Transductive Regression with Local and Global Consistency for Image Super-Resolution

Xianming Liu[1], Debin Zhao[1], Ruiqin Xiong[2], Siwei Ma[2], Wen Gao[1,2], Huifang Sun[3]
[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[2]School of Electronic Engineering & Computer Science, Peking University, Beijing, China
[3]Mitsubishi Electric Research Laboratories, Cambridge, MA, USA

Abstract

In this paper, we propose a novel image super-resolution algorithm, referred to as interpolation based on transductive regression with local and global consistency (TRLGC). Our algorithm first constructs a set of local interpolation models which can predict the intensity labels of all image samples, and a loss term will be minimized to keep the predicted labels of available low-resolution (LR) samples sufficiently close to the original ones. Then, all of the losses evaluated in local neighborhoods are accumulated together to measure the global consistency on all samples. Furthermore, a graph-Laplacian based manifold regularization term is incorporated to penalize the global smoothness of intensity labels, such smoothing can alleviate the insufficient training of the local models and make them more robust. Finally, we construct a unified objective function to combine together the accumulated loss of the locally linear regression, square error of prediction bias on the available LR samples and the manifold regularization term, which could be solved with a closed-form solution as a convex optimization problem. In this way, a transductive regression algorithm with local and global consistency is developed. Experimental results on benchmark test images demonstrate that the proposed image super-resolution method achieves very competitive performance with the state-of-the-art algorithms.

I. Introduction

Image super-resolution, which is the art of rescaling a low-resolution (LR) image to a high-resolution (HR) version, has become a very active research area in image processing [1-8]. Image super-resolution is born not only in the importance of enhancing resolution of images, such as in the fields of satellite remote sensing and consumer electronics, but also using image super-resolution to understand the validity of different image models in inverse problems

Considering the underlying image models during interpolation, most of image super-resolution algorithms can be categorized as global or local ones. A global algorithm trains the interpolation model using the whole image sample set, while a local algorithm aims to train the model by using only useful local information. The representative global methods are those based on classical data-invariant linear filters, such as bilinear, bicubic [2]. These methods have a relatively low complexity, but suffer from the inability to adapt to varying pixel structures, which result in blurred edges and annoying artifacts. The local algorithms usually result in better empirical results since it is hard to find a unified model with a good predictability for the whole image sample set. In the literature, some local learning methods have been proposed with great success. Li and Orchard [4] propose to adapt the interpolation based on the geometric duality between the LR and the HR covariance. In [5], Zhang and Wu propose to partition the local neighborhood of each missing sample into two oriented subsets in orthogonal directions, and then fuse the

1068-0314/11 $26.00 © 2011 IEEE
DOI 10.1109/DCC.2011.24

directional interpolation results by minimum mean square-error estimation. Takeda *et al.* propose to use kernel regression as an effective tool for interpolation in image processing [6]. Recently, Zhang and Wu propose the named SAI algorithm [7], which learns and adapts varying scene structures using a locally linear regression model, and interpolates the missing pixels in a group by a soft-decision manner.

Image super-resolution is an ill-posed problem. The key to this task is the prior assumption about the properties that the intensity labels should have over the sample set. One common assumption is *intensity consistency*, which means: (1) nearby pixels are likely to have the same or similar intensity labels; and (2) pixels on the same structure (manifold) are likely to have the same or similar intensity labels. Note that the first assumption is local, and the second one is global. Accordingly, it is a smart idea to consider both local and global information contained in the image sample set during learning.

From a machine learning perspective, the available LR image pixels can be regarded as labeled samples and the missing HR image pixels as unlabeled ones. What image super-resolution does is to learn *latent models* of the image sample set in a supervised manner. In all local methods mentioned above, interpolation models are learned only according to the labeled samples in a local neighborhood and then mapped to the missing HR samples to perform inference. During this procedure, unlabeled samples are left out and the information hidden in them are not sufficiently explored. Inspired by the success of semi-supervised learning [8-12], it is reasonable to leverage both labeled and unlabeled data to achieve better predictions.

In this paper, we propose a novel image super-resolution algorithm, referred to as interpolation based on transductive regression with local and global consistency (TRLGC). The basic idea is to predict the intensity label of a data point according to its local neighborhood in a linear way, and then uses a global optimization to ensure robust predictions. Specially, in each neighborhood, an optimal model is estimated via regularized locally linear regression. With this model, the intensity labels of all samples in the neighborhood can be predicted. A loss term will be minimized to keep the predicted labels of available LR samples sufficiently close to the original ones. Then, all of the losses evaluated in local neighborhoods are accumulated together to measure the global consistency on the label and unlabeled data. Furthermore, a graph-Laplacian based manifold regularization term is incorporated to penalize the global smoothness of intensity labels, such smoothing can alleviate the insufficient training of the local models and make them more robust. Finally, we propose a unified loss function to combine together the global loss of the locally linear regression, square error of intensity labels of the available LR samples and the manifold regularization term, which could be solved with a closed-form solution as a convex optimization problem. In this way, a transductive regression algorithm with local and global consistency is developed for image super-resolution.

The rest of this paper is organized as follows. Section II presents the framework of proposed image super-resolution method. Section III illustrates the algorithm details and gives some discussion about the proposed algorithm. Experimental results are provided in Section IV. Section V concludes the paper.

II. The Framework of Proposed Scheme

The image super-resolution problem could be defined as follows: given an image sample set including n pixel points $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l, \mathbf{x}_{l+1}, \cdots, \mathbf{x}_n\} \in \Re^2$, the first l

points are the available LR samples with intensity labels $\{y_i\}_{i=1}^{l}$, the task is to infer the intensity labels $\{y_i\}_{i=l+1}^{n}$ of the remaining $n-l$ missing HR samples in $\{\mathbf{x}_i\}_{i=l+1}^{n}$. By the fact that unlabeled samples are given beforehand and no other test samples will ever be considered, this is a *transductive learning* problem.

A. The Global Principle

Given a set of training sample-label pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, the global principle used for selecting a good interpolation function f is to minimize the following *global structural risk*:

$$\mathcal{J}_g = \sum_{i=1}^{l} \mathcal{L}(y_i, f(\mathbf{x}_i, \mathbf{w})) + \lambda ||f||_F^2, \tag{1}$$

where $\mathcal{L}(\cdot,\cdot)$ is the loss function (*e.g.*, square loss in least square regression), $||f||_F$ is the induced norm of f in the functional space F (*e.g.*, F can be a reproducing kernel Hilbert space (RKHS) induced by some kernel k). Clearly, Eq.(1) is a supervised manner, which only exploits labeled samples to train the interpolation function. Motivated by the great success of semi-supervised learning, it is more reasonable to explore additional discrimination information hidden in unlabeled samples to train a good f.

From a geometric perspective, there is a probability distribution p to generate image samples. The available LR samples are $(\mathbf{x}, y)$ pairs generated according to $p(\mathbf{x}, y)$, the rest missing HR samples are simply drawn according to the marginal distribution $p(\mathbf{x})$ of p. It is usually assumed [12] that there is a specific relationship between $p(\mathbf{x})$ and $p(y|\mathbf{x})$. In another word, if two points $\mathbf{x}_1$ and $\mathbf{x}_2$ are close in the intrinsic geometry of $p(\mathbf{x})$, the conditional distribution $p(y|\mathbf{x}_1)$ and $p(y|\mathbf{x}_2)$ should be similar, *i.e.*, $p(y|\mathbf{x})$ should vary smoothly along the geodesics in the intrinsic geometry of $p(\mathbf{x})$. Accordingly, we define a general geometric framework for model learning, which seeks a global optimal interpolation function f by minimizing the following objective function:

$$\mathcal{R}_g = \sum_{i=1}^{l} \mathcal{L}(y_i, f(\mathbf{x}_i, \mathbf{w})) + \gamma_A ||f||_F^2 + \gamma_I ||f||_I^2, \tag{2}$$

where the additional penalty term $||f||_I^2$ reflects the intrinsic geometric information of the marginal distribution $p(\mathbf{x})$. $||f||_I^2$ can be approximated by

$$||f||_I^2 = \sum_{i,j} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 \mathbf{W}_{ij} = \mathbf{f}^{\mathrm{T}} \mathbf{L} \mathbf{f}, \tag{3}$$

where $\mathbf{W}_{ij}$ is the edge weight in the data adjacency graph which reflects the affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$, $\mathbf{f} = (f(\mathbf{x}_1), \cdots, f(\mathbf{x}_n))^T$ is the intensity label vector of all samples. And $\mathbf{L} = \mathbf{D} - \mathbf{W} \in \Re^{n \times n}$ is the graph Laplacian where $\mathbf{D}$ is a diagonal degree matrix with $\mathbf{D}(i,i) = \sum_j \mathbf{W}_{ij}$.

B. The Local Principle

Actually, as pointed out by [13], it is usually not easy to find an unique function which holds good predictability in the entire data space. But it is much easier to seek some functions that are capable of producing good predictions on some specified regions of the input space. Accordingly, we resort to the local learning model to improve the accuracy

of predictions. More specially, for each data point $\mathbf{x}_i \in \mathbf{X}$, we consider the linear affine transformation function $f_i(\cdot; \mathbf{w}_i, b_i)$ defined as follows:

$$f_i(\mathbf{x}_i; \mathbf{w}_i, b_i) = \mathbf{w}_i^T \Phi(\mathbf{x}_i) + b_i, \tag{4}$$

where $\mathbf{w}_i$ and b_i are the *weight vector* and *bias* of the linear estimator, $\Phi(\mathbf{x}_i) \in \Re^{d\times 1}$ is the intensity label vector of four 8-connected neighboring samples of $\mathbf{x}_i$; $f_i(\mathbf{x}_i)$ is the estimated intensity label of $\mathbf{x}_i$.

To compute the model parameters, we split the whole input space into c local neighborhoods and formulate model learning as a set of optimization problems. It is usually more effective to minimize the following local cost function for each neighborhood $\mathcal{N}_i (1 \leq i \leq c)$:

$$\mathcal{J}_l^i = \sum_{j=1}^{l} \mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) \mathcal{L}(y_j, f_i(\mathbf{x}_j; \mathbf{w}_i, b_i)) + \gamma_A ||\mathbf{w}_i||_F^2, \tag{5}$$

where $\mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i)$ is a local kernel centered at $\mathbf{c}_i$ with width ε_i.

In this way, we define a function f with parameter $(\mathbf{w}_i, b_i)$ for each local region $\mathcal{N}_i$ centered at $\mathbf{c}_i$, that is, we define c local interpolation functions $\{f_i\}_{i=1}^c$. Now it is natural to add together the losses estimated on all of the c neighborhoods, and the *total local structural risk* is

$$\mathcal{J}_l = \sum_{i=1}^{c} \mathcal{J}_l^i. \tag{6}$$

Now let us return to the semi-supervised learning scenario, which aims to learn from both labeled and unlabeled data samples. In some bridge regions (*e.g.*, regions connecting different texture or objects), the number of label points is usually not enough to train a robust predictor. To solve this problem, we propose to introduce a set of pseudo labels $\{f_1, f_2, \cdots, f_n\}$ playing the same role as in some Bayesian methods such that f_i directly determines the final estimated label of $\mathbf{x}_i$. Then we can redefine the total local structural risk as

$$\mathcal{J}_l = \sum_{i=1}^{c} \sum_{j=1}^{n} \mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) \mathcal{L}(f_j, f_i(\mathbf{x}_j; \mathbf{w}_i, b_i)) + \gamma_A ||\mathbf{w}_i||_F^2. \tag{7}$$

In this way, interpolation models are trained locally using $\{\mathbf{x}_i, f_i\}_{i=1}^n$. Note that by minimizing $\mathcal{J}_l$ we can obtain the optimal $\{f_i\}_{i=1}^n$ and $\{(\mathbf{w}_i, b_i)\}_{i=1}^c$.

C. Local and Global Consistency

Recalling the graph-Laplacian regularization framework introduced before, we may also expect f_i to have some geometrical properties. More concretely, we hope $\{f_i\}_{i=1}^n$ to be sufficiently smooth with respect to the intrinsic data graph. Ultimately, we can construct a unified objective function which uses both labeled and unlabeled data and achieves local and global consistency:

$$\mathcal{R}_l = \sum_{j=1}^{l} \mathcal{L}(y_j, f_j) + \lambda \cdot \left(\sum_{i=1}^{c} \sum_{j=1}^{n} \mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) \mathcal{L}(f_j, f_i(\mathbf{x}_j; \mathbf{w}_i, b_i)) + \gamma_A ||\mathbf{w}_i||_F^2 \right) + \gamma_I \mathbf{f}^T \mathbf{L} \mathbf{f}, \tag{8}$$

where the first term is called as *prediction loss*, and its minimization will cause f_i sufficiently close to y_i on the labeled data points. The second term is called as *local*

structural loss, and its minimization will cause f to the have the desired properties as we minimize Eq.(7) regarding f_i as the label of $\mathbf{x}_i$. These two terms punish the predictability and complexity of the local prediction functions, which are therefore called as local regularization. The third term is called as *manifold regularization term*, which penalizes the smoothness of the intensity labels over the entire data graph and is thus referred as global regularization.

III. Implementation Details

In the previous section, we introduce the global and local principal for image super-resolution, and construct a unified framework to perform transductive regression with local and global consistency. To derive a practical image interpolation algorithm, we should define a proper local kernel $\mathcal{K}$ and derive an efficient solution for the objective function defined in Eq.(8). In the following, let us take these two issues into account.

A. Patch-based Bilateral Kernel Weights

In the proposed framework, local kernel weights play a very important role. We hope local kernel weights could provide the prior with the flexibility to model explicitly the local salient features of an image, and could efficiently handle the statistical outliers in transductive regression.

Some efforts in other image processing tasks have been initiated in this direction. *bilateral filter* [14] is proposed to combine gray levels based on both their geometric closeness and their photometric similarity. The bilateral weights can be represented by the following equation:

$$\mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) = \frac{1}{C_i} \exp\left\{-\frac{||\mathbf{x}_j - \mathbf{c}_i||^2}{\varepsilon_i^2}\right\} \exp\left\{-\frac{||y(\mathbf{x}_j) - y(\mathbf{c}_i)||^2}{\varepsilon_i^2}\right\}, \varepsilon_i > 0, \qquad (9)$$

where C_i is the normalization factor. The underlining idea of the bilateral filtering is to do the smoothing according to pixels not only close in the space domain, but also close in feature domain as well, thus the edge sharpness is preserved by avoiding the cross edge smoothing. Bilateral filter performs well in presence of moderate noise, but the comparison of the grey level or color values at a single pixel is no more robust when these values get noisier. This drawback is overcome by the *non-local-means* algorithm [15], in which each weight is proportional to the similarity between the local neighborhood of the pixel being processed and the neighborhood corresponding to other image pixels. The non-local-means weight is defined as follows:

$$\mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) = \frac{1}{C_i} \exp\left\{-\frac{G \cdot ||SW(\mathbf{x}_j) - SW(\mathbf{c}_i)||^2}{\varepsilon_i^2}\right\}, \varepsilon_i > 0, \qquad (10)$$

where G is a Gaussian kernel used to take into account the distance between the central pixel and other pixels in the patch, and $SW(\mathbf{x})$ represents the pixel patch whose components are intensity values of pixels in the similarity window centered on $\mathbf{x}$. This patch comparison permits a reliable similarity measure involving pixels which can fall far away from each other.

Image priors in product form are very attractive since they have the ability to enforce simultaneously many properties on an image. In this paper, we combine the edge-preserving property of bilateral filter and the robust property of non-local-means weight to design

efficient local kernel weights, which are called *patch-based bilateral kernel weights* as define in Eq. (11).

$$\mathcal{K}(\mathbf{x}_j - \mathbf{c}_i, \varepsilon_i) = \frac{1}{C_i} \exp\left\{-\frac{||\mathbf{x}_j - \mathbf{c}_i||^2}{\varepsilon_i^2}\right\} \exp\left\{-\frac{G \cdot ||SW(\mathbf{x}_j) - SW(\mathbf{c}_i)||^2}{\varepsilon_i^2}\right\}, \varepsilon_i > 0. \tag{11}$$

For the data adjacency graph in manifold regularization term, the affinity weight $\mathbf{W}_{ij}$ can be computed in the same way.

B. Optimizing the Objective Function

Now return to the total local loss defined in Eq. (7). With the local kernel $\mathcal{K}$ defined above and the loss function $\mathcal{L}(\cdot, \cdot)$ defined as square loss, the total local loss can be further formulated as

$$\mathcal{J}_l = \sum_{i=1}^{c} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \theta(\mathbf{x}_i, \mathbf{x}_j)(\mathbf{w}_i^T \Phi(\mathbf{x}_j) + b_i - f_j)^2 + \gamma_A ||\mathbf{w}_i||^2, \tag{12}$$

where $\theta(\mathbf{x}_i, \mathbf{x}_j)$ is the non-local-means part of $\mathcal{K}$, and $\mathcal{N}(\mathbf{x}_i)$ represents the local neighborhood centered at $\mathbf{x}_i$. Similarly, the local structural loss in each neighborhood can be rewritten as

$$\mathcal{J}_l^i = \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)} \theta(\mathbf{x}_i, \mathbf{x}_j)(\mathbf{w}_i^T \Phi(\mathbf{x}_j) + b_i - f_j)^2 + \gamma_A ||\mathbf{w}_i||^2. \tag{13}$$

Let

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{\Phi}_i^T & \mathbf{1} \\ \sqrt{\gamma_A} \cdot \mathbf{I}_d & \mathbf{0} \end{bmatrix}, \mathbf{\Phi}_i = [\Phi(\mathbf{x}_{i_1}), \Phi(\mathbf{x}_{i_2}), \cdots, \Phi(\mathbf{x}_{i_{n_i}})], \widehat{\mathbf{f}}_i = [f_{i_1}, f_{i_2}, \cdots, f_{i_{n_i}}, \mathbf{0}^T]^T, \tag{14}$$

where $\mathbf{x}_{i_j}$ is the j-th neighbor of $\mathbf{x}_i$, n_i is the cardinality of $\mathcal{N}(\mathbf{x}_i)$, $\mathbf{I}_d$ is the $d \times d$ identity matrix, $\mathbf{1} = [1, 1, \cdots, 1]^T \in \Re^{n_i \times 1}$ and $\mathbf{0}$ is a $d \times 1$ zero vector, Eq. (13) can be formulated in the matrix form as

$$\mathcal{J}_l^i = \left(\mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widehat{\mathbf{f}}_i\right)^T \cdot \mathbf{V} \cdot \left(\mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widehat{\mathbf{f}}_i\right), \tag{15}$$

where $\mathbf{V} = diag\left(\theta(\mathbf{x}_i, \mathbf{x}_{i_1}), \theta(\mathbf{x}_i, \mathbf{x}_{i_2}), \cdots, \theta(\mathbf{x}_i, \mathbf{x}_{i_{n_i}}), 1, \cdots, 1\right) \in \Re^{(n_i+d)\times(n_i+d)}$. $\mathbf{V}$ is a diagonal matrix, so the above equation can be further formulated as

$$\begin{aligned} \mathcal{J}_l^i &= \left(\mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widehat{\mathbf{f}}_i\right)^T \left(\mathbf{V}^{\frac{1}{2}}\right)^T \cdot \mathbf{V}^{\frac{1}{2}} \left(\mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widehat{\mathbf{f}}_i\right) \\ &= \left(\mathbf{V}^{\frac{1}{2}} \mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \mathbf{V}^{\frac{1}{2}} \widehat{\mathbf{f}}_i\right)^T \left(\mathbf{V}^{\frac{1}{2}} \mathbf{G}_i \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \mathbf{V}^{\frac{1}{2}} \widehat{\mathbf{f}}_i\right). \end{aligned} \tag{16}$$

Let $\widetilde{\mathbf{G}_i} = \mathbf{V}^{\frac{1}{2}} \mathbf{G}_i$ and $\widetilde{\mathbf{f}_i} = \mathbf{V}^{\frac{1}{2}} \widehat{\mathbf{f}}_i$, the above equation becomes

$$\mathcal{J}_l^i = \left(\widetilde{\mathbf{G}_i} \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widetilde{\mathbf{f}_i}\right)^T \left(\widetilde{\mathbf{G}_i} \begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix} - \widetilde{\mathbf{f}_i}\right). \tag{17}$$

To derive the optimal transformation parameters $(\mathbf{w}_i, b_i)$, we take the derivative of the loss function $\mathcal{J}_l^i$ with respect to $(\mathbf{w}_i, b_i)$ and set the derivative to 0, then the optimal solution can be represented by

$$\begin{bmatrix} \mathbf{w}_i \\ b_i \end{bmatrix}^* = (\widetilde{\mathbf{G}_i}^T \widetilde{\mathbf{G}_i})^{-1} \widetilde{\mathbf{G}_i}^T \widetilde{\mathbf{f}_i} \tag{18}$$

With this solution, the total structural loss defined in Eq. (12) becomes

$$\mathcal{J}_l = \sum_i \mathcal{J}_l^i = \sum_i \widetilde{\mathbf{f}_i}^T \widehat{\mathbf{G}_i}^T \widehat{\mathbf{G}_i} \widetilde{\mathbf{f}_i}, \tag{19}$$

where $\widehat{\mathbf{G}_i} = \mathbf{I} - \widetilde{\mathbf{G}_i}(\widetilde{\mathbf{G}_i}^T \widetilde{\mathbf{G}_i})^{-1} \widetilde{\mathbf{G}_i}^T$. For $\widehat{\mathbf{G}_i}$, we have the following theorem.

Theorem 1. $\widehat{\mathbf{G}_i}$ is a orthogonal projection matrix

According to the property of orthogonal projection matrix, $\mathcal{J}_l$ can be rewritten as

$$\mathcal{J}_l = \sum_i \mathcal{J}_l^i = \sum_i \widetilde{\mathbf{f}_i}^T \widehat{\mathbf{G}_i} \widetilde{\mathbf{f}_i}. \tag{20}$$

We split the matrix $\widehat{\mathbf{G}_i}$ into four blocks after the n_i-th row and column:

$$\widehat{\mathbf{G}_i} = \begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{C}_i & \mathbf{D}_i \end{bmatrix}, \tag{21}$$

where $\mathbf{A}_i \in \Re^{n_i \times n_i}$. Let $\mathbf{f}_i = [f_{i_1}, f_{i_2}, \cdots, f_{i_{n_i}}]^T$ and $\mathbf{K} = diag\left(\theta(\mathbf{x}_i, \mathbf{x}_{i_1}), \theta(\mathbf{x}_i, \mathbf{x}_{i_2}), \cdots, \theta(\mathbf{x}_i, \mathbf{x}_{i_{n_i}})\right) \in \Re^{n_i \times n_i}$, then

$$\widetilde{\mathbf{f}_i} = \mathbf{V}^{\frac{1}{2}} \widehat{\mathbf{f}_i} = \begin{bmatrix} \mathbf{K}^{\frac{1}{2}} & \\ & \mathbf{I}_d \end{bmatrix} \begin{bmatrix} \mathbf{f}_i \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{K}^{\frac{1}{2}} \mathbf{f}_i \\ \mathbf{0} \end{bmatrix}. \tag{22}$$

According to Eq. (21) and Eq. (22), we can derive

$$\widetilde{\mathbf{f}_i}^T \widehat{\mathbf{G}_i} \widetilde{\mathbf{f}_i} = [\mathbf{f}_i^T \mathbf{K}^{\frac{1}{2}} \quad \mathbf{0}^T] \begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{C}_i & \mathbf{D}_i \end{bmatrix} \begin{bmatrix} \mathbf{K}^{\frac{1}{2}} \mathbf{f}_i \\ \mathbf{0} \end{bmatrix} = \mathbf{f}_i^T \mathbf{K}^{\frac{1}{2}} \mathbf{A}_i \mathbf{K}^{\frac{1}{2}} \mathbf{f}_i = \mathbf{f}_i^T \widehat{\mathbf{A}_i} \mathbf{f}_i. \tag{23}$$

For $\widehat{\mathbf{A}_i}$, we have the following theorem:

Theorem 2.

$$\widehat{\mathbf{A}_i} = \mathbf{I}_{n_i} - \left(\Psi^T \mathbf{F}_i^{-1} \Psi + \frac{\Psi^T \mathbf{F}_i^{-1} \Psi \Upsilon \Upsilon^T \Psi^T \mathbf{F}_i^{-1} \Psi}{1-e} - \frac{\Psi^T \mathbf{F}_i^{-1} \Psi \Upsilon \Upsilon^T}{1-e} - \frac{\Upsilon \Upsilon^T \Psi^T \mathbf{F}_i^{-1} \Psi}{1-e} + \frac{\Upsilon \Upsilon^T}{1-e} \right), \tag{24}$$

where $\Psi^T = \mathbf{K}^{\frac{1}{2}} \mathbf{\Phi}_i^T$, $\Upsilon = \mathbf{K}^{\frac{1}{2}} \mathbf{1}$, and $\mathbf{F}_i = \Psi \Psi^T + \gamma_A \mathbf{I}_d$, $e = \Upsilon^T \Psi^T \mathbf{F}_i^{-1} \Psi \Upsilon$.

Let $\mathbf{f} = [f_1, f_2, \cdots, f_n]^T \in \Re^{n \times 1}$ be the intensity vector of all samples, $\widehat{\mathbf{f}} = [\mathbf{f}_1^T, \mathbf{f}_2^T, \cdots, \mathbf{f}_c^T] \in \Re^{(n_i \times c) \times 1}$ be the concatenated label vector , and defining the selection matrix $\mathbf{S}$ which is a 0-1 matrix with $\mathbf{S}_{ij} = 1$ if $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_i)$, we can get $\widehat{\mathbf{f}} = \mathbf{S}\mathbf{f}$. We further define the block-diagonal matrix

$$\mathbf{G} = \begin{pmatrix} \widehat{\mathbf{A}_1} & & 0 \\ & \ddots & \\ 0 & & \widehat{\mathbf{A}_c} \end{pmatrix} \in \Re^{(n_i \times c) \times (n_i \times c)},$$

then Eq. (20) can be rewritten as

$$\mathcal{J}_l = \mathbf{f}^T \mathbf{S}^T \mathbf{G} \mathbf{S} \mathbf{f}. \tag{25}$$

Let $\mathbf{M} = \mathbf{S}^T \mathbf{G} \mathbf{S}$, so finally Eq. (12) can be rewritten as

$$\mathcal{J}_l = \mathbf{f}^T \mathbf{M} \mathbf{f}. \tag{26}$$

Therefore, we can derive the final formulation of the loss function

$$\mathcal{R}_l = (\mathbf{f} - \mathbf{y})^T \mathbf{J} (\mathbf{f} - \mathbf{y}) + \lambda \mathbf{f}^T \mathbf{M} \mathbf{f} + \gamma_I \mathbf{f}^T \mathbf{L} \mathbf{f}, \tag{27}$$

where $\mathbf{y}$ records the intensity labels of the labeled image samples and $\mathbf{J} \in \Re^{n \times n}$ is a diagonal matrix whose diagonal elements are one for labeled samples and zero for unlabeled data. By taking $\partial \mathcal{R}_l / \partial \mathbf{f} = 0$, we can derive a closed-form solution as

$$\mathbf{f} = 2 \left(2\mathbf{J} + \lambda (\mathbf{M} + \mathbf{M}^T) + 2\gamma_I \mathbf{L} \right)^{-1} \mathbf{J} \mathbf{y}. \tag{28}$$

C. Discussion

In our method, the first term in Eq. (27) is related to intensity label fitting, which means that the predicted intensity labels of the LR image samples should not change too much from the original ones. The second term is the sum of local loss on all neighborhoods, which means the predicted intensity labels should not change too much between neighboring samples. The third term is the graph-Laplacian based manifold regularization term, which keeps the intensity labels globally smooth and can alleviate the insufficient training of the local models and make them more robust. Meanwhile, intensity label information is propagated from labeled samples to unlabeled ones through Laplacian graph, and finally a global optimal propagation is achieved.

It is much easier to understand our method from a mixed-regularization perspective. That is, the last two terms of Eq. (27) can both be reviewed as regularization terms with different regularization matrices, one is derived from local learning and the other is derived from global geometric. Different types of regularization matrices may better reveal different (maybe complementary) information and thus could provide a more accurate predictor.

IV. Experimental Results

In this section, experimental results are presented to demonstrate the advantage of the proposed TRLGC algorithm. For thoroughness and fairness of our comparison study, we exploit some widely used images as test ones. Fig. 1 lists the used seven sample images in our experiments. Our algorithm is compared with some representative work in the literature. More specifically, seven approaches are included in our comparative study: (1) bicubic interpolation [2], (2) locally-adaptive zooming algorithm (LAZA) [3], (3) new edge-directed interpolation (NEDI) [4], (4) DFDF [5], (5) kernel regression (KR) for image reconstruction [6] (6) SAI interpolation [7], (7) our approach.

Following the same setting as SAI, we downsample these HR images by a factor of two in both row and column dimensions to get the corresponding LR images, from which the original HR images are reconstructed by the proposed and competing methods. Since the original HR images are known in the simulation, we can compare the interpolated results with the true images, and measure the objective and subjective quality of those interpolated images. Table I tabulates the objective quality comparison with respect to

Fig. 1: Seven sample images in the test set.

TABLE I: The PSNR(dB)and SSIM results comparison of seven interpolation algorithms

Image	Bicubic		LAZA		NEDI		DFDF		KR		SAI		TRLGC	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Airplane	30.17	0.9119	30.17	0.9122	28.74	0.9117	30.53	0.9144	29.11	0.9049	30.72	0.9177	**30.88**	**0.9183**
Tower	39.84	0.9882	38.43	0.9875	39.85	0.9905	39.69	0.9892	40.28	0.9897	41.49	0.9919	**41.73**	**0.9940**
Girl	31.54	0.7651	31.84	0.7809	31.95	0.7884	31.81	0.7741	31.92	0.7775	31.77	0.7702	**32.15**	**0.7890**
Peppers	27.58	0.9105	27.51	0.9050	27.21	0.9082	27.83	0.9152	27.02	0.9122	27.66	0.9200	**27.99**	**0.9208**
Flowers	25.44	0.6759	25.65	0.6850	25.70	0.6929	25.74	0.6844	25.79	0.6848	25.96	0.6973	**26.19**	**0.7015**
Door	32.23	0.8617	32.28	0.8618	32.22	0.8587	32.27	0.8607	32.20	0.8505	32.46	0.8643	**32.60**	**0.8660**
Splash	33.65	0.9293	33.45	0.9281	33.38	0.9290	33.79	0.9256	33.38	0.9245	33.54	0.9298	**33.99**	**0.9320**
Average	31.49	0.8632	31.34	0.8658	31.29	0.8685	31.67	0.8662	31.39	0.8634	31.94	0.8702	**32.17**	**0.8745**

PSNR of the seven different methods when applied to the seven test images of Fig. 1. It can be observed that for all instances the proposed TRLGC algorithm consistently works better than other methods. Compared with global methods, such as bicubic, the proposed method can significantly improve the objective quality of generated HR images. The average gains is 0.68dB. Our method also outperforms the edge detection based local methods, such as LAZA and FDI, for which the average gains are 0.83dB and 0.5dB respectively. By exploiting labeled and unlabeled samples together and keeping local and global consistency in transductive regression, our method leads to a significant performance benefits compared with kernel regression interpolation. The gain is 0.78dB in terms of average PSNR. NEDI and SAI are both based on auto-regression model. Our method is more effective compared with them. The TRLGC method can improve 0.88dB and 0.23dB with respect to average PSNR compared with NEDI and SAI respectively.

PSNR can measure the intensity difference between two images, but it may fail to describe the visual perception quality of the image. How to evaluate the visual quality of an image is a very difficult problem and an active research topic. The SSIM index proposed in [16] is one of the most commonly used measures for image visual quality assessment. We further use SSIM to measure the visual quality of these interpolation algorithms. From Table I, it could be seen that TRLGC again achieves the highest average SSIM scores among the competing methods. It means our method can keep the image structure more wonderful.

V. Conclusion

In this paper, we presented an efficient image super-resolution algorithm based on transductive regression with local and global consistency. Our method is novel in two

aspects: (1) both labeled and unlabeled data are explored in the process of model learning. Such a transductive manner is particularly crucial for some bridge regions (*e.g.*, regions connecting different texture or objects) and can further boost the performance; (2) local and global consistency is achieved during regression, which can make the predictor more robust. These two aspects can be cast into a unified optimization framework, which can be efficiently solved with a closed-form solution. Experimental results on benchmark test images demonstrate that the proposed method achieves very competitive interpolation performance with the state-of-the-art interpolation algorithms.

ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundations of China (60736043, 61073083, 60803068), in part by the Major State Basic Research Development Program of China (973 Program 2009CB320904)

REFERENCES

[1] R. Xiong and W. Gao, "Tanner Graph Based Image Interpolation," IEEE Data Compression Conference 2010 (DCC2010), pp. 376-385, Snowbird, Utah, USA, March 24-26, 2010

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

[3] S. Battiato, G. Gallo, and F. Stanco. "A locally-adaptive zooming algorithmfor digital images," Image and Vision Computing, 20:805-812, 2002.

[4] X. Li and M. T. Orchard, "New edge-directed interpolation," IEEE Transactions on Image Processing, vol. 10, no. 10, pp. 1521-1527, Oct. 2001.

[5] L. Zhang and X. Wu, "Image interpolation via directional filtering and data fusion," IEEE Transactions on Image Processing, vol. 15, no. 8, pp.2226-2238, Augest 2006.

[6] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," IEEE Transactions on Image Processing, vol. 16, No. 2, pp. 349-366, February 2007.

[7] X. Zhang, X. Wu, "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," IEEE Transactions on Image Processing, vol.17, No. 6, pp.887-896, 2008

[8] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," In The 20th International Conference on Machine Learning (ICML), 2003.

[9] Fei Wang, "A general learning framework using local and global regularization," Pattern Recognition 43(9): 3120-3129 (2010)

[10] C. Cortes, M. Mohri, "On Transductive Regression," NIPS 2006: 305-312 June 2008.

[11] M. Wu and B. Scholkopf, "Transductive Classification via Local Learning Regularization," Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007), 628-635.

[12] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from examples," J. Mach. Learn. Res., vol. 7, pp. 2399-2434, 2006.

[13] V. Vapnik (1995), The Nature of Statistical Learning Theory, Springer, NY.

[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in Proceedings of the IEEE International Conference on Computer Vision, pages 839-846, 1998.

[15] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Modeling Simulation, vol. 4,no. 2, pp. 490-530, 2005.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, Apr. 2004.