

RESOLUTION-INVARIANT IMAGE REPRESENTATION FOR CONTENT-BASED ZOOMING

Jinjun Wang, Shenghuo Zhu, Yihong Gong

NEC Laboratories America, Inc. Cupertino, CA 95014, USA
{jjwang,zsh,ygong}@sv.nec-labs.com

ABSTRACT

This paper presents a novel Resolution-Invariant Image Representation (\mathcal{R} IIR) framework, and applies it for Content-Based Zooming (CBZ) applications. We explain how to generate a multi-resolution bases set, from which the learned image representation can be resolution-invariant. This provides the key technology to support the continues image up-scaling task for the CBZ applications, which existing example-based resolution enhancement approaches cannot handel, or simply 2-D image interpolation algorithm cannot give satisfactory image quality for. We discuss two clustering based methods to construct the bases set. Experimental results show that, both the two methods give good image quality, and the proposed \mathcal{R} IIR framework outperforms existing methods in various aspects.

Index Terms— Image representation, Content-based Zooming, Super-Resolution

1. INTRODUCTION

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 Zooming (CBZ) 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 [1], user attention [2], etc. In this way the viewing process becomes less tedious because the display area is more effectively utilized.

While many existing CBZ works focus on solving the optimal scale, another important issue is how to generate the magnified image region with satisfactory quality. Simple 2-D image interpolation methods cannot recover the missing high-frequency components, and hence their qualities are limited. The super-resolution (SR) idea was introduced by [3] who reconstructed the high-resolution (HR) image by recovering additional information from multiple low-resolution (LR) images using signal processing techniques. However, it cannot

accomplish the single-image resolution enhancement task.

In recent years, the reported example-based SR methods were able to handel the single-frame resolution enhancement problem by learning the co-occurrence prior between HR and LR image patches or coefficients [4], and processing the LR input along with appropriate smoothness constraint [5, 6, 7, 8]. However, in typical example-based SR approaches, 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. In CBZ applications, the image content factors are usually continues, hence continues-scaling capability is necessary for CBZ. Although models at finer granularity can be built for existing example-based SR approaches, solving the optimal parameters for individual models during the magnification process requires huge computation. For this reason, existing example-based SR approaches are not capable.

In this paper, we introduce a novel Resolution-Invariant Image Representation (\mathcal{R} IIR) framework to address the CBZ problem. The \mathcal{R} IIR framework is based on multi-resolution bases set, therefore it is very scalable. Besides, the \mathcal{R} IIR presents better modeling capacity than traditional example-based SR methods [5, 8]. Finally but the more importantly, the \mathcal{R} IIR is resolution-invariant. It can be used to scale the image to various scale by using respective resolution bases. This makes it extremely suitable for CBZ applications.

The remainder of the paper is organized as follows: Section 2 introduces the image representation model in typical resolution enhancement approaches; Section 3 presents our \mathcal{R} IIR framework, and Section 4 applies it to the CBZ application. Experimental results are listed in Section 5, and Section 6 concludes the paper and discussed some future work.

2. IMAGE REPRESENTATION

It is assumed that [5, 7, 9, 10], any HR image $I = I^h + I^m + I^l$ (Fig. 1. a) consists of a high frequency layer (denoted as I^h , Fig. 1. b), a middle frequency layer (I^m , Fig. 1. c), and a low frequency layer (I^l , Fig. 1. d). The down-graded LR image $\bar{I} = I^m + I^l$ results from discarding the high frequency components. Since I^h is independent of I^l [5], estimating I^h requires maximizing $Pr(I^h|I^m)$. Typical example-based SR approach work as follows: First an example patch pair

set $\mathcal{S} = \{\mathcal{S}^m, \mathcal{S}^h\}$ is extracted from the training images. $\mathcal{S}^m = \{\mathbf{p}_i^m\}_{i=1}^N$ and $\mathcal{S}^h = \{\mathbf{p}_i^h\}_{i=1}^N$ represent the middle frequency and the high frequency bases respectively. Each element \mathbf{p}_i^m is the column expansion of a square image patch from \mathbf{I}^m , and \mathbf{p}_i^h is from the corresponding patch in \mathbf{I}^h . Then for a given LR input, the middle frequency patches $\{\mathbf{y}^m\}$ are extracted, and the missing $\{\mathbf{y}^h\}$ can be estimated using the co-occurrence patterns in \mathcal{S} . Methods using Bayesian belief propagation [5], Graph-Cut [9], and Gibbs Sampling [7] to solve a parametric Markov Network to obtain an MAP estimation of \mathbf{y}^h , or Local-Linear-Embedding (LLE) method [8] to learn an optimal linear combination of several candidate patches to represent \mathbf{y}^h , have been reported. However, these methods are either slow or have limited modeling capacity.

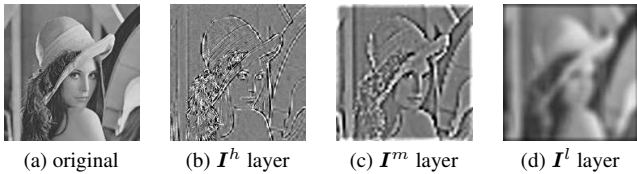


Fig. 1. Image frequency layers

In this paper, we adopted the idea similar to [8] and [10]. To elaborate, define an $D^m \times N$ matrix \mathbf{P}^m to represent the middle frequency patch set \mathcal{S}^m where D^m is the size of each expanded patch (\mathbf{p}_i^m). Since \mathcal{S}^m is typically over-complete, an L1 regularization term is further imposed such that each \mathbf{y}^m is represented using only a few base elements. Hence we seek to find the optimal representation $\{\mathbf{x}^*\}$ by

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}^m - \mathbf{P}^m \mathbf{x}\|^2 + \gamma |\mathbf{x}|. \quad (1)$$

In this paper, $\{\mathbf{x}^*\}$ will be regarded as the image representation. With $\{\mathbf{x}^*\}$, $\{\mathbf{y}^{h*}\}$ can be computed by

$$\{\mathbf{y}^{h*}\} = \{\mathbf{P}^h \mathbf{x}^*\}, \quad (2)$$

where \mathbf{P}^h is the $D^h \times N$ matrix representation of base \mathcal{S}^h , with D^h the size of each \mathbf{p}_i^h .

3. RESOLUTION-INVARIANT REPRESENTATION

3.1. Generating multi-resolution bases set

The base \mathcal{S} created in previous section has single-scale up capacity (defined by the resolution relationship between \mathbf{I} and $\bar{\mathbf{I}}$), while for CBZ applications, continues-scaling is necessary. Hence this subsection discusses how to generate a multi-resolution bases set $\mathcal{S} = \{\mathcal{S}_u^m, \mathcal{S}_u^h\}$, $u = 1, \dots, U$. To start with, for each training image \mathbf{I} , we try to obtain its LR version $\bar{\mathbf{I}}$ by first down-sampling \mathbf{I} to $1/U$ scale, then up-sampling it back to the original size. In this way the high frequency layer \mathbf{I}^h is discarded in $\bar{\mathbf{I}}$. Similar to section 2, N image patch pairs can be extracted from the \mathbf{I}^m and \mathbf{I}^h layers respectively, and we denote the obtained base as $\mathcal{S}_U =$

$\{\mathcal{S}_U^m, \mathcal{S}_U^h\} = \{\mathbf{p}_{i,U}^m, \mathbf{p}_{i,U}^h\}_{i=1}^N$. The subscript U is used to indicate that \mathcal{S}_U is from the U^{th} resolution (\mathbf{I} and \mathbf{I}_U are interchangeable thereafter).

Next, we want to obtain similar bases $\{\mathcal{S}_1, \dots, \mathcal{S}_{U-1}\}$ but from different resolution version of \mathbf{I} . To elaborate, for the u^{th} set \mathcal{S}_u , $u = 1, \dots, U-1$, we shrink \mathbf{I} to u/U size (denoted as \mathbf{I}_u), down-sample \mathbf{I}_u to $1/u$ size then up-sample back to obtain $\bar{\mathbf{I}}_u$. Then we can accordingly extract $\mathcal{S}_u = \{\mathcal{S}_u^m, \mathcal{S}_u^h\} = \{\mathbf{p}_{i,u}^m, \mathbf{p}_{i,u}^h\}_{i=1}^N$. Note that the order of the base elements are arranged in the way 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. With these above steps, the obtained bases set $\mathcal{S} = \{\mathcal{S}_u^m, \mathcal{S}_u^h\}$, $u = 1, \dots, U$ has multi-scale up capacity. However, the key advantage of \mathcal{S} is that, the optimal representation $\{\mathbf{x}^*\}$ solved by Eq. (1) is now resolution-invariant. Hence the same $\{\mathbf{x}^*\}$ solved for \mathcal{S}_u can also be applied for \mathcal{S}_v , $u \neq v$. The next subsection gives the proof.

3.2. Resolution-invariant property

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

$$\bar{\mathbf{I}}_u = \mathbf{I}_U \downarrow_u \otimes \mathbf{G}_u^m, \quad (3)$$

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

$$\begin{aligned} \mathbf{I}_u^m &= \bar{\mathbf{I}}_u - \bar{\mathbf{I}}_u \otimes \mathbf{G}_u^l \\ &= \mathbf{I}_U \downarrow_u \otimes \mathbf{G}_u^m - \mathbf{I}_U \downarrow_u \otimes \mathbf{G}_u^m \otimes \mathbf{G}_u^l, \end{aligned} \quad (4)$$

where \mathbf{G}_u^l simulates a Gaussian blurring kernel to discard the middle frequency layer. Let \mathbf{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 \mathbf{p}_u^m , and \mathbf{y}_u^m be the middle frequency component of an input patch \mathbf{y}_u . With Eq. (4) we can have

$$\mathbf{P}_u^m = \mathbf{P}_U \downarrow_u \otimes \mathbf{G}_u^m - \mathbf{P}_U \downarrow_u \otimes \mathbf{G}_u^m \otimes \mathbf{G}_u^l, \quad (5)$$

and

$$\mathbf{y}_u^m = \mathbf{y}_U \downarrow_u \otimes \mathbf{G}_u^m - \mathbf{y}_U \downarrow_u \otimes \mathbf{G}_u^m \otimes \mathbf{G}_u^l. \quad (6)$$

Taking Eq. (5) and Eq. (6) into Eq. (1), we can get the optimal representation by

$$\mathbf{x}_u^* = \underset{\mathbf{x}_u}{\operatorname{argmin}} \|\mathbf{y}_U - \mathbf{P}_U \mathbf{x}_u\|^2 + \gamma |\mathbf{x}_u|, \quad (7)$$

From Eq. (7), it can be seen that the optimal representation \mathbf{x}_u^* is irrelevant of u , which means that if the different resolution versions of the same image are related by Eq. (3), then the optimal representation learned from any resolution level can also be applied for another level. We call this the *Resolution-Invariant Image Representation* (\mathfrak{R} IIR) property, $\{\mathbf{x}^*\}$ an \mathfrak{R} IIR, and the multi-resolution bases sets \mathcal{S} an \mathfrak{R} IIR set. This \mathfrak{R} IIR framework provides the key technique for our presented CBZ application, as discussed in the next section.

4. CONTENT-BASED ZOOMING

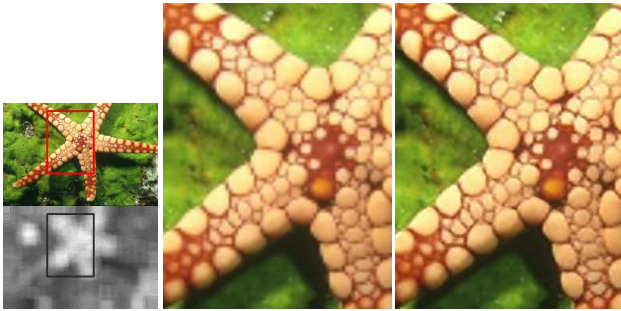
4.1. Continues image scaling using \mathcal{S}

To perform continues image scaling, first a multi-resolution bases set is built as explained in previous section. Then a new base at the required scale can be generated on the fly using interpolation method. To illustrate, 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

$$\mathbf{p}_{i,v} = w_{u,v}\tilde{\mathbf{p}}_{i,u} + (1 - w_{u,v})\tilde{\mathbf{p}}_{i,u+1}, \quad (8)$$

where $\tilde{\mathbf{p}}_{i,u}$ is the patch interpolated from scale u , and $\tilde{\mathbf{p}}_{i,u+1}$ is interpolated from scale $u + 1$. The weight $w_{u,v} = (1 + \exp((v - u - 0.5) * \tau))^{-1}$ where in our implementation, $\tau = 10$. Then with \mathcal{S}_v , and the previously obtained $\{\mathbf{x}^*\}$, the original image can be scaled to $\times v$ resolution.

Now doing CBZ is straightforward. As illustrated in Fig. 2: First, the user navigates to the desired image region (highlighted by the rectangle region), and the optimal scale v is computed based on the content of the region. Then using the stored representations $\{\mathbf{x}^*\}$ and \mathcal{S}_v , the magnified image region can be generated (Fig. 2 middle column). Note that to speed up processing, only patches inside the highlighted region need to be processed.



src region: 74×99 pixels, dest region: 240×320 pixels, scale: $\times 3.25$

Fig. 2. Illustration of Content-Based Zooming (left: the original image and the local complexity map; middle: BiCubic interpolation; right: our results)

4.2. Reducing the redundancy of bases set \mathcal{S}

Although the sampled bases set \mathcal{S} in subsection 3.1 is able to learn the \mathcal{Y} IR, \mathcal{S} is not compact, and thus solving Eq. (1) requires lengthy computation. We want to learn a more compact bases set $\hat{\mathcal{S}} = \{\hat{\mathcal{S}}_u^m, \hat{\mathcal{S}}_u^h\}_{u=1}^U$ to replace \mathcal{S} . Each $\hat{\mathcal{S}}_u^m = \{\mathbf{B}_u^m\}$ or $\hat{\mathcal{S}}_u^h = \{\mathbf{B}_u^h\}$ contains M elements, and $M < N$.

In this paper, we uses clustering method to compress \mathcal{S} . Due to the nature of images, there are many very similar patches in \mathcal{S} , as well as many unique patches. While most clustering methods could group similar patches together, they handel the unique patches differently. In this paper, we attempted both the KMean and the DBScan algorithms. The

KMean algorithm tries to minimize total intra-cluster variance, hence it will absorb unique patches into a nearby cluster. On the other side, the DBScan algorithm estimates the density distribution of the data to find a suitable number of clusters, hence it detects the unique patches as outliers, and can keep them as clusters. Our strategy to use both the two clustering algorithms is depicted in Alg. 4.1. The ‘‘Cluster’’ function in Line 1 can be either KMean or DBScan.

Algorithm 4.1 Redundancy reducing method

input: $\{\mathbf{P}_u^m, \mathbf{P}_u^h\}, u = 1, \dots, U, \epsilon, r$

output: $\{\mathbf{B}_u^m, \mathbf{B}_u^h\}, u = 1, \dots, U$

- 1: label \leftarrow Cluster(\mathbf{P}_r^m)
 - 2: $M \leftarrow$ length(unique(label))
 - 3: **for** $u = 1$ to U **do**
 - 4: $\mathbf{B}_u^m \leftarrow D_u^m \times M$ random matrix, $\mathbf{B}_u^h \leftarrow D_u^h \times M$ random matrix.
 - 5: **for** $i = 1$ to M **do**
 - 6: idx \leftarrow find(label=i),
 - 7: $\mathbf{B}_u^m(:, i) \leftarrow$ mean($\mathbf{P}_u^m(:, idx)$),
 - 8: $\mathbf{B}_u^h(:, i) \leftarrow$ mean($\mathbf{P}_u^h(:, idx)$).
 - 9: **end for**
 - 10: **end for**
-

5. EXPERIMENTAL RESULTS

In subsection 3.2, the \mathcal{Y} IR property has been theoretically proven. However, the image down-grading process in realistic imaging system may not be simply modeled by Eq. (3), hence in the first experiment we wanted to examine how much the \mathcal{Y} IR property holds for real images. First, we generated a five-level multi-resolution bases set \mathcal{S} with around 6000 patch pairs in each level. Next, for each testing image, we extracted around 2000 patches from every resolution level as well, and five versions were used as that in \mathcal{S} . Then we solved Eq. (1) to get their representations $\{\mathbf{x}_{j,u}\}_{j=1}^{2000}, u = 1, \dots, 5$ at each resolution level respectively. If the \mathcal{Y} IR property holds, $\mathbf{x}_{j,u}$ should be very similar to $\mathbf{x}_{j,v}, u \neq v$. Hence we computed the overall similarity between every two elements in $\mathbf{x}_j = \{\mathbf{x}_{j,1}, \mathbf{x}_{j,2}, \mathbf{x}_{j,3}, \mathbf{x}_{j,4}, \mathbf{x}_{j,5}\}$. 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. (1), to get the ‘‘Low sparsity’’ and ‘‘High sparsity’’ profiles. We also tested different number of clusters for both the K-Mean and DBScan methods. As can be seen from Fig. 3, the minimal similarity score is greater than 0.44, and the maximal score reaches almost 0.8. The results shows the high similarities between x_u and x_v , which proves that the \mathcal{Y} IR property holds to a considerable extent for real images. The scores are relatively lower at higher redundancy level because Eq. (1) may not select similar but not exactly the same base elements for recon-

struction. When such redundancy is removed, the similarity becomes significantly higher.

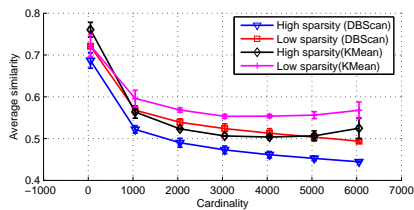


Fig. 3. Similarity between image representations

Next to evaluate the quality of processed images, we magnified 30 testing images and compared the results with two existing example-based SR methods, “KNN” [5] and “LLE” [8], two functional interpolation SR methods, “Enhance” [11] and “Soft edge” [12], and a sparse representation method using non- \mathcal{R} IIR base, “SC” [10], in terms of PSNR score over BiCubic interpolation. As can be seen from Fig. 4, in all the scales, both the “SC” method and our \mathcal{R} IIR method achieved the best results. This is because both the methods utilized a similar representation model as in Eq. (1). However, the representation learned in the “SC” method is not resolution-invariant. Besides, compared to the DBScan algorithm, the bases generated by the KMean algorithm gives slightly better results.

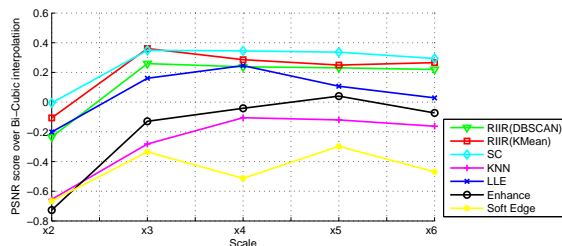


Fig. 4. Comparison of magnified image quality

Finally to evaluate the performance of \mathcal{R} IIR framework for CBZ applications, we designed a subjective user study targeted at a QCIF (320×240 pixels) display size which is typical for most PDA/cell phones. The interface is based on a software simulator running on a Dell Precision 650 PC with 3.2Ghz CPU and 2G memory. 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 Fig. 2. Then an optimal scale factor is calculated based on the content of the selected region, and the \mathcal{R} IIR method was used to magnify the selected region to the QCIF screen. Ten people were asked to rate the results in comparison to BiCubic interpolation, and our \mathcal{R} IIR technique won a large margin.

Note that although solving Eq.(1) is expensive, when $\{x^*\}$ is obtained, reconstructing the magnified image only requires matrix-vector multiplication in Eq.(2). During the above mentioned subjective test, the average processing time is 0.8 second.

6. CONCLUSION AND FUTURE WORK

We proposed the *Resolution-Invariant Image Representation* (\mathcal{R} IIR) framework and applied it for Content-based Zooming. Experimental results show that the \mathcal{R} IIR method improves the quality of magnified image over existing methods in various aspects. The future work includes the following issues: first, in addition to image magnification, we will attempt the \mathcal{R} IIR framework to improve image shrinking quality; second, we will examine additional strategies to learn a more compact bases set with uncompromised or even improved reconstruction accuracy; and third, we will evaluate the \mathcal{R} IIR property in other image representations which can be solved faster, or requires less storage space. This has obvious the importance for coding, streaming, personalization, etc applications.

7. REFERENCES

- [1] P. Chiu, K. Fujii, and Q. Liu, “Content based automatic zooming: Viewing documents on small displays,” *ACM MM’08*, pp. 817–820, 2008.
- [2] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *PAMI*, pp. 1254–1259, 1998.
- [3] R. Tsai and T. Huang, “Multiframe image restoration and registration,” *JAI Press*, pp. 317–339, 1984.
- [4] J. Sun, H. Tao, and H. Shum, “Image hallucination with primal sketch priors,” *CVPR’03*, pp. 729–736, 2003.
- [5] W. Freeman, E. Pasztor, and O. Carmichael., “Learning low-level vision,” *IJCV*, no. 1, pp. 25–47, 2000.
- [6] D. Capel and A. Zisserman, “Super-resolution from multiple views using learnt image models,” *CVPR’01*, pp. 627–634, 2001.
- [7] Q. Wang, X. Tang, and H. Shum, “Patch based blind image super resolution,” *ICCV’05*, pp. 709–716, 2005.
- [8] H. Chang, D. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” *CVPR’04*, 2004.
- [9] U. Mudenagudi, et al, “Super resolution using graph-cut,” *ACCV’06*, pp. 385–394, 2006.
- [10] J. Yang, J. Wright, T. Huang, and M. Yi, “Image super-resolution as sparse representation of raw image patches,” *CVPR’08*, 2008.
- [11] Jinjun Wang and Yihong Gong, “Fast image super-resolution using connected component enhancement,” *ICME’08*, 2008.
- [12] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, “Soft edge smoothness prior for alpha channel super resolution,” *CVPR’07*, pp. 1–8, 2007.