

Robust Web Image/Video Super-Resolution

Zhiwei Xiong, Xiaoyan Sun, *Member, IEEE*, and Feng Wu, *Senior Member, IEEE*

Abstract—This paper proposes a robust single-image super-resolution method for enlarging low quality web image/video degraded by downsampling and compression. To simultaneously improve the resolution and perceptual quality of such web image/video, we bring forward a practical solution which combines adaptive regularization and learning-based super-resolution. The contribution of this work is twofold. First, we propose to analyze the image energy change characteristics during the iterative regularization process, i.e., the energy change ratio between primitive (e.g., edges, ridges and corners) and nonprimitive fields. Based on the revealed convergence property of the energy change ratio, appropriate regularization strength can then be determined to well balance compression artifacts removal and primitive components preservation. Second, we verify that this adaptive regularization can steadily and greatly improve the pair matching accuracy in learning-based super-resolution. Consequently, their combination effectively eliminates the quantization noise and meanwhile faithfully compensates the missing high-frequency details, yielding robust super-resolution performance in the compression scenario. Experimental results demonstrate that our solution produces visually pleasing enlargements for various web images/videos.

Index Terms—Adaptive regularization, compression artifacts removal, energy change ratio, learning-based super-resolution (SR), primitive/nonprimitive field, web image/video.

I. INTRODUCTION

WITH the Internet flourishing and the rapid progress in hand-held photographic devices, image and video are becoming more and more popular on the web, due to their rich content and easy perception. Consequently, image search engines and online video websites have experienced an explosion of visits during the past few years. However, limited by the network bandwidth and server storage, most web image/video exists in a low quality version degraded from the source. The most common degradations are downsampling and compression. Downsampling exploits the correlation in the spatial domain while compression further exploits the correlation in the frequency and temporal (for video) domains. Quality degradation greatly lowers the required bandwidth and storage, making the access to web image/video practical and convenient. But these benefits are obtained at the expense of impairing the perceptual experience of users, as degradation inevitably leads to

information loss, which behaves as various artifacts in the resulting image/video, e.g., blurring, blocking and ringing.

There is a large demand for improving the perceptual quality of web image/video, among which the resolution enhancement, also known as super-resolution (SR), is an especially important issue and attracts a lot of attention. SR refers to the techniques achieving high-resolution (HR) enlargements of pixel-based low-resolution (LR) image/video. Basically, there are two kinds of SR, according to the amount of LR images utilized: multi-image SR, which requires several LR images of the same scene to be aligned in subpixel accuracy, and single-image SR, which generates a HR image from a unique source.

SR has many applications in the real world. Take image search engines for example: once a query is entered, a large number of images need to be returned simultaneously, and the results are first displayed in their LR forms (often called “thumbnails”). Users then need to click on the thumbnail to get its original HR version. Nevertheless, sometimes it is frustrating that the source image is removed or the server is temporarily unavailable. Single-image SR, at this moment, can save users from the bother of linking to every source if only an enlarged preview is desired.

Previous work on single-image SR can be roughly divided into four categories: interpolation-based [1]–[4], reconstruction-based [5], [6], classification-based [7] and learning-based [8]–[17]. Despite great diversity in implementation, these methods have a common premise that the LR image is only degraded by downsampling. This is not always true in the web environment, where compression is widely adopted. For image search engines, compression helps reduce the thumbnail size by up to 50% without obvious perceptual quality loss when presented in the LR form. But now if SR (any of the above) is directly performed, compression artifacts will be magnified out and the perceptual quality of resulting HR images will be poor.

On the other hand, multi-image SR has been used to enlarge video for a long time [5], [18]–[22], and corresponding techniques for compressed video SR have also been reported in literature [23]–[25]. These methods generally assume a prior distribution of the quantization noise and then integrate this knowledge into a Bayesian SR framework, or use the quantization bounds to determine convex sets which constrain the SR problem. In practical applications, however, the compression artifacts caused by the quantization noise are largely dependent on the video content and difficult to be modeled with an explicit distribution. Moreover, since the performance of Bayesian SR heavily depends on the accuracy of frame registration and motion estimation, these methods are not capable of reconstructing high frequency details of dynamic videos that contain fast and complex object motions.

In this paper, we present a practical solution which combines adaptive regularization with learning-based SR to simulta-

Manuscript received June 07, 2009; revised February 17, 2010; First published March 15, 2010; current version published July 16, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kenneth K. M. Lam.

Z. Xiong is with University of Science and Technology of China, Hefei, 230027, China. This work was done while he was with Microsoft Research Asia, Beijing, 100081, China (e-mail: xzw@mail.ustc.edu.cn).

X. Sun and F. Wu are with Microsoft Research Asia, Beijing, 100081, China (e-mail: xysun@microsoft.com; fengwu@microsoft.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2010.2045707

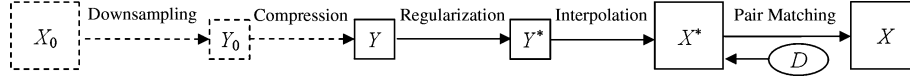


Fig. 1. Flowchart of our compressed image super-resolution scheme.

neously improve the resolution and perceptual quality of compressed image/video. A straightforward implementation of this idea has been reported in our previous work [26], where the regularization strength is determined by the JPEG compression quality parameter (QP), followed by learning-based pair matching to further enhance the high-frequency details in the interpolated image. This simple yet effective combination gives perceptually high quality SR results for compressed thumbnail images. To further improve the robustness of such an approach, we propose a more solid criterion for the adaptive regularization control in this work, based on the convergence property of the image energy change ratio between primitive and nonprimitive fields during iterative regularization. By appropriately locating the turning point where regularization loses its efficacy in distinguishing primitive components from compression artifacts, the pair matching accuracy in learning-based SR can be steadily and greatly improved. In this way, the quantization noise is effectively eliminated while the missing high-frequency details are faithfully compensated. Moreover, the proposed single-image SR method can be directly applied into compressed video SR, by introducing certain interframe interactions on the regularization strength and simple spatial-temporal consistency optimization, as reported in our latest work [37]. Different from conventional methods, our solution does not require any specific assumption on the quantization noise or the object motion, which greatly extends its scope of application in the web environment.

The rest of this paper is organized as follows. Section II formulates the single-image SR problem in the compression scenario and briefly introduces the regularization and learning-based SR techniques used in our scheme. The adaptive regularization control is elaborated on in Section III. Section IV extends the proposed method into compressed video SR. Experimental results are presented in Section V, and Section VI concludes the paper.

II. COMPRESSED IMAGE SUPER-RESOLUTION

A. Problem Formulation

An overview of our single-image SR scheme in the compression scenario is shown in Fig. 1. Suppose X_0 is an original HR image, it is first downsampled with a low-pass filter g (mostly isotropic Gaussian) to form an LR measurement Y_0

$$Y_0 = (g * X_0) \downarrow^\alpha \quad (1)$$

where $\downarrow^\alpha$ is a decimation operator with scaling factor α . Y_0 is then compressed, resulting in a degraded LR measurement Y

$$Y = Y_0 + E_Q \quad (2)$$

where E_Q represents the quantization error introduced by compression in the spatial domain.

Y is the actual input of our SR system. This system consists of three modules: PDE regularization, bicubic interpolation and learning-based pair matching. Regularization is first performed on Y to get an artifacts-relieved LR image Y^*

$$Y^* = f^N(Y) \quad (3)$$

where $f(\cdot)$ denotes the PDE regularization functional and the superscript N represents the total iteration number of regularization, which determines the regularization strength. Y^* is then upsampled with scaling factor β to get an intermediate HR result X^*

$$X^* = (h * Y^*) \uparrow^\beta \quad (4)$$

where h stands for the bicubic interpolation filter. The final HR image X is obtained after learning-based pair matching from X^* and a prepared database D . The maximum a posteriori probability (MAP) estimate of X can be expressed as

$$\hat{X} = \arg \max_X p(X|X^*, D) \quad (5)$$

B. Learning-Based Pair Matching

Single-image SR aims to obtain a HR reconstruction X from a LR measurement Y . For learning-based approaches, a set of examples organized in a database D are utilized in the online reconstruction process. These examples usually exist as co-occurring patch pairs $\{x_k, y_k\}_{k=1}^K$ extracted from training images at two different resolution levels. The basic idea of using examples in SR is that natural images are special signals occupying only a vanishingly small fraction of the high dimensional image space. Therefore, high-frequency details that do not exist in Y can be “stolen” from D through *pair matching*, i.e., given an LR patch from the input measurement, seek in the database for similar LR examples, and their corresponding HR counterparts can then be used for the reconstruction as they provide high-frequency details that fit the input measurement.

In our scheme, the primitive-based hallucination method proposed in [12] is adopted for pair matching, for which image primitives (edges, ridges and corners) are represented by examples and pair matching is only applied to the primitive components in images. The superiority of this method is twofold. First, human observations are especially sensitive to image primitives when going from LR to HR. Second, the primitive manifold is of intrinsic lower dimensionality compared with raw image patch manifolds as used in [10], [13], [14] and can be more effectively represented by examples. (Please refer to [12] for detailed procedures of primitive example generation.)

Generally speaking, learning-based pair matching exploits the correspondence between image signals at two different resolution levels, whereas another kind of degradation—compression—is seldom considered in previous works. One may sug-

gest directly involving compression in preparing the examples. Unfortunately, this implementation will heavily lower the pair matching accuracy as the quantization noise, unlike the high frequency components lost in downsampling, is difficult to be effectively represented by examples. The underlying reason is that compression corrupts the primitive pattern (or other feature patterns) contained in examples, and thus the correspondence between them. As an alternative, we propose to keep the database away from compression while introducing regularization on the compressed LR measurement.

C. PDE Regularization

Among various available regularization techniques, anisotropic PDE's [27]–[30] are considered to be one of the best, due to their ability to smooth data while preserving visually salient features in images. A brief restatement of PDE regularization is given below. Suppose I is a 2D scalar image, the PDE regularization can be formulated as the juxtaposition of two *oriented 1D heat flows* along the gradient direction η and its orthogonal, named the isophote direction ξ (as ξ is everywhere tangent to the isophote lines in the image), with corresponding weights c_η and c_ξ

$$\frac{\partial I}{\partial t} = c_\xi \frac{\partial^2 I}{\partial \xi^2} + c_\eta \frac{\partial^2 I}{\partial \eta^2}, \quad \eta = \frac{\nabla I}{\|\nabla I\|}, \quad \xi = \eta^\perp \quad (6)$$

where $\|\nabla I\| = \sqrt{I_x^2 + I_y^2}$ denotes the image gradient magnitude. The choice of c_η and c_ξ is not determinate; only certain properties need to be satisfied. In this paper, we use the following weights as suggested in [30]

$$c_\eta = \frac{1}{1 + \|\nabla I\|^2}, \quad c_\xi = \frac{1}{\sqrt{1 + \|\nabla I\|^2}} \quad (7)$$

This is one possible choice inspired from the *hyper-surface formulation* of the scalar case [31]. The PDE in (6) can be equivalently written as

$$\frac{\partial I}{\partial t} = \text{trace}(\Theta \mathbf{H}) \quad (8)$$

where $\mathbf{H}$ is the *Hessian matrix* of I and Θ is an anisotropic 2×2 tensor defined as

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}, \quad \Theta = c_\xi \xi \xi^T + c_\eta \eta \eta^T \quad (9)$$

Θ gives the exact smoothing geometry performed by the PDE, which can be viewed as a thin ellipsoid with the major axis perpendicular to the gradient direction, as shown in Fig. 2.

In a numerical scheme, the input measurement is regularized iteratively. In the n th iteration, there is

$$f(I) : I_n = I_{n-1} + \lambda \Delta I_n, \quad I_0 = I, \quad n = 1, 2, \dots, N \quad (10)$$

where λ is a positive constant controlling the updating step, N is the total iteration number, and the image intensity change velocity ΔI_n can be calculated from (8) based on the spatial discretization of the gradients and the Hessians.

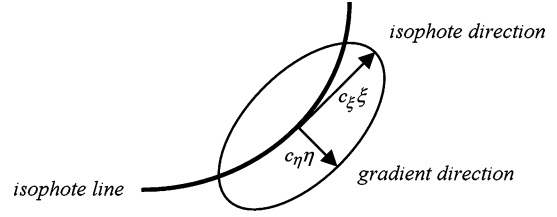


Fig. 2. PDE smoothing geometry. It can be viewed as a juxtaposition of two oriented 1D heat flows along the gradient direction η and the isophote direction ξ , with corresponding weights c_η and c_ξ .

As pointed out in [30], regularization PDE's generally do not converge toward a very interesting solution. Most of the time, the image obtained at $n \rightarrow \infty$ is constant, corresponding to an image without any variations. Therefore, the total iteration number of regularization N is often manually determined according to the image degradation level. In the scenario of compressed image SR, one question arising here is that how to adaptively control this regularization strength N to maximize the posterior in (5), given prefixed pair matching database D and PDE regularization functional f . In other words, can we find the turning point where compression artifacts are effectively eliminated while primitive components are still well preserved?

III. ADAPTIVE REGULARIZATION CONTROL

A. Energy Change During Regularization

To obtain appropriate regularization strength that well balances artifacts removal and primitive preservation, we propose to investigate the image energy change characteristics during the iterative regularization process. For this purpose, an image is first divided into primitive field and nonprimitive field. This partition can be determined by the orientation energy edge detection [32]. Suppose Γ is a bitmap storing detected edge pixel locations in I

$$\Gamma(u, v) = \begin{cases} 1, & (u, v) \text{ is an edge pixel} \\ 0, & \text{else.} \end{cases} \quad (11)$$

The primitive field (PF) is defined as

$$P(u, v) = \begin{cases} 1, & \Gamma(i, j) = 1, (i, j) \in N_\rho(u, v) \\ 0, & \text{else} \end{cases} \quad (12)$$

where $N_\rho(u, v)$ refers to a ρ th order neighborhood of (u, v)

$$N_\rho(u, v) = \{(i, j) : (u - i)^2 + (v - j)^2 \leq \rho\} \quad (13)$$

Correspondingly, the nonprimitive field (NPF) is defined as

$$Q(u, v) = \overline{P(u, v)} = 1 - P(u, v) \quad (14)$$

Fig. 3 illustrates the PF and NPF partition with $\rho = 1$.

After the n th iteration of regularization, the image energy change in PF and NPF can be calculated as

$$\begin{aligned} \Delta E_n^P &= \sum_u \sum_v |\Delta I_n(u, v)|^2 P(u, v) \\ \Delta E_n^Q &= \sum_u \sum_v |\Delta I_n(u, v)|^2 Q(u, v) \end{aligned} \quad (15)$$

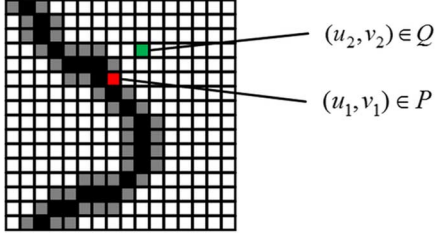


Fig. 3. Partition of primitive and nonprimitive fields with $\rho = 1$. Black edge pixels and their gray neighbors constitute the primitive field while the remaining white pixels constitute the nonprimitive field. Pixel (u_1, v_1) is in the primitive field and pixel (u_2, v_2) is in the nonprimitive field.

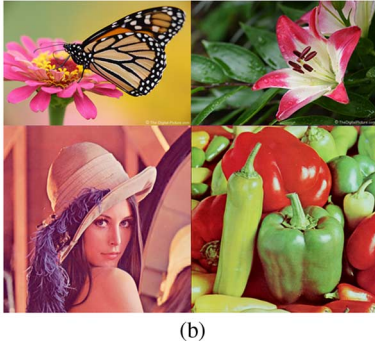
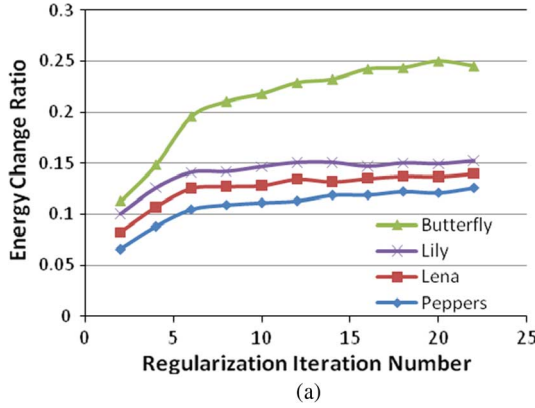


Fig. 4. (a) Image energy change ratio between PF and NPF during iterative regularization, and (b) test images.

where ΔI_n is the image intensity change in (10). We denote the energy change ratio between PF and NPF as

$$r_n = \frac{\Delta E_n^P}{\Delta E_n^Q}. \quad (16)$$

Fig. 4(a) gives several practical results of the $r \sim n$ curve on the test images shown in Fig. 4(b). The test images are first 1/3 downsampled from the original and then compressed by JPEG with QP set to 60. It can be seen all those $r \sim n$ curves converge after a few iterations. To obtain a general distribution pattern of the $r \sim n$ convergence speed, we use several video sequences (each containing 1500 frames) randomly downloaded from Youtube [35] for test, instead of gathering a large number of images. As shown in Fig. 5, most frames require less than 15 iterations, which validates the convergence property of the $r \sim n$ curve in the web environment.

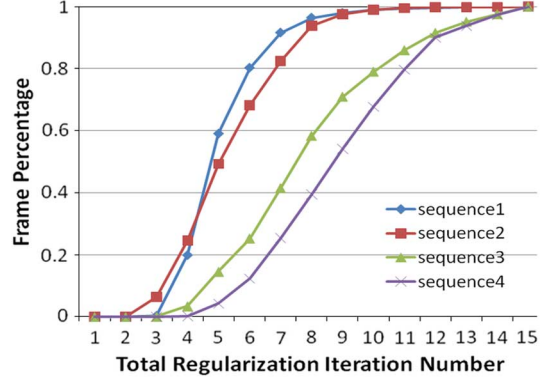


Fig. 5. Distribution of total regularization iteration number for web videos.

The convergence property of the $r \sim n$ curve can be figured out intuitively from the edge-preserving nature of PDE regularization. In the earlier stage, the energy change in NPF is more intensive than that in PF, so r increases with n . When regularization is performed to a certain stage, PDE loses its efficacy in distinguishing salient features from images, and r then remains at a stable level. Further, when ρ is small, there is a large probability that compression artifacts such as ringing and blocking appear in NPF. So PDE regularization removes compression artifacts first.

According to the energy change characteristics during PDE regularization, we can now determine appropriate regularization strength to maximize the posterior in (5), by locating the turning point where the $r \sim n$ curve tends to converge. At this time, artifacts removal and primitive preservation in a compressed image are best balanced. In practice, we stop the regularization at the N th iteration when

$$r_N - r_{N-1} < \mu R, \quad R = \max_n (r_n - r_{n-1}), \quad 0 < \mu < 1 \quad (17)$$

where μ is a constant and R represents the fastest increasing speed of r .

B. Pair Matching Accuracy

To demonstrate the necessity and effectiveness of adaptive regularization in compressed image SR with learning-based methods, we then investigate the pair matching accuracy under three different circumstances, i.e., without compression, with compression but no regularization and with both compression and adaptive regularization. We denote them as p_0 , p_1 and p_2 .

To compare the pair matching accuracy, we use a Receiver Operating Characteristic (ROC) curve to demonstrate the tradeoff between match error and hit rate. We define the match error as

$$e = \frac{\|x - x'\|_2^2}{\|x\|_2^2} \quad (18)$$

where x denotes the real missing HR primitive patch and x' is the HR example found through pair matching from the existent LR primitive patch y . For a given match error e , the hit rate h is the percentage of test data whose match error is less than e .

Fig. 6 presents three ROC curves based on the pair matching results of 50,000 primitive patches over 100,000 pairs of trained

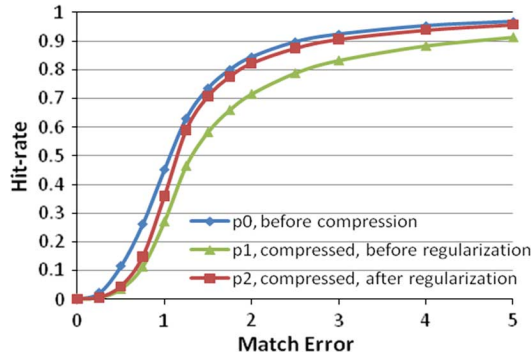


Fig. 6. ROC curves of pair matching accuracy. 50,000 primitive patches are tested over 100,000 trained examples.

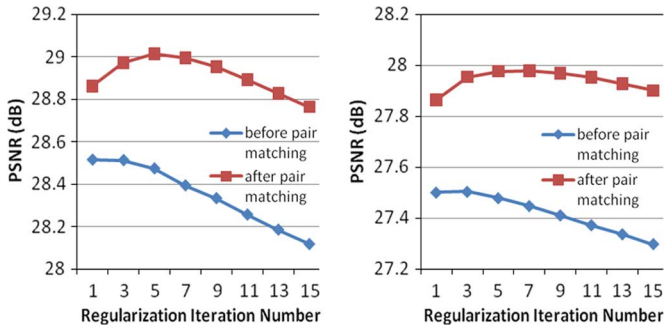


Fig. 7. PSNR curves against the iteration number of regularization, before and after pair matching. Test images are taken from Fig. 4(b). Left: *Lena*, and right: *Peppers*.

examples. The test data is sampled from images in Fig. 4(b), whereas the training images are shown in Fig. 11. As can be observed from p_0 and p_1 , when compression is involved, the pair matching accuracy degrades heavily. On the other hand, with adaptive regularization, p_2 is higher than p_1 at any match error and quickly approaches p_0 , which indicates that the proposed adaptive regularization steadily and greatly improves the pair matching accuracy in compressed image SR.

To further verify the adaptivity criterion in our scheme, in Fig. 7 we plot two PSNR curves against the iteration number of regularization, with and without pair matching. It can be observed that, though PDE regularization removes artifacts (and thus improves the perceptual quality), it gradually lowers the PSNR, as we don't enforce any fidelity constraints. However, after pair matching, the PSNR curve pattern changes. The peak value appears near the iteration number when the corresponding $r \sim n$ curve converges, if compared with Fig. 4(a).

C. Revisit of Training Set

In the above discussion, we assume a prefixed pair matching database D , mainly to verify the adaptivity criterion in the regularization step. Nevertheless, for learning-based approaches, the choice of training set is still an important issue. Therefore, we also analyze how the choice of training set influences the performance of our method.

Since the primitive pair matching mainly exploits the correspondence between LR and HR primitive components, the training images are generally required to have rich and diverse primitive patterns. However, we would like to point out that, as

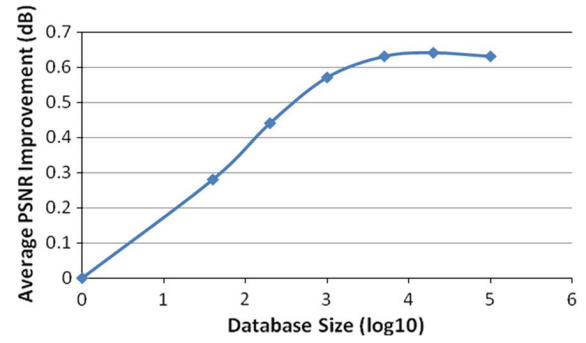


Fig. 8. PSNR improvement between directly interpolated images and those with our proposed SR under different database sizes. Test images are taken from Fig. 4(b).

the primitive pattern is a low-level vision feature, the SR performance only depends on the number of distinct primitive patterns in the trained database, instead of the structural similarity between training images and test images. Moreover, once the number of primitive patterns increases to a certain value, the SR performance will remain stable.

To observe the effect on the size of database, we conduct another experiment. During the experiment, we gradually increase the size of database from 0 (no pair matching) to 100,000 pairs of distinct primitive patterns (extracted from a training set of 16 Kodak images [36] shown in Fig. 11), and then measure the average PSNR improvement between directly interpolated compressed images and those with our proposed SR. The result is shown in Fig. 8, where the PSNR improvement tends to be saturated when the database exceeds a certain size. Note that the minimum database size required for a stable SR performance *without compression* can be larger than that indicated in Fig. 8, because compression, as well as regularization, reduces the number of distinct primitive patterns in the input images.

D. Discussion and Summarization

Compressed image SR is a practical problem in the web application of single-image SR, but has rarely been investigated before. In this subsection, we would like to supplement some intuition on why learning-based pair matching and PDE regularization are combined to solve this problem.

It is a natural idea to conceive a two-step strategy to address this problem, first relieving compression artifacts and then performing common SR. However, choosing an effective combination is nontrivial. Since there is actually no way to eliminate quantization noise without impairing high-frequency image details, it requires the subsequent SR to have a strong ability to recover the weakened information. As mentioned in the introduction, there are several categories of single-image SR methods, among which learning-based pair matching shows its superiority when the integration of subclass priors, i.e., trained examples, is more powerful than a generic smoothness prior as used in other approaches. Besides the impressive results obtained in domain-specific applications (e.g., face, text [8], [9]), primitive-based hallucination [12] further shows appealing performance for generic image SR, which has been validated in the following works [15]–[17]. Therefore, we choose the primitive pair matching as our SR method.

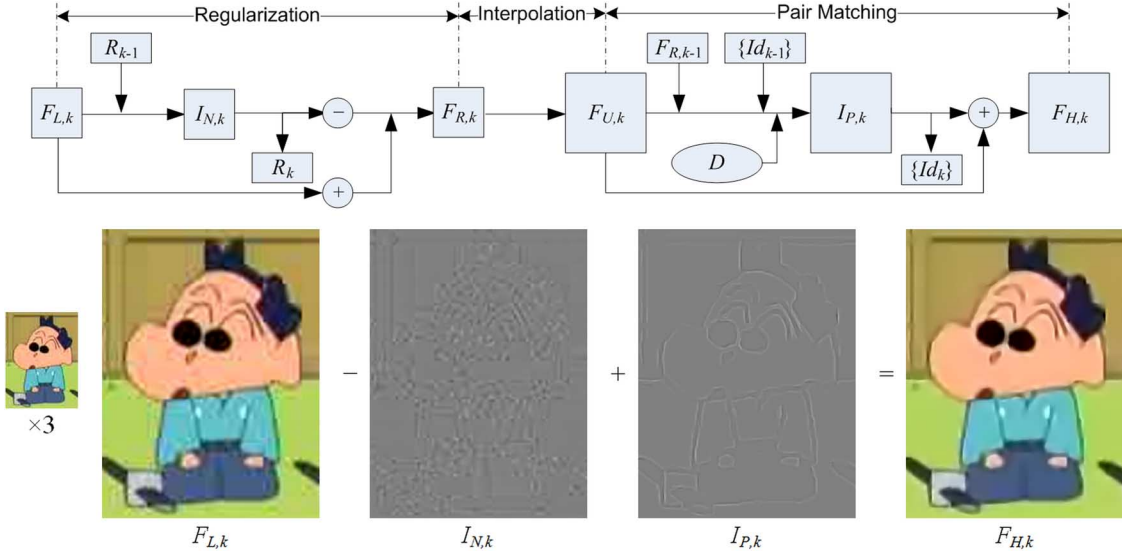


Fig. 9. Framework of our solution for compressed video super-resolution. The input frame $F_{L,k}$ and eliminated noise image $I_{N,k}$ are presented in their interpolated version in the example for better visualization. Note $I_{N,k}$ is only the luminance component of the noise image.

As learning-based SR is usually performed at patch level and relatively noise-sensitive, we then prefer a global algorithm for compression artifacts removal to avoid local inconsistency (i.e., adjacent regions should be stably smoothed). This is the first reason we use PDE regularization, due to its global smoothing property. On the other hand, a new challenge emerges from the combinative method. That is, as stated all along this section, how to adaptively control the regularization strength to best exploit the capability of these two somewhat contradictive techniques (one tends to smooth and the other tends to enhance). This is the second reason we choose PDE regularization, due to its progressive smoothing property.

To summarize this section, we give a more informative description of our algorithm for compressed image SR as follows:

Input: compressed low resolution image Y

Output: enhanced high resolution image X

Begin

1. Upsample Y to X_0^* through bicubic interpolation.
2. Find a PF/NPF partition (P, Q) of X_0^* through the orientation energy edge detection [32].
3. Perform iterative PDE regularization on Y :
 - a) After each iteration, upsample the regularized image Y_n^* to X_n^* , $n = 1, 2, \dots$ through bicubic interpolation;
 - b) Calculate the image energy change $(\Delta E_n^P, \Delta E_n^Q)$ between X_n^* and X_{n-1}^* based on the PF/NPF partition (P, Q) ;
 - c) Calculate the energy change ratio between PF and NPF as $r_n = \Delta E_n^P / \Delta E_n^Q$, and also record the maximum slope of r_n as $R = \max_n (r_n - r_{n-1})$, $n \geq 2$;
 - d) If $r_n - r_{n-1} < \mu R$, stop regularization, and keep the total iteration number N .

4. Extract LR primitive patches from X_N^* and find corresponding HR primitive patches from a prepared database D through pair matching speeded up by the approximate nearest neighbor (ANN) tree searching [33].
5. Add the HR primitive patches back to X_N^* to form the final HR image X , where the compatibility of neighboring HR primitive patches is enforced by averaging the pixel values in overlapped regions.

End

IV. COMPRESSED VIDEO SUPER-RESOLUTION

A. Framework

Since the above introduced single-image SR method does not require frame registration or motion estimation, it can be directly applied into the compressed video SR in a frame-by-frame style. By integrating certain interframe interactions on the regularization strength and simple spatio-temporal coherency constraints, our scheme is competent for the SR task of web videos with dynamic content and different degradation levels.

The framework of our solution is shown in Fig. 9. Similar to that of image SR, it consists of three steps. First, a k th frame $F_{L,k}$ from an LR video is divided into PF and NPF and iterative PDE regularization is performed on $F_{L,k}$, during which the energy change velocities in both PF and NPF are recorded. When the ratio of these two velocities converges (judged by a parameter R_k , which is also influenced by that of the previous frame R_{k-1}), regularization stops and the accumulated noise image $I_{N,k}$ is subtracted from $F_{L,k}$, resulting in an artifacts-relieved frame $F_{R,k}$. Then, $F_{R,k}$ is upsampled to the desired resolution through bicubic interpolation. Last, the primitive components in the interpolated frame $F_{U,k}$ are enhanced with learning-based pair matching. Meanwhile, the temporal consistency is enforced by referring to the previous interpolated frame $F_{U,k-1}$ and its pair matching indices $\{Id_{k-1}\}$. Adding the primitive enhancing

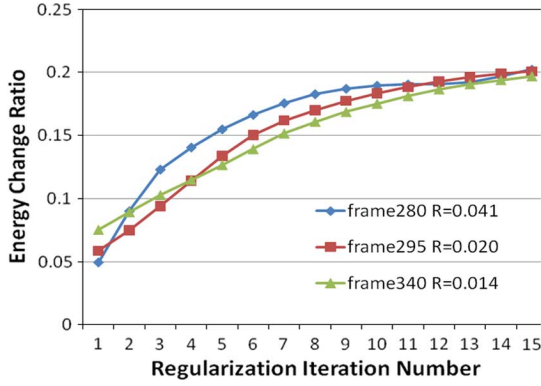


Fig. 10. Energy change characteristics in consecutive frames. Frames with heavier degradation exhibit slower $r \sim n$ convergence speed. Test frames are shown in Fig. 15.

image $I_{P,k}$ back to $F_{U,k}$, the final HR frame $F_{H,k}$ is generated. A practical example is given in Fig. 9 to visualize this framework.

B. Interframe Interaction

The regularization strength control elaborated in Section III can adapt to different degradation levels due to quantization within a single image/frame. However, in a video sequence with fast motion or scene switch, compression artifacts in consecutive frames could greatly vary due to inaccurate interframe prediction, even when the quantization levels are set to be the same. The adaptive regularization should also take these circumstances into consideration.

For two frames with similar content, the one with heavier degradation requires higher regularization strength, and this adaptivity is mainly reflected by the parameter R . One can easily find in Fig. 10 that frames with heavier degradation have smaller R (refer to Fig. 15 for the test frames), which means the convergence speed of the $r \sim n$ curve is inversely proportional to the degradation level in consecutive frames.

To further improve the adaptivity of regularization, we record R_{k-1} of the $(k-1)$ th frame and for the k th frame, R_k is calculated as

$$R_k = \frac{R_{k,0}^2}{R_{k-1}}, \quad k = 1, 2, \dots \quad (19)$$

where $R_{k,0}$ is the initial quantity measured from the current frame. If $R_{k,0} < R_{k-1}$, it suggests the degradation in the current frame is more severe than that in the previous frame. Then $R_{k,0}$ is further diminished to increase the regularization strength of the current frame (according to (17), reducing R will increase N), and vice versa. In this way, the regularization strength can also adapt to the variable degradation between consecutive frames caused by fast motion or scene switch, making the quality improvement on the whole video sequence more stable.

C. Spatio-Temporal Coherency Optimization

After the adaptive regularization, compression artifacts in each frame are effectively reduced while primitive components are still well preserved. Primitives in the interpolated frame are then enhanced with learning-based pair matching. The main



Fig. 11. Training images (1536 × 1024 pixels). 100,000 pairs of primitive patches are extracted from these images.

problem when applying this step to video is, how to make the enhanced primitives temporally consistent to avoid flicker, especially for sequences with slight motion. To solve this problem, we propose to optimize the spatio-temporal coherency with a simple yet effective constraint.

We first define two terms for the convenience of narration. Let y_{ik} denote an LR primitive patch extracted from the location i of the k th interpolated frame, and x_{ik} represent the enhancing HR primitive patch corresponding to y_{ik} . In the temporal domain, we record the pair matching indices in the $(k-1)$ th frame. Then, for the k th frame, each y_{ik} is compared with y_{ik-1} in the same position of the previous interpolated frame (in case y_{ik-1} exists). If y_{ik} is judged the same as y_{ik-1} , i.e., the sum of absolute difference (SAD) is smaller than a given threshold, the pair matching index of y_{ik-1} is directly assigned to y_{ik} ; or else a new pair matching for y_{ik} is conducted in the database.

In the spatial domain, since some pair matching results are derived from the previous frame and others are generated from the current frame, the compatibility of learned enhancing patches should be optimized. Specifically, for each y_{ik} that cannot use the index from the previous frame and a new pair matching is required, we take the first M pair matching results $\{x_{ik,1}, x_{ik,2}, \dots, x_{ik,M}\}$ as candidates for x_{ik} , and the optimum one is found by

$$\hat{n} = \arg \min_{1 \leq m \leq M} d(x_{ik,m}, x_{jk}) \quad (20)$$

where x_{jk} is the previous selected enhancing patch in the current frame in raster-scan order, and function d measures the SAD in the overlapped region of two patches. In summary, the enhancing patch x_{ik} corresponding to y_{ik} can be denoted as

$$x_{ik} = \begin{cases} x_{ik-1}, & \frac{\|y_{ik} - y_{ik-1}\|_2^2}{\|y_{ik-1}\|_2^2} < \varepsilon \\ x_{ik,\hat{n}}, & \text{otherwise} \end{cases} \quad (21)$$

where x_{ik-1} is the selected enhancing patch in the location i of the $(k-1)$ th frame, and ε is a small threshold. Finally, the primitive enhancing image is generated by assembling all enhancing patches, where pixel values in the overlapped regions are averaged.

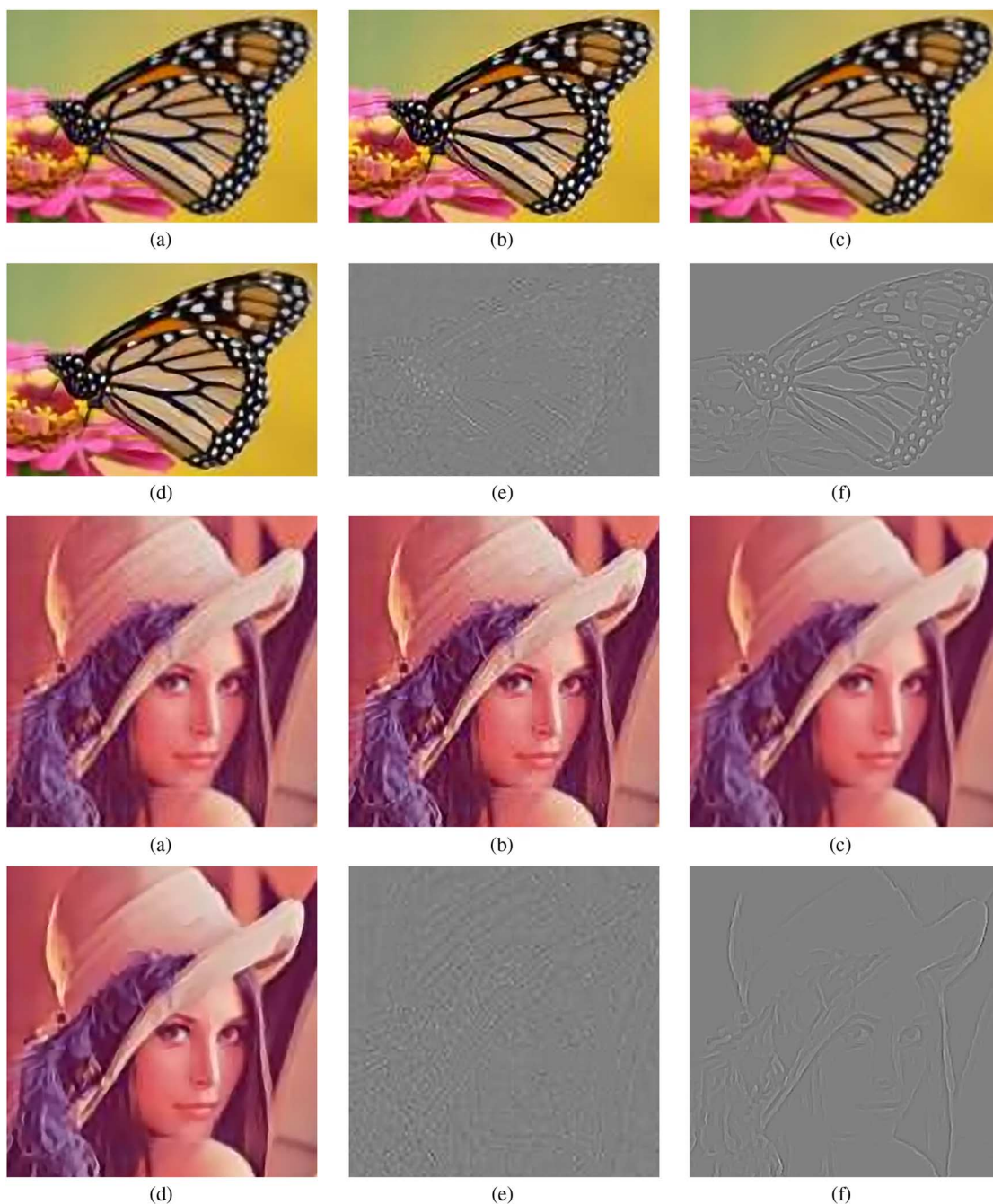


Fig. 12. Super-resolution results of offline images “Butterfly” and “Lena”. (a) Bicubic interpolation, (b) learning-based pair matching, (c) PDE regularization, (d) our approach, (e) luminance component of eliminated noise image after adaptive regularization, and (f) primitive enhancing image after pair matching.

V. EXPERIMENTAL RESULTS

A. Image Results

We test our SR scheme on both offline images degraded by designated downsampling and compression from the sources, as well as compressed thumbnail images on the web. For the offline test images, we use Gaussian filter for downsampling with decimation factor $\alpha = 3$, and JPEG for compression with QP = 60. For the test images on the web, the downsampling process is totally unknown while the compression format is still JPEG, but with unknown QP's. The PDE updating step $\lambda = 5.0$, the neighborhood order of primitive field $\rho = 1$, the regularization

strength control parameter $\mu = 0.2$, and the upsampling factor $\beta = 3$. A 16M record database consisting of 100,000 pairs of 9×9 sized primitive patches is used in pair matching. These examples are trained from 16 representative natural images shown in Fig. 11. For color images, regularization is performed on both the luminance and chrominance components, for which compression is applied. Learning-based pair matching, however, is only performed on the luminance component, as human observers are more sensitive to the luminance change in images when going through LR to HR.

Fig. 12 gives the SR results of two offline images obtained through several methods, including bicubic interpolation,

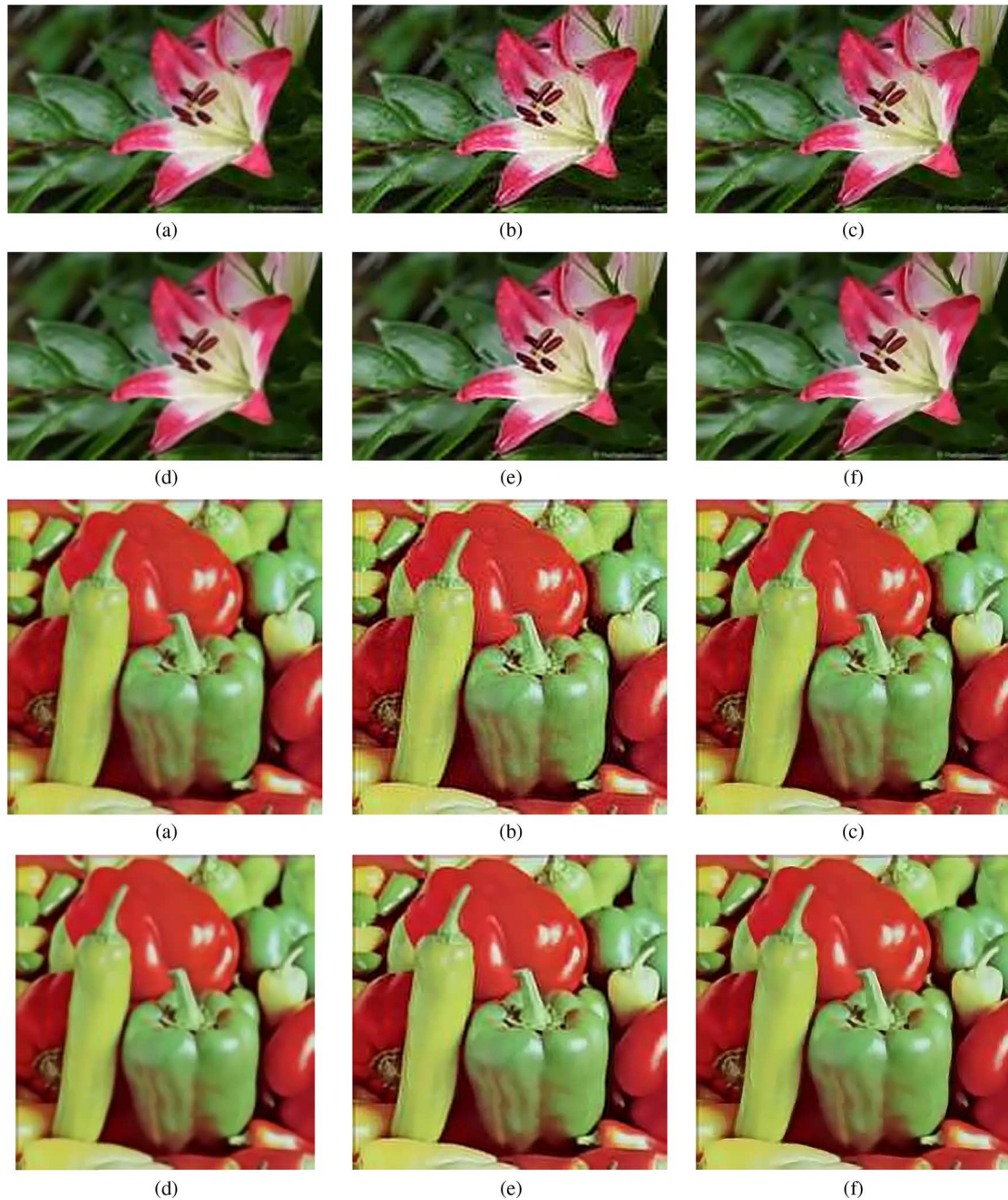


Fig. 13. Super-resolution results of offline images “Lily” and “Peppers”. (a) Directional interpolation, (b) backprojection, (c) learning-based pair matching, (d) directional interpolation after adaptive regularization, (e) backprojection after adaptive regularization, and (f) learning-based pair matching after adaptive regularization.

learning-based pair matching, PDE regularization and our proposed approach. Compared with bicubic interpolation, regularization effectively reduces the compression artifacts while pair matching well compensates the high frequency details, as demonstrated in the eliminated noise image and the primitive enhancing image. However, neither single regularization nor single pair matching generates satisfactory SR results. Taking the advantages of the two techniques, our combinative approach restores visually pleasing HR images from the compressed LR measurements. Note that this combination is nontrivial, but with adaptive regularization control as an essential coupling,

which guarantees the pair matching accuracy in the learning process.

To further verify that the close coupling between regularization and SR, especially learning-based pair matching, is truly required, we also combine our adaptive regularization with non-learning-based SR techniques, e.g., directional interpolation [4] and backprojection [5]. Experimental results are presented in Fig. 13 and Table I. It can be seen that, firstly, the perceptual quality of all SR results is improved after adaptive regularization. For backprojection and pair matching, the PSNR is also improved. (Single interpolation doesn’t see a PSNR gain as PDE

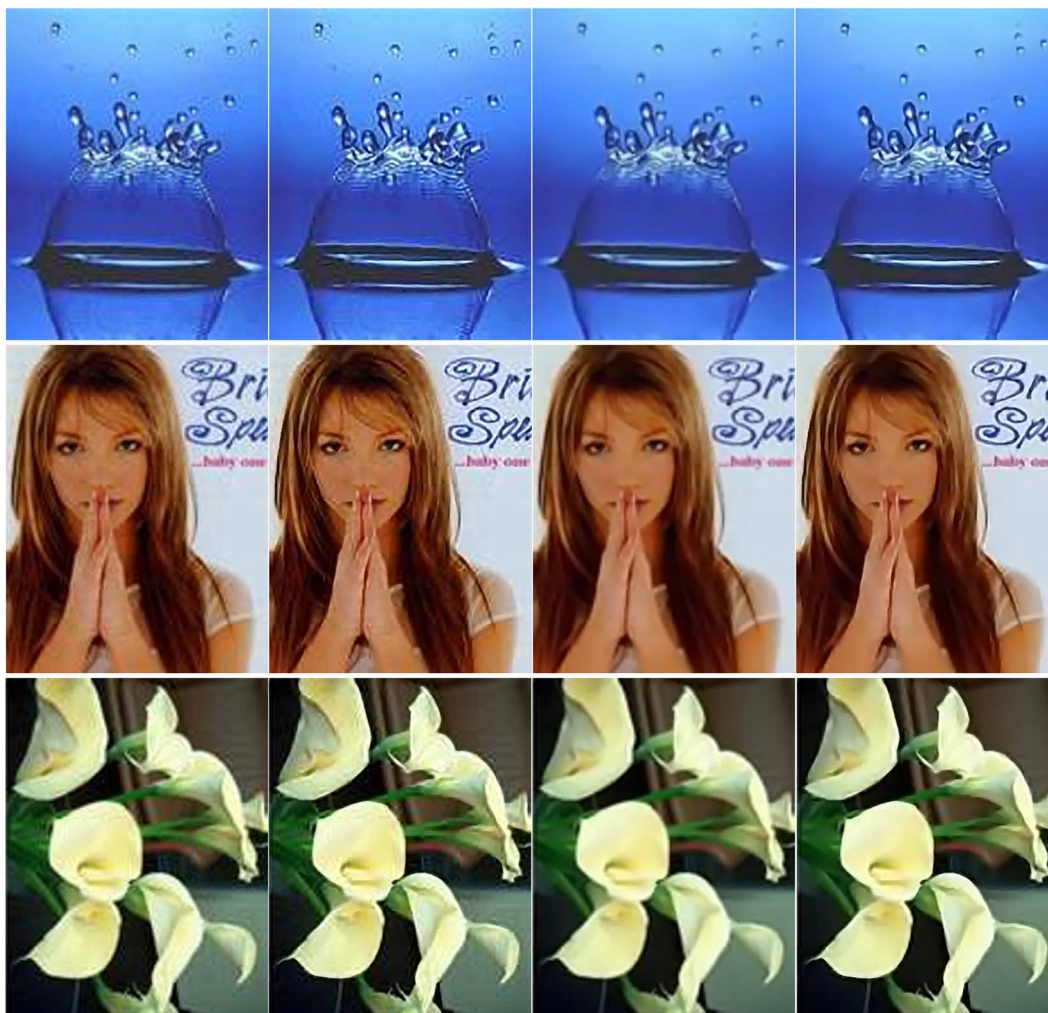


Fig. 14. Super-resolution results of thumbnail web images. From left to right: bicubic interpolation, learning-based pair matching, PDE regularization and our approach.

TABLE I
PSNR (DB) CORRESPONDING TO FIG. 13

image	(a)	(b)	(c)	(d)	(e)	(f)
Lily	27.44	27.55	28.06	27.23	27.96	28.20
Peppers	27.21	27.28	27.82	27.05	27.70	27.96

regularization lowers the PSNR). Therefore, in the compressed image SR scenario, the integration of adaptive regularization is necessary. Secondly, learning-based pair matching, among the tested SR techniques, achieves both the best perceptual quality and objective quality results, which in turn indicates the effectiveness of our proposed combination.

In Fig. 14 we present some SR results of thumbnail web images from *Bing Image Search* [34]. One can easily observe a distinct perceptual quality improvement with our method over bicubic interpolation, PDE regularization and learning-based pair matching.

The computational complexity of our solution is not high. Although the pair matching step is relatively time-consuming, it can be greatly speeded up by the ANN tree searching algorithm

[33]. On the other hand, the database size we used is small compared with that generally required in learning-based SR without compression, and it can be even smaller for real-time application according to Fig. 8. The run time of our algorithm is tested on a Pentium IV 3.0G PC, and it is able to upscale a thumbnail image sized 160×160 in less than 1 second on average. Therefore, this technique can serve as a useful online enlarge-preview tool for image search engines.

B. Video Results

Our solution for compressed video SR is tested on a variety of web videos downloaded from *YouTube* [35]. They are generally in a 320×240 resolution but with different degradation levels. We perform a uniform $3 \times$ SR on them, still using the above database. In the pair matching stage, the candidate number of the enhancing patch $M = 16$, and the SAD threshold $\varepsilon = 0.05$.

Fig. 15 shows three frames extracted from a super-resolved web cartoon video. This result demonstrates the effectiveness of our solution in three aspects. First, the total iteration number of regularization, as enclosed in the caption, is appropriately dependent on the degradation level of each frame. Second, the primitive enhancing images preserve both temporal and spatial

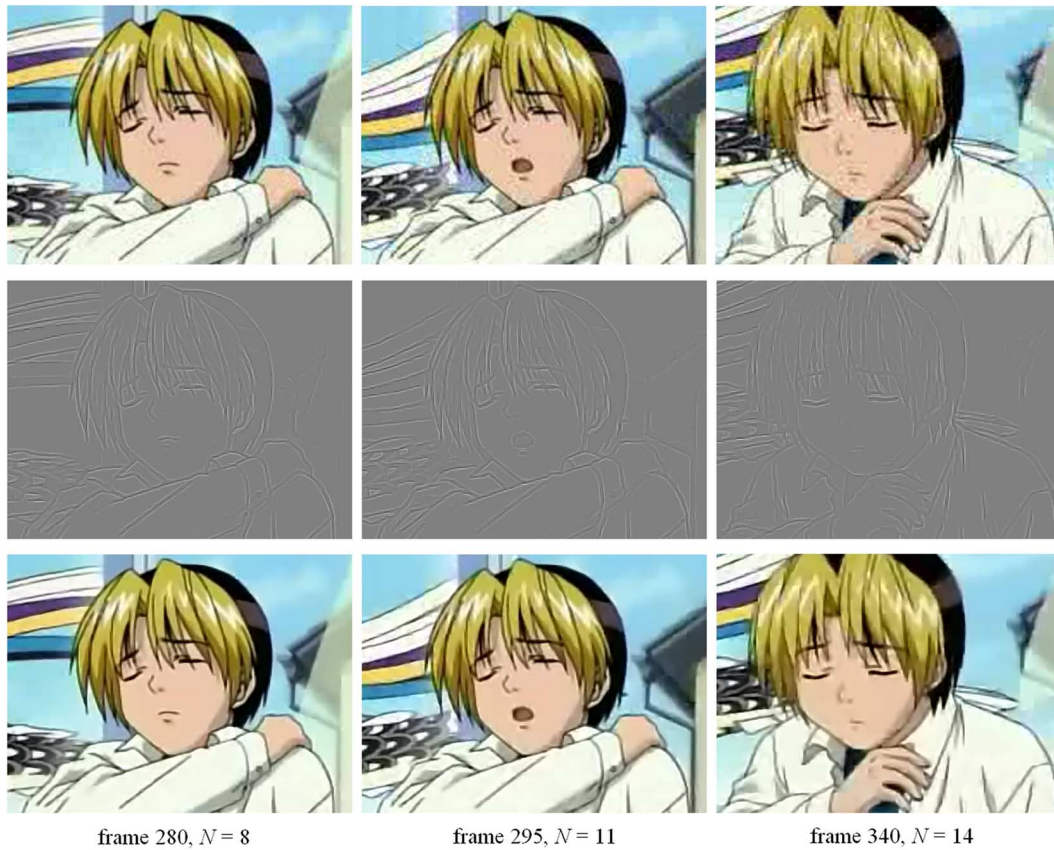


Fig. 15. Super-resolution result of a web video. Top: bicubic interpolation, middle: primitive enhancing images, and bottom: our approach.



Fig. 16. Super-resolution result of an image with rich texture regions. Left: bicubic interpolation, and right: our approach.

consistency due to coherency optimization. Last, the combination of adaptive regularization and learning-based pair matching steadily improve the perceptual quality of directly interpolated videos, even when severe compression artifacts and fast motions are presented. (Please see the electronic version for better visualization.)

C. Applicability Discussion

In general, our method is able to restore HR images/videos from compressed LR measurements with different content and degradation levels. However, it still has certain limitations in application. For images with rich texture regions, neither PDE regularization nor learning-based primitive enhancing works well. In this case, our method may not give significant perceptual quality improvement. Fig. 16 shows an example.

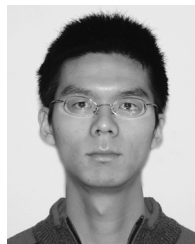
In addition, the upsampling factor in implementation needs not to be exactly the same as that in the database. The performance will not be impacted much if these two factors are close to each other. Only the case when the upsampling factor in implementation is much smaller than that in the database should be avoided (too many high-frequency details may be added and the resulting image may look noisy). We suggest a database with upsampling factor of 3 can deal with most cases.

VI. CONCLUSION

In this paper, we present a robust single-image SR method in the compression scenario, which is competent for simultaneously increasing the resolution and perceptual quality of web image/video with different content and degradation levels. Our method combines adaptive PDE regularization with learning-based pair matching to eliminate the compression artifacts and meanwhile best preserve and enhance the high-frequency details. This method can be naturally extended to video with certain interframe interaction and simple spatio-temporal coherency optimization. Experimental results, including both offline and online tests, validate the effectiveness of our method. Due to its robust performance and low complexity, our solution provides a practical enlarge-preview tool for thumbnail web images, especially those provided by image search engines; it may also be applied to video resizing for online video websites, in case more powerful computational resources (e.g., GPU) are available.

REFERENCES

- [1] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 12, pp. 1153–1160, Dec. 1981.
- [2] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *Proc. IEEE Int. Conf. Image Processing*, 1996, vol. 3, pp. 707–710.
- [3] L. Xin and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.
- [4] Z. Xiong, X. Sun, and F. Wu, "Fast directional image interpolator with difference projection," in *Proc. IEEE Int. Conf. Multimedia & Expo*, 2009, pp. 81–84.
- [5] M. Irani and S. Peleg, "Motion analysis for image enhancement: Resolution, occlusion and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, pp. 324–335, Dec. 1993.
- [6] B. S. Morse and D. Schwartzwald, "Image magnification using level-set reconstruction," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 333–340.
- [7] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Optimal image scaling using pixel classification," in *Proc. IEEE Int. Conf. Image Processing*, 2001, pp. 864–867.
- [8] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 9, pp. 1167–1183, Sep. 2002.
- [9] C. Liu, H. Y. Shum, and C. S. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local non-parametric model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2001, pp. 192–198.
- [10] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. IEEE Int. Conf. Computer Vision*, 1999, pp. 1182–1189.
- [11] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar.–Apr. 2002.
- [12] J. Sun, N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2003, pp. 729–736.
- [13] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2004, pp. 275–282.
- [14] J. Yang, J. Wright, Y. Ma, and T. Huang, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.
- [15] W. Fan and D. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–7.
- [16] L. Ma, Y. Zhang, Y. Lu, F. Wu, and D. Zhao, "Three-tiered network model for image hallucination," in *Proc. IEEE Int. Conf. Image Processing*, 2008, pp. 357–360.
- [17] Z. Xiong, X. Sun, and F. Wu, "Image hallucination with feature enhancement," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops*, 2009, pp. 2074–2081.
- [18] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.
- [19] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1064–1076, Aug. 1997.
- [20] N. R. Shah and A. Zakhori, "Resolution enhancement of color video sequences," *IEEE Trans. Image Process.*, vol. 8, no. 6, pp. 879–885, Jun. 1999.
- [21] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object based high resolution image reconstruction from low resolution video," *IEEE Trans. Image Process.*, vol. 6, no. 10, pp. 1446–1451, Oct. 1997.
- [22] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochrome and color video using motion compensation," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 278–287, Feb. 2001.
- [23] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 4, pp. 217–226, Apr. 2002.
- [24] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of compressed video using transform-domain statistics," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 33–43, Jan. 2004.
- [25] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 898–911, Jul. 2004.
- [26] Z. Xiong, X. Sun, and F. Wu, "Super-resolution for low quality thumbnail images," in *Proc. IEEE Int. Conf. Multimedia & Expo*, 2008, pp. 181–184.
- [27] G. Aubert and P. Kornprobst, "Mathematical problems in image processing: Partial differential equations and the calculus of variations," in *Applied Mathematical Sciences*. New York: Springer-Verlag, Jan. 2002.
- [28] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
- [29] J. Weickert, *Anisotropic Diffusion in Image Processing*. Stuttgart, Germany: Teubner-Verlag, 1998.
- [30] D. Tschumperle and R. Deriche, "Vector-valued image regularization with PDEs: A common framework for different applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 4, pp. 506–517, Apr. 2005.
- [31] G. Aubert and P. Kornprobst, "Mathematical problems in image processing: Partial differential equations and the calculus of variations," in *Applied Math. Sciences*. New York: Springer-Verlag, Jan. 2002.
- [32] P. Perona and J. Malik, "Detecting and localizing edges composed of steps, peaks and roofs," in *Proc. IEEE Int. Conf. Computer Vision*, 1990, pp. 52–57.
- [33] D. Mount and S. Arya, Ann: Library for Approximate Nearest Neighbor Searching [Online]. Available: <http://www.cs.umd.edu/mount/ANN/>
- [34] [Online]. Available: <http://www.bing.com/images?FORM=Z9LH3>
- [35] [Online]. Available: <http://www.youtube.com/>
- [36] [Online]. Available: <http://www.kodak.com/digitalImaging/samples/imageIntro.shtml>
- [37] Z. Xiong, X. Sun, and F. Wu, "Web cartoon video hallucination," in *Proc. IEEE Int. Conf. Image Processing*, 2009, pp. 3941–3944.



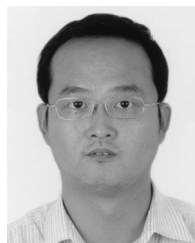
Zhiwei Xiong received the B.S. degree in electronic engineering from University of Science and Technology of China (USTC), Hefei, China, in 2006. He is now pursuing the Ph.D. degree in electronic engineering at USTC.

He has been a research intern in Microsoft Research Asia since 2007, where his research concentrates on image/video compression, image/video processing and computer vision. He is particularly interested in the inverse problems such as image super-resolution, regularization, and restoration.



Xiaoyan Sun (M'04) received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1997, 1999, and 2003, respectively.

She joined Microsoft Research Asia, Beijing, China, as an Associate Researcher in 2003 and has been a Researcher since 2006. She has authored or co-authored over 30 conference and journal papers and submitted several proposals and contributed techniques to MPEG-4 and H.264. Her research interests include video/image compression, video streaming, and multimedia processing.



Feng Wu (M'99–SM'06) received the B.S. degree in electrical engineering from the University of Xi'an Electrical Science and Technology, Xi'an, China, in 1992, and the M.S. and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1996 and in 1999, respectively.

He joined Microsoft Research Asia, Beijing, China, as an Associate Researcher in 1999 and was promoted to Lead Researcher in 2006. He has played a major role in Internet Media Group to develop scalable video coding and streaming technologies.

He has authored or co-authored over 100 papers in video compression and contributed some technologies to MPEG-4 and H.264. His research interests include video and audio compression, multimedia transmission, and video segmentation.