Image Restoration via Multi-Prior Collaboration

Feng Jiang[1],[2], Shengping Zhang[1], Debin Zhao[1] and S.Y. Kung[2]

[1]School of Computer Science, Harbin Institute of Technology
[2]School of Electrical engineering, Princeton University

Abstract. This paper proposes a novel multi-prior collaboration framework for image restoration. Different from traditional non-reference image restoration methods, a big reference image set is adopted to provide the references and predictions of different popular prior models and accordingly further guide the subsequent multi-prior collaboration. In particular, the collaboration of multi-prior models is mathematically formulated as a ridge regression problem. Due to expensive computation complexity of handling big reference data, scatter-matrix-based kernel ridge regression is proposed, which achieves high accuracy while low complexity. Additionally, an iterative pursuit is further proposed to obtain refined and robust restoration results. Five popular prior methods are applied to evaluate the effectiveness of the proposed multi-prior collaboration framework. Compared with the state-of-the-art image restoration approaches, the proposed framework improves the restoration performance significantly.

1 Introduction

Mathematically, image restoration aims to reconstruct the original high quality image $\boldsymbol{u}$ from its observed degradation $\boldsymbol{y}$, which is a typical ill-posed linear inverse problem and can be generally formulated as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{u} + \boldsymbol{n} \ , \tag{1}$$

where $\boldsymbol{u} \in \mathbb{R}^d$ and $\boldsymbol{y} \in \mathbb{R}^d$ are lexicographically stacked representations of the original image and the degraded image, respectively, $\boldsymbol{H} \in \mathbb{R}^{d\times d}$ is a matrix representing a non-invertible linear degradation operator and $\in \mathbb{R}^d$ is usually additive Gaussian white noise. When $\boldsymbol{H}$ is an identity matrix, the problem becomes image denoising [1]; when $\boldsymbol{H}$ is a blur operator, the problem becomes image deblurring [2]; when $\boldsymbol{H}$ is a mask, that is, $\boldsymbol{H}$ is a diagonal matrix whose diagonal entries are either 1 or 0, the problem becomes image inpainting [3]; when $\boldsymbol{H}$ is a set of random projections, the problem becomes compressive sensing [4, 5]. To cope with the ill-posed nature of image restoration, one type of scheme in the literature employs image prior knowledge for regularizing the solution to the following minimization problem

$$\arg\min_{\boldsymbol{u}} \frac{1}{2}\|\boldsymbol{H}\boldsymbol{u} - \boldsymbol{y}\|_2^2 + \lambda\Psi(\boldsymbol{u}) \ , \tag{2}$$

where $\frac{1}{2}\|\boldsymbol{H}\boldsymbol{v}-\boldsymbol{y}\|_2^2$ is the ℓ_2-norm based data-fidelity term, $\Psi(\boldsymbol{u})$ denotes image prior (also called the regularization term) and λ is the regularization parameter. In fact, the above regularization-based minimization can be strictly derived from Bayesian inference with prior knowledge in image generation models [6]. Many optimization approaches for regularization-based image inverse problems have been developed [7–9].

It has been widely recognized that image prior knowledge plays a critical role for image restoration approaches. Therefore, designing effective regularization terms to reflect intrinsic image prior models is at the core of image restoration. Classical regularization terms utilize local structural patterns and are built on the assumption that images are locally smooth except at edges. Several representative work in the literature includes half quadrature formulation [10], Mumford-Shah (MS) model [11] and total variation (TV) models [12]. These regularization terms demonstrate high effectiveness in preserving edges and recovering smooth regions. However, they usually smear out image details and cannot deal well with fine structures since they only exploit local statistics, neglecting nonlocal statistics of nature images.

In recent years, the most significant nonlocal statistics in image processing is perhaps the nonlocal self-similarity exhibited by natural images. The nonlocal self-similarity depicts the repetitiveness of higher level patterns (e.g., textures and structures) globally positioned in images. A representative work is the popular nonlocal means (NLM) [13], which takes advantage of this image property to conduct a type of weighted filtering for denoising tasks. This simple weighted approach is quite effective in generating sharper image edges and preserving more image details. Later, inspired by the success of nonlocal means (NLM) denoising filter, a series of nonlocal regularization terms have been proposed [14], which can be roughly divided into two categories according to their formulations. The first one is directly derived from NLM [15], since nonlocal filtering can essentially be understood as a quadratic regularization based on a nonlocal graph, as detailed for instance in the geometric diffusion framework in [16]. The other one goes one step further to solve general inverse problems by incorporating nonlocal graph into traditional regularization terms, such as nonlocal total variation (NL/TV) [17] and nonlocal Mumford-Shah (NL/MS) [18]. Due to the utilization of self-similarity prior by adaptive nonlocal graph, nonlocal regularization terms produce superior performance over the local ones, with the ability of preserving sharper image edges and more image details [19]. Nonetheless, there are still plenty of image details and structures that cannot be recovered accurately. The reason is that the above nonlocal regularization terms depend on the weighted graph, while it is inevitable that the weighted manner gives rise to disturbance and inaccuracy [20].

Vision information in natural images is extremely complex. Although many statistical priors have been explored from distinctive viewpoints, each prior model has its shortcomings especially for a specific applications. Intuitively, it is natural to consider to combine the existing prior models and let them collaborate to result in a better solution. However, from point of view of theory, is difficult to

propose a general collaboration framework that combines multiple prior models. Some previous work incorporates two or more prior models into a regularization-based framework for image restoration [21–23]. In [21], the local total variation model and nonlocal adaptive 3-D sparse representation model are combined to solve image restoration from partial random samples in spatial domain. In CS-MRI (Compressive Sensing Magnetic Resonance Imaging) models, the linear combination of total variation and wavelet sparse regularization is popular [22, 23].

The existing methods of combining multiple prior models mainly add the prior knowledge as a regularization term in the objective function to be optimized. Although some improvements have been reported, the complex objective function with pre-defined weights of different priors and the high computation complexity in the optimization process restrict the regularization methods from being a general framework. The regularization terms depend on the weighted graph, however it is inevitable that the weighted manner gives rise to disturbance and inaccuracy. In this paper, a novel and general multi-prior collaboration framework is proposed as shown in Fig. 1. Instead of merging multi-prior in a traditional predefined regularization term, we evaluated the potential of the prior models on current degraded images as well as their coupling dynamically. The collaboration of multiple priors is mathematically formulated as a ridge regression problem. Due to the computation complexity of dealing with big reference data, scatter-matrix-based kernel ridge regression (KRR) is proposed. Compared with the traditional KRR, scatter-matrix-based KRR achieves relative low computation cost and ensures accuracy and robustness at the same time.

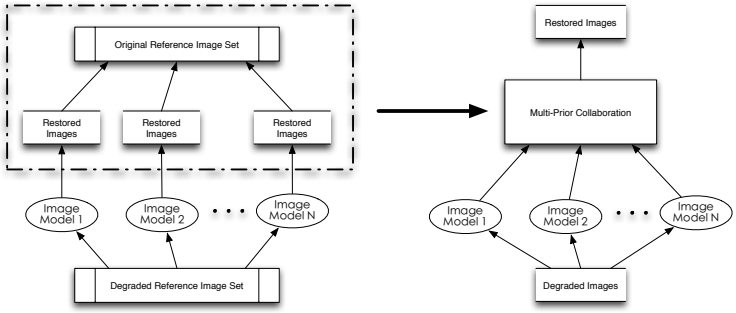

Fig. 1. The proposed multi-prior collaboration framework.

2 Kernel Ridge Regression based Multi-Prior Collaboration Model

Given the prior model set $\mathbb{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_M\}$ and a degraded image $\boldsymbol{I}'$, we can obtain the restored images $\{\boldsymbol{I}'_{r\oplus 1}, \ldots, \boldsymbol{I}'_{r\oplus M}\}$ with M prior models respectively. let $\boldsymbol{I}$ denote the non-degraded image corresponding to the degraded image $\boldsymbol{I}'$. It is reasonable to assume the residual $\boldsymbol{I} - \boldsymbol{I}'$ has a linear relationship with the innovation brought by each prior model $\{\boldsymbol{I}'_{r\oplus 1} - \boldsymbol{I}', \ldots, \boldsymbol{I}'_{r\oplus M} - \boldsymbol{I}'\}$, which can be expressed as

$$\boldsymbol{I} - \boldsymbol{I}' = [\boldsymbol{I}'_{r\oplus 1} - \boldsymbol{I}', \boldsymbol{I}'_{r\oplus 2} - \boldsymbol{I}', \ldots, \boldsymbol{I}'_{r\oplus M} - \boldsymbol{I}']\boldsymbol{\omega} = \Gamma\boldsymbol{\omega} \tag{3}$$

where $\Gamma = [\boldsymbol{I}'_{r\oplus 1} - \boldsymbol{I}', \ldots, \boldsymbol{I}'_{r\oplus M} - \boldsymbol{I}'] \in \mathbb{R}^{d\times M}$ and $\boldsymbol{\omega} \in \mathbb{R}^M$ is the weight vector for the restoration which can be obtained by a regression process. The final restoration image $\hat{\boldsymbol{I}}$ can be obtained as

$$\hat{\boldsymbol{I}} = \boldsymbol{I}' + \Gamma\boldsymbol{\omega} \ . \tag{4}$$

In the multi-prior collaboration, the training data, i.e., the reference images, are used to predict the restoration performance with respect to the current degraded image. Ideally, a reference images set with the similar content and category can evaluate the restoration capability of the different priors when applied to the current degraded image.

Define $\mathcal{R} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_N\}$ is the set of reference nature images with the similar content and category with the current degraded image $\boldsymbol{I}'$, and their corresponding degraded images consist of the set of reference degraded images $\mathcal{R}' = \{\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_N\}$. For each degraded image $\boldsymbol{R}'_n \in \mathbb{R}^d$, a restored result $\boldsymbol{R}'_{n\oplus m}$ can be obtained with prior $\mathcal{P}_m$. Obviously, according to Eq. 3, the reference images should satisfy

$$\begin{bmatrix} \boldsymbol{R}'_{1\oplus 1} - \boldsymbol{R}'_1 & \ldots & \boldsymbol{R}'_{1\oplus M} - \boldsymbol{R}'_1 \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}'_{N\oplus 1} - \boldsymbol{R}'_N & \ldots & \boldsymbol{R}'_{N\oplus M} - \boldsymbol{R}'_N \end{bmatrix} \boldsymbol{\omega} = \boldsymbol{X}^\top\boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{R}_1 - \boldsymbol{R}'_1 \\ \vdots \\ \boldsymbol{R}_N - \boldsymbol{R}'_N \end{bmatrix} = \boldsymbol{y} \ , \tag{5}$$

where $\boldsymbol{X} \in \mathbb{R}^{M\times\hat{N}}$ is the sample matrix and $\hat{N} = d\times N$ is the number of samples.

A classic solution to overcome the above over-fitting problem in the regression analysis is the so-called ridge regularization, also known as Tikhonov regularization, in which a ridge penalty term is imposed on the objective function so as to keep the regressor coefficients under control. This leads to the classic ridge regression method. Given a finite training dataset, the objective of a linear regressor is to minimize the following cost function

$$E_{RR}(\boldsymbol{\omega}) = \sum_{i=1}^{\hat{N}} \epsilon_i^2 + \rho\|\boldsymbol{\omega}\|^2 \ , \tag{6}$$

where $\epsilon_i = \boldsymbol{\omega}^\top\boldsymbol{x}_i - y_i$ and $\boldsymbol{x}_i$ is the i-th column of matrix $\boldsymbol{X}$ and y_i is the i-th element of vector $\boldsymbol{y}$ and ρ is the ridge parameter. The principle of ridge

regression lies in the incorporation of a penalty term into the loss function to control the regressor coefficients. Being adjustable by the ridge parameter ρ, the penalty term can effectively mitigate the over-fitting problem and thus enhance the robustness of the learned classifier.

2.1 Kernel Based Approach: Learning Models in Empirical Space

With a kernel function $\phi : \mathbb{R}^M \rightarrow \mathbb{R}^J$ mapping a sample $\boldsymbol{x}_i$ from the M-dimensional space to the J-dimensional space, the objective is to find the decision vector $\boldsymbol{u} \in \mathbb{R}^J$ and the threshold b such that

$$\min_{u,b} E_{KRR}(\boldsymbol{u}, b) = \min_{u,b} \{\sum_{i=1}^{\hat{N}} \epsilon_i^2 + \rho \|\boldsymbol{u}\|_2^2\} \ , \tag{7}$$

where

$$\epsilon_i = \boldsymbol{u}^\top \phi(\boldsymbol{x}_i) + b - y_i \ , \forall i = 1, \ldots, N \ . \tag{8}$$

In matrix notation,

$$E_{KRR}(\boldsymbol{u}, b) = \|\boldsymbol{\Phi}^\top \boldsymbol{u} + be - \boldsymbol{y}\|_2^2 + \rho \|\boldsymbol{u}\|_2^2 \ . \tag{9}$$

where $\boldsymbol{\Phi} = [\phi(\boldsymbol{x}_1), \ldots, \phi(\boldsymbol{x}_{\hat{N}})] \in \mathbb{R}^{J \times \hat{N}}$. The zero-gradient point of $E_{KRR}(\boldsymbol{u}, b)$ with respect to $\boldsymbol{u}$ can be obtained as

$$\frac{\partial E_{KRR}(\boldsymbol{u}, b)}{\partial \boldsymbol{u}} = 2\boldsymbol{\Phi}(\boldsymbol{\Phi}^\top \boldsymbol{u} + be - \boldsymbol{y}) + 2\rho \boldsymbol{u} = 0 \ . \tag{10}$$

By Eq. 10, we have

$$\boldsymbol{u} = -\rho^{-1} \boldsymbol{\Phi}(\boldsymbol{\Phi}^T \boldsymbol{u} + be - \boldsymbol{y}) \ . \tag{11}$$

Thus, there exists an $\hat{N}$-dimensional vector $\boldsymbol{a}$ such that $\boldsymbol{u} = \boldsymbol{\Phi}\boldsymbol{a}$. This establishes the validity of LSP, *i.e.* $\boldsymbol{u} \in \text{span}[\boldsymbol{\Phi}]$. The knowledge of the LSP is instrumental for the actual solution for $\boldsymbol{u}$. More exactly, on plugging the LSP, $\boldsymbol{u} = \boldsymbol{\Phi}$ *boldsymbola* into Eq. 9, we obtain

$$E'_{KRR}(\boldsymbol{a}, b) = \|\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\boldsymbol{a} + be - \boldsymbol{y}\|_2^2 + \rho \boldsymbol{a}^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi}\boldsymbol{a} \tag{12}$$

$$= \|\boldsymbol{K}\boldsymbol{a} + be - \boldsymbol{y}\|_2^2 + \rho \boldsymbol{a}^\top \boldsymbol{K}\boldsymbol{a} \ . \tag{13}$$

where $\boldsymbol{K} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \in \mathbb{R}^{\hat{N} \times \hat{N}}$. The zero-gradient point of $E'_{KRR}(\boldsymbol{a}, b)$ with respect to $\boldsymbol{a}$ leads to

$$\begin{bmatrix} \boldsymbol{K} + \rho \boldsymbol{I} & \boldsymbol{e} \\ e^\top & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ b \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \ , \tag{14}$$

2.2 Scatter-matrix based approach: Multi-prior Models Collaboration in the Intrinsic Space

By Eq. 10, we have the following optimal decision vector

$$\boldsymbol{u} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \rho\boldsymbol{I})^{-1}\boldsymbol{\Phi}(\boldsymbol{y} - b\boldsymbol{e}) \tag{15}$$

$$= (\boldsymbol{S} + \rho\boldsymbol{I})^{-1}\boldsymbol{\Phi}(\boldsymbol{y} - b\boldsymbol{e}) \ . \tag{16}$$

where $\boldsymbol{S} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\top \in \mathbb{R}^{J\times J}$. The first-order gradient of $E'_{KRR}(\boldsymbol{u}, b)$ with respective to b leads to

$$\frac{\partial E'_{KRR}(\boldsymbol{u}, b)}{\partial b} = \boldsymbol{e}^\top\boldsymbol{\Phi}^\top\boldsymbol{u} + b\boldsymbol{e}^\top\boldsymbol{e} - \boldsymbol{e}^\top\boldsymbol{y} \tag{17}$$

$$= \boldsymbol{e}^\top\boldsymbol{\Phi}^\top\boldsymbol{u} + \hat{N}\times b - \boldsymbol{e}^\top\boldsymbol{y} \tag{18}$$

Let the first-order gradient be zero, the solution for KRR may be derived from the matrix system

$$\begin{bmatrix} \boldsymbol{S} + \rho\boldsymbol{I} & \boldsymbol{\Phi}\boldsymbol{e} \\ \boldsymbol{e}^\top\boldsymbol{\Phi}^\top & \hat{N} \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ b \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}\boldsymbol{y} \\ \boldsymbol{e}^\top\boldsymbol{y} \end{bmatrix} \ , \tag{19}$$

We have so far independently derived the optimal KRR solutions both for the intrinsic space and for the empirical-space formulation. Logically, both approaches should lead to the same solutions. However, as an alternative verification, we shall now provide an algebraic method to directly connect the two solutions. For the kernel-matrix-based solution (Eq. 20), note that $\boldsymbol{K} = \boldsymbol{\Phi}^\top\boldsymbol{\Phi}$, we obtain

$$\begin{bmatrix} \boldsymbol{e}^\top & -\rho \end{bmatrix} \begin{bmatrix} \boldsymbol{K} + \rho\boldsymbol{I} & \boldsymbol{e} \\ \boldsymbol{e}^\top & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ b \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}^\top & -\rho \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \tag{20}$$

$$\begin{bmatrix} \boldsymbol{e}^\top\boldsymbol{K} + \boldsymbol{e}^\top\rho\boldsymbol{I} - \rho\boldsymbol{e}^\top & \boldsymbol{e}^\top\boldsymbol{e} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ b \end{bmatrix} = \boldsymbol{e}^\top\boldsymbol{y} \tag{21}$$

$$\begin{bmatrix} \boldsymbol{e}^\top\boldsymbol{K} & \boldsymbol{e}^\top\boldsymbol{e} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ b \end{bmatrix} = \boldsymbol{e}^\top\boldsymbol{y} \tag{22}$$

$$\boldsymbol{e}^\top\boldsymbol{K}\boldsymbol{a} + \boldsymbol{e}^\top\boldsymbol{e}b = \boldsymbol{e}^\top\boldsymbol{y} \tag{23}$$

$$\boldsymbol{e}^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\boldsymbol{a} + \hat{N}b = \boldsymbol{e}^\top\boldsymbol{y} \tag{24}$$

$$\boldsymbol{e}^\top\boldsymbol{\Phi}^\top\boldsymbol{u} + \hat{N}\boldsymbol{b} = \boldsymbol{e}^\top\boldsymbol{y} \tag{25}$$

By Woodbury's matrix identity [24], we have

$$(\boldsymbol{K} + \rho\boldsymbol{I})^{-1} = \rho^{-1}\boldsymbol{I} - \rho^{-1}\boldsymbol{\Phi}^T(\rho\boldsymbol{I} + \boldsymbol{S})^{-1}\boldsymbol{\Phi} \tag{26}$$

and it follows that

$$\boldsymbol{\Phi}(\boldsymbol{K} + \rho\boldsymbol{I})^{-1} = \rho^{-1}\boldsymbol{\Phi} - \rho^{-1}\boldsymbol{S}(\rho\boldsymbol{I} + \boldsymbol{S})^{-1}\boldsymbol{\Phi} = (\rho\boldsymbol{I} + \boldsymbol{S})^{-1}\boldsymbol{\Phi} \ . \tag{27}$$

We can obtain

$$\boldsymbol{u} = \boldsymbol{\Phi}\boldsymbol{a} = \boldsymbol{\Phi}(\boldsymbol{K} + \rho\boldsymbol{I})^{-1}(\boldsymbol{y} - b\boldsymbol{e}) = (\rho\boldsymbol{I} + \boldsymbol{S})^{-1}\boldsymbol{\Phi}(\boldsymbol{y} - b\boldsymbol{e}) \ . \tag{28}$$

On pre-multiplying the above by $(\rho \boldsymbol{I} + \boldsymbol{S})$, we arrive at

$$(\rho \boldsymbol{I} + \boldsymbol{S})\boldsymbol{u} = \boldsymbol{\Phi y} - \boldsymbol{\Phi e b} \tag{29}$$

By combining Eq. 25 and Eq. 29, which is equivalent to the formulation given in Eq. 19 derived for the original space.

3 Iterative pursuit degraded image restoration

In order to obtain more refined restoration results, an iterative pursuit is provide to guide the restoration process. This process is based on the following assumptions: the reference big data simulate the restoration process and their performance will give a reliable prediction of the current restoration, which will provide the restoration for each single prior models. Firstly, in each iterative step each image in reference image set is restored with prior models respectively and the results and the corresponding residual with the ground truth is then used to get the multi prior model corporation rule; then, each image in the reference degraded set is further updated with the same rule. A simple illustration is shown as Fig. 2. In each step of the iterative process, for each reference image $\boldsymbol{R}'_n$, M restored images $\{\boldsymbol{R}'_{n\oplus 1}, \ldots, \boldsymbol{R}'_{n\oplus M}\}$ are obtained, then projected to $\{\widetilde{\boldsymbol{R}}'_{n\oplus 1}, \ldots, \widetilde{\boldsymbol{R}}'_{n\oplus J}\}$ where $\widetilde{\boldsymbol{R}}'_{n\oplus j}(m,n) = \phi_j(\boldsymbol{R}'_{n\oplus 1}(m,n), \ldots, \boldsymbol{R}'_{n\oplus M}(m,n))$. The ground truth $\boldsymbol{R}_n$ is also projected into $\{\widetilde{\boldsymbol{R}}_{n\oplus 1}, \ldots, \widetilde{\boldsymbol{R}}_{n\oplus J}\}$ where $\widetilde{\boldsymbol{R}}_{n\oplus j}(m,n) = \phi_j(\boldsymbol{R}_{n\oplus 1}(m,n), \ldots, \boldsymbol{R}_{n\oplus M}(m,n))$.

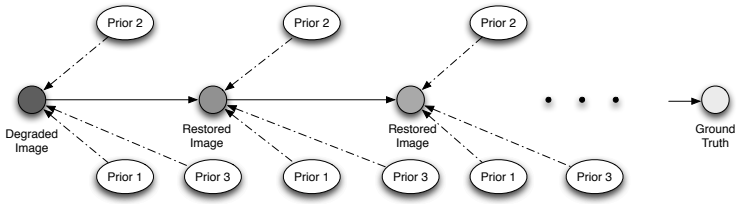

Fig. 2. An illustration of the proposed multi-prior collaboration framework.

In the situation of linear regression, for each prior model $\mathcal{P}_m$, it has N restoration results when applied to the set $\mathcal{R}'$. The divergent of PSNR of these N restored images reflects the prediction certainty of $\mathcal{P}_m$. Supposed for degraded images $\mathcal{R}'$, $\mathcal{P}_m$ gets the prediction certainty, i.e., the variation of Peak Signal to Noise Ratio (PSNR) as σ_m^2, then in the linear collaboration, the total certainty can be expressed as $\sum_{m=1}^{M} \omega_m^2 \sigma_m^2$ Since a prior model performs restoration differently for different images, we hope the restored results of each prior model has least divergent.

Similarly, in the nonlinear situation, the prediction certainty can be introduced into Eq. 7 and resulting the following objective function

$$E'_{KRR}(\boldsymbol{u}, b) = \sum_{i=1}^{\hat{N}} \epsilon_i^2 + \rho\|\boldsymbol{u}\|_2^2 + \lambda \sum_{j=1}^{J} u_j^2 \sigma_j^2 \ , \tag{30}$$

where $\sigma_j^2 = var(PSNR(\widetilde{\boldsymbol{R}}'_{1\oplus j}, \widetilde{\boldsymbol{R}}_{1\oplus j}), \ldots, PSNR(\widetilde{\boldsymbol{R}}'_{N\oplus j}, \widetilde{\boldsymbol{R}}_{N\oplus j}))$.

In matrix notation,

$$E'_{KRR}(\boldsymbol{u}, b) = \|\boldsymbol{\Phi}^\top \boldsymbol{u} - \boldsymbol{y}\|_2^2 + \rho\|\boldsymbol{u}\|_2^2 + \boldsymbol{u}^\top \boldsymbol{U} \boldsymbol{u} \ . \tag{31}$$

where $\boldsymbol{U} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2)$ is a diagonal matrix.

The zero-gradient point of $E'_{KRR}(\boldsymbol{u}, b)$ with respect to $\boldsymbol{u}$ can be obtained as

$$\frac{\partial E_{KRR}(\boldsymbol{u}, b)}{\partial \boldsymbol{u}} = 2\boldsymbol{\Phi}(\boldsymbol{\Phi}^\top \boldsymbol{u} - \boldsymbol{y}) + 2(\rho \boldsymbol{I} + \boldsymbol{U}) = 0 \ . \tag{32}$$

The solution is

$$\boldsymbol{u} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \rho \boldsymbol{I} + \boldsymbol{U})^{-1} \boldsymbol{\Phi}(\boldsymbol{y} - b\boldsymbol{e}) = (\boldsymbol{S} + \rho \boldsymbol{I} + \boldsymbol{U})^{-1} \boldsymbol{\Phi}(\boldsymbol{y} - b\boldsymbol{e}) \ . \tag{33}$$

The detailed procedure of the proposed image restoration is summarized in Algorithm 1.

4 Experiments

To evaluate the performance of the proposed framework, we take 5 classical priors: TV, BM3D [25], FOE [26], KSVD [27], BLSGSM [28]. The total variation (TV) model [29] has been quite successful used in many aspects of the image processing. TV model favors the piecewise smoothness, and it represents a local prior. In this approach, TV model serve as a regularizer in an unconstrained convex minimization problem, and the parameter before TV model controls the tradeoff between the regularity and fidelity terms. And the Block-matching 3D (BM3D) [25] algorithm is based on the phenomenon that a patch (block) tends to recur in a natural image. Denoising process is mainly that grouping similar 2D image blocks into 3D data arrays then through the procedure of collaborative filtering. Obviously, this method is based on a nonlocal prior. The Fields of Experts(FOE) [26] that extends traditional Markov Random Field (MRF) models by learning potential functions over extended pixel neighborhoods is a classic method, The approach develop a framework for learning generic, expressive image priors that capture the statistics of natural scenes. And the approach of KSVD [27] taken is based on sparse representation over the trained overcomplete dictionary. In the sparse representation modeling the choice of dictionary is an important issue, and there has been much effort in learning dictionaries from a set of example image patches [27]. In the algorithm of BLS-GSM [28], the authors describe the method for removing noise from digital images, based on a statistical model of the coefficients of an overcomplete multi-scale oriented basis.

Algorithm 1: The proposed multi-prior collaboration and degraded image restoration

Input : corrupted image $\boldsymbol{I}'$, M prior models $\mathbb{P} = \{\mathcal{P}_1, \ldots, \mathcal{P}_M\}$, reference non-degraded images set $\mathcal{R} = \{\boldsymbol{R}_1, \boldsymbol{R}_2, \ldots, \boldsymbol{R}_N\}$, λ, ρ
Output: Final restored image
Initialization: obtain reference degraded image set $\mathcal{R}_d = \{\boldsymbol{R}'_1, \boldsymbol{R}'_2, \ldots, \boldsymbol{R}'_N\}$
repeat
  for *each* $\mathcal{P} \in \mathbb{P}$ **do**
    Restore I' with $\mathcal{P}_m$ and obtain I'_m
  end
  Obtain $\phi_{I'} = \phi(I'_1 - I', I'_2 - I', \cdots, I'_M - I') \in \mathbb{R}^{d \times J}$
  for *each* $\boldsymbol{R}'_n \in \mathcal{R}_d$ **do**
    for *each* $\mathcal{P} \in \mathbb{P}$ **do**
      Restore $\boldsymbol{R}'_n$ with $\mathcal{P}_m$ and obtain $\boldsymbol{R}'_{n\oplus m}$
    end
    Obtain $\phi_{R'_n} = \phi(\boldsymbol{R}'_{n\oplus 1} - \boldsymbol{R}'_n, \boldsymbol{R}'_{n\oplus 2} - \boldsymbol{R}'_n, \cdots, \boldsymbol{R}'_{n\oplus M} - \boldsymbol{R}'_n) \in \mathbb{R}^{d \times J}$
    project $\{\boldsymbol{R}'_{n\oplus 1}, \ldots, \boldsymbol{R}'_{n\oplus M}\}$ to $\{\widetilde{\boldsymbol{R}}'_{n\oplus 1}, \ldots, \widetilde{\boldsymbol{R}}'_{n\oplus J}\}$ with
    $\widetilde{\boldsymbol{R}}'_{n\oplus j}(m,n) = \phi_j(\boldsymbol{R}'_{n\oplus 1}(m,n), \ldots, \boldsymbol{R}'_{n\oplus M}(m,n))$
    Translate $\boldsymbol{R}_n$ to $\{\widetilde{\boldsymbol{R}}_{n\oplus 1}, \ldots, \widetilde{\boldsymbol{R}}_{n\oplus J}\}$ with
    $\widetilde{\boldsymbol{R}}_{n\oplus j}(m,n) = \phi_j(\boldsymbol{R}_{n\oplus 1}(m,n), \ldots, \boldsymbol{R}_{n\oplus M}(m,n))$
  end
  Obtain $\boldsymbol{U}$ and $\boldsymbol{\Phi} = (\phi_{R'_1}^\top, \phi_{R'_2}^\top, \ldots, \phi_{R'_N}^\top)^\top$
  Compute $\boldsymbol{u}$ using Eq. 33
  Update $\boldsymbol{I}' = \boldsymbol{u}^\top \phi_{I'} + \boldsymbol{I}'$
  for *each* $\boldsymbol{R}'_n \in \mathcal{R}_d$ **do**
    Update $\boldsymbol{R}'_n = \boldsymbol{u}^\top \phi_{R'_n} + \boldsymbol{R}'_n$
  end
until *Stopping criterion is satisfied*

In experimental settings, the type and intensity of noise is known, that is, the noise model is determined. So almost all the proposed algorithms rely on some explicit or implicit assumptions about the true(noise free) image. The experimental results of TV, BM3D, KSVD, BLSGSM and FOE are all generated by the original authors' codes, with the corresponding parameters manually optimized.

4.1 Linear Collaboration of multi-prior models

Because of the fact that the content of the image affect restoration effect, eight training (and test) images are come from a similar certain class of textured images. These eight images are shown in Figure 3.

Table 1 lists PSNR results of TV model and BM3D model collaboration. All the images are degraded by additive Gaussian noise $\sigma = 12$. The coefficients of each image is obtained by leave-one-out cross validation. That means,in order to obtain the coefficient of an image we train the other 7 images.

Fig. 3. Images used for evaluation and reference set (grey, 512×512).

Fig. 4. comparison of Gaussian noise removal.

Table 1. TV and BM3D priors collaboration performance (PSNR).

Priors	***img1***	***img2***	***img3***	***img4***	***img5***	***img6***	***img7***	***img8***
TV	32.36	32.14	32.73	30.39	30.42	30.78	31.00	32.56
BM3D	32.89	32.37	33.26	30.79	30.84	31.03	31.52	33.06
Coef_of TV	0.1299	0.1280	0.1325	0.1362	0.1369	0.1352	0.1383	0.1331
Coef_of BM3D	0.7735	0.7802	0.7701	0.7768	0.7755	0.7780	0.7720	0.7698
The proposed	33.01	32.60	33.35	30.99	31.03	31.25	31.70	33.15
Gain	0.12	0.23	0.09	0.20	0.19	0.22	0.18	0.09

In the Table 1, the second and third rows represent the restoration results from TV and BM3D models respectively. The fourth and fifth rows represent the coefficients of each image. The sixth row reports the performance of the collaboration of the TV and BM3D with the proposed method. Generally, BM3D method can get a better result than TV model in this image training set, so we list the gain between BM3D result and combination result in seventh row. From Table 1, we could find that the coefficients of TV part are in the range of (0.1280, 0.1383) and the coefficients of BM3D part are in the range of (0.7698, 0.7802). The coefficients are relatively stable.

4.2 Nonlinear Collaboration of multi-prior models

One of the distinct difference of our method is the introduction of reference data in the image restoration process. The proposed multi-prior collaboration framework is expected a better performance on a bigger reference data set. To verify the learning efficiency of nonlinear collaboration of multi-prior models with big data, we use the INRIA Holiday dataset [29], which has 1491 images in total. One advantage of this data set is that there are multiple images in the same scene captured at different viewpoints and focal lengths, which ensure the contained images have the similar content and category. 8 images are selected as test images and the rest are used as reference images. These same scene and category images are used to simulate the similar images obtained form a big training data for the degraded images. All the test images are degraded by additive Gaussian noise $\sigma = 12$. We used iterative pursuit restoration and TRBF3 kernel in this experiment and the results are shown as Table 2. Nine configurations are used to evaluate our proposed multi-prior model collaboration.

Table 2. TV and BM3D priors collaboration performance (PSNR).

	Priors	***img1***	***img2***	***img3***	***img4***	***img5***	***img6***	***img7***	***img8***	***Gain***
1	T&F	32.65	32.35	32.97	30.77	30.79	30.97	31.40	32.80	0.29
2	T&K	32.12	32.48	33.08	30.93	30.95	31.16	31.54	32.92	0.31
3	T&BL	32.86	32.45	33.11	30.81	30.83	31.07	31.52	32.95	0.14
4	BM&F	32.98	32.58	33.33	30.98	31.02	31.23	31.69	33.14	0.15
5	BM&K	33.39	32.98	33.73	31.38	31.42	31.64	32.10	33.54	0.55
6	BM&BL	33.15	32.72	33.46	31.11	31.14	31.36	31.83	33.27	0.28
7	F&K	32.86	32.52	33.13	31.05	31.06	31.22	31.66	32.98	0.39
8	F&BL	32.98	32.56	33.24	30.97	30.99	31.19	31.67	33.08	0.27
9	K&BL	32.89	32.46	33.12	30.90	30.92	31.12	31.58	32.96	0.18

T: TV prior; **K**: K-SVD prior; **BL**: BLS-GSM prior;
BM: BM3D prior. **Training data**: INRIA Holiday dataset

Obviously in all the configuration, the collaborated prior models considerably outperforms the unique prior restoration in all the cases, with the highest PSNR, achieving the average PSNR improvements over the performance of unique prior adopted is 0.55 dB. With the adopted of big reference data set, the proposed scatter-matrix-based KRR have a significant improvement in the collaboration of multi-prior. An average of 0.28 DB improvement is obtain in the image restoration.

In the configuration of BLS_GSM prior and K-SVD prior, FoE prior and K-SVD prior, the proposed method get the most significant improvement 0.55 and 0.39 DB. It can be seen that BLS_GSM and K-SVD, FOE and KSVD models have less coupled, while in the configuration of TV and BLS_GSM, only 0.14 DB is got, it implied that there were prior coupled in these two prior models.

If a TRBF3 kernel is adopted, the scatter-matrix-based KRR has an overwhelming speed advantage over direct KRR. While the conventional KRRs com-

plexity is N^3, the complexity of the scatter-matrix-based KRR is $\max\{J^3, NJ^2\}$. This represents a major saving in computational time. This shows that when N is huge, the learning complexity becomes very costly both for direct KRR. In contrast, the scatter-matrix based KRR has a complexity with a growth rate proportional to N. It allows a super-saving when N becomes enormously large. It effectively overcomes the curse of empirical degree.

The large-scale database of images plays important role in the quality of reconstructed images. For an input image, if we cannot find highly correlated images in the cloud, the scheme cannot output a satisfying quality reconstruction. Although in our experiment, we used a predefined reference data set, significant improvement is achieved compared with the unique prior. A refined data set with more similar content is expected to further improve the restoration performance.

5 Conclusions

In this paper we present a novel framework for image restoration by combining multiple prior models. An iterative procedure is adopted in which the reference data are used for evaluation and prediction in each step. Different prior models are assessed based on their performance on reference data dynamically and scatter-matrix-based KRR is adopted to collaborate the multi-prior models with liner computational complexity with the size of reference data set, which applies to the bid data situation. Our proposed method not only provides an effective way to collaborate multiple prior models, but also demonstrates the coupling relationship of the exiting prior models. It provides a metric for the further prior model design. Besides, the scatter-matrix-based KRR shows significant capability of solving large-scale image restoration while with relative low computation complexity.

Acknowledgement. This work was supported in part by the Major State Basic Research Development Program of China (973 Program 2015CB351804) and the National Natural Science Foundation of China under Grant No. 61272386, 61100096 and 61300111.

References

1. Bioucas-Dias, J., Figueiredo, M.: A new twIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image Processing **16** (2007) 2992–3004
2. Dong, W., Zhang, L., Shi, G., Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Transactions on Image Processing **20** (2011) 1838–1857
3. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research **11** (2010) 19–60
4. Donoho, D.L.: Compressed sensing. IEEE Transactions on Information Theory **52** (2006) 1289–1306

5. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory **52** (2006) 489–509
6. Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381** (1996) 607–609
7. Eckstein, J., Bertsekas, D.: On the douglas rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55** (1992) 293–318
8. Yin, W., Osher, S., Goldfarb, D., Darbo, J.: Bregman iterative algorithms for L1-minimization with applications to compressed sensing. SIAM Journal on Imaging Sciences **1** (2008) 142–168
9. Setzer, S.: Operator splittings, bregman methods and frame shrinkage in image processing. International Journal of Computer Vision **92** (2011) 265–280
10. Geman, D., Reynolds, G.: Constrained restoration and the recovery of discontinuitie. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 367–383
11. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics **42** (1989) 577–685
12. Chan, R., Dong, Y., Hintermuller, M.: An efficient two-phase L1-TV method for restoring blurred images with impulse noise. IEEE Transactions on Image Processing **19** (2010) 1731–1739
13. Efros, A.A., Leung, T.K.: Texture synthesis by non-parametric sampling. In: ICCV. Volume 2. (1999) 1022–1038
14. Buades, A., Coll, B., Morel, J.M.: Image enhancement by non-local reverse heat equation. CMLA Technical Report **22** (2006)
15. Zhang, J., Liu, S., Xiong, R., Ma, S., Zhao, D.: Improved total variation based image compressive sensing recovery by nonlocal regularizatio. In: IEEE International Symposium on Circuits and Systems. (2013) 2836–2839
16. Coifman, R.R., Lafon, S., Lee, A.B., Maggioni, M., Nadler, B., Warner, F., Zucker, S.W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. In: Proceedings of the National Academy of Sciences. Volume 102. (2005) 7426–7431
17. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. CMLA Technical Report **23** (2007)
18. Jung, M., Bresson, X., Chan, T.F., Vese, L.A.: Nonlocal mumford-shah regularizers for color image restoration. IEEE Transactions on Image Processing **20** (2011) 1583–1598
19. Kindermann, S., Osher, S., Jones, P.: Deblurring and denoising of images by nonlocal functionals. Multiscale Modeling and Simulation **4** (2005) 1091–1115
20. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM Journal on Imaging Sciences **3** (2010) 253–276
21. Zhang, J., Xiong, R., Ma, S., Zhao, D.: High-quality image restoration from partial random samples in spatial domain. In: VCIP. (2011) 1–4
22. Ma, S., Yin, W., Zhang, Y., Chakraborty, A.: An efficient algorithm for compressed mr imaging using total variation and wavelets. In: CVPR. (2008) 1–8
23. Chen, C., Huang, J.: Compressive sensing mri with wavelet tree sparsity. In: NIPS. (2012) 1124–1132
24. Woodbury, M.A.: Inverting modi?ed matrices. Princeton University (1950)

25. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. Image Processing, IEEE Transactions on **16** (2007) 2080–2095
26. Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Volume 2., IEEE (2005) 860–867
27. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. Signal Processing, IEEE Transactions on **54** (2006) 4311–4322
28. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixtures of gaussians in the wavelet domain. Image Processing, IEEE Transactions on **12** (2003) 1338–1351
29. Jegou, H., Douze, M., Schmid, C.: Inria holidays dataset (2008)