Single Image Super-resolution using Deformable Patches

Yu Zhu[1], Yanning Zhang[1], Alan L. Yuille[2]
[1]School of Computer Science, Northwestern Polytechnical University, China
[2]Department of Statistics, UCLA, USA
zhuyu1986@mail.nwpu.edu.cn, ynzhang@nwpu.edu.cn, yuille@stat.ucla.edu

Abstract

We proposed a deformable patches based method for single image super-resolution. By the concept of deformation, a patch is not regarded as a fixed vector but a flexible deformation flow. Via deformable patches, the dictionary can cover more patterns that do not appear, thus becoming more expressive. We present the energy function with slow, smooth and flexible prior for deformation model. During example-based super-resolution, we develop the deformation similarity based on the minimized energy function for basic patch matching. For robustness, we utilize multiple deformed patches combination for the final reconstruction. Experiments evaluate the deformation effectiveness and super-resolution performance, showing that the deformable patches help improve the representation accuracy and perform better than the state-of-art methods.

1. Introduction

Single image *super-resolution* (SR) [4, 8, 9, 11, 12, 23] is a technology that recovers a *high-resolution* (HR) image from one *low-resolution* (LR) input image. It is more ill-posed than SR on the image sequence [5, 14] since there is no interlaced sampling information between frames for single image SR. A key point in single image SR problem is what extra information or prior could be used for estimating the HR details. The most common SR method is analytical interpolation based on simple smoothness assumption. Moreover, more sophisticated priors, *e.g.* the edge statistics priors [6, 17], are also exploited in SR literature.

Recent progresses show that the image patches exhibit promising ability to express a variety of local structures [9, 16, 22, 25]. By using patches, the example-based SR approaches estimate HR details by seeking for the most similar one [9] or the best linear combination of them [4, 12, 23]. Another research direction [8, 11] utilizes the self-similarity based on the fact that local image structures tend to repeat within and across the scales.

An inevitable difficulty in SR is the correspondence ambiguity between HR and LR patches. In other words, we may find several different HR patches corresponding to the same LR patch, regardless of what prior is applied. This may lead to the artifacts or blurring textures. In example-based SR, a trivial solution is to make the dictionary large enough to cover as many visual patterns as possible. But this makes the patches correspondence even more ambiguous. To address this problem, an alternative method is the joint learning of HR/LR patch dictionary. This leads to a more compact dictionary [12, 21, 23], however, the problem still remains due to inherent large ambiguities.

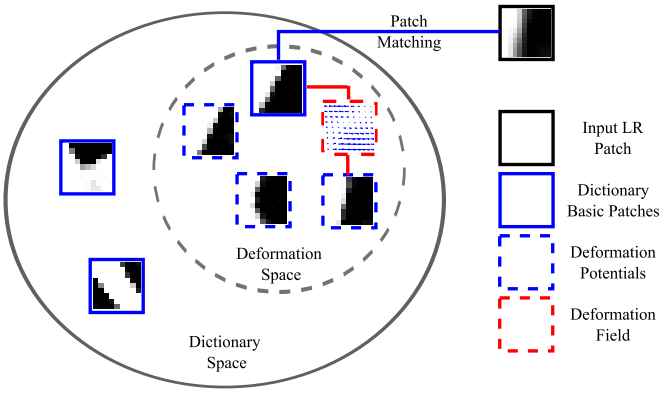


Figure 1. The idea of deformable patches. The dictionary may not contain the patches in dictionary space. But the basic patch can be deformed to a potential patch to fit the input LR patch. Thus the dictionary can express more patterns using the finite basic patches.

If we allow one or more HR patches to deform to match the LR patches, it becomes more likely to find the true HR patches among the deformed versions of basic patches in dictionary. On basis of this idea, we use patches as a deformation field rather than a fixed vector. As shown in Figure 1, it can represent a bundle of deformed variants, making the dictionary capable of covering more visual patterns. The deformation allows continuous warping of basic patch, with rotation and translation transforms as the particular cases, which potentially corresponds to a manifold of image patch subspace in practice. The deformation field is similar

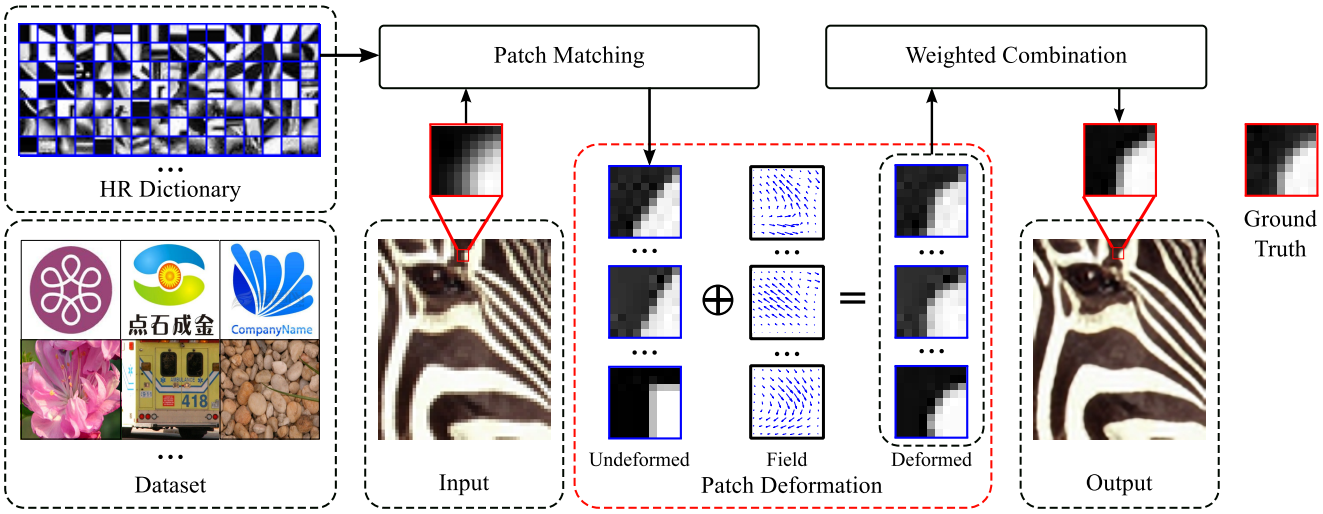


Figure 2. Overview of the proposed method. The input LR image is interpolated to the HR image size and cropped into LR patches. For each LR patch, we choose the best basic patches via deformation similarity. After being deformed, these patches are weighted combined. Here we select 3 patches from real experiment for illustration. Note that he super-resolved result is very similar to the ground truth.

to that arising when modeling optical flow [7, 13], but it has not been used for patch modeling or super-resolution.

In this paper, we propose a novel deformable-patch-based method for single image SR, aiming to improve performance by exploiting a more expressive dictionary. Figure 2 illustrates the framework of our method. The main contributions of this paper are summarized as below:

1. We propose a deformable patches model for single image SR problem, making the dictionary more expressive.
2. We develop an effective patch matching strategy to select the best basic patch for deformation, based on a deformation cost between the LR input and HR patch.
3. We extend our deformation model of single patch to a weighted combination of several deformed candidates for more robust and reliable HR estimation.

2. Related Work

For single image SR, the most popular methods are bilinear and bicubic interpolations based on the "smoothness" assumption, which is simple but easily leads to the artifacts and blurring effect around the image discontinuities such as edges and corners. In contrast, the edge statistics prior is more sophisticated and effective. Representative work includes Fattal [6], Sun *et al*. [17] and Tai *et al*. [18]. Nevertheless, a few of parameters are far too insufficient to handle more complex cases in an image. Meanwhile, gradient cue is very sensitive to the noise. Recent studies show that image structures tend to repeat themselves within and across scales. On basis of this observation, many HR details can be recovered from self-examples instead of the external database. Glasner *et al*. [11] and Freedman [8] show that the self-examples can be helpful in the case of discontinuous structures. But for uniform textures, the false edges tend to occur.

Example-based SR methods usually use a universal set of example patches to predict the missing high frequency details. For a reliable HR details prediction, Freeman *et al*. [9] proposed a MRF method solved by belief propagation to impose neighborhood consistency constraints. Another way is to make the learned relationship (*e.g*. couple dictionary) more generative and compact. Neighbor embedding based method[3, 4, 10] is inspired by LLE algorithm from manifold learning. Under the assumption of manifold local consistency for HR and LR patches, the HR details is predicted based on linear combination of its K neighbors estimated by corresponding LR neighbors. Similarly, Yang *et al*. [22, 23] use sparse representation for corresponding HR/LR dictionary elements with shared coefficients, leading to a compact and powerful dictionary. As the extension, He *et al*. [12] use beta process for sparse coding, allowing a mapping function between HR and LR coefficients. Nevertheless, all of these methods mentioned above use patches as a fixed vectors. This requires an extremely large dictionary to cover the input patch structures or linear combination components. Another work Ye [24] is related to ours. They use deformable patches for digit image recognition. But the deformation in their paper is actually a rigid affine transformation, *i.e*. scaling, rotating and translating at given interval to form the new patches. There is no degradation as that in SR problem.

Moreover, our work is also related to classical optical flow approaches. Under the smoothness assumption, Horn-

Schunck method [13] exploits a first order Taylor series expansion to model the flow field. We follow the similar but different way. The deformation in our work is imposed on the patches at different resolution rather than of between two adjacent image frames in optical flow. And also, our application is different, *i.e*., we deform patches to give them the ability to appear in different shapes, hence making the dictionary more expressive.

3. Deformable Patches for Super-resolution

In this section, we present a deformable patch model for super-resolution and develop the algorithm to obtain the solution.

For single image super-resolution, the LR patch $\boldsymbol{Y}$ is a blurred and downsampled version of the HR patch $\boldsymbol{X}$:

$$\boldsymbol{Y} = \boldsymbol{DHX} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{D}$ is the downsampling matrix, $\boldsymbol{H}$ the blurring matrix, and $\boldsymbol{n}$ is the noise term. All the patches here are vectorized for matrix representation.

The degradation gives the fundamental constraint that the estimated HR patch should be consistent with the LR input via degrading. The deformable patch is under the same constraint in our deformation model.

3.1. Deformation on Single Patch

3.1.1 Deformation Model

We start to present our model with the premise that we have got a basic HR patch $\boldsymbol{B}_h$ for deformation. Our mission is to deform the patch to fit the observed LR input. In Section 3.2 we will elaborate on how to choose the appropriate patch from the dictionary. Note that $\boldsymbol{B}_h$ is also used for denoting the intermediate result of HR patch estimation since we solve the problem via alternative iteration (see Section 3.1.2).

Given the basic HR patch $\boldsymbol{B}_h$, we normalize it first and then formulate the final HR patch $\boldsymbol{B}_r$ via deformation as follow:

$$\boldsymbol{B}_r = \alpha\phi(\boldsymbol{B}_h) + \beta \tag{2}$$

Here we consider two type of deformation: the local warp $\phi(\boldsymbol{B}_h)$ along x and y dimension and the intensity transformation by contrast α and mean value β. In this paper, firstly we focus on the local warp $\phi(\boldsymbol{B}_h)$ and ignore α and β by normalizing the patches. Then we estimate them separately after we get the local warped patch.

For the local warp function $\phi(\boldsymbol{B}_h)$, we model the deformation in the horizontal direction $\boldsymbol{u}$ and vertical direction $\boldsymbol{v}$ separately, *i.e*. the *deformation field* $\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{y})$. In later notation, we ignore the grid index $\boldsymbol{x}$ and $\boldsymbol{y}$ for simplicity. Now the explicit form of ϕ is as follow:

$$\phi(\boldsymbol{B}_h) = \boldsymbol{B}_h(\boldsymbol{x} + \boldsymbol{u}, \boldsymbol{y} + \boldsymbol{v}) \tag{3}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the image grid indices. Obviously, within a small patch, large deformation field is not reasonable, so the assumption of slow deformation field can be applied naturally. Under this assumption, ϕ has the following form via first order Taylor expansion:

$$\begin{aligned}\phi(\boldsymbol{B}_h) &\approx \boldsymbol{B}_h + \boldsymbol{B}_{hx} \circ \boldsymbol{u} + \boldsymbol{B}_{hy} \circ \boldsymbol{v} \\ &= \boldsymbol{B}_h + diag(\boldsymbol{B}_{hx})\boldsymbol{u} + diag(\boldsymbol{B}_{hy})\boldsymbol{v}\end{aligned} \tag{4}$$

where the operator $\circ$ denotes point-wise multiplication. $\boldsymbol{B}_{hx}$ and $\boldsymbol{B}_{hy}$ are the derivatives of $\boldsymbol{B}_h$ along the x and y dimensions respectively. The point-wise multiplication is equal to the matrix multiplication using diagonal matrix. Note that all the patches and $\boldsymbol{u}$, $\boldsymbol{v}$ are their vectorized version here and later.

Taking the degradation Eq.(1) into account, we form the energy function to be minimized as the total of error term E_{error} and prior term E_{prior}:

$$\begin{aligned}E(\boldsymbol{u}, \boldsymbol{v}) &= ||\boldsymbol{DHB}_r - \boldsymbol{P}_l||^2 + \psi(\boldsymbol{u}, \boldsymbol{v}) \\ &= \underbrace{||\boldsymbol{P}_d + \boldsymbol{P}_x\boldsymbol{u} + \boldsymbol{P}_y\boldsymbol{v}||^2}_{E_{\text{error}}} + \underbrace{\psi(\boldsymbol{u}, \boldsymbol{v})}_{E_{\text{prior}}}\end{aligned} \tag{5}$$

where

$$\begin{aligned}\boldsymbol{P}_d &= \boldsymbol{DHB}_h - \frac{\boldsymbol{P}_l - \beta}{\alpha} \\ \boldsymbol{P}_x &= \boldsymbol{DH}diag(\boldsymbol{B}_{hx}) \\ \boldsymbol{P}_y &= \boldsymbol{DH}diag(\boldsymbol{B}_{hy})\end{aligned} \tag{6}$$

In the error term E_{error}, $\boldsymbol{P}_d$ denotes the difference between the normalized LR patch $\boldsymbol{P}_l$ and the degraded version of the HR patch $\boldsymbol{B}_h$. Here we give initialized α and β using the standard deviation and mean value of $\boldsymbol{P}_l$. This is reasonable when $\boldsymbol{B}_h$ is normalized beforehand. $\boldsymbol{P}_x$ and $\boldsymbol{P}_y$ denotes the degradation and point-wise multiplication process imposed on $\boldsymbol{u}$ and $\boldsymbol{v}$. Minimizing error term E_{error} follows the basic constraint that the deformed and degraded HR patch should be consistent with the input LR patch.

The prior term $E_{\text{prior}} = \psi(\boldsymbol{u}, \boldsymbol{v})$ is the motion prior to regularize the deformation filed. Plenty of research work is about that[13, 15]. When choosing a prior for the patch deformation field, we consider the slowness and smoothness as well as the deformation flexibility.

The slowness prior is related to the intensity of deformation field $(\boldsymbol{u}, \boldsymbol{v})$, while the smoothness prior has the form of the first order or second order derivatives of $(\boldsymbol{u}, \boldsymbol{v})$. Then we get the following prior form:

$$\begin{aligned}\psi(\boldsymbol{u}, \boldsymbol{v}) = \mu(||\boldsymbol{u}||_2^2 + ||\boldsymbol{v}||_2^2) + \lambda(||\nabla\boldsymbol{u}||_2^2 + ||\nabla\boldsymbol{v}||_2^2) \\ + \eta(||\nabla^2\boldsymbol{u}||^2 + ||\nabla^2\boldsymbol{v}||^2)\end{aligned} \tag{7}$$

where ∇ and ∇^2 denote the gradient and Laplace operator respectively. μ, λ and η is the regularization constant to control the contribution of the prior components.

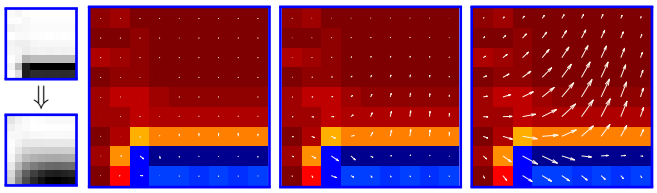


Figure 3. A deformation field example via different single prior. From left to right: the basic HR patch and the input low patch, the deformation field on $\mu = 0.1$, $\lambda = 0.1$ and $\eta = 0.1$ respectively

We give an example of the deformation field regularized by three different priors. Figure 3 indicates that using slowness prior individually leads to a rather low intensity of the deformable filed. With regard to the first order derivative prior, if the ideal warping is not shift-like, this prior also suppresses the deformation field intensity to reduce the possible change within the neighborhood. By contrast, the field using second order derivative prior has similar trend with first order, but it is more flexible and natural. So in this paper, we choose $\mu = 0$, $\lambda = 0$ and $\eta = 0.1$.

By applying the above prior, the energy function Eq. (5) resembles the objective function in optical flow. The difference is that we estimate the patch deformation in different scales connected by the degradation $\boldsymbol{D}$ and $\boldsymbol{H}$ rather than the two adjacent frames. With the help of deformation, we can estimate the HR details more precisely.

3.1.2 Optimizing for Energy Function

In this subsection, we solve the minimization problem of Eq.(5). After the normalization of the basic and input patches, there are two variables $\boldsymbol{B}_d = \phi(\boldsymbol{B}_h)$ and $(\boldsymbol{u}, \boldsymbol{v})$ to estimate. In our algorithm, they are updated alternatively until convergence.

First we calculate the deformation field $(\boldsymbol{u}, \boldsymbol{v})$. Given the basic HR patch $\boldsymbol{B}_h$, the minimization of Eq.(5) is a quadratic problem under the L_2 norm regularization. Here $\boldsymbol{B}_h$ denotes the deformed patch $\boldsymbol{B}_d^{k-1}$ in the k-th iteration or the HR patch from the dictionary in the 1-st iteration, For simplicity, we make the following notation:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \quad \boldsymbol{G} = \begin{bmatrix} \boldsymbol{P}_x & \boldsymbol{P}_y \end{bmatrix}$$

$$\boldsymbol{\Gamma} = \mu + \lambda \begin{bmatrix} -\nabla^2 & 0 \\ 0 & -\nabla^2 \end{bmatrix} + \eta \begin{bmatrix} (\nabla^2)^2 & 0 \\ 0 & (\nabla^2)^2 \end{bmatrix}$$

Then Eq.(5) has the form of:

$$E(\boldsymbol{M}) = ||\boldsymbol{P}_d + \boldsymbol{G}\boldsymbol{M}||^2 + \boldsymbol{M}^\top \boldsymbol{\Gamma} \boldsymbol{M} \tag{8}$$

Let the derivative of E be zero, we can easily get the optimized deformation field:

$$\boldsymbol{M} = -(\boldsymbol{G}^\top \boldsymbol{G} + \boldsymbol{\Gamma})^{-1} \boldsymbol{G}^\top \boldsymbol{P}_d \tag{9}$$

For the similar problem, Horn-Schunck method[13] gives an iterative solution via Euler-Lagrange Equation. It is necessary for the optical flow estimation on overall image pixels because the closed-form solution involves an extremely large matrix inversion which is impossible for computation. However, our deformation occur within small local patches *e.g.* 7×7 patches. Then $\boldsymbol{G}^\top \boldsymbol{G} + \boldsymbol{\Gamma}$ is just a matrix of the size 98×98. So it is feasible to get the direct solution by Eq. (9).

After obtaining the motion filed $(\boldsymbol{u}, \boldsymbol{v})$, the deformed patches $\boldsymbol{B}_d$ can be estimated according to Eq. (4).

3.1.3 Estimation of α and β

The algorithm described in Section 3.1.2 is imposed on the normalized version of the patches. In this section, we give the estimation of α and β in Eq. (2). Here we suppose that α and β are both scalars, to prevent the model from being more complicated and ill-posed. By minimizing the difference between the degraded version of HR patch and input LR patch, we have:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} ||\boldsymbol{P}_l - \alpha \boldsymbol{D}\boldsymbol{H}\boldsymbol{B}_d - \beta||^2 \tag{10}$$

This can be minimized by the pseudo inversion method to get the least square estimation:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (\boldsymbol{A}^\top \boldsymbol{A}) \boldsymbol{A}^\top \boldsymbol{P}_l \tag{11}$$

where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{D}\boldsymbol{H}\boldsymbol{B}_d & \mathbf{1} \end{bmatrix}$, $\mathbf{1}$ denotes the all-1 column vector with the same dimension of the degraded version of $\boldsymbol{B}_d$. Then, via Eq. (2), we can get the final single HR patch.

3.2. Patch Matching Strategy

In the previous section, we present our basic idea that how we deform a basic HR patch to fit the LR input. In this section, we are ready to elaborate on how to select the best basic patch from the dictionary for a specific LR input.

For an arbitrary patch in the HR dictionary, we measure its *deformation similarity*, *i.e.* the ability to fit the LR input, instead of measuring the similarity of raw intensity or gradient features. Intuitively, If a basic HR patch in the dictionary is *easy* to deform to the input patch, two principles ought to be followed: 1) The deformed patch should be consistent with the LR input after degradation. 2) The deformation field $(\boldsymbol{u}, \boldsymbol{v})$ should be small and simple to conform the assumption of Taylor expansion in Eq. (4). These two principles are also followed by energy function Eq. (5). So naturally, we use the minimization of the energy function as the deformation similarity for HR basic patch matching.

In Section 3.1.2, we give the closed-form solution of Eq. (5) when calculating the deformation field. The deformation similarity is defined as the minimized energy in the

first iteration. By substitute the solution Eq. (9), we have:

$$Sim(\boldsymbol{B}_h, \boldsymbol{P}_l) = \boldsymbol{P}_d^\top \boldsymbol{P}_d - \boldsymbol{P}_d^\top \boldsymbol{G}(\boldsymbol{G}^\top \boldsymbol{G} + \boldsymbol{\Gamma})^{-1} \boldsymbol{G}^\top \boldsymbol{P}_d \tag{12}$$

For each input LR patch, we traverse all the HR patches to find the best patch for the HR estimation:

$$\boldsymbol{B}_h = \arg\min_{\boldsymbol{B}_h} Sim(\boldsymbol{B}_h, \boldsymbol{P}_l) \tag{13}$$

Now we get the explicit form for the deformable field and the deformation similarity. Note that the degradation factors $\boldsymbol{D}$ and $\boldsymbol{H}$ are still unknown. They exist in the form of $\boldsymbol{H}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{H}$, and $\boldsymbol{H}^\top \boldsymbol{D}^\top \boldsymbol{P}_l$. Normally, $\boldsymbol{D}$ and $\boldsymbol{H}$ are large cyclic matrix, and $\boldsymbol{H}$ is related to what blurring kernel we use, which is very complicated. Here we use bicubic downsampling for $\boldsymbol{D}\boldsymbol{H}$ and bicubic upsampling followed by back-projection [2] for $\boldsymbol{H}^\top \boldsymbol{D}^\top$. Therefore in our algorithm, the LR patches is the enhanced bicubic upscaled version, of the same size as HR patches. The LR dictionary is also prepared by $\boldsymbol{H}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{H}$ process on HR patches.

3.3. Combination of Deformed Patches

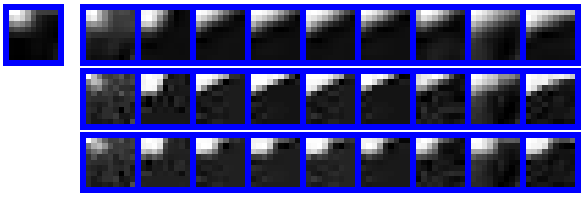


Figure 4. Several candidates selected from the dictionary by deformation similarity. The input LR patch is on the left. From top to bottom: the LR version, HR version and the deformed HR version.

The deformation similarity presented by the previous section selects the best HR patch flexibly. Nevertheless, it also allows improper patches to win. Although we are deforming the HR patches, we see only the degraded version due to the degradation factors $\boldsymbol{D}$ and $\boldsymbol{H}$. Figure 4 give an example of a winner during one matching precess. We can see that the LR version of the winner matches the input patch well, but other suboptimal patches show that its HR version is not likely to be the best match. Another problem is that both the input patch and the raw patches from the dictionary contain noise more or less. So the reliable HR estimation can not be obtained by the single patch. Instead, to make the estimation more precisely, we perform deformation on each of the M best HR patches $\{\boldsymbol{B}_{di}\}_i^M$ and combine the results by a weighted average:

$$\boldsymbol{B}_d^k = \sum_{i=1}^{M} \omega_i^k \boldsymbol{B}_{di}^k \tag{14}$$

where for $N \times N$ HR patches, $k = \{1, ..., N \times N\}$ indexes each pixels within the patch. In the other word, we assign different weight configuration for each pixel. The weights have the form $\omega_i^k = \frac{1}{Z} \exp(-(\boldsymbol{B}_{di}^k - \mu_k)^2 / 2\sigma_k^2)$ with Z the normalization factor. μ_k and σ_k^2 are the mean value and variance of all the M pixels in k-th position of each patch.

4. Experimental Results

In this section, we evaluate our algorithm via the reconstruction precision and visual quality. In both cases, our algorithm can achieve better results than the competing methods in literatures.

4.1. Dictionary and Experiments Setting

We start from the random selection of the high-resolution images to form the HR dictionary. It is common to use the natural image datasets and randomly select enormous patch pairs as in [9, 12, 20, 22, 23]. However, not all the patches in HR images have fine details due to camera focusing. An image may contain clear focused foreground object but blurring background. Furthermore, dense texture details are not necessarily captured by the HR/LR patch pairs because they lose the details more easily during degradation and then their HR details tend to form false textures, artifacts and blurring. Therefore the useful HR patches from dataset are edges, corners and the structures that are still remarkable in LR images.

Based on that, we combine the natural image dataset with the logo dataset, in order to make the dictionary cover both sharp edge patterns as well as the natural textures. The combined dataset consists of 28 logo images and 34 natural images. Some of the examples are shown in Figure 2. Finally we randomly select a number of HR/LR patch pairs from the dataset. The raw patches extracted from the dataset are pruned by eliminating the smooth patches with the LR variance less than 10. LR part of the dictionary is used for deformation field and deformation similarity calculation (See Section 3.2).

In the experiments, the patch size is 7×7 and the regularization constant is $\eta = 0.1$. In the patch matching step, we choose $M = 9$ deformed patches for the weighted combination. The input image is scaled to the HR dimensions by bicubic interpolation followed by back-projection[2]. The experiments are conducted on $3\times$ and $4\times$ super-resolution. For $4\times$ case, we do $2\times$ upscaling twice. For color images, super-resolution is done on Y channel in YCbCr color space, and the other two channels are upscaled by bicubic interpolation.

We evaluate the deformation effectiveness in term of PSNR, High PSNR stands for good performance. If the patches are set overlapped, the overlapped areas are averaged for final result. However the averaging process leads to blurring inevitably. So we incorporate the non-local method[1] and back projection[2] as post processing step, as other work[12, 23] does.

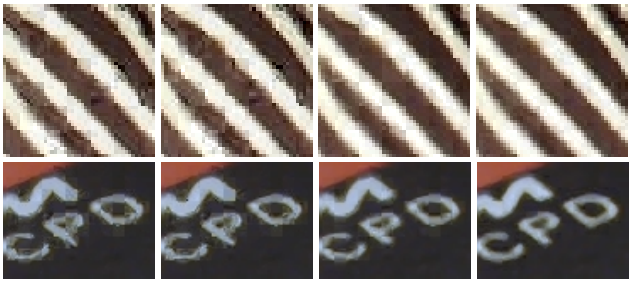


Figure 5. An intuitive deformation effect after applying deformation model and weighted combination (3×, dictionary size 30000 and overlap 0). From left to right: undeformed patch(UP), deformed patch(DP), weighted combined undeformed patch(UP+W) and deformed patches plus weighted combination (DP+W).

4.2. Evaluations on Deformation

To evaluate the performance of deformable patches, we first conduct the experiments on the test images, with the pixels non-overlapped, in order to make comparison on the single patch representation ability of deformable patches. We take 2 dictionary learning methods (Sparse Coding Dictionary Learning, SCDL [23] and Beta Process Joint Dictionary Learning, BPJDL [12]) as competitors. The dictionary size is chosen as 1024 in each case. In the end of the section, we compare the final super-resolved result on the test images with both self-similarity based methods (Glansner *et al*. [11], Freedman[8]) and dictionary learning methods (SCDL[23], BPJDL[12]).

The first experiment demonstrates the performance on whether the deformation or weighted combination is used. We compared the results when using single undeformed patches (UP), deformed patches (DP), weighed combined undeformed patches (UP+W) and deformed patches plus weighted combination (DP+W). As shown in Figure 5, it is remarkable that the deformed patches along the edges are more consistent with the neighborhood. Via the multi-patch weighted combination, the texture is much more natural, less of jaggy and noise. Table 1 shows that performance improves a lot by using proposed method, indicating that by the help of the deformation and weighted combination, the reconstruction accuracy improves a lot in terms of PSNR.

Another experiment is conducted on different dictionary sizes. Two dictionary based methods (SCDL [23] and BPJDL [12]) are evaluated as competitors. Figure 6 shows the comparison on lena and zebra image. From the figure, our result is superior to the other two competitors in most cases. It is worthy to point out that our method can achieve good performance even using smaller dictionary. Note that the performance of our method is more stable as the size of the dictionary decreases. When using dictionary smaller than 10000, our method performs similarly to the dictionary learning methods that use the dictionary of size more than 50000. The comparison validates the ability of proposed method that it makes the finite dictionary more expressive.

Table 1. PSNR(dB) after applying deformation model and weighted combination (3×, dictionary size 30000 and overlap 0). UP: undeformed patch, DP: deformed patch, W: weighted combination.

Image	UP	DP	UP+W	DP+W
lena	29.3259	29.6104	30.4901	**30.8557**
zebra	22.5464	23.0735	24.4748	**25.0028**
cameraman	24.539	24.7518	25.4206	**25.6633**
oldman	27.914	28.1613	29.2855	**29.6448**
child	28.1921	28.4638	29.5217	**29.8793**

Table 2 compares the final results on the five images for testing. From the table, self-similarity based methods[8, 11] achieve lower PNSR than the other method, because they focus on the edge enhancements more than reconstruction. Overall BPJDL[12] performs better than SCDL[23] via the introduction of the mapping matrix between the low/high sparse coding coefficients. Finally, the proposed method shows that the weighted combined deformable patches achieve better performance than the state of art methods.

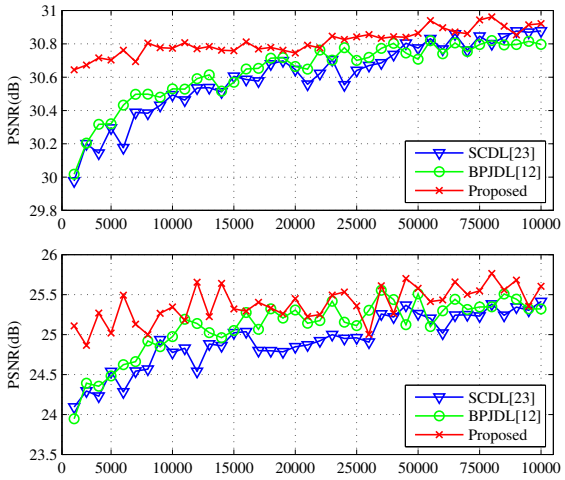


Figure 6. The lena and zebra image reconstruction PSNR(dB) when using the dictionaries of different size (3×, overlap 0). We choose dictionary size ranging from 2000 to 100000 at the interval of 1000 under 25000 and at the interval of 5000 above 25000.

4.3. Evaluations on Visual Quality

In this section, we compare the proposed method with the recent representative work on single image super-resolution[8, 11, 12, 20, 23] in terms of visual quality. The work of Glasner *et al*. [11] and Freedman [8] are methods based on self-similar examples, while Yang *et al*. [23] and He *et al*. [12] use sparse coding for HR details estimation within the same framework that we use. We also compare Yang *et al*.'s another work [20] that exploits in-place examples for super-resolution.

Figure 7 demonstrates the super-resolution results by 4×

Table 2. PSNR(dB) of the final super-resolved test images (3×, dictionary size 30000 and overlap 6)

Image	Bicubic	Glasner [11]	Freedman [8]	SCDL [23]	BPJDL [12]	Proposed
lena	30.0986	30.3197	30.6928	31.6493	31.6755	**31.7536**
zebra	23.6214	25.7724	26.8935	27.2387	27.5010	**27.7907**
cameraman	25.1935	25.9155	25.5409	26.2110	26.2032	**26.2221**
oldman	29.4678	28.9615	30.2424	30.5824	30.6059	**30.6666**
child	29.3479	29.4468	29.7914	30.9166	30.9433	**30.9467**

on "chip" and "child" images. As shown in the figure, Freedman [8] successfully preserves the edges though a little blurring around it. However, many false edges occur within the digits, making the true edge ambiguous. The same effect occurs around the child's iris. The sparse coding method [12, 23] generates more natural edges, but it is hard to avoid the blurring and artifacts ,*e.g*. the pupil, since the dictionary is fixed and support the patch space finitely. The in-place example regression [20] can deal with the edge better, but still lead too much blurring effect. The method does not recover the true details in some areas, *e.g*. the right corner of digit 9. Sparse coding based methods[12, 23] can recover some key structure such as the child's pupil, but still they bring in severe blurring and noticeable artifacts around the edges. By comparison, our method preserves the edge better and can generate more natural textures. This can be seen from the edge of the digits as well as the sharp reconstructed structure around the child's iris. Moreover, the shape of the child's pupil is well recovered.

Figure 8 shows more results on the natural images. The motorbike image is from the PASCAL dataset, which is contaminated severely by the JPEG compression. Our method handles the case well. Note that the areas around the tire and the damper are more free of jaggy and noise. The flower image is from Berkeley Segmentation Dataset. Overall, [8, 11] can enhance the edges, but some structure such as the spots on the beetles are not well recovered (*e.g*. the shape distorts and false edges occur). Sparse coding based methods[12, 23] remain faithful to the shape, but again bring much blurring to the results. The lena image shows that the proposed method can recover the edges better than the others. The details are much sharper, with less noticeable artifacts. These experiments show that our method performs better than the state of art methods.

5. Conclusion and Future Work

In this paper, we proposed a single image super-resolution method using deformable patches. By considering each patch as a deformable field rather than a fixed vector, the patch dictionary is more expressive. We also apply minimized energy based deformation similarity and weighted combination to make the final HR patch estimation both flexible and reliable. For future work, we will study the deformable patch ability for various of texture *e.g*. logo, animal, flowers, people, cars *et al*. Moreover, It is a worthy investigation to develop our method to handle more complex cases in the real video sequences. In addition, we are going to extend our method to the dictionary learning method rather than simple patch selection. Another plan is to combine the video-frame related information, using the techniques similar to 3DSKR[19] and Non-local Mean[16][25].

Acknowledgement The authors would like to thank George Papandreou and Jun Zhu for their useful comments. This work was supported by Chinese Scholarship Council, grants NSF of China (61231016, 61301193, 61303123, 61301192), NPU-FFR-JCT20130109 and NIH: 5RO1EY022247-03, ONR N000014-10-1-0933.

References

[1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005.

[2] D. P. Capel. *Image Mosaicing and Super-resolution*. PhD thesis, University of Oxford, 2001.

[3] T.-M. Chan, J. Zhang, J. Pu, and H. Huang. Neighbor embedding based super-resolution algorithm through edge detection and feature selection. *PR Letters*, 30(5):494–502, 2009.

[4] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *CVPR*, pages 275–282, 2004.

[5] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Fast and robust multi-frame super-resolution. *IEEE TIP*, 13:1327–1344, 2003.

[6] R. Fattal. Image upsampling via imposed edge statistics. In *SIGGRAPH*, 2007.

[7] D. Fleet and Y. Weiss. *Optical Flow Estimation*. Springer, 2005.

[8] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 30(2):12:1–12:11, 2011.

[9] W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *Computer Graphics and Applications*, 22(2), 2002.

[10] X. Gao, K. Zhang, D. Tao, and X. Li. Image super-resolution with sparse neighbor embedding. *IEEE TIP*, 21(7):3194–3205, 2012.

[11] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, pages 349–356, 2009.

[12] L. He, H. Qi, and R. Zaretzki. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *CVPR*, pages 345–352, 2013.

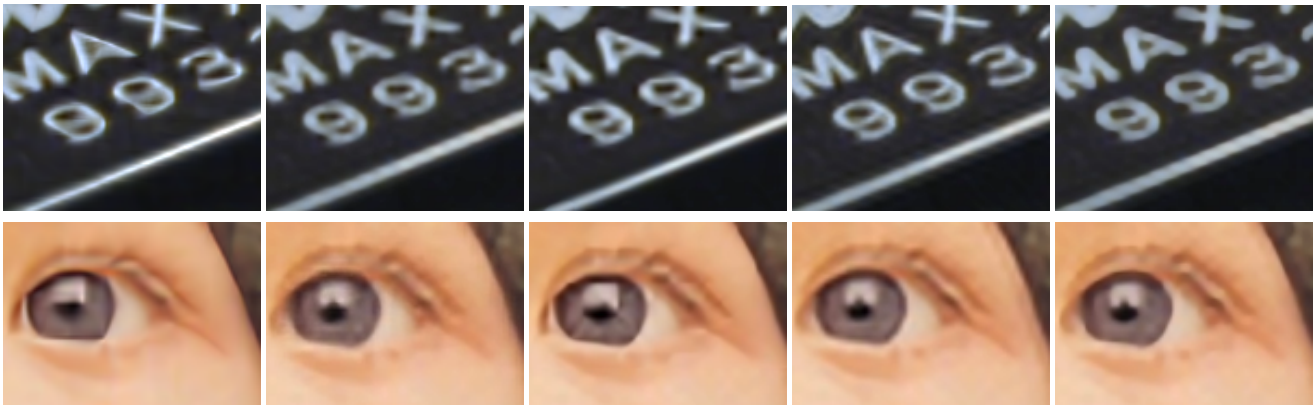
[13] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[14] C. Liu and D. Sun. A bayesian approach to adaptive video super resolution. In *CVPR*, pages 209–216, 2011.

[15] H. Lu, T. Lin, A. L. F. Lee, L. A. Vese, and A. L. Yuille. Functional form of motion priors in human motion perception. In *NIPS*, pages 1495–1503, 2010.

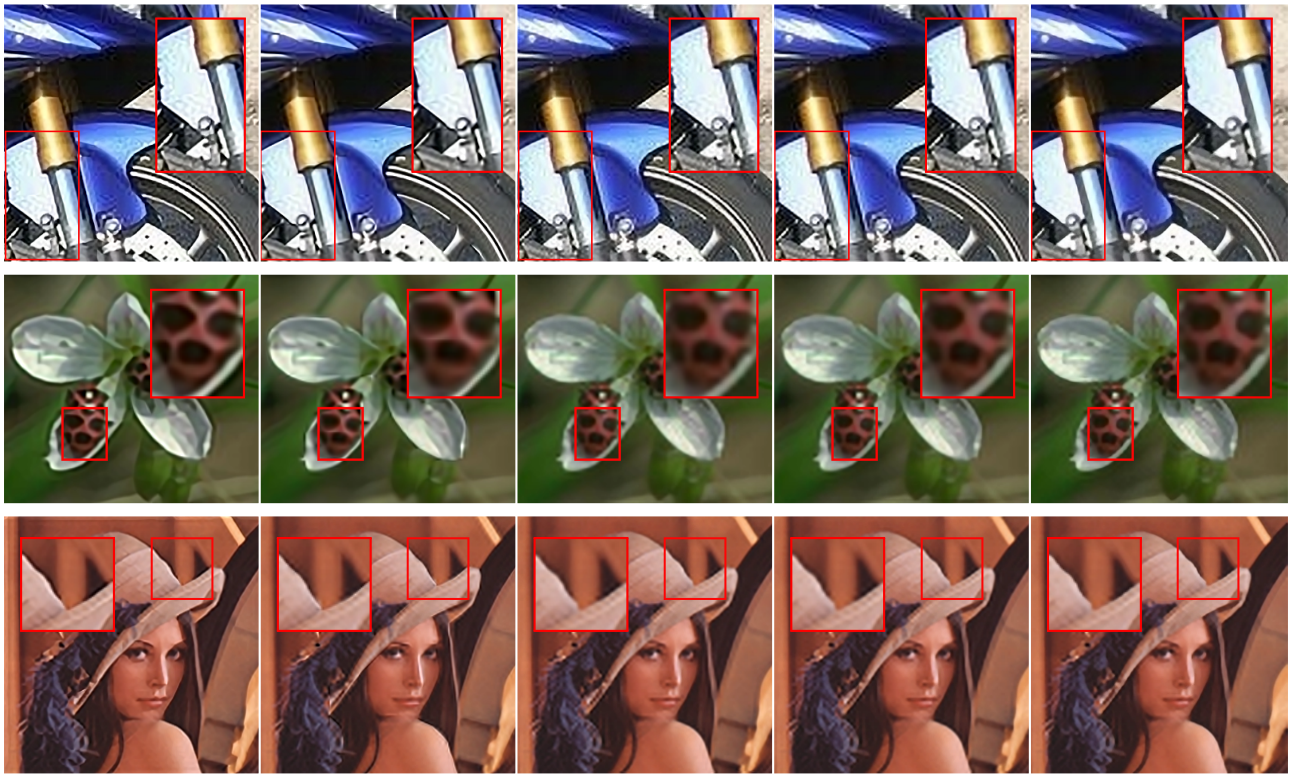
[16] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE TIP*, 18(1):36–51, 2009.

[17] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *CVPR*, 2008.



Freedman [8] SCDL [23] In-place example [20] BPJDL[12] Proposed

Figure 7. Super-resolution results by $4\times$ on "chip" and "child". Our method can preserve edge better and estimate textures more naturally. The effect is better viewed in zoomed PDF.



Glasner [11] Freedman [8] SCDL [23] BPJDL[12] Proposed

Figure 8. Super resolution results by $3\times$ (motorbike, flower and lena). Our method makes a robust estimation and generates appropriate edges, both sharp and natural. The effect is better viewed in zoomed PDF.

[18] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin. Super resolution using edge prior and single image detail synthesis. In *CVPR*, pages 2400–2407, 2010.

[19] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *IEEE TIP*, 18(9):1958–1975, 2009.

[20] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *CVPR*, pages 1059–1066, 2013.

[21] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *IEEE TIP*, 21(8):3467–3478, 2012.

[22] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, 2008.

[23] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010.

[24] X. Ye and A. Yuille. Learning a dictionary of deformable patches using gpus. In *ICCV Workshops*, pages 483–490, 2011.

[25] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang. Non-local kernel regression for image and video restoration. In *ECCV*, pages 566–579, 2010.