IOSR Journal of Electronics and Communication Engineering (IOSRJECE)
ISSN: 2278-2834 Volume 2, Issue 3 (July-Aug 2012), PP 07-18
www.iosrjournals.org

Visual Quality Improvement of an Image/Video in Web Application

[1]N.Sravani,[2]N. Ramanjaneyulu
[1]*P.G.Student, Dept. of ECE, RGMCET, Nandyal, A.P, India.*
[2]*Associate Professor, Dept. of ECE, RGMCET, Nandyal, A.P, India.*

Abstract: *This project proposes a robust single-image super-resolution method for enlarging low quality web image/video degraded by down sampling and compression. To 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. The suggested approach has to be developed using matlab tool.*

Index Terms: *Adaptive regularization, learning-based super-resolution (SR), artifacts, down sampling.*

I. Introduction

With the Internet flourishing and the rapid progress inhand-held photographic devices, image and video arebecoming more and more popular on the web, due to their richcontent and easy perception. Consequently, image search enginesand online video websites have experienced an explosionof visits during the past few years. However, limited by thenetwork bandwidth and server storage, most web image/videoexists in a low quality version degraded from the source. Themost common degradations are down sampling and compression.Downsampling exploits the correlation in the spatial domainwhile compression further exploits the correlation in thefrequency and temporal (for video) domains. Quality degradationgreatly lowers the required bandwidth and storage, makingthe access to web image/video practical and convenient. Butthese benefits are obtained at the expense of impairing the perceptualexperience of users, as degradation inevitably leads toinformation loss, which behaves as various artifacts in the resultingimage/video, e.g., blurring, blocking and ringing. There is a large demand for improving the perceptual qualityof web image/video, among which the resolution enhancement, also known as super-resolution (SR), is an especially importantissue and attracts a lot of attention. SR refers to the techniquesachieving high-resolution (HR) enlargements of pixelbasedlow-resolution (LR) image/video. Basically, there are twokinds of SR, according to the amount of LR images utilized:multi-image SR, which requires several LR images of the samescene 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 imagesearch engines for example: once a query is entered, a large number of images need to be returned simultaneously, andthe results are first displayed in their LR forms (often called"thumbnails"). Users then need to click on the thumbnailto get its original HR version. Nevertheless, sometimes it isfrustrating that the source image is removed or the server istemporarily unavailable. Single-image SR, at this moment, cansave users from the bother of linking to every source if only anenlarged preview is desired.Previous work on single-image SR can be roughly dividedinto four categories: interpolation-based [1]–[4], reconstruction-based [5], [6], classification-based [7] and learning-based[8]–[17]. Despite great diversity in implementation, thesemethods have a common premise that the LR image is onlydegraded by down sampling. This is not always true in the webenvironment, where compression is widely adopted. For imagesearch engines, compression helps reduce the thumbnail sizeby up to 50% without obvious perceptual quality loss whenpresented in the LR form. But now if SR (any of the above) isdirectly performed, compression artifacts will be magnified outand the perceptual quality of resulting HR images will be poor.On the other hand, multi-image SR has been used to enlargevideo for a long time [5], [18]–[22], and corresponding techniquesfor compressed video SR have also been reported inliterature [23]–[25]. These methods generally assume a priordistribution of the quantization noise and then integrate thisknowledge into a Bayesian SR framework, or use the quantizationbounds to determine convex sets which constrain the SRproblem. In practical applications, however, the compression artifactscaused by the quantization noise are largely dependenton the video content and difficult to be modeled with an explicitdistribution. Moreover, since the performance of Bayesian SRheavily depends on the accuracy of frame

registration and motionestimation, these methods are not capable of reconstructinghigh frequency details of dynamic videos that contain fast andcomplex object motions.

In this paper, we present a practical solution which combinesadaptive regularization with learning-based SR to simulta-neouslyimprove 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.

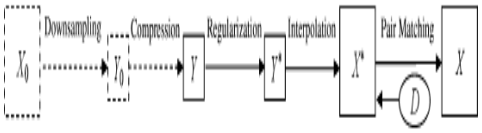


Fig. 1. Flowchart of our compressed image super-resolution scheme.

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 interframeinteractions on the regularization strength and simple spatial-temporal consistency optimization, as reported in our latestwork [37]. Different from conventional methods, our solutiondoes not require any specific assumption on the quantizationnoise 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 down sampled with a low-pass filter$g$(mostly isotropic Gaussian) to form an LR measurementY_0

$$Y0 = (g * X0) \downarrow \alpha \quad (1)$$

Where$\downarrow \alpha$ is a decimation operator with scaling factor Y_0isthen compressed, resulting in a degraded LR measurement$Y$

$$Y = Y0 + E_Q \quad (2)$$

where$E_Q$where represents the quantization error introduced by compression in the spatial domain.$Y$is the actual input of our SR system. This system consistsof 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^*

$$Y * = f\mathrm{N}(\mathrm{Y}) \quad (3)$$

where f(.) denotes the PDE regularization functional and the superscript N represents the total iteration number of regularization, which determines the regularization strength. is thenupsampled with scaling factorβ to get an intermediate HR resultX^*

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

where stands for the bicubic interpolation filter. The final HRimage X is obtained after learning-based pair matching fromX^* and a prepared database$D$. The maximum a posterior probability(MAP) estimate of Xcan 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* froma LR measurement Y . For learning-based approaches, a set of examples organized in a database*D* are utilized in the onlinereconstruction process. These examples usually exist as co-occurring patch pairs $\{x_k,y_k\}$=1extracted from training imagesat two different resolution levels. The basic idea of using examplesin SR is that natural images are special signals occupyingonly a vanishingly small fraction of the high dimensional imagespace. Therefore, high-frequency details that do not exist *Y*incan be "stolen" from*D* through *pair matching*, i.e., given anLR patch from the input measurement, seek in the database forsimilar LR examples, and their corresponding HR counterpartscan then be used for the reconstruction as they provide high-frequencydetails that fit the input measurement.In our scheme, the primitive-based hallucination method proposedin [12] is adopted for pair matching, for which imageprimitives (edges, ridges and corners) are represented by examplesand pair matching is only applied to the primitive componentsin images. The superiority of this method is twofold. First,human observations are especially sensitive to image primitiveswhen going from LR to HR. Second, the primitive manifold is ofintrinsic lower dimensionality compared with raw image patchmanifolds as used in [10], [13], [14] and can be more effectivelyrepresented by examples. (Please refer to [12] for detailed proceduresof primitive example generation.)Generally speaking, learning-based pair matching exploitsthe correspondence between image signals at two different resolutionlevels, whereas another kind of degradation—compression—is seldom considered in previous works. One may suggest 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 down sampling, is difficult to be effectivelyrepresented by examples. The underlying reason is thatcompression corrupts the primitive pattern (or other feature patterns)contained in examples, and thus the correspondence betweenthem. As an alternative, we propose to keep the databaseaway from compression while introducing regularization on thecompressed LR measurement.

C. PDE Regularization

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

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

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

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

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

$$\frac{\partial I}{\partial t} = \mathrm{trace}(\boldsymbol{\Theta}\mathbf{H}) \qquad (8)$$

WhereH is the *Hessian matrix*of *I*and $\theta$is an anisotropic 2X2 tensor defined as

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

θgives 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 th iteration, there is

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

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

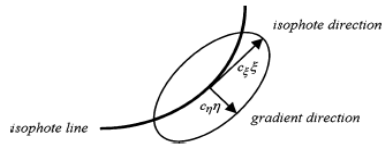


Fig. 2. PDE smoothing geometry. It can be viewed as a juxtaposition of two oriented 1D heat flows along the gradient direction η and the isophote direction ξ, with corresponding weights c_η and c_ξ.

As pointed out in [30], regularization PDE's generally do notconverge toward a very interesting solution. Most of the time,the image obtained at $n \to \infty$ is constant, corresponding toan image without any variations. Therefore, the total iterationnumber of regularization is$N$ often manually determined accordingto the image degradation level. In the scenario of compressedimage SR, one question arising here is that how to adaptivelycontrol this regularization strength$N$ to maximize theposterior in (5), given prefixed pair matching database$D$andPDE regularization functional. Inother words, can we find theturning point where compression artifacts are effectively eliminatedwhile 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$T$is a bitmap storing detected edge pixel locations in$I$

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

The primitive field (PF) is defined as

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

where$Np(u,v)$ refers to a ρth order neighborhood of(u,v)

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

Correspondingly, the nonprimitive field (NPF) is defined as

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

Fig. 3 illustrates the PF and NPF partition with $= 1$.
After the nth iteration of regularization, the image energy change in PF and NPF can be calculated as

$$\Delta E_n^P = \sum_u \sum_v |\Delta I_n(u,v)|^2 P(u,v)$$
$$\Delta E_n^Q = \sum_u \sum_v |\Delta I_n(u,v)|^2 Q(u,v) \qquad (15)$$

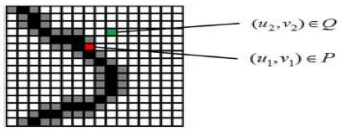
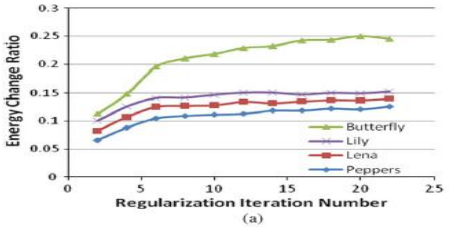


Fig. 3. Partition of primitive and nonprimitive fields with $\rho = 1$. Black edge pixels and their gray neighbors constitute the primitive field while the remaining white pixels constitute the nonprimitive field. Pixel (u_1, v_1) is in the primitive field and pixel (u_2, v_2) is in the nonprimitive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



(a)

(b)

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

where$\Delta In$is the image intensity change in (10).We denote theenergy change ratio between PF and NPF as

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

Fig. 4(a) gives several practical results of the $r\sim n$curve onthe test images shown in Fig. 4(b). The test images are first 1/3downsampled from the original and then compressed by JPEGwith QP set to 60. It can be seen all those $r\sim n$ curves convergeafter a few iterations. To obtain a general distribution pattern tern of the$r\sim n$ convergence speed, we use several video sequences(each containing 1500 frames) randomly downloadedfrom *Youtube* [35] for test, instead of gathering a large numberof images. As shown in Fig. 5, most frames require less than15 iterations, which validates the convergence property of the$r\sim n$curve in the web environment.

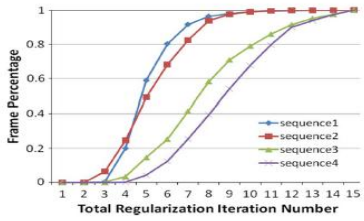


Fig. 5. Distribution of total regularization iteration number for web videos.

The convergence property of the $r\sim n$ curve can be figuredout intuitively from the edge-preserving nature of PDE regularization.In the earlier stage, the energy change in NPF is moreintensive than that in PF, so increases with .When regularizationis performed to a certain stage, PDE loses its efficacy indistinguishing salient features from images, and then remainsat a stable level. Further, when is small, there is a large probabilitythat compression artifacts such as ringing and blockingappear in NPF. So PDE regularization removes compression artifactsfirst.

According to the energy change characteristics during PDE regularization, we can nowdetermine 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 Nth iteration when

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

where $\mu$is a constant and Rrepresents the fastest increasingspeed of r.

B. Pair Matching Accuracy

To demonstrate the necessity and effectiveness of adaptiveregularization in compressed image SR with learning-basedmethods, we then investigate the pair matching accuracy under three different circumstances,i.e., without compression, withcompression but no regularization and with both compressionand adaptive regularization. We denote them as ,and .To compare the pair matching accuracy, we use a Receiver Operating Characteristic (ROC)curve to demonstrate thetradeoff between match error and hit rate. We define the matcherror as

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

where$x$ denotes the real missing HR primitive patch and x' isthe HR example found through pair matching from the existent LR primitive patch . For a given match error , 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

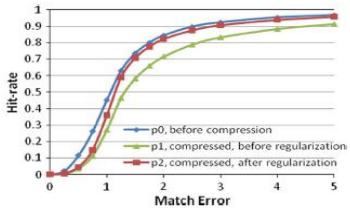


Fig. 6. ROC curves of pair matching accuracy. 50,000 primitive patches are tested over 100,000 trained examples.

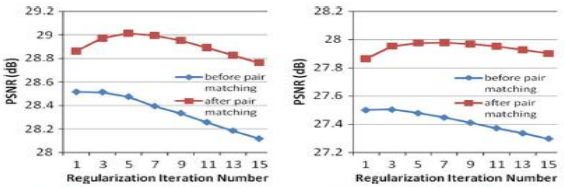


Fig. 7. PSNR curves against the iteration number of regularization, before and after pair matching. Test images are taken from Fig. 4(b). Left: *Lena*, and right: *Peppers*.

examples. The test data is sampled from images in Fig. 4(b),whereas the training images are shown in Fig. 11. As can beobserved from $\rho_0$and ρ_1, when compression is involved, the pairmatching accuracy degrades heavily. On the other hand, withadaptive regularization,$\rho_2$is higher than$\rho_1$at any match errornand quickly approachesρ_0 , which indicates that the proposedadaptive regularization steadily and greatly improves the pairmatching accuracy in compressed image SR.To further verify the adaptivity criterion in our scheme, inFig. 7 we plot two PSNR curves against the iteration numberof regularization, with and without pair matching. It can be observedthat, though PDE regularization removes artifacts (andthus improves the perceptual quality), it gradually lowers thePSNR, as we don't enforce any fidelity constraints. However,after pair matching, the PSNR curve pattern changes. The peakvalue 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 matchingdatabase, mainly to verify the adaptivity criterion in the regularizationstep.Nevertheless,forlearning-based approaches, thechoice of training set is still an important issue. Therefore, wealso analyze how the choice of training set influences the performanceof our method.Since the primitive pair matching mainly exploits the correspondencebetween LR and HR primitive components, thetraining images are generally required to have rich and diverseprimitive patterns. However, we would like to point out that, as

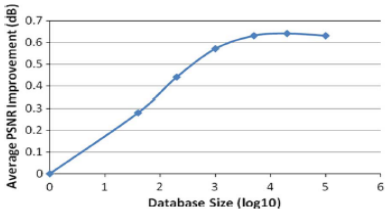


Fig. 8. PSNR improvement between directly interpolated images and those with our proposed SR under different database sizes. Test images are taken from Fig. 4(b).

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 interpolatedcompressed images and those with our proposed SR. The resultis shown in Fig. 8, where the PSNR improvement tendsto be saturated when the database exceeds a certain size. Notethat 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, reducesthe number of distinct primitive patterns in the input images.

D. Discussion and Summarization

Compressed image SR is a practical problem in the web applicationof single-image SR, but has rarely been investigatedbefore. In this subsection, we would like to supplement someintuition on why learning-based pair matching and PDE regularizationare combined to solve this problem.

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

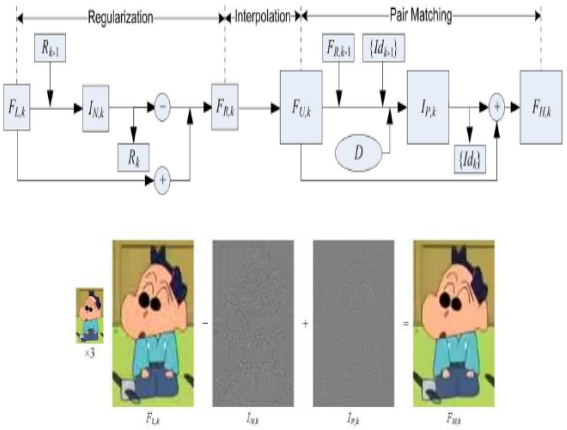


Fig 9. Framework of our solution for compressed video super-resolution. The input frame $F_{L,k}$ and eliminated noise image $I_{N,k}$ are presented in their interpolated version in the example for better visualization. Note $I_{N,k}$ is only the luminance component of the noise image.

As learning-based SR is usually performed at patch level andrelatively 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 $X0$*through bicubic interpolation.
2. Find a PF/NPF partition $(P, Q)$of$X0$through the orientation energy edge detection [32].
3. Perform iterative PDE regularization on :
   a) After each iteration, upsample the regularizedimage Yn* toXn*,$n = 1,2,$ .... through bicubicinterpolation;
   b) Calculate the image energy change(ΔEn^{P} ,ΔEn^{Q})between Xn * and $Xn-1$ *based on the PF/NPFpartition(P, Q) ;
   c) Calculate the energy change ratio between PF andNPF as rn =($\Delta En^{P}/\Delta En^{Q}$) and also record themaximum slope of r_n as$R = \max(rn - r_{n-1}), n \geq 2$;
   d) If $rn - r_{n-1} < \mu R$,stop reguralisation and keep theiteration number N.
4. Extract LR primitive patches from $Xn*$ and findcorresponding HR primitive patches from a prepareddatabase$D$ through pair matching speeded up by theapproximate nearest neighbor (ANN) tree searching [33].
5. Add the HR primitive patches back to $Xn*$to form thefinal HR image$X$,where the compatibility of neighboringHR primitive patches is enforced by averaging the pixelvalues in overlapped regions.

End

IV. Compressed Video Super-Resolution

A. Framework

Since the above introduced single-image SR method doesnot require frame registration or motion estimation, it can bedirectly applied into the compressed video SR in a frame-byframe 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,aKth frame$F_{L,K}$from an LR video is divided into PF and NPF and iterativePDE regularization is performed on $F_{L,K}$during which the energychange velocities in both PF and NPF are recorded. Whenthe ratio of these two velocities converges (judged by a parameterR_K, which is also influenced by that of the previous frameR_{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 resolutionthrough bicubicinterpolation.Last, the primitive componentsin the interpolated frame, $F_{U,K}$are enhanced with learning-basedpair matching. Meanwhile, the temporal consistency is enforcedby referring to the previous interpolated frame, $F_{U,K-1}$ and itspair matching indices$\{Id_{k-1}\}$. Adding the primitive enhancing

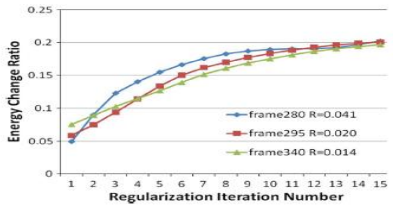


Fig. 10. Energy change characteristics in consecutive frames. Frames with heavier degradation exhibit slower $r \sim n$ convergence speed. Test frames are shown in Fig. 15.

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 thisframework.

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 $r \sim n$of the 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 kth frame,R_k and for the th frame, is calculated as

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

Where $R_{k,0}$is the initial quantity measured from the currentframe. If $R_{k,0} < R_{k-1}$,it suggests the degradation in thecurrent frame is more severe than that in the previous frame.Then $R_{k0}$is further diminished to increase the regularizationstrength of the current frame (according to (17), reducing Rwill increase N),and vice versa.In this way, the regularizationstrength can also adapt to the variable degradation betweenconsecutive frames caused by fast motion or scene switch,making the quality improvement on the whole video sequencemore stable.

C. Spatio-Temporal Coherency Optimization

After the adaptive regularization, compression artifacts ineach 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 mainproblem when applying this step to video is how to make the enhanced primitives

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

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 iof the kth interpolated framex_{ik}, and represent the enhancing HR primitive patch corresponding toy_{ik}.In the temporal domain,we record the pair matching indices in the (k-1)th frame.Then, for the th frame, each $y_{ik}$is compared with y_{ik-1}inthe 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 ofabsolute difference (SAD) is smaller than a given threshold, thepair matching index of y_{ik-1} is directly assigned toy_{ik};or elsea new pair matching for$y_{ik}$is conducted in the database.

In the spatial domain, since some pair matching resultsare derived from the previous frame and others are generatedfrom the current frame, the compatibility of learned enhancingpatches should be optimized. Specifically, for each y_{ik} thatcannot use the index from the previous frame and a new pairmatching is required, we take the first M pair matching results { $x_{ik},1,x_{ik},2,\ldots,x_{ik},M$}as candidates for x_{ik}, and theoptimum one is found by

$$\hat{m} = \arg\min_{1 \le m \le M} d(x_{ik,m}, x_{jk}) \qquad (20)$$

wherex_{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

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

where x_{ik-1} is the selected enhancing patch in the location I of the (k-1)th frame, and 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.

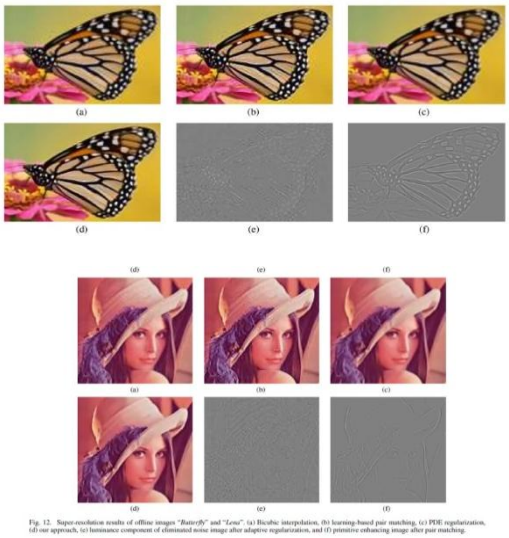


Fig. 12. Super-resolution results of offline images "*Butterfly*" and "*Lena*". (a) Bicubic interpolation, (b) learning-based pair matching, (c) PDE regularization, (d) our approach, (e) luminance component of eliminated noise image after adaptive regularization, and (f) primitive enhancing image after 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 =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 of9x9 sized primitive patches is used in pair matching. These examplesare trained from 16 representative natural images shownin Fig. 11. For color images, regularization is performed on boththe luminance and chrominance components, for which compressionis applied. Learning-based pair matching, however, isonly performed on the luminance component, as human observersare more sensitive to the luminance change in imageswhen going through LR to HR.

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

learning-based pair matching, PDE regularization and ourproposed 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 nonlearning- 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.Forbackprojection and pair matching, the PSNR is also improved.(Single interpolation doesn't see a PSNR gain as PDE

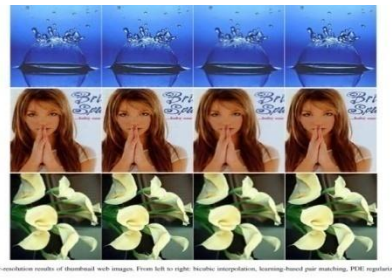


Table I
Psnr (Db) Corresponding To Fig. 13

IMAGE	(a)	(b)	©	(d)	(e)	(f)
lily	27.44	27.55	28.06	27.23	27.96	28.2
peppers	27.21	27.28	27.82	27.05	27.7	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 qualityand objective quality results, which in turn indicates the effectivenessof 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 160X160 in less than 1 second on average. Therefore, this technique can serve as a useful online enlarge-previewtool 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 320X240 resolution but with different degradation levels. We perform a uniform 3XSR 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/videosfrom compressed LR measurements with different content anddegradation levels. However, it still has certain limitations in application.For images with rich texture regions, neither PDE regularizationnor learning-based primitive enhancing works well.In this case, our method may not give significant perceptualquality 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 methodin 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 solutionprovides a practical enlarge-preview tool for thumbnail webimages, especially those provided by image search engines; itmay also be applied to

video resizing for online video websites,in case more powerful computational resources (e.g., GPU) areavailable.

References

[1] R. G. Keys, "Cubic convolution interpolation for digital imageprocessing,"*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," *IEEETrans. Image Processing*, vol. 10, no. 10, pp. 1521–1527, Oct. 2001.

[4] Z. Xiong, X. Sun, and F.Wu, "Fast directional image interpolator withdifference 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 usinglevel-set reconstruction," in *Proc. IEEE Conf. Computer Vision andPattern Recognition*, 2001, pp. 333–340.

[7] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Optimal imagescaling using pixel classification," in *Proc. IEEE Int. Conf. ImageProcessing*, 2001, pp. 864–867.

[8] S. Baker and T. Kanade, "Limits on super-resolution and how to breakthem," *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 hallucinatingfaces: Global parametric model and local non-parametricmodel," 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 superresolution,"*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 withprimal sketch priors," in *Proc. IEEE Conf. Computer Vision and PatternRecognition*, 2003, pp. 729–736.

[13] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution throughneighbor embedding," in *Proc. IEEE Conf. Computer Vision andPattern Recognition*, 2004, pp. 275–282.

[14] J. Yang, J. Wright, Y. Ma, and T. Huang, "Image super-resolution assparse representation of raw image patches," in *Proc. IEEE Conf. ComputerVision and Pattern Recognition*, 2008, pp. 1–8.

[15] W. Fan and D. Yeung, "Image hallucination using neighbor embeddingover visual primitive manifolds," in *Proc. IEEE Conf. Computer Visionand Pattern Recognition*, 2007, pp. 1–7.

[16] L. Ma, Y. Zhang, Y. Lu, F. Wu, and D. Zhao, "Three-tiered networkmodel 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 RecognitionWorkshops*, 2009, pp. 2074–2081.

[18] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolutionframes 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 videoreconstruction with arbitrary sampling lattices and nonzero aperturetime," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1064–1076, Aug.1997.

[20] N. R. Shah and A. Zakhor, "Resolution enhancement of color videosequences," *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 highresolution image reconstruction from low resolution video," *IEEETrans. Image Process.*, vol. 6, no. 10, pp. 1446–1451, Oct. 1997.

[22] B. C. Tom and A. K. Katsaggelos, "Resolution enhancement of monochromeand 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 stilland video reconstruction from MPEG coded video," *IEEE Trans. CircuitsSyst. Video Technol.*, vol. 12, no. 4, pp. 217–226, Apr. 2002.

[24] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Super-resolutionreconstruction 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, "Bayesianesolution enhancement of ompressed video," *IEEE T*

[26] Z. Xiong, X. Sun, and F.Wu, "Super-resolution for low quality thumbnailimages," 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 regularizationwith PDEs: A common framework for different applications," *IEEETrans. 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 ofsteps, peaks and roofs," in *Proc. IEEE Int. Conf. Computer Vision*,1990, pp. 52–57.

[33] D. Mount and S. Arya, Ann: Library for Approximate NearestNeighbor 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.