



Learning-based image restoration for compressed images

Lin Ma^{a,b,*}, Debin Zhao^b, Wen Gao^c

^a Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

^b School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

^c National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing, China

ARTICLE INFO

Article history:

Received 24 February 2011

Accepted 16 May 2011

Available online 6 June 2011

Keywords:

Image restoration

Compression artifacts

Primitive

ABSTRACT

In this paper, we propose a novel learning-based image restoration scheme for compressed images by suppressing compression artifacts and recovering high frequency (HF) components based upon the priors learnt from a training set of natural images. The JPEG compression process is simulated by a degradation model, represented by the signal attenuation and the Gaussian noise addition process. Based on the degradation model, the input image is locally filtered to remove Gaussian noise. Subsequently, the learning-based restoration algorithm reproduces the HF component to handle the attenuation process. Specifically, a Markov-chain based mapping strategy is employed to generate the HF primitives based on the learnt codebook. Finally, a quantization constraint algorithm regularizes the reconstructed image coefficients within a reasonable range, to prevent possible over-smoothing and thus ameliorate the image quality. Experimental results have demonstrated that the proposed scheme can reproduce higher quality images in terms of both objective and subjective quality.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In order to accommodate with the bandwidth of the Internet and the storage space, image and video compression schemes are highly demanded. In most image and video coding standards, block-based discrete cosine transform (BDCT) coding has prevailed, which aims at reducing the inter-pixel statistical redundancy. However, for the sake of achieving higher compression ratio, BDCT together with the coarse quantization gives rise to the discontinuity of intensities between adjacent blocks which is named as blocking artifacts. Truncating the high frequency (HF) BDCT coefficients would also result in ringing artifacts

around the edges. Consequently, the subjective quality of the compressed image is unpleasant and image restoration for ameliorating the image quality is necessary.

The compression artifacts can be suppressed in transform domain (e.g. DCT [1–3], overcomplete wavelet representation (OWR) [4–6]), spatial domain [7–10,12–14], or the combinations [15–21]. For DCT domain deblocking, direct manipulations on DCT coefficients can alleviate the artifacts before the images are fully decoded, which results in lower computations. In particular, Zeng [1] models the blocking artifact as 2-D step edge and suppresses it by applying the zero-masking scheme. A signal adaptive filtering scheme for reducing the blocking artifacts is proposed in Ref. [2], which considers the masking effect of HVS, adaptive weighting mechanism and quantization constraint. Liu and Bovik [3] propose to change a step edge into a slop one through modifying certain DCT coefficients to alleviate the blocking artifacts. Moreover, OWR is employed for deblocking. In Ref. [4], blocking artifacts are removed in wavelet domain by exploiting cross-scale correlations among wavelet coefficients and protecting the edge information. Liew et al. [5]

* Corresponding author at: Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong. Tel.: +852 26098255; fax: +852 26035558.

E-mail addresses: lma@ee.cuhk.edu.hk (L. Ma), dbzhao@jdl.ac.cn (D. Zhao), wgao@jdl.ac.cn (W. Gao).

¹ Part of this work was done while Lin Ma was with Microsoft Research Asia as a Research Intern.

propose to suppress blocking and ringing artifacts by analyzing statistical characteristic of block discontinuities as well as behavior of wavelet coefficients across scales for different image features.

Deblocking in spatial domain is proposed by applying spatial-adaptive filtration [7–10]. Apostolopoulos and Jayant [7] propose filtering pixel-by-pixel along the block boundaries to identify and reduce both the blocking and mosquito noise. Kim et al. [8] select two separate filtering modes to process the pixel around the block boundary according to its behavior. In Ref. [9], 1-D horizontal and vertical low-pass filtering for suppressing blocking artifacts and 2-D signal-adaptive filtering for removing ringing artifacts are utilized for postprocessing low bit-rate compressed videos. Based on the non-local denoising algorithms [11], postfiltering in shifted widows (PSW) of image blocks is proposed in Refs. [12–14], which suppresses blocking artifacts by averaging coefficients of neighboring image blocks in the shifted windows.

On the other hand, many researches regarded image compression as a distortion process and put forward iterative algorithms for restoring the original images. The projection onto convex sets (POCS) algorithms [15,20,22–25] represent the prior information of the original as convex sets, and they converge in the intersection of all the sets through iterating projections. The most commonly used convex sets are quantization constraint sets (QCS) and smoothness constraint sets (SCS). In Ref. [15], the deblocking algorithm presents the iterative procedure based on QCS and SCS to restore the coded image to its original artifact-free form. Yang et al. [20,23,24] propose to recover images by incorporating local statistical properties, human perceptual characteristics, and new family of directional SCS based on lineally modeling of image edge structure. Park and Kim [25] narrow down QCS to form the narrow quantization constraint sets (NQCS) for restoring images of higher quality.

However, all of the above traditional postprocessing or restoration algorithms may not recover the HF components [28], which have been discarded during the quantization step of compression. Recently, learning-based image restoration schemes have been proposed to reconstruct a high-quality image by introducing the learned HF information from pre-designed codebooks into the degraded low-quality image. Sheppard et al. [26] introduce nonlinear interpolative vector quantization (NLIVQ) into image restoration. It performs nonlinear restoration of diffraction-limited images concurrently with quantization. Based on NLIVQ, a blind image restoration method [27] is proposed by estimating the HF information of a given blurred image from its low-frequency (LF) information based on the designed multiple codebooks. Liaw et al. [28] propose to restore the image based on the classified vector quantization (CVQ), which employs a codebook to transform the compressed image into a set of indices, and decodes the indices to enhance the compressed image based upon another corresponding codebook. Actually, all of these existed learning-based image restoration schemes share the same assumption as image super-resolution [29,30], which is that the degraded image patch can be employed as

the index of the proper HF image patch in the learnt codebook for restoring the image. However, it is an ill-posed problem. Since the number of the degraded image patches is always of smaller amount, whereas the number of HF image patches is of larger amount, one degraded image patch will be mapped to more than one HF patches. This situation often occurs in the quantization step of image/video compression. Many original image/video patches will be quantized into the same degraded patches. Consequently, one degraded image patch can be mapped to many original patches.

Several approaches have been proposed to solve this problem for image super-resolution and restoration. Freeman et al. [29,30] model the LF and HF patches as a Markov network and employ Bayesian belief propagation to find a local maximum of the posterior probability. Liu et al. [31] propose a two-step statistical modeling approach that integrates both a global parametric model and a local non-parametric one. Based on the locally linear embedding (LLE) [32], image super-resolution [34,35] and image restoration for compressed images [38] has been proposed by considering the local geometry during the mapping.

Inspired by image hallucination [33–38], a novel learning-based image restoration scheme is proposed. JPEG compression process is simulated by a degradation model, which comprises signal attenuation and additive Gaussian noise model. First, local filtering process is analyzed and employed to remove the additive Gaussian noise. Second, in order to strengthen the mapping relationship while synthesizing the HF component, a differential image enhancement algorithm is proposed to enforce consistency between primitives to meet the contour smoothness constraint. Third, a Markov-chain model is employed to handle the attenuation process of the degradation model by modeling the relationship between the existent HF primitives in the learnt codebook and the enhanced ones. Finally the quantization constraint algorithm regularizes the DCT coefficients of the restored image within a reasonable range.

The rest of the paper is organized as follows. In Section 2, we firstly introduce the degradation model for simulating JPEG compression process. Subsequently, based on the degradation model, learning-based image restoration scheme is proposed. Detailed information of learning and synthesizing strategy is introduced in Section 3. Experimental results are demonstrated in Section 4. Finally, Section 5 concludes the paper.

2. Proposed image restoration framework

2.1. Image degradation model

As many high-quality images could generate the same degraded images, restoring the original image from the single degraded image is an under-constrained problem. However, if the degradation process could be modeled accurately, image restoration will appear much easier.

Geman et al. [16] proposed a simple yet efficient image degradation model, which have been widely used for image restoration [16,17] and image quality evaluation [18,19]. The degradation model could be viewed as a simple signal

attenuation and additive Gaussian noise model, which is defined by

$$D = G * R + N. \quad (1)$$

Considering the degradation model at pixel or block level, it could be regarded as the processes on the corresponding random fields (RFs), which could be expressed as

$$D = \{G_i * R_i + N_i : i \in I\}, \quad (2)$$

where $R = \{R_i, i \in I\}$ denotes RF of the original signal, $D = \{D_i, i \in I\}$ denotes RF of the corresponding degraded signal, $G = \{G_i, i \in I\}$ is a deterministic scalar attenuation field, $N = \{N_i, i \in I\}$ is a stationary additive zero-mean Gaussian noise RF, and I denotes the set of spatial indices (such as pixel/block position) for the corresponding RFs. Additionally, it is assumed that N is white and independent of R . It is also constrained that the field G varies slowly [18,19], which means that the neighboring attenuation RF appears nearly the same. From Ref. [19], it has been demonstrated that most of the distortion types appeared in real world systems could be roughly described locally by the combination of the attenuation and additive Gaussian noise RFs.

Based on the image degradation model, we would like to restore the artifact-free image from the degraded image, especially the JPEG coded image, which means to generate R from D by removing the additive noise N and reversing the attenuation G .

2.2. Proposed image restoration scheme

The framework of the proposed image restoration scheme is illustrated in Fig. 1, which consists of three steps. The left is the proposed Local Filtering (LOF), which is firstly employed to remove the additive Gaussian noise. The proposed Learning-based Synthesis (LS) is followed to capture the property of the attenuation, and try to recover the lost HF component. The right is Quantization Constraint (QC), which is utilized to model the error introduced by quantization of JPEG compression and

constrain the restored image within the range of quantization errors.

In LOF step, each pixel in the degraded image is generated by Eq. (2). For one pixel D_m (m indicates the pixel location), its neighboring pixels could be denoted as: $\{D_n, n \in \Omega(m)\}$, $\Omega(m)$ is the spatial neighborhood of D_m . Therefore, we can generate the pixel value by removing the additive Gaussian noise by

$$D_m^{LOF} = \sum_{n \in \Omega(m)} \varpi(m, n) D_n, \quad (3)$$

where the family of weights $\{\varpi(m, n), n \in \Omega(m)\}$ depends on the similarity between the neighborhoods of D_m and D_n , and satisfy the usual conditions $0 \leq \varpi(m, n) \leq 1$ and $\sum_{n \in \Omega(m)} \varpi(m, n) = 1$. The Euclidean distance is employed to depict the neighborhood similarity to obtain the weights, which are defined by

$$\varpi(m, n) = (e^{-\frac{(\|v(m) - v(n)\|_2^2)}{h}}) / Z(m), \quad (4)$$

where $v(m)$ is the values of the D_m neighboring pixels, h is a parameter that acts as a degree of filtering, which is simply set as 600. $Z(m)$ is the normalizing constant

$$Z(m) = \sum_{n \in \Omega(m)} e^{-\frac{(\|v(m) - v(n)\|_2^2)}{h}}. \quad (5)$$

Each degraded image pixel D_n in Eq. (3) can be expressed by the degradation model according to Eq. (2)

$$D_m^{LOF} = \sum_{n \in \Omega(m)} \varpi(m, n) \{G_n * R_n + N_n\}. \quad (6)$$

Although the attenuation intensity varies over different image regions, it has been assumed that the neighboring attenuation RF appears very similar [18,19], which we can regard the attenuation intensities $\{G_n, n \in \Omega(m)\}$ are of the same value, denoted as G'_n . Therefore, D_m^{LOF} can be expressed as

$$D_m^{LOF} = \sum_{n \in \Omega(m)} \varpi(m, n) \{G'_n * R_n + N_n\}. \quad (7)$$

In Ref. [11], it has been claimed that under stationary assumptions, for a pixel i , the LOF algorithm can converge to the conditional expectation of i once observed a neighborhood of i . In this case, the stationary conditions

II. PROPOSED IMAGE RESTORATION FRAMEWORK

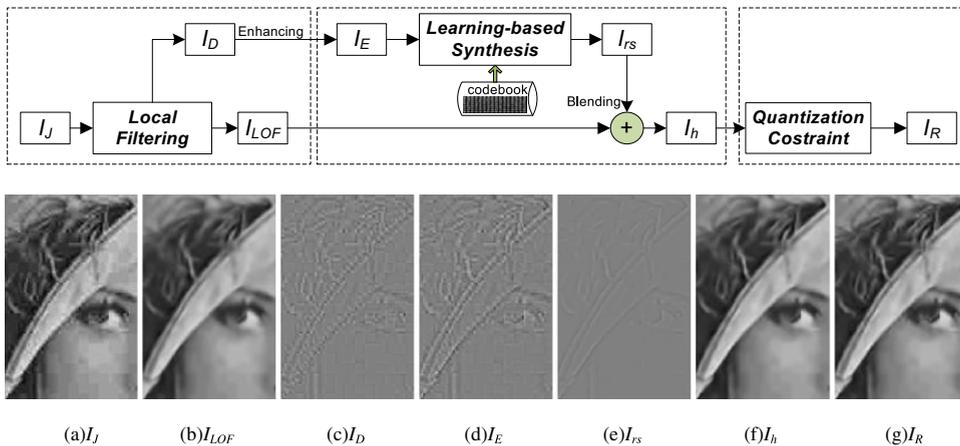


Fig. 1. The proposed image restoration framework.

amount to say that as the size of the image grows we can find many similar blocks for all the details of the image. And according to the general framework in Ref. [51], in the case that an additive white noise model is assumed, it can be shown that the conditional expectation is the function of $D(\Omega(i)\setminus\{i\})$ that minimizes the mean square error with the image. $D(\Omega(i)\setminus\{i\})$ denote the pixel values in the neighborhood of pixel i .

Let V, U, N be random fields such that $V=U+N$, where N is a signal independent white noise. Then the following statements are held:

- a) $E[V(i)|D(\Omega(i)\setminus\{i\}) = v_x] = E[U(i)|D(\Omega(i)\setminus\{i\}) = v_x]$
- b) The expected random variable $E[V(i)|D(\Omega(i)\setminus\{i\})]$ is the function of $D(\Omega(i)\setminus\{i\})$ that minimizes the mean square error: $\min_f E[U(i)-f(D(\Omega(i)\setminus\{i\}))]^2$.

Detailed information about the hypothesis of the theorem and its proof can be found in Refs. [51,11]. In our LOF step, the JPEG coded image D_n can be viewed as V . N_n has been assumed as independent and identically distributed (i.i.d) random variables. Then the original image R_n after the attenuation filter G'_n is regarded as U . With the two aforementioned statements, the proposed LOF is regarded as the approximated function of $D(\Omega(i)\setminus\{i\})$ to minimize the restoration error. Then after the LOF process, the additive Gaussian noise can be removed. D_m^{LOF} in Eq. (7) is simplified as

$$D_m^{LOF} = G'_n * \sum_{n \in \Omega(m)} \varpi(m,n) R_n. \quad (8)$$

Then the filtered pixels D_m^{LOF} compose an image I_{LOF} with Gaussian noise removed. However, n could not be set infinity. The first reason is that the image resolution is finite and the attenuation intensity varies greatly in different local regions. The most important reason is that blurring artifacts will be introduced to I_{LOF} . Therefore, in this paper, the pixels in the 5×5 window centered at pixel m are employed for LOF. By subtracting I_{LOF} from I_j , a differential image I_D is obtained, which comprises most of the additive Gaussian noise. Although LOF is constrained within a local 5×5 window, the weighting process on R_n will still lead to HF information loss. Therefore, there exists some detailed HF components in I_D , which will be helpful for reconstructing HF components in LS step.

For Eq. (8), the LOF on the original image R could be viewed as a convolution process, which could be simplified as

$$D_m^{LOF} = G'_n * \{G_{\varpi(m,n)} * R_n\}. \quad (9)$$

where $G_{\varpi(m,n)}$ is a convolution filter to approximate LOF. As to the whole image, Eq. (9) could be derived as

$$I_{LOF} = G' * \{G_{\varpi} * R\} = \tilde{G} * R, \quad (10)$$

where \tilde{G} could be viewed as the combination of the two attenuation filters G' and G_{ϖ} .

In LS step, the aim is to restore the original image R , given the attenuated image I_{LOF} as well as I_D . During the image hallucination process, the downsampled image could be obtained by blurring and decimation [36], which could also be modeled by Eq. (10). Inspired by recent

progresses on image hallucination [33–38], the learning-based synthesis is proposed to restore the HF information truncated during quantization of JPEG compression, with the assistance of learnt codebook and the differential image I_D . In order to strengthen the mapping relationship between the degraded image patches and the original ones, an enhanced differential image $I_E = I_D + I_P$ is obtained by blending one predictive HF component I_P with I_D to preserve the local consistency of the contours. Then a Markov-chain based inference is employed to replace the image patch in I_E with the existent image patch in the learnt codebook. The synthesized image I_{rs} can be inferred by maximizing posterior probability $pr(H|I_E)$, with $pr(H)$ as the priors

$$I_{rs} = \operatorname{argmax}(pr(H|I_E)). \quad (11)$$

Detailed information about how to learn the priors $pr(H)$ and how to synthesize the image I_{rs} will be presented in Section. 3.

By blending I_{rs} with I_{LOF} , a high-quality image I_h is generated. However, similar with learning-based image hallucination, irregularities will be introduced to I_h [38], because of the weak mapping relationship between low-dimensional degraded primitive and the high-dimensional original one. In order to overcome the problem, during the QC step, the property of quantization error is described and DCT coefficients of the constructed image are regularized within a reasonable range.

During JPEG compression process, coefficients are quantized after BDCT, which introduces the blocking and ringing artifacts. Let B_R denotes DCT coefficients of the original 8×8 image block, B_D is the corresponding coefficients after quantization, Q is the quantization matrix, and B_{rs} is the restored coefficients after LS step. Therefore, the quantization error QE_B of the block could be obtained by

$$\begin{aligned} QE_B(i,j) &= B_R(i,j) - B_D(i,j) \\ &= B_R(i,j) - \operatorname{round}\left(\frac{B_R(i,j)}{Q(i,j)}\right) \times Q(i,j), \end{aligned} \quad (12)$$

where (i,j) denotes the DCT subband, and round operator converts its parameter into an integer. Eq. (12) implies that the true value of the original image block lies in the well-defined range, namely the quantization constraint set (QCS) $S_{QC}(i,j)$ [23–25]

$$S_{QC}(i,j) = \{\hat{B}_R(i,j) : |\hat{B}_R(i,j) - B_D(i,j)| \leq Q(i,j)/2\}. \quad (13)$$

The upper and lower bounds of $S_{QC}(i,j)$ for the compressed image block can be expressed as

$$\begin{cases} \hat{B}_R^{up}(i,j) = B_D(i,j) + \xi \cdot Q(i,j)/2 \\ \hat{B}_R^{low}(i,j) = B_D(i,j) - \xi \cdot Q(i,j)/2 \end{cases} \quad (14)$$

Normally, ξ is set equal to 1. However, as the narrow property of the QCS has been researched in Ref. [25], $\xi=0.6$ is recommended by Refs. [12,13] for improving the performance.

In order to remove the irregularities introduced and restore the true original image block, the restored block B_{rs} in I_{rs} after LS should satisfy: $B_{rs}(i,j) \in S_{QC}(i,j)$. Then the resulting restored image I_r could be generated by projecting

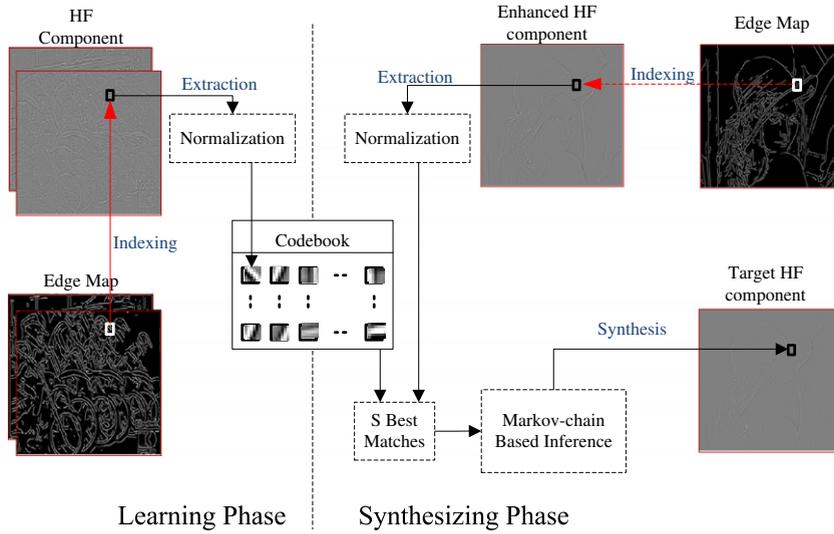


Fig. 2. Learning and synthesizing strategy.

B_{rs} onto the pre-defined QCS according to

$$B_{rs}(i,j) = \begin{cases} \hat{B}_R^{up}(i,j), & \text{if } B_{rs}(i,j) > \hat{B}_R^{up}(i,j) \\ \hat{B}_R^{low}(i,j), & \text{if } B_{rs}(i,j) < \hat{B}_R^{low}(i,j) \\ B_{rs}(i,j), & \text{otherwise} \end{cases} \quad (15)$$

3. Learning and synthesizing strategy

What we learned as the priors for image restoration are the image primitives (image patches lying on image contours [33]). They are the elements which describe the intensity variations in images and their local geometries in Marr's vision [39]. The flowchart of learning and synthesizing strategy is illustrated in Fig. 2. In learning phase, the primitives extracted from the HF component of the image are normalized and trained to build the codebook. In the synthesizing phase, the degraded image primitives are firstly enhanced to strengthen the mapping relationship. Then, S best matched primitives are retrieved from the codebook for each enhanced one. Finally, the Markov-chain based inference is employed to synthesize the target HF components.

3.1. Learning phase

From Mumford's pattern theory [40], the extracted image primitive P can be derived from the latent pattern L with several global geometric and photometric transformations, i.e. scaling, orientation, translation and illumination, which can be formulated as

$$P = cD_t D_s D_o L + d, \quad (16)$$

where c denotes the contrast, d is the bias for illumination, and D_s , D_o and D_t are scaling, orientation and translation transformation matrices, respectively. And the latent pattern L consists of the local transformation information of the primitive, such as the curvature and the intensity variations.



Fig. 3. Examples of the clustered primitives of size 7×7 .

Therefore, with the assumption that the relationship between primitives is independent of contrast and bias transformation, the dimensionality of primitives could be reduced. And as each primitive extracted lies on the contour, the translation matrix D_t can be approximated by the identity matrix. With the independent ingredients removed, the extracted primitive P can be normalized as P'

$$P' = \frac{1}{c} D_t^{-1} (P - d) = D_s D_o L. \quad (17)$$

Therefore, what we need to learn is the latent pattern L with scaling and orientation transformations. Compared with non-primitive image patches, the normalization eliminates the effects of illumination and translation, which greatly reduces the dimensionality. Consequently, the normalized primitives are easier for clustering [41] and more representative for mapping. Furthermore, the enhanced LBG (ELBG) algorithm [42] is proposed for clustering the normalized primitives to discard the noisy or unimportant primitive patterns. Some examples of the clustered primitives are illustrated in Fig. 3.

In order to evaluate the representative ability of the normalized primitives, The Receiver Operating Characteristics (ROC) curve [33] is used to demonstrate the relationship between hit rate and match error. Given a match error e , the hit rate h_r denotes the percentage of the test data whose match errors are less than e . The match error $Er(s)$ of the given sample s is defined by the metric $Er(s) = \|s - s'\|_2 / \|s\|_2$ between s and the nearest sample s' in the trained codebook. Therefore, the higher the hit rate, the better the training codebook.

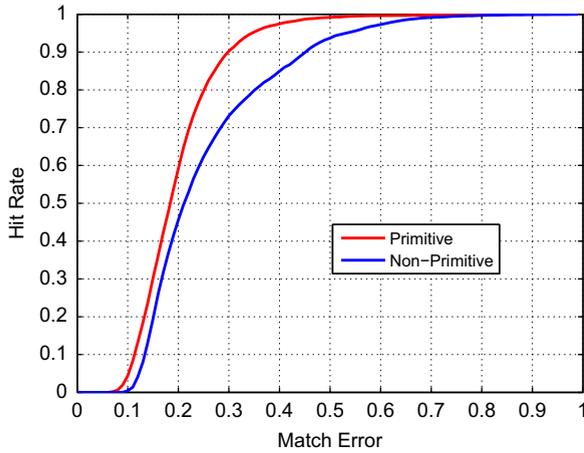


Fig. 4. ROC curves of mapping accuracies over primitives and non-primitives.

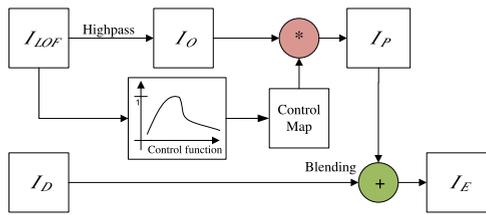


Fig. 5. The proposed enhancement scheme.

The ROC curves of primitives and non-primitives, which are selected randomly from the image, are presented in Fig. 4 based upon the mapping accuracy of 10^4 test data over the trained codebook with size of 10^5 . It can be observed that primitives lead to higher hit rate than non-primitives under the same match error. Therefore, the primitives are lower-dimensional and more representative than non-primitives, which make the mapping of primitives more reasonable and reliable.

3.2. Synthesizing phase

The mapping relationship between the degraded primitive and the original ones has been weakened, due to the compression. Therefore, the differential image I_D is firstly enhanced to obtain the HF enhanced component in Fig. 2. Then after normalization, the Markov-chain model is utilized for synthesizing the target HF component based on the retrieved S best matched primitives.

3.2.1. Differential image enhancement

Fig. 5 illustrates the proposed enhancement algorithm to enforce the local consistency along the contours of the differential image I_D . Similar to Refs. [43,45], control function [44] is employed to depict the local activities of the image I_{LOF} and generate the predictive HF component I_P . Consequently, the enhanced HF image I_E is obtained by

$$I_E = I_D + I_P = I_D + CI_O, \quad (18)$$

where C is the control map which is obtained by performing the control function C_f on I_{LOF} , I_O is the HF component

of I_{LOF} . Control function C_f is defined according to

$$C_f(x,y) = \frac{LA(x,y) + a}{k(LA(x,y))^2 + l}$$

$$a = \frac{C_0 LA_0}{2(1-C_0)}, \quad l = \frac{LA_0}{2(1-C_0)}, \quad k = \frac{1}{2LA_0}, \quad (19)$$

where (x,y) indicates the pixel position of the image, $LA(x,y)$ is the local activity descriptor, with LA_0 and C_0 denoting the two parameters to adjust the control function. With careful definitions of the two parameters, only some selected detailed information which relates to the contour regions is enhanced, whereas the other regions remain unaffected. By combining I_P together with I_D , enhanced HF image I_E is constructed with local compatibilities between primitives enforced, which greatly resembles the HF information of the original image. As to the local activity $LA(x,y)$, the magnitude of Sobel operator is employed to reflect the intensity variance over local regions.

Furthermore, average root mean square (ARMS) error is employed to demonstrate the efficiency of the proposed enhancement scheme. Given an underlying image primitive in enhanced HF image I_E or non-enhanced HF image I_D , ARMS error between the given primitive and the best matched one which is retrieved from the codebook is calculated. Some typical images coded by JPEG with different compression ratios (in terms of bpp) are utilized to show the ARMS errors. As depicted in Table 1, with the proposed enhancement, HF image primitives of lower average ARMS errors in the same codebook could be retrieved, in comparison with the ones inferred from the non-enhancing HF image. Therefore, the proposed enhancement method could help reconstruct more faithful HF component.

3.2.2. Markov-chain based inference

Based on the enhanced HF component, the Markov-chain based algorithm is employed to synthesize the target HF component, which considers the mapping accuracy γ from the enhanced image primitive to the existent one in the learnt codebook, and the neighbor compatibility φ between the neighboring primitives in the synthesized image I_{rs} .

Based upon the learnt codebook, I_{rs} could be inferred by a linear sum of a number of N learned primitives $\{P_n, n=1, \dots, N\}$. The underlying enhanced differential

Table 1
ARMS error comparison of typical images for enhancement.

Image	Number of image patches	Enhancement ARMS	Non-enhancing ARMS
LENA (0.307 bpp)	17,254	7.19	7.58
BOAT (0.460 bpp)	22,716	8.52	8.95
PEPPERS (0.309 bpp)	14,764	7.24	7.70
BABOON (0.342 bpp)	38,783	8.77	9.09
BARBARA (0.444 bpp)	27,907	10.36	11.01

Table 2
PSNR comparison of different image restoration schemes.

Image	LENA			PEPPERS			BARBARA			BABOON							
	0.189	0.244	0.307	0.363	0.248	0.309	0.365	0.465	0.228	0.338	0.444	0.537	0.256	0.342	0.457	0.620	
Methods																	
JPEG	28.24	30.41	31.95	32.96	30.18	31.58	32.47	33.61	23.86	25.70	27.05	28.25	21.52	22.47	23.43	24.51	
Xiong [4]	29.24	31.05	32.30	33.03	30.45	31.45	32.05	32.76	23.99	25.15	25.80	26.25	21.86	22.59	23.25	23.96	
Yang [24]	28.99	31.03	32.42	33.35	30.65	31.93	32.75	33.80	24.25	26.10	27.42	28.60	21.86	22.76	23.69	24.71	
Chen [2]	29.14	31.16	32.40	33.28	30.88	31.91	32.70	33.87	24.52	26.11	27.12	28.07	21.97	22.90	23.67	24.49	
Zhai [12]	29.23	31.25	32.68	33.57	30.76	32.05	32.85	33.86	24.43	26.15	27.51	28.69	21.98	22.88	23.76	24.79	
Liew [5]	29.55	31.44	32.73	33.55	31.16	32.29	33.03	33.99	24.50	26.05	27.38	28.56	21.99	22.80	23.65	24.68	
Proposed	29.60	31.66	33.07	33.94	31.42	32.65	33.43	34.36	24.78	26.62	28.15	29.42	22.04	22.95	23.89	24.88	

primitives in I_E are $\{P_n^E, n=1, \dots, N\}$. In order to maintain the consistency along the contour, primitives are first grouped into M contours $Ec = \{Ec_m, m=1, \dots, M\}$. The joint posterior $pr(H|I_E)$ in Eq. (11) can be approximated by the results of each contour

$$pr(H|I_E) = pr(Ec|I_E) \approx \prod_m pr(Ec_m|I_E). \quad (20)$$

The synthesis process of each contour Ec_m can be viewed as a first-order Markov-chain model

$$pr(Ec_m|I_E) \propto \prod_i^{n_m} \gamma(P_i, P_i^E) \prod_i^{n_m-1} \varphi(P_i, P_{i+1}), \quad (21)$$

where P_i^E is the i th enhanced differential primitive on the contour Ec_m , P_i is the primitive to be inferred from the learnt codebook and n_m is the total number of primitives on Ec_m .

Given an enhanced differential primitive P_i^E , normalization is first performed to obtain the normalized version P_i^E and its contrast information c_i^E . Then S best match normalized primitives $\{P_i(s), s=1, \dots, S\}$ together with their contrast information $\{c_i(s), s=1, \dots, S\}$ are retrieved from the codebook. The scaling factors $c_i^E/c_i(s)$ compensate the contrast differences between P_i^E and $P_i(s)$. Therefore, the S best match primitives with contrast correction are $\{P_i(S) = (c_i^E/c_i(s)) \cdot P_i(s), s=1, \dots, S\}$. With the S best matched primitives, the Markov-chain based inference will reconstruct the expected HF component.

In the proposed scheme, the mapping accuracy function γ is defined by the Euclidean distance, whereas neighbor compatibility function φ is defined by

$$\varphi(P_i, P_{i+1}) = \exp(-d(P_i, P_{i+1})/\sigma_d^2), \quad (22)$$

where $d(P_i, P_{i+1})$ is the Sum Squared Difference (SSD) of the overlapped region between P_i and P_{i+1} , σ_d is a tuning variance. The optimal MAP solution of Eq. (22) can be obtained by performing the Belief Propagation (BP) [46] algorithm. The fact that BP converged to a solution of the Markov-chain very quickly led us to believe that simpler operations may suffice. Freeman et al. [30] proposed a one-pass algorithm, in which only the compatibilities for neighboring primitives are computed, which are already selected, typically from above and to the left, in raster-scan order processing. Therefore, with the properly pre-structure the trained codebook, we only need to search the nearest neighbor codeword to the given input primitives in the learned codebook.

4. Experimental results

The proposed scheme is tested on various images compressed by the JPEG standard [48]. Several typical images, namely, LENA, PEPPERS, BARBARA and BABOON (512×512) are utilized for testing. The images can be categorized into two groups. LENA and PEPPERS, which are smooth images, mostly concentrate on LF DCT coefficients, whereas BARBARA and BABOON contain more detailed information such as texture, which results in high percentage of HF DCT coefficients.

The learnt codebook is built by extracting the primitives from a training set of 24 Kodak images [47]. As we



Fig. 6. Subjective quality comparison on LENA at 0.244 bpp. Top row from left to right: part of original LENA; JPEG-coded LENA (30.41 dB); Xiong et al.'s result (31.05 dB); Yang et al.'s result (31.02 dB). Bottom row from left to right: Chen et al.'s result (31.16 dB); Zhai et al.'s result (31.25 dB); Liew et al. result (31.44 dB); the proposed result (31.66 dB).

only pay attention to the luminance component of the image, which is more sensitive to human vision system than the chromatic components, the primitives are extracted from the luminance component of the natural image. Each primitive is represented by a 7×7 image patch with the central pixel on the contour. And after the ELBG clustering process, the total number of the learnt primitives in the codebook is about 60,000.

The objective quality of the restoration image is evaluated by the Peak Signal-to-Noise Ratio (PSNR). The higher the PSNR is, the smaller the difference between the restored image and the original image. We have compared the proposed scheme with several other well known image deblocking methods. Chen et al.'s algorithm [2] is based on the idea of modifying DCT coefficients to reduce blocking artifacts. Xiong et al.'s [4] and Liew et al.'s [5] methods are based on OWE and modification on wavelet coefficients. Yang et al. [24] utilize POCS algorithm (with 5 iterations in the test). Zhai et al. [12] employs PSW for deblocking. The PSNR performance comparison is depicted in Table 2. The highest two PSNR results for each column are emphasized by italicizing the values. We can observe that the proposed scheme performs the best among all the tabulated methods. Greater improvements can be obtained for both smooth and texture images. For LENA compressed at $\text{bpp}=0.244$, and BARBARA coded at $\text{bpp}=0.537$, about 1.2 dB gain is obtained compared with the JPEG coded images.

Parts of the restored LENA image (0.244 bpp) obtained from different methods are illustrated in Fig. 6. The deblocked result of Xiong et al.'s is a little blurry, while the blocking artifacts could still be perceived in the results of Yang et al.'s, Xiong et al.'s and Zhai et al.'s results. For the deblocked results from Chen et al.'s and Liew et al.'s method, although the blocking artifacts have been alleviated, the ringing artifacts still exist, especially the region of LENA's hat brims. This is the visual results of the Gibbs effect, which is caused by the loss of HF information. However in our restored image, the HF information truncated during quantization is recovered with the learnt codebook. Therefore, the visual ringing artifacts are removed and the blocking artifacts are alleviated. Furthermore, Fig. 7 has demonstrated the efficiencies of the different steps in the proposed method. The first column is JPEG coded image, the second column is the process image after LOF, the third column is the processed image after LS, while the last column is the processed image after QC regularization. It can be observed that each process can help to improve the image quality in terms of PSNR. Meanwhile the proposed scheme can generate images with high visual qualities. For the QC regularization step, images with different contents present different improvements of image qualities. For example, QC regularization improved PSNR by 0.6 dB for the BARBARA at 0.537 bpp, while only 0.05 and 0.2 dB improvement is obtained for the PEPPERS at

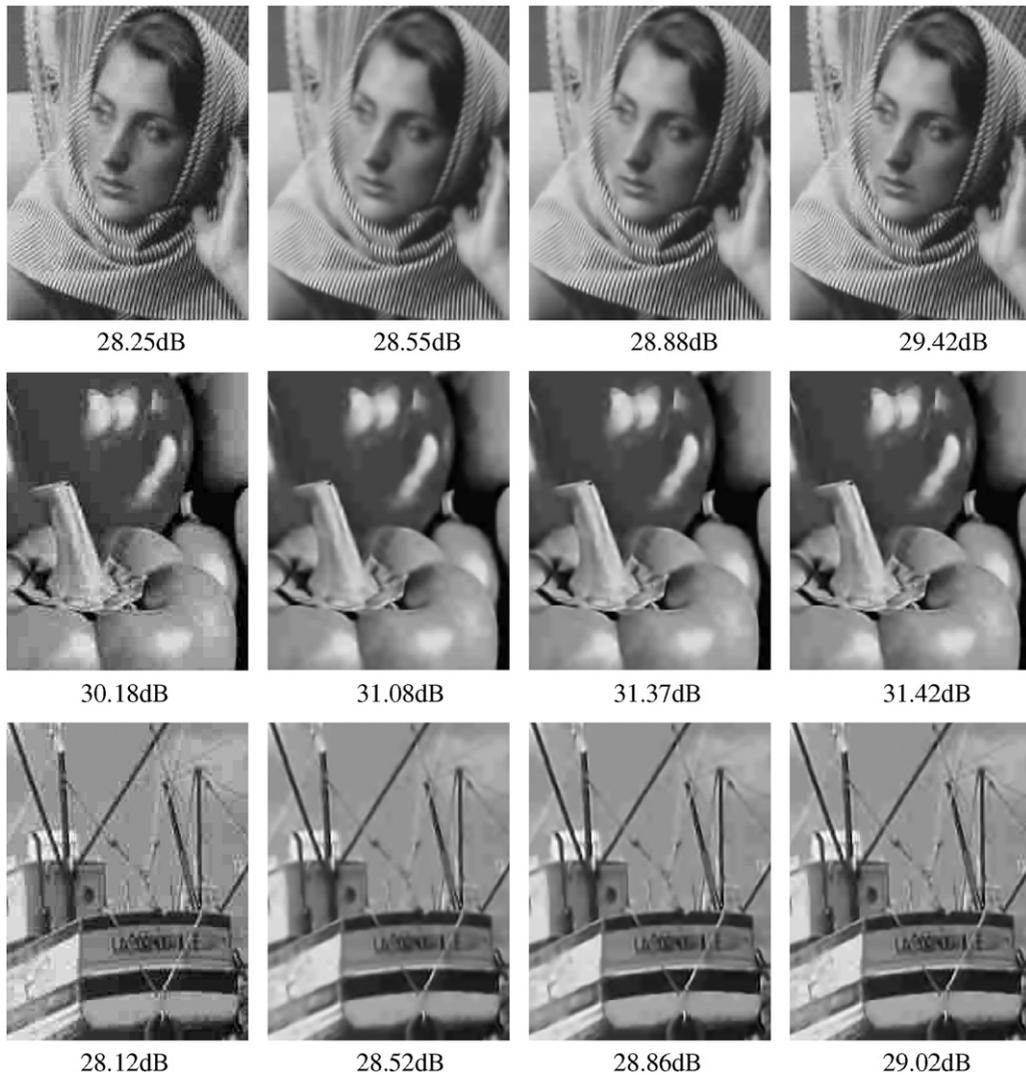


Fig. 7. Restored images by the proposed scheme. Top row: BARBARA at 0.537 bpp; middle row: PEPPERS at 0.248 bpp; bottom row: BOAT at 0.291 bpp. The first column is the JPEG coded image; the second column is the processed image after LOF; the third column is the processed image after LS; and the last column is the resulting restored image after QC regularization.

0.248 bpp and the BOAT at 0.291 bpp, respectively. The reason may be that BARBARA contains more HF components compared with BOAT and PEPPERS. And the QC regularization can help to reconstruct more faithful HF components after the projection.

Furthermore, we have provided the restoration results of the images/frames downloaded from the Internet, which are real webcam quality images. The experimental results are demonstrated in Fig. 8. For the first two rows, the images are coded by JPEG, which are processed by the LOF, the LS, and the QC step. For the following four rows, the video frames of the video clips are downloaded from the YOUTUBE [52]. As we cannot easily obtain the information about the video codec and quantization parameter, only the first two steps, namely the LOF and LS, are employed for restoring the video frames. Since the original images of the real webcam quality ones are not available, the corresponding PSNR values cannot be

provided. However, it can be observed that the visual quality can be significantly improved, even with only the first two steps of the proposed method. In the future, the authors will focus on the integration of the QC step into the video restoration scheme. Detailed information about the experimental results can be referred to Fig. 8.

There are several aspects for further investigation. Recently, classified vector quantization (CVQ) has been introduced into learning-based image restoration [28]. With proposed CVQ, the codebooks are divided into sub-classes, according to the illumination, type and orientation. Better performance can be expected when the pre-learned codebooks are carefully designed and structured considering the local activities, i.e. edge directions. Moreover, recent progresses in sparse representation of images have been achieved, particularly in image super-resolution [49]. If we approach the restoration problem from the perspective of compressed sensing [50], sparse representations can be



Fig. 8. Restored images/frames by the proposed scheme. Left column is the coded image/frame; right column is the restored image/frame by the proposed method. The top two rows are the images downloaded from the Internet; the bottom four rows are the frames of the video clips, which are downloaded from the YOUTUBE [52].

correctly recovered from the degraded signal under mild conditions. Therefore, higher-quality restored images could be generated because of the precisely description of the discarded HF component.

5. Conclusions

In this paper, a novel learning-based image restoration method for compressed images is proposed. Image primitives are extracted to construct the codebook for restoration, due to its high representative and significant sensitivity to HVS. Based upon our approach, blocking artifacts are alleviated and ringing artifacts are removed, meanwhile the HF components are recovered. The experimental results have demonstrated that our proposed scheme outperforms the other deblocking methods in terms of subjective and objective quality. With further incorporating with CVQ or new sparse representation methods, greater performances are expected, which can be utilized to suppress the compression artifacts in still images and video sequences.

Acknowledgement

This work was supported by the National Science Foundation of China: 60736043 and the Major State Basic Research Development Program of China (973 Program 2009CB320905). The authors are grateful to Dr. Feng Wu, Yonghua Zhang and Zhiwei Xiong for their valuable discussions. Thanks to Mr. Renqi Zhang for sharing the codes for the comparison.

References

- [1] B. Zeng, Reduction of blocking effect in DCT-coded images using zero-masking techniques, *Signal Process.* 79 (no. 2) (Dec. 1999) 205–211.
- [2] T. Chen, H.R. Wu, B. Qiu, Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts, *IEEE Trans. Circuits Syst. Video Technol.* 11 (5) (May. 2001) 594–602.
- [3] S. Liu, A.C. Bovik, Efficient DCT-domain blind measurement and reduction of blocking artifacts, *IEEE Trans. Circuits Syst. Video Technol.* 12 (12) (Dec. 2002) 1139–1149.
- [4] Z. Xiong, M.T. Orchard, Y.Q. Zhang, A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations, *IEEE Trans. Circuits Syst. Video Technol.* 7 (2) (Apr. 1997) 433–437.
- [5] A.W.C. Liew, H. Yan, Blocking artifacts suppression in block-coded images using overcomplete wavelet representation, *IEEE Trans. Circuits Syst. Video Technol.* 14 (4) (Apr. 2004) 433–437.
- [6] N.C. Kim, I.H. Jang, D.H. Kim, W.H. Hong, Reduction of blocking artifact in block-coded images using wavelet transform, *IEEE Trans. Circuits Syst. Video Technol.* 8 (3) (Jun. 1998) 253–257.
- [7] J.G. Apostolopoulos, N.S. Jayant, Postprocessing for very low bit-rate video compression, *IEEE Trans. Image Process.* 8 (8) (Aug. 1999) 1125–1129.
- [8] S.D. Kim, J. Yi, M. Kim, J.B. Ra, A deblocking filter with two separate modes in block-based video coding, *IEEE Trans. Circuits Syst. Video Technol.* 9 (1) (Feb. 1999) 156–160.
- [9] H.W. Park, Y.L. Lee, A postprocessing method for reducing quantization effects in low bit-rate moving picture coding, *IEEE Trans. Circuits Syst. Video Technol.* 9 (1) (Feb. 1999) 161–171.
- [10] A.S. Al-Fohoum, A.M. Reza, Combined edge crispness and statistical differencing for deblocking JPEG compressed images, *IEEE Trans. Image Process.* 10 (9) (Sep. 2001) 1288–1298.
- [11] A. Buades, B. Coll, J.M. Morel, A non-local algorithm for image denoising, in: *Proceedings of the Computer Vision Pattern Recognition*, vol. 2, Jun. 2005, pp. 60–65.
- [12] G. Zhai, W. Zhang, X. Yang, W. Lin, Y. Wu, Efficient image deblocking based on postfiltering in shifted windows, *IEEE Trans. Circuits Syst. Video Technol.* 18 (1) (Jan. 2008) 122–126.
- [13] G. Zhai, W. Zhang, X. Yang, W. Lin, Y. Wu, Efficient deblocking with coefficient regularization, shape-adaptive filtering, and quantization constraint, *IEEE Trans. Multimedia* 10 (5) (Aug. 2008) 735–745.
- [14] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, Shifted window based filtering for alleviating blocking artifacts, in: *Proceedings of the IEEE Workshop on Signal Processing Systems*, Oct. 2008, pp. 289–294.
- [15] A. Zakhor, Iterative procedure for reduction of blocking effects in transform image coding, *IEEE Trans. Circuits Syst. Video Technol.* 2 (1) (Mar. 1992) 91–95.
- [16] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. PAMI* no. 6 (Nov. 1984) 721–741.
- [17] S. Osher, L.I. Rudin, E. Fatemi., Nonlinear total variation based noise removal algorithms, *Physica D* 27 (60) (1992) 259–268.
- [18] H.R. Sheikh, A.C. Bovik, G. Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, *IEEE Trans. Image Process.* 14 (12) (Dec. 2005) 2117–2128.
- [19] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (Feb. 2006) 430–444.
- [20] Y. Yang, N.P. Galatsanos, Removal of compression artifacts using projections onto convex sets and line process modeling, *IEEE Trans. Image Process.* 6 (10) (Oct. 1997) 1345–1357.
- [21] T.P. O'Rourke, R.L. Stevenson, Improved image decompression for reduced transform coding artifacts, *IEEE Trans. Circuits Syst. Video Technol.* 5 (6) (Mar. 1995) 490–499.
- [22] J.J. Zou, H. Yan, A deblocking method for BDCT compressed images based on adaptive projections, *IEEE Trans. Circuits Syst. Video Technol.* 15 (3) (Mar. 2005) 430–435.
- [23] Y. Yang, N.P. Galatsanos, A.K. Katsaggelos, Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images, *IEEE Trans. Circuits Syst. Video Technol.* 3 (6) (Dec. 1993) 421–432.
- [24] Y. Yang, N.P. Nikolas, P. Galatsanos, Projection-based spatially adaptive reconstruction of block-transform compressed images, *IEEE Trans. Image Process.* 4 (7) (Jul. 1995) 896–908.
- [25] S.H. Park, D.S. Kim, Theory of projection onto the narrow quantization constraint set and its application, *IEEE Trans. Image Process.* 8 (10) (Oct. 1999) 1361–1373.
- [26] D.G. Sheppard, A. Bilgin, M.S. Nadar, B.R. Hunt, M.W. Marcellin, A vector quantizer for image restoration, *IEEE Trans. Image Process.* 7 (1) (Jan. 1998) 119–124.
- [27] R. Nakagaki, A.K. Katsaggelos, A VQ-based blind image restoration algorithm, *IEEE Trans. Image Process.* 12 (9) (Sep. 2003) 1044–1053.
- [28] Y. Liaw, W. Lo, J.Z. Lai, Image restoration of compressed image using classified vector quantization, *Pattern Recognition* 35 (2002) 329–340.
- [29] W.T. Freeman, E. Pasztor, O. Caemichael, Learning low-level vision, *Int. J. Comput. Vis.* (Nov. 2000) 25–47.
- [30] W.T. Freeman, T.T. Jones, E.C. Pasztor, Example-based super-resolution, *IEEE Comput. Graph Appl.* 22 (Mar. 2002) 56–65.
- [31] C. Liu, H.Y. Shum, C. Zhang, A two-step approach to hallucinating faces: global parametric model and local non-parametric model, in: *Proceedings of the Computer Vision Pattern Recognition*, vol. 1, 2001, pp. 192–198.
- [32] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embeddings, *Science* 290 (5500) (Dec. 2000) 2323–2326.
- [33] J. Sun, N.N. Zheng, H. Tao, H.Y. Shum, Image hallucination with primitive sketch priors, in: *Proceedings of the Computer Vision Pattern Recognition*, Jun. 2003.
- [34] H. Chang, D.Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: *Proceedings of the Computer Vision Pattern Recognition*, vol. 1, Jul. 2004, pp. 275–282.
- [35] W. Fan, D.Y. Yeung, Image hallucination using neighbor embedding over visual primitive manifolds, in: *Proceedings of the Computer Vision Pattern Recognition*, Jun. 2007, pp. 1–7.
- [36] Z. Xiong, X. Sun, F. Wu, Image hallucination with feature enhancement, in: *Proceedings of the Computer Vision Pattern Recognition*, 2009, pp. 2074–2081.
- [37] L. Ma, Y. Zhang, Y. Lu, F. Wu, D. Zhao, Three-tiered network model for image hallucination, in: *Proceedings of the International Conference on Image Processing*, Oct. 2008, pp. 357–360.
- [38] L. Ma, F. Wu, D. Zhao, W. Gao, S. Ma, Learning-Based Image Restoration for Compressed Image through Neighboring Embedding, Springer-Heidelberg, Berlin, Dec. 2008, pp. 279–286.
- [39] D. Marr, Vision: a computational investigation into the human representation and processing of visual information, 1982.

- [40] D. Mumford, Pattern theory: a unifying perspective, *Perception as Bayesian Inference* (1996) 25–62.
- [41] F. Wu, X. Sun, Image compression by visual pattern vector quantization (VPVQ), *Data Compression Conf.* (2008) 123–131.
- [42] G. Patane, M. Russo, The enhanced LBG algorithm, *Neural Networks* 14 (Nov. 2001) 1219–1237.
- [43] H. Greenspan, C.H. Anderson, S. Akber, Image enhancement by nonlinear extrapolation in frequency space, *IEEE Trans. Image Process.* 9 (6) (Jun. 2000) 1035–1048.
- [44] B.W. Jeon, R.H. Park, S. Yang, Resolution enhancement by prediction of the high-frequency image based on the Laplacian pyramid, *EURASIP J. Appl. Signal Process.* 2006 (2006).
- [45] G. Tamponi, A. Polesel, A rational unsharp masking technique, *J. Electron. Imaging* 7 (Apr. 1998) 333–338.
- [46] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann Publication, Sep. 1988.
- [47] [Online]. Available: <<http://r0k.us/graphics/kodak>>.
- [48] W.B. Pennebaker, J.L. Mitchell, *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993.
- [49] J. Yang, J. Wright, Y. Ma, T. Huang, Image super-resolution as sparse representation of raw image patches, in: *Proceedings of the Computer Vision Pattern Recognition*, 2008, pp. 1–8.
- [50] E.J. Candes, Compressive sampling, in: *Proceedings of the International Congress of Mathematicians*, 2006.
- [51] G.G. Roussas, Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl.* 36 (1990) 107–116.
- [52] [Online]. Available: <<http://www.youtube.com>>.