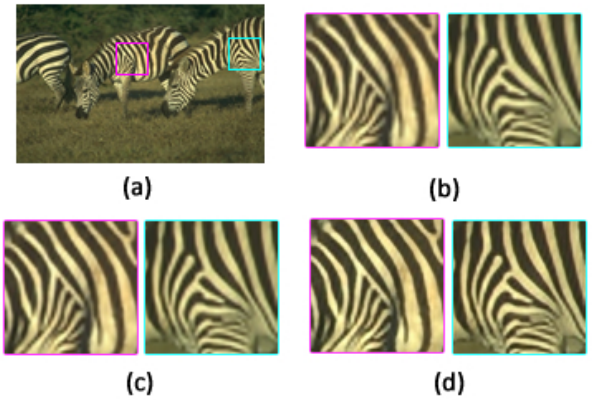

Fig. 7. Upsampled results of the "BDS_16068" image, (a) the LRI and two selected close-up areas, (b) close-ups with bicubic interpolation, (c) close-ups with PCANet based method, (d) close-ups with the MMFRTN.

100000 local patches are randomly chosen from the training set to compute the MMFRTN feature/HR residual pairs. Total 1200 coupled-dictionary atoms are then clustered from these 100000 MMFRTN feature/HR residual pairs.

We compare the proposed method with some other methods on four testing image sets, i.e., an image set consists of some selected natural images, and three image sets used in other literatures of "Set 14" [43], "Set 5" [44], and "B100" [45]. The first image set contains 5 natural images selected from the Web and other 24 images randomly selected form the Berkeley image dataset [34] [2]. "Set 14" [43] and "Set 5" [44] contain 14 and 5 commonly used images respectively for upsampling evaluation. "B100" [45] consists of 100 testing images selected from the Berkeley image dataset.

For color image, it is firstly converted from RGB to YUV. The proposed method and other methods in the comparison are then applied only on Y (intensity) component, and bicubic interpolation is used for U and V components. In our experiments, the input LRIs are obtained by downsampling the original HRIs with bicubic interpolation and the LRIs are then upsampled to their original size with different methods in the comparison respectively. The upsampling factors in our experiment are set as 2, 3, and 4.

B. Compared methods

In this paper, we compare the MMFRTN based upsampling with the traditional bicubic interpolation, and three example-based methods of the Chang's LLE [36], the ScSR [17] and the ASDS [19]. We also apply the PCANet to image upsampling with the same framework as the proposed MMFRTN based method. In PCANet based upsampling, we also utilize the local feature patches to take place of the global histogram feature. The parameters of the PCANet are set according to [29].

C. Experimental results

Fig.6 compares the upsampled results with the proposed MMFRTN based method and traditional bicubic interpolation on different images. Obviously, the proposed method obtains sharper and clearer edges.

Fig. 7 illustrates the comparison of bicubic interpolation, PCANet based upsampling and the proposed method on an image selected from Berkeley dataset. (In this paper, images selected from Berkeley dataset are denoted as "BDS_xxx".) We can find that bicubic interpolation causes blur and jaggy edges. The PCANet based upsampling recovers sharper edges than bicubic interpolation, but there are still slightly jaggy artifacts around the thin lines. The proposed method reproduces sharper and smoother edges than PCANet based method. This also demonstrates the proposed MMFRTN is more effective to extract the local line-like feature.

Fig.8 shows the upsampled results of "BDS_12003" image with different methods for 2X magnification. By comparing the edges marked in the red squares, we can get the following findings. Firstly, the bicubic interpolation reproduces blurry edges and texture. Secondly, the example-based methods of the LLE, the ScSR, the ASDS, the PCANet based method, and the MMFRTN based method can recover sharper edges than bicubic by means of extra HR examples. Comparing to these state-of-the-art example-based methods, the proposed method reproduces finer and clearer sharp edges. Fig. 9 illustrates 2X magnification results of "BDS_15004" image with these methods. We can get similar findings to Fig.8 by comparing the sharp edge areas marked with the red squares. The proposed MMFRTN still recovers sharper and clearer

TABLE I
PSNRS FOR THE UPSAMPLED IMAGE (2X) WITH DIFFERENT METHODS (DB)

Images	Bicubic	LLE	PCANet	ScSR	ASDS	Ours
Lenna	35.42	36.69	36.77	36.75	36.18	**37.25**
Zebra	26.11	28.45	28.61	28.52	28.70	**29.40**
Face	34.63	35.32	35.39	35.36	34.66	**35.76**
Girl	34.85	35.36	35.42	35.37	34.52	**35.72**
Pirate	33.29	34.35	34.44	34.40	33.81	**34.90**
BDS_16068	27.26	28.50	28.57	28.49	28.53	**29.62**
BDS_2018	26.44	27.62	27.69	27.64	27.40	**28.12**
BDS_15004	27.44	28.50	28.71	28.58	28.40	**29.26**
BDS_22093	26.73	27.49	27.63	27.54	27.42	**28.06**
BDS_23025	31.29	31.96	32.11	32.00	31.95	**32.59**
BDS_24004	25.60	26.18	26.27	26.21	26.24	**26.61**
BDS_25098	30.17	31.62	31.70	31.65	31.65	**32.39**
BDS_35058	39.73	40.83	40.88	40.83	39.69	**41.27**
BDS_35070	36.35	38.02	38.12	38.09	37.37	**38.47**
BDS_12003	31.11	32.39	32.54	32.46	31.81	**32.92**
BDS_67079	28.42	29.20	29.37	29.23	28.78	**29.57**
BDS_138078	29.09	30.19	30.27	30.25	30.10	**30.70**
BDS_302003	32.93	35.07	35.21	35.16	34.64	**35.64**
BDS_3063	36.09	37.98	38.09	38.00	37.29	**38.86**
BDS_3096	39.63	42.41	42.52	42.47	41.32	**42.82**
BDS_12074	33.82	35.28	35.41	35.34	34.76	**35.89**
BDS_35010	29.78	31.71	31.86	31.72	31.09	**32.63**
BDS_42049	31.65	34.33	34.44	34.36	33.79	**35.58**
BDS_56028	26.50	27.54	27.71	27.61	27.42	**28.42**
BDS_76002	28.16	29.00	29.11	29.03	28.71	**29.54**
BDS_86000	27.92	28.97	29.09	29.02	28.82	**29.55**
BDS_101085	25.74	26.32	26.42	26.33	26.24	**26.82**
BDS_102061	28.14	28.71	28.85	28.76	28.54	**29.43**
BDS_108082	30.13	30.80	30.97	30.82	30.53	**31.54**
Average	30.84	32.10	32.21	32.13	31.74	**32.74**
Average Gain	1.90	0.64	0.53	0.61	1.00	

[2]Image datasets and source code of the MMFRTN can be downloaded from the following website: http://zycv.890m.com/zyProjectPages/MMFRTN.html

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

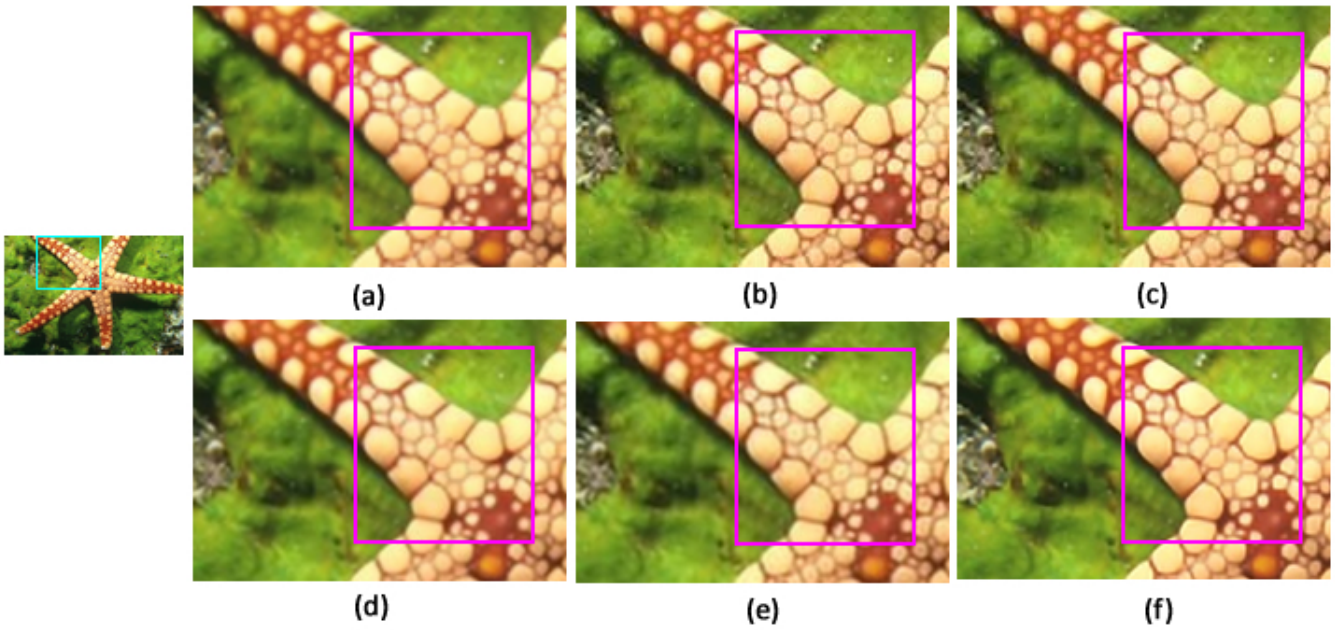

Fig. 8. Upsampled results of the "BDS_12003" image, (a) with bicubic interpolation, (b) with the LLE [36], (c) with the PCANet [29], (d) with the ScSR [17], (e) with the ASDS [19], (f) with the MMFRTN.

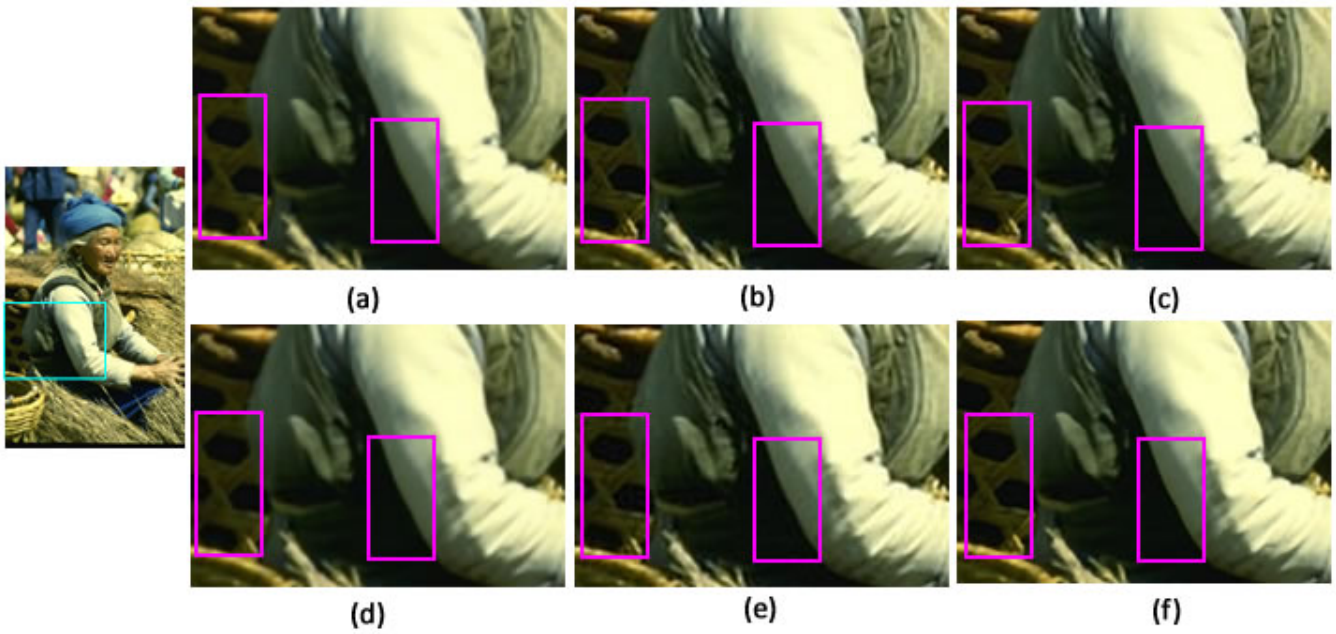

Fig. 9. Upsampled results of the "BDS_15004" image, (a) with bicubic interpolation, (b) with the LLE [36], (c) with the PCANet [29], (d) with the ScSR [17], (e) with the ASDS [19], (f) with the MMFRTN.

edges than these methods.

As aforementioned, simple gradient-based feature cannot well model complex edges due to its limited discriminative capability. Fig. 10 illustrates some close-ups of zigzag edges upsampled by the ScSR, the ASDS, and the MMFRTN. The first- and second-order derivative is used as the local feature in the ScSR and the ASDS. This feature is simply obtained by calculating the grayscale difference along vertical and horizontal directions and it cannot accurately characterize the complex local edges. By comparing the zigzag edge area marked with the red circles, we find that the ScSR and the ASDS reproduce some unnatural black spots, while the proposed MMFRTN can recover sharp and natural edges.

Fig. 11 compares the 3X magnification results of "monarch" image with different example-based methods. The example-based methods of the Chang's method, the ScSR and the ASDS can recover sharper edges than the bicubic interpolation, and the MMFRTN reproduces sharper and clearer edges

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

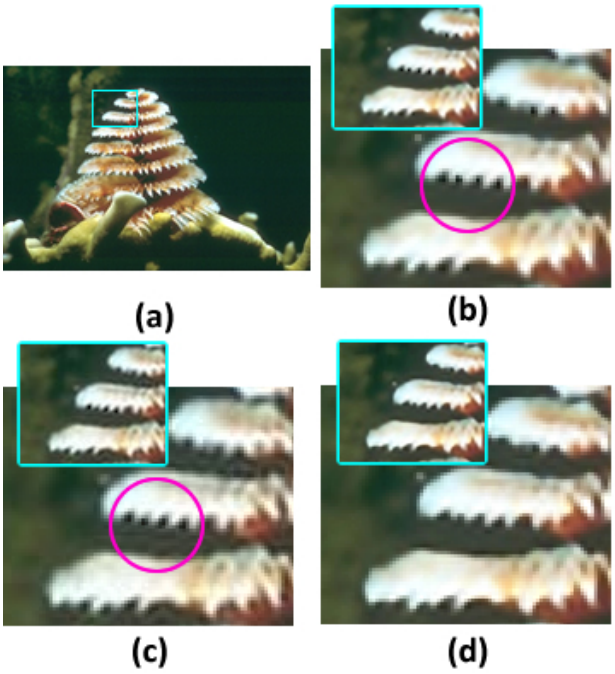

Fig. 10. Close-ups of upsampled results of the "BDS_12074" image (2X). (a) close-up area, (b) close-ups of upsampled result with ScSR [17], (c) close-ups of upsampled result with ASDS [19], (d) close-ups of upsampled result with the MMFRTN.

than these example-based methods.

Fig. 12 shows upsampled results of "zebra" image for 4X magnification. By comparing the streaks, it can be seen that the MMFRTN reconstructs sharper and clearer edges than other example-based methods. In order to visually compare the acutance of different results, we further zoom up edges and mark the width of the edge with two red arrows, as illustrated in Fig. 12.

Fig. 13 illustrates the upsampled result of the MMFRTN with large upsampling factor. The interpolated image is extremely blurred when the upsampling factor increases to 8, while the proposed MMFRTN still recovers much sharper results.

The 2X magnification results of different example-based methods on some natural images selected from web and Berkeley dataset are listed in Table I. The last row shows the average PSNR gains of the MMFRTN based method over other methods. We can see that the PSNR values of the proposed method are higher than other example-based methods.

Table II lists the average PSNR values of 2X, 3X, and 4X magnification on three image datasets, i.e., "Set14", "Set5", and "B100". We can get the following findings from Table II. Firstly, the MMFRTN still produces better results than the PCANet on these image datasets, and the proposed feature is thus more effective for image upsampling. Secondly, by comparing results of the Chang's LLE method and the MMFRTN based method, we find that the proposed MMFRTN feature can characterize local edges much better than traditional gradient-based feature. Finally, the proposed method always achieves higher average PSNR values than other methods for various upsampling factors and image datasets.

VI. Conclusion

In this paper, we proposed a multi-level modified finite Radon transform (MMFRTN) to extract the local line-like feature. The proposed MMFRTN consisted of three stages, i.e., the modified finite Radon transform (MFRT) based multi-level filtering stage, the local binary nonlinear mapping stage, and the local feature pooling stage. Then we presented a MMFRTN based image upsampling method by means of the MMFRTN feature and locally linear embedding based reconstruction. Experimental results demonstrated the effectiveness of the proposed method over some state-of-the-art methods.

Acknowledgment

The authors would like to sincerely thank A. Giachetti, Q. Shan, W. S. Dong, J. C. Yang, and T. H. Chan for sharing the source codes of the ICBI, the Shan's, the ASDS, the ScSR, and the PCANet.

References

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

[2] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049-1075, Nov. 1999.

[3] D. C. Van Essen, C. H. Anderson, and D. J. Felleman, "Information processing in the primate visual system: An integrated systems perspective," *Science*, vol. 255, no. 5043, pp. 419-423, 1992.

[4] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong "Soft edge smoothness prior for alpha channel super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1-8.

[5] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521-1527, Oct. 2001.

[6] R. Fattal, "Image upsampling via impose edge statistics," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007, Art. ID 95.

[7] J. Sun, Z. Xu, and H. Y. Shum, "Image super-resolution using gradient profile prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1-8.

[8] L. Wang, S. Xiang, G. Meng, H. Wu, and C. Pan, "Edge-directed single image super-resolution via adaptive gradient magnitude self-interpolation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1289-1299, Aug. 2013.

[9] D. Su and P. Willis, "Image interpolation by pixel-level data-dependent triangulation," *Comput. Graph. Forum*, vol. 23, no. 2, pp. 189-201, 2004.

[10] Q. Wang and R. K. Ward, "A new orientation-adaptive interpolation method," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 889-900, Apr. 2007.

[11] C. Zwart and D. Frakes, "Segment Adaptive Gradient Angle Interpolation," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 2960-2969, Aug. 2013.

[12] X. Liu, D. Zhao, R. Xiong, S. Ma, W. Gao, and H. Sun, "Image interpolation via regularized local linear regression," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3455-3469, Dec. 2011.

[13] Q. Wang, R. Ward, and J. Zou, "Contrast enhancement for enlarged images based on edge sharpening," in *Proc. IEEE Int. Conf. Image Proces.*, Sep. 2005, vol. 2, pp. 1-4.

[14] Q. Shan, Z. Li, J. Jia, et al, "Fast image/video upsampling," *ACM Transactions on Graphics*, vol. 27, 153, 2008.

[15] A. Giachett and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2760-2768, Oct. 2011.

[16] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *Int. J. Comput. Vision*, vol. 40, no. 1, pp. 25-47, Jun. 2000.

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

TABLE II
AVERAGE PSNR VALUES (UPSAMPLING FACTORS ARE 2, 3, AND 4) ON THREE IMAGE DATASETS ("SET14", "SET5", AND "B100") WITH DIFFERENT METHODS (DB)

	Set14			Set5			B100		
	2X	**3X**	**4X**	**2X**	**3X**	**4X**	**2X**	**3X**	**4X**
Bicubic	30.36	27.67	26.06	33.64	30.38	28.42	29.35	27.17	25.95
LLE	31.19	27.94	26.32	35.17	31.14	29.32	30.17	27.54	26.17
ScSR	31.21	28.01	26.57	35.38	31.23	29.43	30.32	27.74	26.23
ASDS	31.15	27.91	26.94	34.85	31.02	29.54	30.19	27.65	26.32
PCANet	31.37	28.19	26.81	35.43	31.16	29.53	30.43	27.67	26.39
MMFRTN	**32.39**	**28.86**	**27.13**	**35.84**	**31.47**	**29.89**	**30.85**	**28.17**	**26.72**

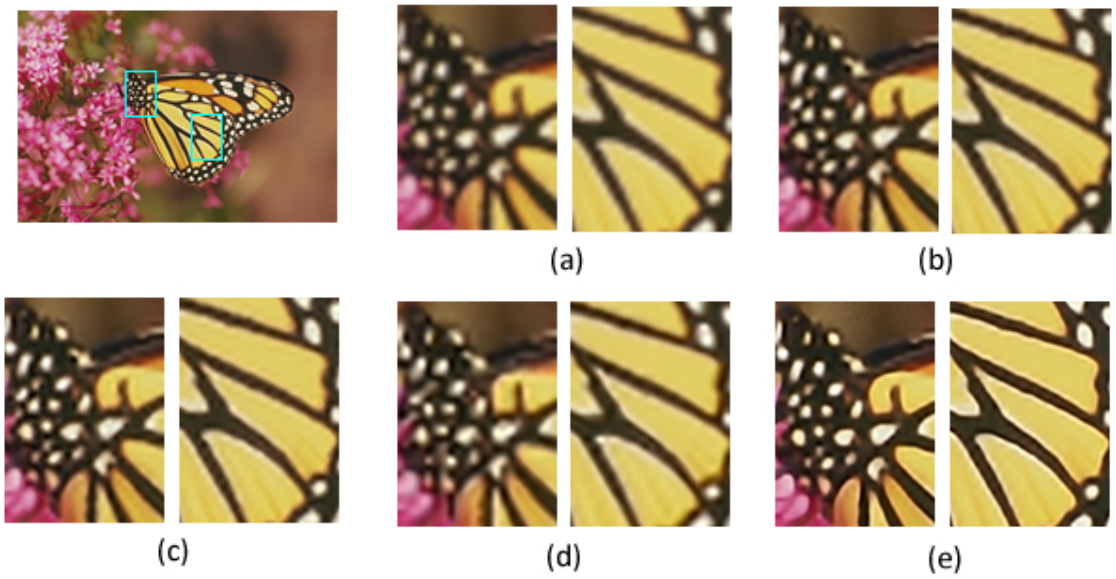

Fig. 11. Upsampled results of the "monarch" image (3X), (a) with bicubic interpolation, (b) with the LLE [36], (c) with the ScSR [17], (d) with the ASDS [19], (e) with the MMFRTN.

[18] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, "Bilevel sparse coding for coupled feature spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2360-2367.

[19] W. Dong, D. Zhang, G. Shi, X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838-1857, Jul. 2011.

[20] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 349-356.

[21] G. Freedman, R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, pp. 12-23, 2011.

[22] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, Jul. 2002.

[23] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 1999.

[24] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun, 2005, pp. 886-893.

[25] G. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527-1554, 2006.

[26] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural network," *Adv. Neural Inform. Process. Syst.*, 2012.

[27] K. Kavukcuoglu, P. Sermanet, et al, "Learning convolutional feature hierarchies for visual recognition," *Adv. Neural Inform. Process. Syst.*, 2010.

[28] J. Bruna, and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Ma.ch Intell.*, vol. 35, no. 8, pp. 1872-1886, Aug. 2013.

[29] T. H. Chan, K. Jia, S. Gao, et al, "PCANet: A Simple Deep Learning Baseline for Image Classification," *arXiv preprint arXiv:1404.3606*, 2014.

[30] D. S. Huang, W. Jia, and D. Zhang, "Palmprint verification based on principal lines," *Pattern Recognition*, vol. 41, no. 4, pp. 1316-1328, 2008.

[31] F. Matus, J. Flusser, "Image Representations via a Finite Radon Transform," *IEEE Trans. Pattern Anal. Ma.ch Intell.*, vol. 15, no. 10, pp. 996-1006, Oct. 1993.

[32] Y. Zhao, R. Wang, W. Wang, and W. Gao, "High Resolution Local Structure-Constrained Image Upsampling," *IEEE Trans. Image Process.*, vol.24, no.11, pp.4394-4407, Nov. 2015

[33] T. Michaeli and M. Irani, "Nonparametric blind super-resolution," in Proc. *IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 945-952.

[34] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2001, pp. 416-423.

[35] L. Wang, H. Wu, and C. Pan, "Fast image upsampling via the displacement field," *IEEE Trans. Image Process.*, vol.23, no. 12, pp. 5123-5135, Dec. 2014.

[36] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

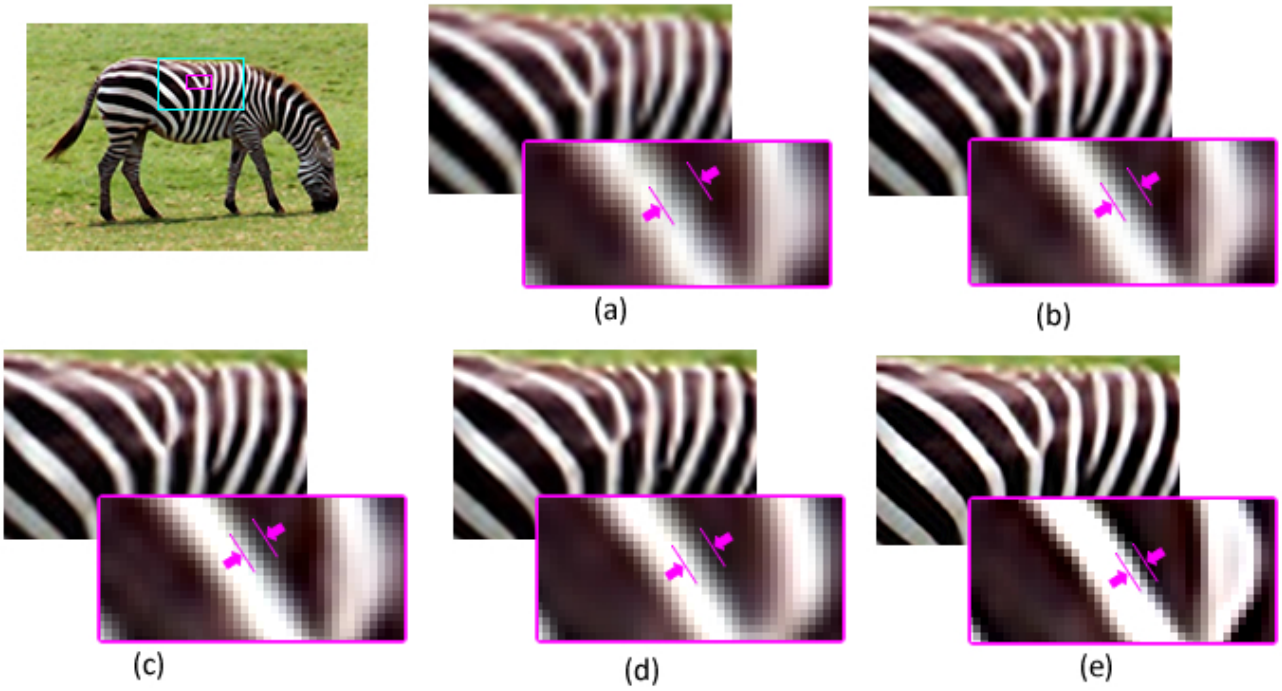

Fig. 12. Upsampled results of the "zebra" image (4X), (a) with bicubic interpolation, (b) with the LLE [36], (c) with the ScSR [17], (d) with the ASDS [19], (e) with the MMFRTN.

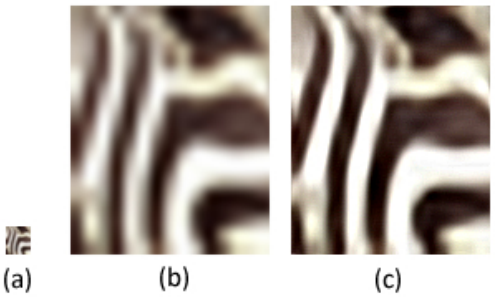

Fig. 13. Upsampled results of 8X magnification, (a) input LRI, (b) with bicubic interpolation, (c) with the MMFRTN.

neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, pp. 275-282.

[37] J. Sun, N. N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 729-736.

[38] M. Turkan, D. Thoreau, and P. Guillotel, "Optimized neighbor embeddings for single-image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 645-649.

[39] M. Turkan, D. Thoreau, and P. Guillotel, "Iterated neighbor embeddings for image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 3887-3891.

[40] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 345 - 352.

[41] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Trans. Image Process.*, vol.23, no. 6, pp. 2569-2581, Jun. 2014.

[42] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol.22, no. 4, pp. 1620-1630, Apr. 2013.

[43] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *Curv. Surfaces*, pp. 711- 730, 2010.

[44] M. Bevilacqua, A. Roumy, C. Guillemot, M. Morel, "Low-complexity single image super-resolution based on nonnegative neighbor embedding," in *Proc. British Machine Vis. Conf.*, 2012, pp.1-10

[45] R. Timofte, V. D. Smet, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920- 1927.

[46] H. Xu, G. Zhai, and X. Yang, "Single image super-resolution with detail enhancement based on local fractal analysis of gradient," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1740-1754, Oct. 2013.

Yang Zhao received the B.E. and the Ph.D. degree from department of automation, University of Science and Technology of China, in 2008 and 2013.

From September 2013, he is a Postdoctoral Fellow with the School of Electronic and Computer Engineering, Peking University Shenzhen Graduate School, China.

His research interests include image processing and pattern recognition.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2015.2504731, IEEE Transactions on Circuits and Systems for Video Technology

Ronggang Wang (M'12) received his Ph.D. Degree from the Institute of Computing Technology, Chinese Academy of Sciences. He was a Research stuff in Orange (France telecom) Labs from 2006 to 2010. He is currently an Associate Professor with Peking University Shenzhen Graduate School.

His research interest is on video coding and processing. He has done many technical contributions to ISO/IEC MPEG and China AVS. He leaded the MPEG Internet Video Coding (IVC) standard, served as MPEG IVC AHG co-chair since 2012, and severed as AVS implementation sub-group co-chair since 2015. He has authored more than 50 papers in international journals and conferences, and held more than 40 patents.

Wenmin Wang received the B.E degree in Computer Application from Harbin University of Science and Technology in 1983, then received the M.E. and the Ph.D. degrees in Computer Architecture from Harbin Institute of Technology, China, respectively in 1986 and 1989.

From 1992, he had 18 years of oversea industrial experience in Japan and America. He came back academia form the end of 2009, currently is a professor with the School of Electronic and Computer Engineering of Peking University, China.

His current research interests include video analysis, multimedia retrieval, artificial intelligence and machine learning.

Wen Gao (M'92-SM'05-F'09) received the Ph.D. degree in electronics engineering from the University of Tokyo, Japan, in 1991.

He is a Professor of computer science with Peking University, China. Before joining Peking University, he was a Professor of computer science with the Harbin Institute of Technology from 1991 to 1995, and a Professor with the Institute of Computing Technology of Chinese Academy of Sciences. He has published extensively including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. He served on the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON AUTONOMOUSMENTAL DEVELOPMENT, the *EURASIP Journal of Image Communications*, the *Journal of Visual Communication and Image Representation*. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ICME and ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations.