Image Primitive Coding and Visual Quality Assessment

Jian Zhang[1,*], Siwei Ma[2], Ruiqin Xiong[2], Debin Zhao[1], and Wen Gao[2]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R. China
{jzhangcs,dbzhao}@hit.edu.cn
[2] National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, P.R. China
{swma,rqxiong,wgao}@pku.edu.cn

Abstract. In this work, we introduce a new content-adaptive compression scheme, called image primitive coding, which exploits the input image for training a dictionary. The atoms composed of the learned dictionary are named as image primitives. The coding performance between the learned image primitives and the traditional DCT basis is compared, and demonstrates the potential of image primitive coding. Furthermore, a novel concept, entropy of primitives (EoP), is proposed for measuring image visual information. Some very interesting results about EoP are achieved and analyzed, which can be further studied for visual quality assessment.

Keywords: image coding, image primitive, visual information, visual quality assessment (VQA).

1 Introduction

With the accelerative growth of high-performance computers and electronic equipments, a great progress of image coding standards have been achieved. One of the important compression techniques is known as transform coding, which decomposes the image over a dictionary and provides compact image representation to obtain compression.

As we know, transform-based coding techniques generally make an assumption that the dictionary is fixed and is built in both the encoder and decoder. For example, in the JPEG [1] and JPEG2000 [2] compression standards, the dictionary considered is the DCT or wavelet, respectively.

The Sparseland model is an emerging and powerful method to describe signals based on the sparsity and redundancy of their representations [3] [4]. Obtaining an overcomplete dictionary from a set of signals allows us to represent them as a sparse linear combination of dictionary atoms. Pursuit algorithms are then used for signal decomposition.

---

* This work was supported in part by Major State Basic Research Development Program of China (973 Program, 2009CB320903), National Science Foundation (60833013, 61121002) and National High-tech R&D Program of China (863 Program, SS2012AA010805).

W. Lin et al. (Eds.): PCM 2012, LNCS 7674, pp. 674–685, 2012.

As a matter of fact, the predetermined fixed dictionaries are targeted at general-purpose image compression. For a specific application, by utilizing content-specific dictionaries optimized for a specific class of images, compression schemes have been demonstrated to acquire substantial gains over fixed dictionaries. For instance, in [5], the authors propose an algorithm for facial image compression by exploiting a sparse approximation of the image patches over a set of pre-trained dictionaries. The task-aware compression method is shown to achieve a dramatic improvement over JPEG2000 for facial imagery. Recently, an algorithm based on iteration-tuned dictionaries (ITDs) for a specific class of images has also been proposed to encode the input image patches in [6], which is able to outperform JPEG and JPEG2000 convincingly for facial images.

It is obvious to see that the main drawback of the task-specific approaches is their loss of generality, only restricting them to encoding a specific class of images for which a suitable dictionary has been pre-learned. In this work, we introduce a new content-adaptive compression scheme, called image primitive coding, which exploits the input image for training a dictionary. The atoms composed of the learned dictionary are named as image primitives. We compare the coding performance between the learned image primitives and the traditional DCT basis, which demonstrate the potential of image primitive coding. Furthermore, a novel concept, entropy of primitives (EoP), is proposed for measuring image visual information. Some very interesting results about EoP are achieved and analyzed.

This paper is organized as follows: In section 2 we introduce the scheme of image primitive coding and sparse coding, and provide the performance between the learned image primitives and the traditional DCT basis in section 3. Based on image primitives, a novel concept, entropy of primitives (EoP), is proposed in section 4. Section 5 shows some very interesting results about EoP. We conclude and discuss some future directions in section 6.

2 Image Primitive Coding

In this section, we will introduce image primitive coding, showing how to achieve image primitives and how to utilize image primitives for coding. Then, a performance comparison of image representation between image primitive and DCT basis is conducted. Some conclusions are also drawn.

2.1 Image Primitive

The scheme of image primitive coding is established on the Sparseland model, which assumes that natural signals, such as images, admit a sparse decomposition over a redundant dictionary. More specifically, given a signal $\mathbf{x} \in \mathbb{R}^n$, this model suggests the existence of a specific dictionary (i.e., a matrix) $_S$ which contains *k* prototype signals, also referred to as atoms. The model assumes that for $\mathbf{x}$, there exists a sparse linear combination of atoms from $\mathbf{D}$ that approximates it well. Put more formally, for $\forall \mathbf{x} \in \mathbb{R}^n$, $\exists \boldsymbol{a} \in \mathbb{R}^k$ such that $\mathbf{x} \approx \mathbf{D}\boldsymbol{a}$ and $\|\boldsymbol{a}\|_0 \ll n$. The notion $\|\bullet\|_0$ is ℓ_0

norm, which counts the number of nonzero elements in a vector. We typically assume $k>n$, implying that the dictionary $\mathbf{D}$ is redundant to $\mathbf{x}$.

An overcomplete dictionary that leads to sparse representations can either be chosen as a pre-specified set of functions or designed by adapting its content to fit a given set of signal examples. In this work, we design an adaptive overcomplete dictionary for input image.

For an input image $\mathbf{X}$, the dictionary learning process starts by partitioning the image into many overlapped patches, which are denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i$, $i = 1, 2, \ldots, \mathrm{N}$. These patches are then collected as training samples. Assuming a local Sparse-Land model on image patches, the K-SVD dictionary training algorithm [7] is applied to the set of patches $\{\mathbf{x}_i\}$, generating a content adaptive dictionary $\mathbf{D}$:

$$\mathbf{D}, \{\boldsymbol{a}_i\} = \underset{\mathbf{D}, \{\boldsymbol{a}_i\}}{\operatorname{argmin}} \sum_k \left\| \mathbf{x}_i - \mathbf{D}\,\boldsymbol{a}_i \right\|_2^2 \quad \text{s.t.} \quad \left\| \boldsymbol{a}_i \right\|_0 < L \quad \forall i. \tag{1}$$

where $\{\boldsymbol{a}_i\}$ are the sparse representation vectors for $\{\mathbf{x}_i\}$. In this paper, the atoms of dictionary $\mathbf{D}$ are named as image primitives. Fig. 1 gives an example of overcomplete DCT basis and a learned dictionary which is trained by 8×8 patches from Image *Lena*.

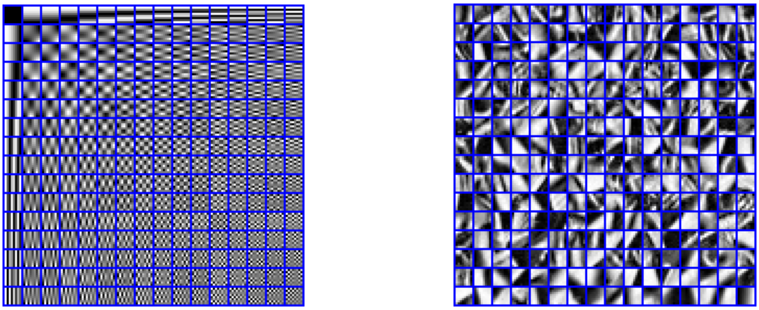

Fig. 1. Left: Overcomplete DCT base dictionary; Right: Dictionary trained over image patches

2.2 Sparse Coding

For a patch $\mathbf{x}_i$, the process of finding its sparse representation vector $\boldsymbol{a}_i$ with respect to a known overcomplete dictionary $\mathbf{D}$ is called sparse coding. As can be seen, owing to the overcompleteness, the null space of $\mathbf{D}$ introduces additional degrees of freedom in the choice of $\boldsymbol{a}_i$, which can be exploited to improve its compressibility. To obtain the sparse representation, sparse coding can be formulated as

$$\boldsymbol{a}_i = \underset{\boldsymbol{a}_i}{\operatorname{argmin}} \left\| \mathbf{x}_i - \mathbf{D}\,\boldsymbol{a}_i \right\|_2^2 \quad \text{s.t.} \quad \left\| \boldsymbol{a}_i \right\|_0 < L \ . \tag{2}$$

Though Problem (2) is NP-hard in general, it can be approximated by a wide range of techniques [4]. In this paper, we adopt orthogonal matching pursuit (OMP) [8] algorithm to solve (2) for its simplicity and efficiency.

3 Comparison between Image Primitive and DCT Basis

In order to prove the validity of image primitive coding scheme, this section gives the coding performance comparison between the image primitive and traditional DCT basis. The comparative setting is as follows. First, the trained dictionary composed of image primitives is produced by the previously mentioned algorithm to the coding image. The size of the trained overcomplete dictionary is set to 256, while the number of traditional DCT basis for 8×8 patches is 64. Then, split the coding image into some non-overlapped patches, and the size of image patch is set to 8×8. Next, for each patch, carry out the process of sparse coding with the trained overcomplete dictionary and traditional DCT basis by OMP. The number of image primitives to represent each patch (denoted by l) is fixed each time with the range from 1 to 10. Finally, the PSNR and SSIM [9] comparison curves with regarding to three test images (shown in Fig. 2) are presented by Figs. 3–5, and the visual reconstruction results of Image *Lena* by the trained overcomplete dictionary and traditional DCT are given in Figs. 6–7.

Fig. 2. Test Images. Left to right: *Lena* (512×512), *Airplane* (512×768), *Peppers* (512×512)

It is clear to see that the reconstructed image quality becomes better and better for the two types of dictionaries, as the number of image primitives to represent each patch l increases. Seen from Figs. 3–5, the values of PSNR and SSIM achieved by image primitives are both higher than those by the traditional DCT basis, especially in the case of low bit rate, i. e., when l is small. From Figs. 6–7, it is obvious that the visual quality of the reconstructed image by image primitives is much better than that by DCT basis with the same value of l. For instance, when $l = 3$, Fig. 6(c) has evident block artifacts, while the block artifacts in Fig. 7(c) are almost invisible.

One important observation from Fig. 7 is when $l > 4$, the reconstructed image is very close to the original image in visual perception. For example, one cannot distinguish between the original Image *Lena* and Fig. 7(f) visually. That means, as $l > 6$, it will add little visual information for human visual system.

Although the results of image primitive coding are encouraging, one key problem is that it requires transmitting the image primitive dictionary along with the compressed data. Thus, mining the trained dictionary structure and compressing the trained dictionary efficiently are very significant, which are also the directions of our future work. Fortunately, recent studies have shown that training sparse dictionary with Sparse K-SVD is possible, where each atom of dictionary is a sparse combination of atoms from pre-specified base, such as DCT or wavelet [10]. This technique allows relatively low-cost transmission of image primitives, thus greatly reducing the number of coding bits.

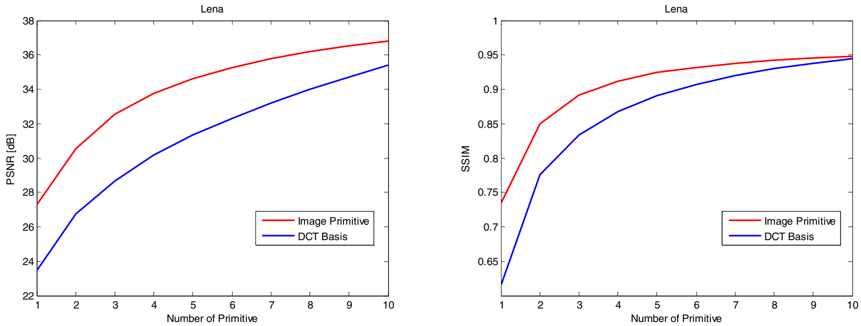

Fig. 3. PSNR and SSIM comparison curves for Image *Lena* with regarding to image primives and DCT basis

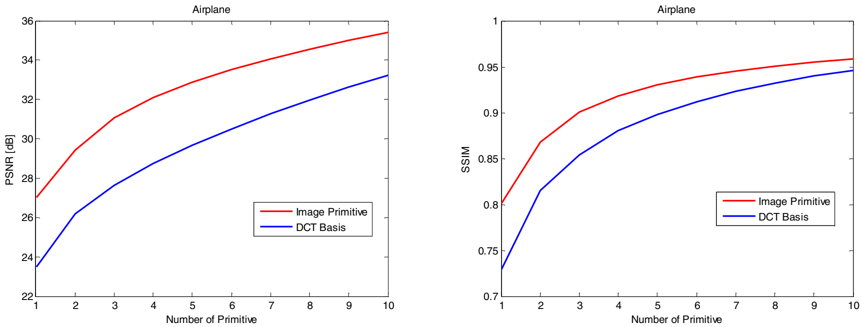

Fig. 4. PSNR and SSIM comparison curves for Image *Airplane* with regarding to image primives and DCT basis

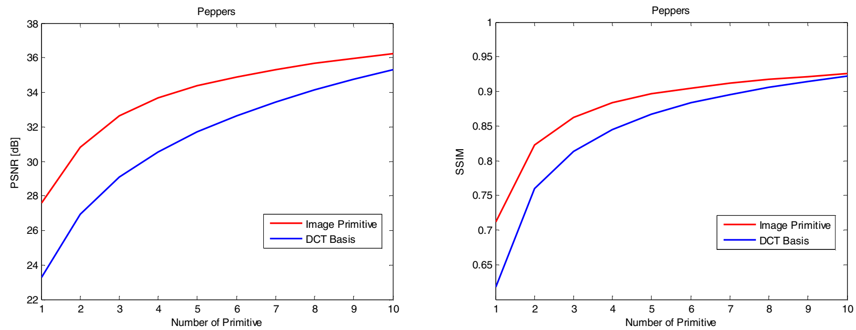

Fig. 5. PSNR and SSIM comparison curves for Image *Peppers* with regarding to image primives and DCT basis

(a) $l = 1$ (b) $l = 2$

(c) $l = 3$ (d) $l = 4$

(e) $l = 5$ (f) $l = 6$

Fig. 6. Visual reconstruction results of Image *Lena* by traditional DCT basis when l equals from 1 to 6. Here, l denotes the number of image primitives to represent each patch.

(a) $l = 1$ (b) $l = 2$

(c) $l = 3$ (d) $l = 4$

(e) $l = 5$ (f) $l = 6$

Fig. 7. Visual reconstruction results of Image *Lena* by its trained image primitives when l equals from 1 to 6. Here, l denotes the number of image primitives to represent each patch.

4 Entropy of Primitive and Visual Quality Assessment

According to the above scheme of image primitive coding, in this section, a novel concept, namely, Entropy of Primitives (EoP) is put forward to measure the amount of image visual information, and some very interesting results about EoP are also provided. Our motivation is as follows. On one hand, a set of image primitives can be learned from an image. On the other hand, the image primitives can also be utilized to approximate the original image. It can be observed that an image with its image primitives have a good corresponding relationship. Therefore, we can measure the amount of visual information of images by the amount of information taken by image primitives. Here, we use the concept of entropy in Shannon theory to describe the amount of information. The details are provided below.

Take Image *Lena* (512×512), for example. There are four steps to calculate its EoP.

Step 1, image primitives are generated by the previously mentioned method. Here, the size of image patch is set to 8×8, and the number of image primitives is set to 256.

Step 2, divide Image *Lena* into non-overlapped image patches, with the patch number equaling $512 \times 512 / 64 = 4096$.

Step 3, for each patch, conduct the process of sparse coding using the trained image primitives, while the number of image primitives (denoted by l) to represent each patch is fixed, e. g., $l = 4$. Thus, the total number of image primitives used for describing the whole image is $tatal = 4096 \times 4 = 16384$. Further, the number of every image primitive used for sparse coding can be calculated, denoted by $num_i, i = 1,2,...,256$. Therefore, the probability of each primitive can be expressed as $p_i = num_i / tatal$.

Step 4, according to Shannon Theory, the entropy of primitives (EoP) for Image *Lena* is written as $\mathrm{EoP} = -\sum_i p_i \log(p_i)$. For instance, when $l = 4$, $\mathrm{EoP} = 4.3183$.

5 Comparison of PSNR, SSIM and EoP

With the concept of EoP, what can we achieve? With the changes of image quality, what are the statistical laws of EoP? Can EoP be exploited to characterize the amount of visual information of an image? The followings will answer the above questions.

When l increases, the quality of reconstruction image becomes better. We can compute PSNR and SSIM for the reconstruction images. Given l, according to the procedures above, we can also compute EoP. Figs. 8–10 present the PSNR, SSIM and EoP results for three test images with respect to the number of image primitives to represent each patch, i. e., l.

We can achieve some very interesting observations from these experimental results.

a. The EoP curves are monotonically increasing with the number of image primitive l and gradually become flat.

b. When l reaches a certain point, 6 in the curve, the EoP value nearly gets to its peak and becomes stable thereafter. This shows that no more visual information could be supplied from the additional image primitive when l is larger than 6, which is in

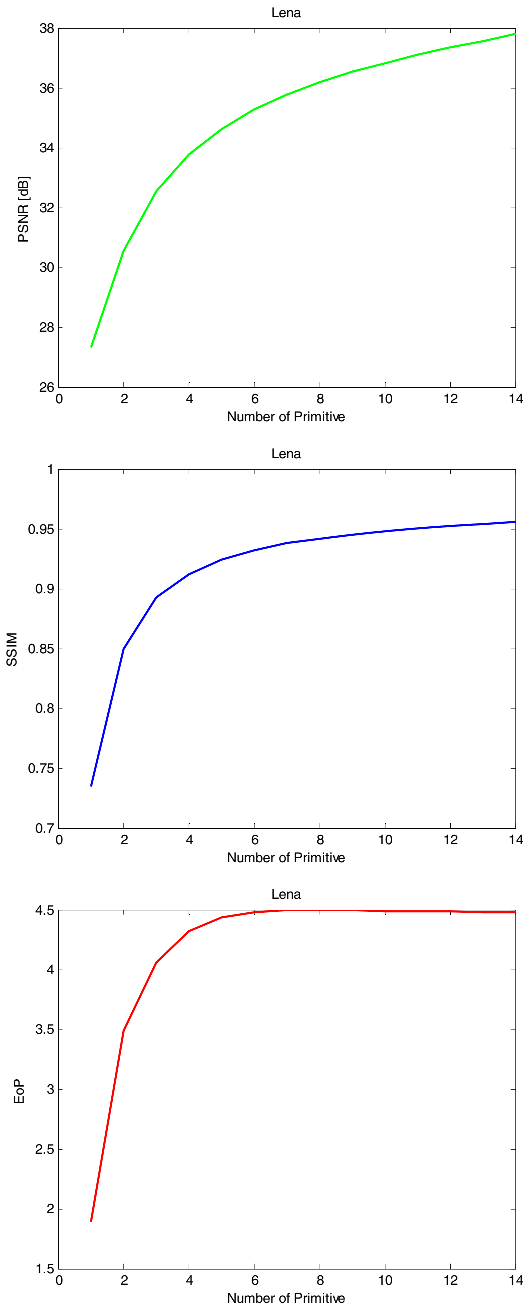

Fig. 8. PSNR, SSIM and EoP results for Image *Lena* with respect to the number of image primitives to represent each patch

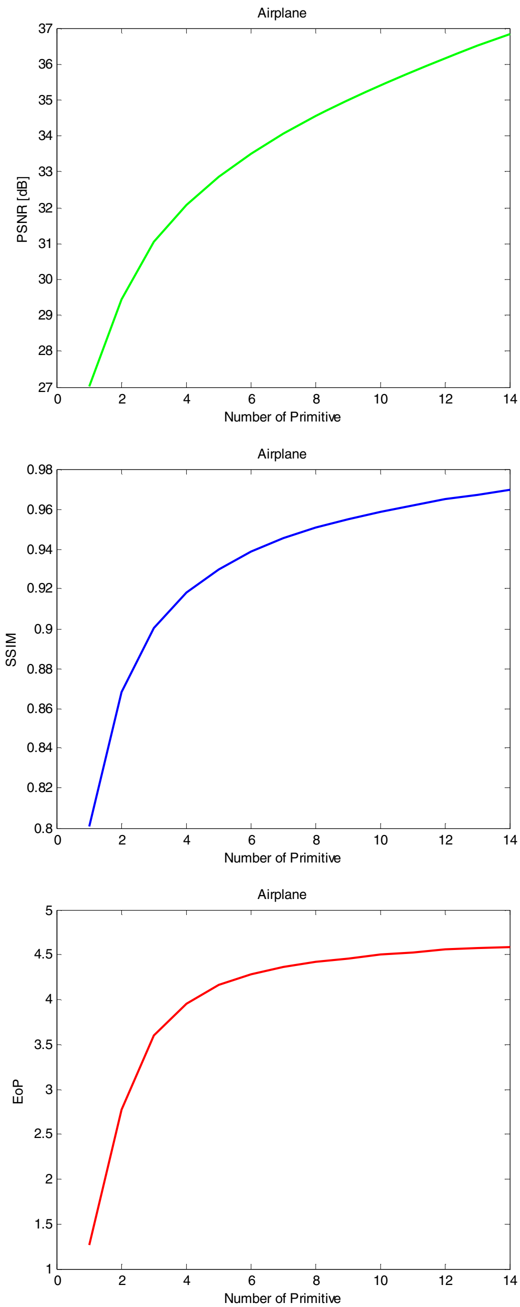

Fig. 9. PSNR, SSIM and EoP results for Image *Airplane* with respect to the number of image primitives to represent each patch

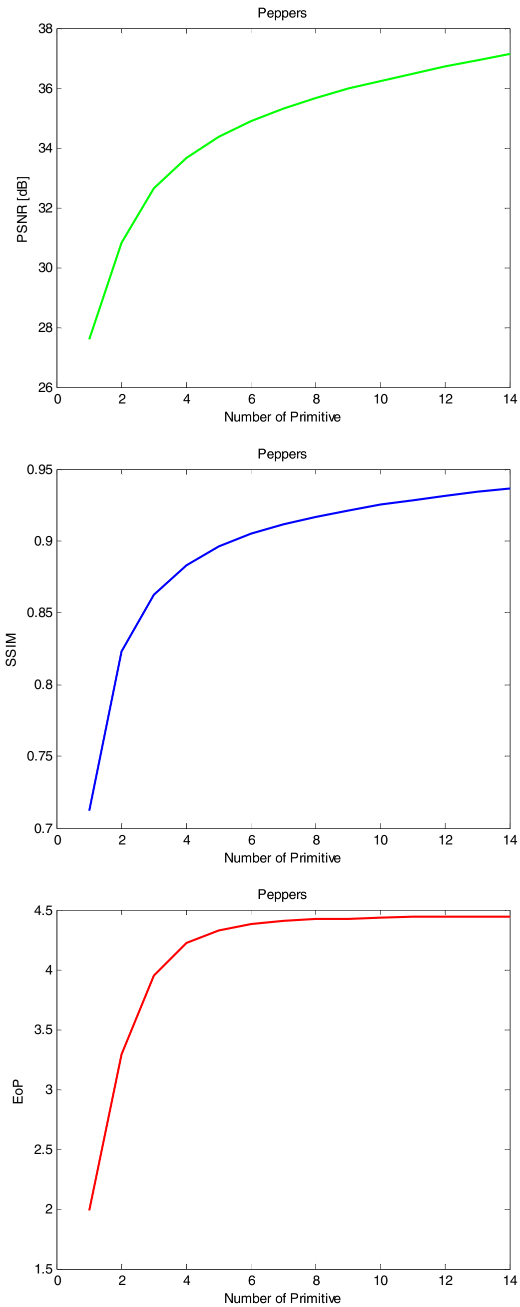

Fig. 10. PSNR, SSIM and EoP results for Image *Peppers* with respect to the number of image primitives to represent each patch

conformity with the foregoing conclusions. We illustrate this observation in Figs. 8–10. By contrast, the PSNR and SSIM curves are still increasing when $l \geq 6$.

Based on these two points, we can say that the proposed concept EoP can evaluate the visual information of images to some extent, which can be further exploited as a criterion for visual quality assessment.

6 Conclusions

In this work, a new content-adaptive compression scheme, called image primitive coding, is introduced and demonstrates the good potentials over the traditional DCT coding. Moreover, a novel concept based on image primitives, namely, entropy of primitives (EoP), is proposed for measuring image visual information. Some very interesting results about EoP are achieved and analyzed, which has a close relationship with visual quality assessment. Future work includes two aspects. For one thing, try to effectively compress image primitives by taking advantage of dictionary structure; for another, it is also very interesting to design a new algorithm for visual quality assessment with the study on EoP.

References

1. Pennebaker, W.B., Mitchell, J.L.: JPEG still image data compression standard. Springer, New York (1993)
2. Taubman, D.S., Marcellin, M.W.: JPEG2000: Image compression fundamentals, standards and practice. Kluwer Academic Publishers, Norwell (2001)
3. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. on Image Processing 15(12), 3736–3745 (2006)
4. Elad, M.: Sparse and redundant representations–From theory to applications in signal and image processing. Springer (2010)
5. Elad, M., Bryt, O.: Compression of facial images using the K-SVD algorithm. Journal of Visual Communication and Image Representation 19(4), 270–283 (2008)
6. Zepeda, J., Guillemot, C., Kijak, E.: Image Compression using the Iteration-Tuned and Aligned Dictionary. In: 36th IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 793–796. IEEE Press (2011)
7. Aharon, M., Elad, M., Bruckstein, A.M.: The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation. IEEE Trans. on Signal Processing 54(11), 4311–4322 (2006)
8. Tropp, J.A., Gilber, A.A.: Signal Recovery from Random Measurements via Orthogonal Matching Pursuit. IEEE Trans. on Information Theory 53(12), 4655–4666 (2007)
9. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. on Image Processing 13(4), 600–612 (2004)
10. Rubinstein, R., Zibulevsky, M., Elad, M.: Double sparsity: learning sparse dictionaries for sparse signal approximation. IEEE Trans. on Signal Processing 58(3), 1553–1564 (2010)