Resolution-Invariant Image Representation and Its Applications

Jinjun Wang, Shenghuo Zhu, Yihong Gong
NEC Laboratories America, Inc.
10080 N. Wolfe Rd., Cupertino, California.
{jjwang, zsh, ygong}@sv.nec-labs.com

Abstract

We present a Resolution-Invariant Image Representation (ЯIIR) *framework in this paper. The* ЯIIR *framework includes the methods of building a set of multi-resolution bases from training images, estimating the optimal sparse resolution-invariant representation of any image, and reconstructing the missing patches of any resolution level. As the proposed* ЯIIR *framework has many potential resolution enhancement applications, we discuss three novel image magnification applications in this paper. In the first application, we apply the* ЯIIR *framework to perform Multi-Scale Image Magnification where we also introduced a training strategy to built a compact* ЯIIR *set. In the second application, the* ЯIIR *framework is extended to conduct Continuous Image Scaling where a new base at any resolution level can be generated using existing* ЯIIR *set on the fly. In the third application, we further apply the* ЯIIR *framework onto Content-Base Automatic Zooming applications. The experimental results show that in all these applications, our* ЯIIR *based method outperforms existing methods in various aspects.*

1. Introduction

Most of the visual data used by digital imaging and display devices are represented as *raster image* with a rectangular grid of pixels. Since images with higher pixel density are desirable in many applications, there is an increasing demand to acquire *high resolution* (HR) raster images from *low resolution* (LR) inputs such as images taken by cell phone cameras, or converting existing standard definition footage into high definition image materials. However, raster images are resolution dependent, and thus cannot scale to an arbitrary resolution without loss of apparent quality. The performance of simple interpolation methods is limited because they poorly recover the missing high-frequency components, and often blur the discontinuities of magnified images.

While the raster image representation is resolution dependent, vector image is a resolution-independent way to represent visual data digitally. The vector representation is scalable because it represents the visual data using geometrical primitives such as points, lines, curves, and shapes or polygon. This capability largely contrasts the deficiency of raster representation. The idea of automatically converting raster image into more scalable vector representation has long been studied. Recently, Ramanarayanan *et al*. [18] vectorized the region boundaries and added to the original raster images to improve sharpness in scaled results; Dai *et al*. [5] represented the local image patches using the background/foreground descriptors and reconstructed the sharp discontinuity between the two; To allow efficient vector representation for multi-colored region with smooth transitions, gradient mesh technique has also been attempted [19]. In addition, commercial raster image vectorization softwares such as [1] are already available. However, vector-based techniques are limited in the visual complexity and robustness. For real images with fine texture or smooth shading, these approaches tend to produce over-segmented vector representations using a large number of irregular regions with flat colors. To illustrate, Figure 1(a) and (b) are vectorized and grown up to $\times 3$ scale using methods in [5] and [1]. The discontinuity artifacts in region boundaries can be easily observed, and the over-smoothed texture regions makes the scaled image watercolor like.

(a) method by [5] (b) method by [1] (c) our method

Figure 1. Comparison of scaling vectorized image and our method

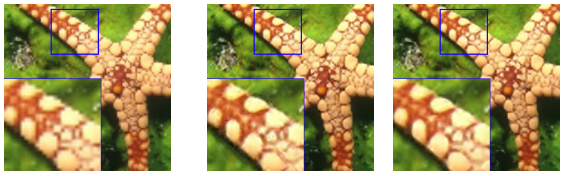
To increase the capacity of vector representation in modeling rich natural scenes, researchers strive to represent images using effective linear transforms. In this way the rich visual information can be encoded into the applied transformation, and the image is vectorized as the coefficients of a small set of bases of the transformation. In simple scenarios, pre-fixed transformations can be used, such as the Bi-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

Linear/BiCubic basis functions in 2-D image interpolation, and the DCT/DWT adopted in JPEG/JPEG-2000 standard, etc. Anisotropic bases such as countourlets [6] have been explicitly proposed to capture different 2-D edge/texture patterns. These techniques usually lead to sparse representation which is very preferable in image/video compression literature [11]. In addition to pre-fixed bases, since natural images typically contain many heterogeneous regions with significantly different geometric structures and/or statistical characteristics, adaptive mixture model representations were also reported. For example, the Bandelets model [7] partitions an image into squared regions according to local geometric flows, and represents each region by warped wavelet bases; the primal sketch model [10] detects the high entropy regions in the image through a sketching pursuit process, and encodes them with multiple Markov random fields. These adaptive representations capture the stochastic image generating process, therefore they are suited for image parsing, recognition and synthesis.

In this paper, we propose a *Resolution-Invariant Image Representation* (ЯIIR) framework by utilizing the adaptive local representation. Instead of using orthonormal bases, such as X-lets or Fourier, we take the advantage of the rich representative power of over-complete bases to represent images in sparse codes. Since we focus on resolution enhancement problems, our proposed representation is based on multi-resolution base set. To this extend, the most related literature is the example-based image super-resolution (SR) area where researchers represent the LR/HR images using LR/HR bases and utilize the co-occurrence prior between the LR and HR representations to reconstruct the SR image [2]. For example, Freeman *et al*. [9] and Qiang *et al*. [22] represented each local region in the LR image using one example LR patch, and applied the co-occurrence prior and global dependence through a parametric Markov network to estimate the HR image representation. Chang *et al*. [3] applied Locally Linear Embedding (LLE) to learn the optimal combination weights of multiple example LR patches to estimate the optimal HR representations. In addition to example patches, representing images in transferred domain, such as edge profile [20], wavelet coefficients [13], or image contourlet [14], has also been examined.

Our proposed model in this paper differs from traditional example-based SR method in the following aspects: Firstly, the bases used by existing example-based SR approaches have only single scale capacity, hence representations learned from different bases are not interchangeable. Our image representation is based on multi-resolution base set, therefore it is more scalable; Secondly, our image representation is global optimal with respect to the applied multi-resolution bases, hence our method gives better results in traditional SR problem, e.g. a $\times 3$ single frame SR problem as shown in Figure 2; Thirdly but the more importantly, we have theoretically and experimentally proved that, our image representation is resolution-invariant. It can be used to scale the image to various scale by using respective resolution bases. This has obvious the importance in many novel resolution enhancement applications that existing method cannot handel well. The experimental results show that, in all the tested applications, our method outperforms existing methods in various aspects.



(a) BiCubic interpolation (b) method by [3] (c) our method

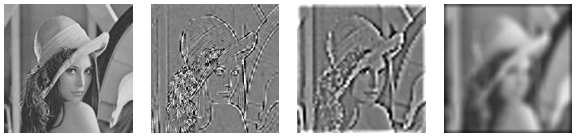
Figure 2. Comparison of scaling using different bases

The rest of the paper is organized as follows: Section 2 revisits the layered image model used by most existing example-based SR methods and the sparse image representation. Section 3 presents our key ЯIIR framework and validates it using real images. Section 4, 5, and 6 introduce three novel image resolution enhancement applications. Section 7 lists the experimental results, and section 8 summarizes the proposed methods and discusses future works.

2. Example-Based Image Representation

2.1. Image layers

Many example-based SR approaches [9, 16, 22, 23] assume that, any HR image $\boldsymbol{I} = \boldsymbol{I}^h + \boldsymbol{I}^m + \boldsymbol{I}^l$ (Figure 3. a) consist of a high frequency layer (denoted as $\boldsymbol{I}^h$, Figure 3. b), a middle frequency layer ($\boldsymbol{I}^m$, Figure 3. c), and a low frequency layer ($\boldsymbol{I}^l$, Figure 3. d). The down-graded LR image $\bar{\boldsymbol{I}} = \boldsymbol{I}^m + \boldsymbol{I}^l$ results from discarding the high frequency components. Hence the example-based approach estimate the $\boldsymbol{I}^h$ layer by maximizing $Pr(\boldsymbol{I}^h|\boldsymbol{I}^m, \boldsymbol{I}^l)$. In addition, since $\boldsymbol{I}^h$ is independent of $\boldsymbol{I}^l$ [9], it is only required to maximize $Pr(\boldsymbol{I}^h|\boldsymbol{I}^m)$.



(a) original (b) $\boldsymbol{I}^h$ layer (c) $\boldsymbol{I}^m$ layer (d) $\boldsymbol{I}^l$ layer

Figure 3. Image frequency layers

A typical example-based SR resolving process works as follows: From training images, an example patch pair set $\mathcal{S} = \{\mathcal{S}^m, \mathcal{S}^h\}$ can be extracted and used as the bases for representation. $\mathcal{S}^m = \{\boldsymbol{p}_i^m\}_{i=1}^N$ and $\mathcal{S}^h = \{\boldsymbol{p}_i^h\}_{i=1}^N$ represent the middle frequency and the high frequency bases respectively. Each element $\boldsymbol{p}_i^m$ is the column expansion

of a square image patch from the middle frequency layer $\boldsymbol{I}^m$, and $\boldsymbol{p}_i^h$ is from the corresponding patch in the high frequency layer $\boldsymbol{I}^h$. The dimensions of $\boldsymbol{p}_i^m$ and $\boldsymbol{p}_i^h$ are $D^m \times 1$ and $D^h \times 1$ respectively, and often $D^m \neq D^h$. Now given an LR input image, the middle frequency patches can be extracted and denoted as $\{\boldsymbol{y}_j^m\}$. The Maximum A Posteriori (MAP) estimation of the missing high frequency components $\{\hat{\boldsymbol{y}}_j^h\}$ that incorporates both the co-occurrence patterns between $\boldsymbol{y}_j^m$ and $\boldsymbol{y}_j^h$ and the neighboring smoothness constraint for $\boldsymbol{y}_j^h$ can be obtained by minimizing [9]:

$$\{\hat{\boldsymbol{y}}_j^{h*}\} = \underset{\{\hat{\boldsymbol{y}}_j\} \subset \mathcal{S}}{\operatorname{argmin}} \sum_j \left(||\boldsymbol{y}_j^m - \hat{\boldsymbol{y}}_j^m||^2 + \lambda ||\hat{\boldsymbol{y}}_j^h - \mathcal{O}(\hat{\boldsymbol{y}}_j^h)||^2 \right), \tag{1}$$

where $\mathcal{O}(\hat{\boldsymbol{y}}_j^h)$ represents the overlapped region in $\hat{\boldsymbol{y}}_j^h$ by neighboring patches. Due to cyclic dependencies, exact inference of Eq. (1) is computationally intractable. Hence approximation techniques, such as Bayesian belief propagation [9], Graph-Cut [16], and Gibbs Sampling [22] have been attempted. These methods all consist of firstly a ranking step to find K nearest elements $\{\boldsymbol{p}_k^m\}_{k=1}^K$ in $\mathcal{S}$ to represent each input patch $\boldsymbol{y}_j^m$, and secondly a searching step to select the winning representation of $\hat{\boldsymbol{y}}_j^h$ from $\{\boldsymbol{p}_k^h\}_{k=1}^K$. However, the overall process is computationally expensive.

Alternative, the LLE method [3] alleviates the smoothness constraint in the overlapped regions by simply taking the average pixel values. In addition, instead of using only one winning element, a linear combination of the K candidates $\hat{\boldsymbol{y}}_j^{m*} = \sum_{k=1}^K x_k^* \boldsymbol{p}_k^m$ is used to represent $\boldsymbol{y}_j^m$. Hence the overall optimization task equals to solving the optimal representation $\boldsymbol{x}_j^*$ for each $\boldsymbol{y}_j^h$ individually, using

$$\boldsymbol{x}_j^* = \underset{\boldsymbol{x}_j}{\operatorname{argmin}} ||\boldsymbol{y}_j^m - \boldsymbol{P}_j^m \boldsymbol{x}_j||^2, \tag{2}$$

where $\boldsymbol{P}_j^m$ is the $D^m \times K$ matrix representation of $\{\boldsymbol{p}_k^m\}_{k=1}^K$. Then $\{\hat{\boldsymbol{y}}_j^{h*}\}$ can be computed by $\{\hat{\boldsymbol{y}}_j^{h*}\} = \{\boldsymbol{P}_j^h \boldsymbol{x}_j^*\}$ where $\boldsymbol{P}_j^h$ is the $D^h \times K$ matrix representation of $\{\boldsymbol{p}_k^h\}_{k=1}^K$.

2.2. Sparse representation

The performance of LLE method is limited by the quality of the K candidates, and hence the solution is non-optimal. To cope with the limitation, searching for the optimal representation $\boldsymbol{x}_j^*$ within the whole example patch set was attempted in [23]. To elaborate, define an $D^m \times N$ matrix $\boldsymbol{P}^m$ to represent the middle frequency patch set $\mathcal{S}^m$. Instead of simply replacing $\boldsymbol{P}_j^m$ in Eq. (2) by $\boldsymbol{P}^m$, since $\mathcal{S}^m$ is typically over-complete, an L1 regularization term is further imposed such that each $\boldsymbol{y}_j^m$ is represented using only a few base elements. Hence we seek to find the optimal representation $\boldsymbol{x}_j^*$ by

$$\boldsymbol{x}_j^* = \underset{\boldsymbol{x}_j}{\operatorname{argmin}} ||\boldsymbol{y}_j^m - \boldsymbol{P}^m \boldsymbol{x}_j||^2 + \gamma |\boldsymbol{x}_j|. \tag{3}$$

Solving Eq.(3) is a standard Sparse-Coding (SC) process [15]. Then $\{\hat{\boldsymbol{y}}_j^{h*}\}$ can be computed by

$$\{\hat{\boldsymbol{y}}_j^{h*}\} = \{\boldsymbol{P}^h \boldsymbol{x}_j^*\}, \tag{4}$$

where $\boldsymbol{P}^h$ is the $D^h \times N$ matrix representation of $\mathcal{S}^h$.

In the above process, although the optimal representation $\boldsymbol{x}_j^*$ is learned from the middle frequency layer by Eq. (3), it can be applied to compute the missing components in the high frequency layer by Eq. (4). Such invariant property can be generalized in Cor. 2.1 below:

Corollary 2.1. *The optimal representation $\boldsymbol{x}^*$ is invariant across different frequency layers given respective bases representing corresponding frequency components.*

Cor. 2.1 is a direct result of the image co-occurrence prior, and has been validated by numerous example-based SR work [9]. Our focus in this paper, however, is to reveal another invariant property between different resolution versions of a same image. This is discussed in the next section.

3. Resolution-Invariant Image Representation

3.1. Generating multi-resolution base set

In this section we want to examine the relation among different layers from different resolution versions of a same image. To start with, we generated a multi-resolution image patch pair set $\mathcal{S}$ with the following steps: First, for each image $\boldsymbol{I}$ in a given dataset, we try to obtain its LR version $\bar{\boldsymbol{I}}$ by first down-sample $\boldsymbol{I}$ to $1/U$ scale, then up-sample it back to the original size, In this way the high frequency layer $\boldsymbol{I}^h$ is discarded in $\bar{\boldsymbol{I}}$. As explained in subsection 2.1, N image patch pairs can be extracted from the $\boldsymbol{I}^m$ and $\boldsymbol{I}^h$ layers respectively, and we denote the obtained set as $\mathcal{S}_U = \{\mathcal{S}_U^m, \mathcal{S}_U^h\} = \{\boldsymbol{p}_{i,U}^m, \boldsymbol{p}_{i,U}^h\}_{i=1}^N$. The subscript U is used to indicate that $\mathcal{S}_U$ is from the U^{th} resolution ($\boldsymbol{I}$ and $\boldsymbol{I}_U$ are interchangeable thereafter).

Next, we want to obtain similar patch pair sets $\{\mathcal{S}_1, ..., \mathcal{S}_{U-1}\}$ but from different resolution version of $\boldsymbol{I}$. To elaborate, for the u^{th} set $\mathcal{S}_u$, $u = 1, ..., U-1$, we shrink $\boldsymbol{I}$ to u/U size (denoted as $\boldsymbol{I}_u$), and down-sample $\boldsymbol{I}_u$ to $1/u$ size then up-sample back to obtain $\bar{\boldsymbol{I}}_u$. Then we can accordingly extract $\mathcal{S}_u = \{\mathcal{S}_u^m, \mathcal{S}_u^h\} = \{\boldsymbol{p}_{i,u}^m, \boldsymbol{p}_{i,u}^h\}_{i=1}^N$. Note that the order of the base elements are specially arranged such that the i^{th} pair in $\{\mathcal{S}_u^m, \mathcal{S}_u^h\}$ and the i^{th} pair in $\{\mathcal{S}_U^m, \mathcal{S}_U^h\}$ are from patches at the same relative location as highlighted by the red center pixels in Figure 4.

As explained above, each base $\mathcal{S}_u$ can be used to do $\times u$ scale image SR. However, in this paper, our intension is to examine the relation among bases from different resolution versions. This is discussed in the next subsection.

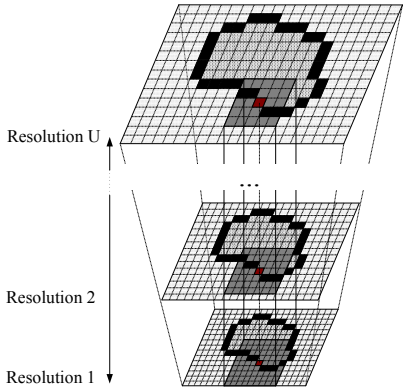


Figure 4. Sampling the base set

3.2. Resolution-invariant property of $\mathcal{S}$

An ideal image down-grading process to obtain $\bar{\boldsymbol{I}}_u$ from $\boldsymbol{I}_U$ can be modeled by

$$\bar{\boldsymbol{I}}_u = \boldsymbol{I}_U \downarrow_u \otimes \boldsymbol{G}_u^m \,, \tag{5}$$

where $\downarrow_u$ denotes the down-sample operator to u/U scale, and $\boldsymbol{G}_u^m$ simulates a Gaussian blurring kernel to discard the high frequency layer. Hence we can have

$$\begin{aligned} \boldsymbol{I}_u^m &= \bar{\boldsymbol{I}}_u - \bar{\boldsymbol{I}}_u \otimes \boldsymbol{G}_u^l \\ &= \boldsymbol{I}_U \downarrow_u \otimes \boldsymbol{G}_u^m - \boldsymbol{I}_U \downarrow_u \otimes \boldsymbol{G}_u^m \otimes \boldsymbol{G}_u^l \,, \end{aligned} \tag{6}$$

where $\boldsymbol{G}_u^l$ simulates a Gaussian blurring kernel to discard the middle frequency layer. Similar to subsection 2.2, let $\boldsymbol{P}_u^m$ be an $D_u^m \times N$ matrix to represent all the elements in base $\mathcal{S}_u^m$, where D_u^m is the dimension of patch $\boldsymbol{p}_u^m$, and $\boldsymbol{y}_u^m$ be the middle frequency component of an input patch $\boldsymbol{y}_u$. With Eq. (6) we can have

$$\boldsymbol{P}_u^m = \boldsymbol{P}_U \downarrow_u \otimes \boldsymbol{G}_u^m - \boldsymbol{P}_U \downarrow_u \otimes \boldsymbol{G}_u^m \otimes \boldsymbol{G}_u^l \,, \tag{7}$$

and

$$\boldsymbol{y}_u^m = \boldsymbol{y}_U \downarrow_u \otimes \boldsymbol{G}_u^m - \boldsymbol{y}_U \downarrow_u \otimes \boldsymbol{G}_u^m \otimes \boldsymbol{G}_u^l \,. \tag{8}$$

As explained in section 2.2, the optimal representation between $\boldsymbol{y}_u^m$ and $\boldsymbol{P}_u^m$ can be obtained by solving Eq. (3). Taking Eq. (7) and Eq. (8) into Eq. (3), we can get

$$\boldsymbol{x}_u^* = \underset{\boldsymbol{x}_u}{\operatorname{argmin}} \, ||\boldsymbol{y}_U - \boldsymbol{P}_U \boldsymbol{x}_u||^2 + \gamma |\boldsymbol{x}_u| \,, \tag{9}$$

From Eq. (9), it can be seen that the optimal representation $\boldsymbol{x}_u^*$ is irrelevant of u. Hence we can generalize the second invariant property for a multi-resolution base set:

Corollary 3.1. *The optimal representation $\boldsymbol{x}^*$ is invariant across different resolution inputs given respective bases representing corresponding resolutions.*

Cor. 3.1 reveals that, if the different resolution versions of the same image are related by Eq. (5), then the optimal representation learned from any resolution version can also be applied for another version. Since the image down-grading process in realistic imaging system may not be simply modeled by Eq. (5), the next subsection validates Cor. 3.1 for real images.

3.3. Validation for realistic imaging system

To examine how much Cor. 3.1 holds for real images, the following experiments examine the similarity of the optimal representation $\boldsymbol{x}$ learned from 5 different resolution versions of a same image (*i.e.* $U = 5$). First, from a training HR image, we generated a multi-resolution patch pair set $\mathcal{S}$ with around 8000 patch pairs in each resolution version. Next, for each testing image, we extracted around 2000 patches from every resolution version as well, and five versions were used as that in $\mathcal{S}$. Then we solved Eq. (3) to get their representations $\boldsymbol{x} = \{\boldsymbol{x}_{j,u}\}_{j=1}^{2000}$, $u = 1, ..., 5$ at each resolution level respectively. If Cor. 3.1 holds, $\boldsymbol{x}_{j,u}$ should be very similar to $\boldsymbol{x}_{j,v}$, $u \neq v$. Hence we computed the overall similarity between every two elements in $\boldsymbol{x}_j = \{\boldsymbol{x}_{j,1}, \boldsymbol{x}_{j,2}, \boldsymbol{x}_{j,3}, \boldsymbol{x}_{j,4}, \boldsymbol{x}_{j,5}\}$ by

$$sim(\boldsymbol{x}_j) = \frac{1}{C_5^2} \sum_{u=1}^{4} \sum_{v=u+1}^{5} cor_{u,v}^j \,,$$

where $cor_{u,v}^j$ is the correlation between $\boldsymbol{x}_{j,u}$ and $\boldsymbol{x}_{j,v}$. Finally, the overall similarity is averaged over the 2000 patched to get a score. To make the experiment more comprehensive, we applied two γ values, $\gamma = 11$ and $\gamma = 25$, for Eq. (3), to get the "Low sparsity" and "High sparsity" profiles. We also tested different redundancy level in the base set by either random removing or using K-Mean clustering methods to reduce the base cardinality from 8000 to until 50. The experiments were repeated 5 times with different training images, and the results are shown in Figure 5. As can be seen from Figure 5, the minimal similarity score is greater than 0.44, and the maximal score reaches almost 0.8. The results proves the high similarities between x_u and x_v from different resolutions, hence Cor. 3.1 still holds to a considerable extent. The reason why the scores are relatively lower when the cardinality is large is due to the nature of images, where there are very similar elements in the base, such that the SC algorithm may not select exactly the same elements for reconstruction. When such redundancy is removed, the similarity between representations becomes significantly higher.

For simplicity, we call Cor. 3.1 the *Resolution-Invariant Image Representation* (ЯIIR) property, $\mathcal{X} = \{\boldsymbol{x}_j^*\}$ an ЯIIR, and the multi-resolution patch pair sets $\mathcal{S}$ an ЯIIR set. With ЯIIR property, it is able to saves the computational expensive SC process for multi-scale resolution enhancement

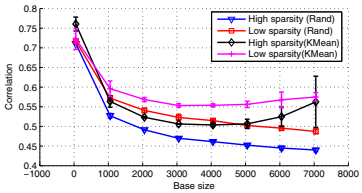


Figure 5. Correlation between different resolution versions

tasks, hence it's very desirable in may important applications. In the following two sections, we apply the ЯIIR framework to three interesting image resolution enhancement applications.

4. Multi-Scale Image Magnification

There are many scenarios where users need different resolution image/video of a same object, *e.g.* viewing an image in a PDF document at different zooming, streaming videos over wireless channels to users with different receiver resolutions, etc. If the original image/video does not meet the resolution requirement, up-sampling methods such as BiCubic interpolation are usually adopted. Although numerous example-based image SR approaches have been reported to achieve better quality than simply 2-D image interpolation (Section 1), they cannot be easily applied for the Multi-Scale Image Magnification (MSIM) problem, because in typical example-based image SR approaches [9], different models are built for different scale factors, and usually these models are not interchangeable, *e.g.* the model computed for $\times 2$ scale SR cannot be used for $\times 3$ SR.

With the ЯIIR method, the MSIM problem can be easily solved. First, the ЯIIR $\mathcal{X}$ can be learned at any resolution level once. Then by applying different ЯIIR sets $\mathcal{S}_u$, the same $\mathcal{X}$ can be used to magnify the image to multiple scales. Hence the testing set $\mathcal{S}$ created in previous section is able to do $\times 1 \sim \times 5$ SR. Finer magnification scales can be achieved by simply increasing the level of resolutions used in creating $\mathcal{S}$. The advantage of the method is that, the magnification process requires only matrix-vector multiplication by Eq. (4), which can be implemented very efficiently. In addition, the ЯIIR set can be stored locally, and the computed ЯIIR can be transmitted together with the image/video document.

The key challenge in applying ЯIIR method for MSIM problem is how to generate a suitable ЯIIR set. By "suitable" we mean that the set is both resolution-invariant and compact. The directly sampled set $\mathcal{S}$ in section 3.1 is inarguably resolution-invariant. However, it's not compact, and thus both searching for the optimal representation and reconstructing the missing high frequency layer require lengthy computation. To cope with the problem, we want to learn a more compact base set $\hat{\mathcal{S}} = \{\hat{\mathcal{S}}_u^m, \hat{\mathcal{S}}_u^h\}_{u=1}^U$ to replace $\mathcal{S}$. Each $\hat{\mathcal{S}}_u^m$ or $\hat{\mathcal{S}}_u^h$ contains M elements, and $M < N$. This is achieved by solving the following optimization problem,

$$\text{minimize}_{\boldsymbol{B},\boldsymbol{X}} \, ||\boldsymbol{P}_u^m - \boldsymbol{B}\boldsymbol{X}||^2 + \gamma \sum_i |\boldsymbol{X}(:,i)| \qquad (10)$$
$$\text{s.t.} \sum_j ||\boldsymbol{B}(:,j)||^2 \leq c, \forall j = 1,..,M \,,$$

where $\boldsymbol{P}_u^m$ represents $\{\boldsymbol{p}_{i,u}^m\}_{i=1}^N$, $\boldsymbol{B}$ is a $D^m \times M$ matrix representing $\hat{\mathcal{S}}_u^m$, $\boldsymbol{X}$ is a $M \times N$ matrix representing $\mathcal{X}$, and u indicates the resolution level used to learn the optimal representation. To solve Eq. (10), an iterative optimization process is applied. First we start from random initialization of $\boldsymbol{B}$. Then we iteratively search for the optimal sparse representations $\boldsymbol{X}$ based on the current values of $\boldsymbol{B}$, and train an optimal base $\boldsymbol{B}$ for the obtained $\boldsymbol{X}$, until converge. Then $\boldsymbol{X}$ is fixed and u is changed between $1 \sim U$ in Eq. (10) to train each $\mathcal{S}_u$.

5. Continuous Image Scaling

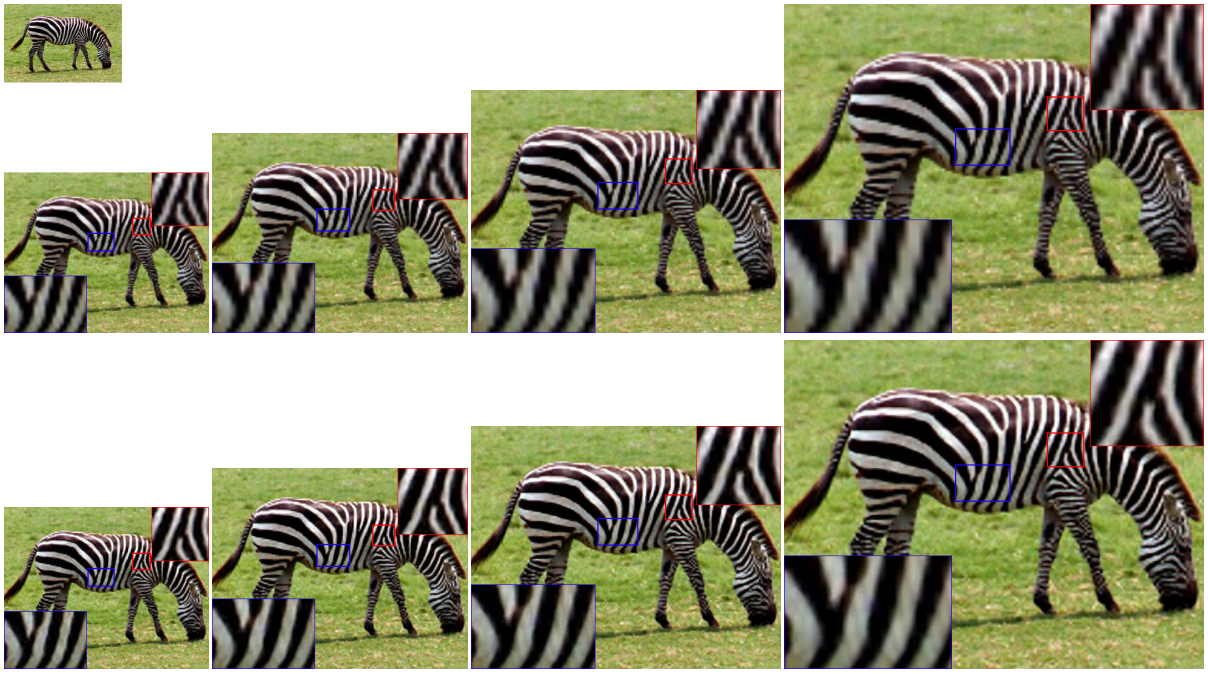
An intuitive extension of MSIM is to perform Continuous Image Scaling (CIS). For this purpose, with traditional example-based SR technique, the users have to store a continuous set of models, or at least at a very fine granularity, which requires huge storage space and computation power. With the ЯIIR method, the user only need to store a limited scale of bases which can then be used to generate a new base at the required scale on the fly. *E.g.*, from the $\mathcal{S} = \{\mathcal{S}_u\}_{u=1}^5$ generated in section 3.1, we can generated a new base at $\times 2.4$ scale using $\mathcal{S}_2$ and $\mathcal{S}_3$. To elaborate, let v be the target scale factor which is between u and $u+1$, the i^{th} element in $\mathcal{S}_v$ can be obtained by

$$\boldsymbol{p}_{i,v} = w_{u,v}\tilde{\boldsymbol{p}}_{i,u} + (1 - w_{u,v})\tilde{\boldsymbol{p}}_{i,u+1} \,, \qquad (11)$$

where $\tilde{\boldsymbol{p}}_{i,u}$ is the patch interpolated from scale u, and $\tilde{\boldsymbol{p}}_{i,u+1}$ is interpolated from scale $u+1$. The weight $w_{u,v} = \big(1 + exp\big((v - u - 0.5) * \tau\big)\big)^{-1}$ where in our implementation, $\tau = 10$. Some example images can be found in Figure 6.

6. Content-Based Automatic Zooming

Viewing visual content on small display devices is a challenging and demanding application nowadays. Existing zoomable User Interfaces (UI) usually provide multiple fixed scales. Hence a typical viewing process is tedious because the user must manually scroll around the image and select different zoom by pressing the scroll and zoom buttons many times. Image Content-Based Automatic Zooming (CBAZ) provides improved viewing experience where zooming can be automatically performed as the user navigates. The optimal scale is determined based on the content in the viewport, *e.g.* complexity, degree of interest, distortion function [4], user attention [12], etc. In this way the



scale: $\times 2.05$ scale: $\times 2.55$ scale: $\times 3.1$ scale: $\times 4.2$

Figure 6. Illustration of Continuous Image Scaling (*top-left: the original image; top: BiCubic interpolation; bottom: our results*)

viewing process becomes faster and less tedious because the display area is more effectively utilized.

While many existing CBAZ works focus on solving the optimal scale, another important issue is how to generate the magnified image region. Since the image content factors are usually continues, the optimal scale factor is also continues. As explained in previous section 5, example-based SR approaches are not capable of such situations. Fortunately, the ЯIIR technique can be utilized. As illustrated in Figure 7: First, the user navigates to the desired image region (highlighted by the center point in the rectangle region), and the optimal scale u is computed based on the content of the region. Then using the stored representations $\{\boldsymbol{x}_j^*\}$ and $\mathcal{S}_u$, the magnified image region can be generated (Figure 7 middle column). Note that to speed up processing, only patches inside the highlighted region need to be estimated.

src region: 114×152 pixels, dest region: 240×320 pixels, scale: $\times 2.1$

src region: 76×102 pixels, dest region: 240×320 pixels, scale: $\times 3.15$

Figure 7. Illustration of Content-Based Automatic Zooming (*left: the original image and the local complexity map; middle: our results; right: BiCubic interpolation*)

7. Experimental Results

The first set of experiments evaluated the performance of the ЯIIR framework for MSIM applications (section 4). To begin with, a ЯIIR set $\mathcal{S}$ was trained. Around 20000 patch pair examples were extracted from some widely used images, *e.g.* "peppers", as input for Eq.(10). Similar to section 3.3, firstly we wanted to examine the ЯIIR property in $\mathcal{S}$. Hence we ranged the cardinality between 50 to 8000 by changing the input value M in Eq.(10), and solved the optimal representations using Eq. (3) for around 30 testing images to examined their averaged similarity score between $\times 2 \sim \times 5$ scales. The results are listed in Figure 8, where it can be seen that the learned base set $\mathcal{S}$ presents good ЯIIR property, as the maximal score reaches almost 0.84, and the

average score is 0.56. Figure 9 plots the first 16 base elements representing the middle and high frequency components at $\times 3$, $\times 4$ and $\times 5$ scales. The high similarity between the corresponding vectors at different scales can be easily observed. Besides, the shapes of these base elements are similar to those found by [17].

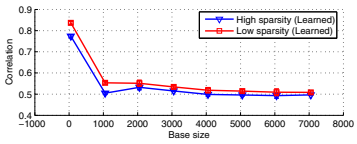


Figure 8. Correlation between different resolution versions

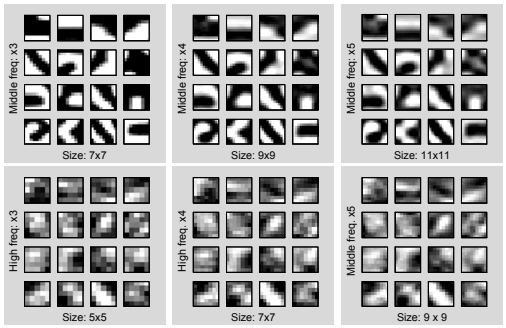


Figure 9. Example learned base elements (*reshaped to be squares*)

Next we wanted to find a suitable cardinality. As explained in section 4, a good base requires both excellent modeling capacity and reduced redundancy. Hence we evaluated the PSNR score over BiCubic interpolation with respect to different cardinalities. According to the results shown in Figure 10, the suitable cardinality falls between $1000 \sim 2000$ as it leads to the highest PSNR score. It is noticed that the improvement by $\times 2$ scale is lowest because the quality of interpolated images by BiCubic method at $\times 2$ scale is already very good. However, when larger scales are used, the performance of BiCubic will decrease.

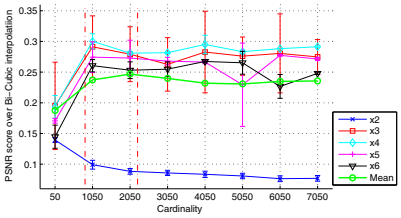


Figure 10. Image quality under different base cardinalities

Finally, we magnified 30 testing images and compared the results with two existing example-based SR methods, "KNN" [9] and "LLE" [3], two functional interpolation SR methods, "Enhance" [21] and "Soft edge" [5], and a sparse representation method based on non-ЯIIR set, "SC" [23], in terms of PSNR score over BiCubic interpolation. As can be seen from Figure 11, in all the cases, both the SC method in [23] and our ЯIIR method achieved the best results which are both sharp and natural, with unperceivable artifacts. This is reasonable because both the two methods utilized the whole bases to learn the optimal sparse image representation. The difference between the two is that, the SC method learns a different sparse representation for different scale factors, while in our ЯIIR method a single, resolution-invariant sparse representation is learned once. Hence our method requires much less computation. In addition, it is only with this resolution-invariant property that the ЯIIR framework can be used in the following two applications discussed in section 5 and section 6.

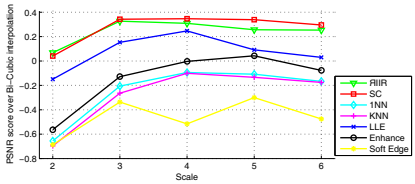


Figure 11. Image quality under different scales

The second experiment demonstrated the CIS application under the ЯIIR framework (section 5). We first generated the ЯIIR base set $\mathcal{S}$ from $\times 1 \sim \times 5$ scales, with step size 0.5, *i.e.* a base is trained at every u and $u + 0.5$ scales, $u = 1, ..., 5$. This would take up 15Mb storage space if the cardinality is selected to be 2000. For each testing image, the ЯIIR is learned at scale $\times 3$. Next we conducted continuous scaling between $\times 1 \sim \times 5$ with step size 0.05. A DELL PRECISION 490 PC (3.2GHz CPU, 2G RAM) was used to conduct a subjective user study where 10 people were asked to compare the image quality with BiCubic interpolation. All of them rated significantly higher scores for our results, and most of them were not aware of the processing delay in generating the magnified images. This results validate the good performance of CIS using ЯIIR method as well as the low computational cost in generating the up-scaled images. Some example images can be found in Figure 6, and a video demo recording the real-time operation of the CIS application is available in the supplemental materials.

The third experiment examined the performance of ЯIIR framework for CBAZ applications (section 6). We targeted at a QCIF (320×240 pixels) display size which is typical for most PDA/cell phones. The provided UI allows the user to navigate to a certain region by clicking on the original image or the local complexity map, as illustrated in Figure 7. Then an optimal scale factor is calculated based on the content of the selected region, and the ЯIIR method was used to magnify the selected region to the QCIF screen. The same 10 people were asked to rate the results in comparison to BiCubic interpolation, and our ЯIIR technique won a large margin. In addition, since only a subset patches need to be reconstructed, the processing delay is even less noticeable.

Some example images can be found in Figure 7, and a video demo recording the real-time operation of the CBAZ application is provided in the supplemental materials.

It is of interest to further discuss the perceptual quality of our zoomed image. A visually pleasing image preserves both sharp edge boundaries and fine texture grains. While the former can be *reconstructed* by using the co-occurrence patterns (Cor. 2.1), the later is usually *synthesized* rather than *reconstructed* [8]. As shown in Figure 12, although our results preserve good edge sharpness, compared to other approaches [8] that superimposed certain level of grains, a texture synthesization step that is irrelevant to the ЯIIR property can be added into our framework to further improve the perceptual image quality.

(a) Bicubic (b) Our method (c) Method by [8]

Figure 12. Texture synthesization capability with $\times 3$ scale

8. Conclusion and Future Work

We proposed the *Resolution-Invariant Image Representation* (ЯIIR) framework based on the idea that an image should have identical representation at various resolution levels. In the framework, we obtain the sparse resolution-invariant representation of any image patch at an input resolution given an ЯIIR base set. With the resolution-invariant representation, we are able to reconstruct the image patch at any resolution level. We discuss three interesting applications: in the first application, we introduced a strategy to built a compact ЯIIR base set and applied it for *Multi-Scale Image Magnification*; in the second application, we proposed to construct a new base at any arbitrary scale from existing ЯIIR base set to allow *Continuous Image Scaling*; and in the third application, we combined the image content-based zooming concept with our ЯIIR framework to perform *Content-Based Automatic Zooming*. Experimental results show that in all these applications, our ЯIIR based method outperforms existing methods in various aspects.

The future work of the research includes the following issues: first, in addition to image magnification, we will investigate the possibility of applying the ЯIIR frameowrk to improve image shrinking quality; second, we will examine additional optimization strategies to learning a more compact base set with uncompromised or even improved reconstruction accuracy; and third, we will evaluate the ЯIIR property in other coding types such that the optimal representation may be learned faster, or requires less storage space. This has obvious the importance for coding, streaming, personalization, etc applications.

References

[1] http://www.vectormagic.com.

[2] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. *CVPR'01*, 2001.

[3] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *CVPR'04*, 2004.

[4] P. Chiu, K. Fujii, and Q. Liu. Content based automatic zooming: Viewing documents on small displays. *ACM MM'08*, 2008.

[5] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong. Soft edge smoothness prior for alpha channel super resolution. *CVPR'07*, 2007.

[6] M. N. Do and M. M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *TIP*, 2005.

[7] L. Erwan and M. Stephane. Sparse geometric image representations with bandelets. *TIP*, 2005.

[8] R. Fattal, M. Agrawala, and S. Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM SIGGRAPH'07*, 2007.

[9] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *IJCV*, 2000.

[10] C. Guo, S. Zhu, and Y. Wu. Towards a mathematical theory of primal sketch and sketchability. *ICCV'03*, 2003.

[11] W. Hong, J. Wright, H. Kun, and M. Yi. Multiscale hybrid linear models for lossy image representation. *TIP*, 2006.

[12] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 1998.

[13] C. Jiji, M. Joshi, and S. Chaudhuri. Single-frame image super-resolution using learned wavelet coefficients. *IJIS-Technology*, 2004.

[14] C. Jiji and C. Subhasis. Single-frame image super-resolution through contourlet learning. *EURASIP Journal on Applied Signal Processing*, 2006.

[15] H. Lee, A. Battle, R. Raina, and A. Ng. Efficient sparse coding algorithms. *Advances in Neural Information Processing Systems, MIT Press*, 2007.

[16] U. Mudenagudi, R. Singla, P. K. Kalra, and S. Banerjee. Super resolution using graph-cut. *ACCV'06*, 2006.

[17] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996.

[18] G. Ramanarayanan, K. Bala, and B. Walter. Feature-based textures. *Proceedings of EGSR'04*, 2004.

[19] J. Sun, H. Tao, and H. Shum. Image hallucination with primal sketch priors. *CVPR'03*, 2003.

[20] J. Sun, Z. Xu, and H. Shum. Image super-resolution using gradient profile prior. *CVPR'08*, 2008.

[21] J. Wang and Y. Gong. Fast image super-resolution using connected component enhancement. *ICME'08*, 2008.

[22] Q. Wang, X. Tang, and H. Shum. Patch based blind image super resolution. *ICCV'05*, 2005.

[23] J. Yang, J. Wright, T. Huang, and M. Yi. Image super-resolution as sparse representation of raw image patches. *CVPR'08*, 2008.