Dictionary Based JPEG Artefact Removal

Petri Luukkonen and Anna Laakkonen
7.5.2015

INTRODUCTION

The Joint Photographic Experts Group (JPEG) [1], [2] method have been over two decades one of the most popular lossy compression for digital images. Because JPEG is lossy, compression introduces high frequency quantization error separately to each individual block, resulting discontinuity across block boundaries. Here is introduced a dictionary based method for JPEG artefact removal. For chosen test image have been applied heavy JPEG compression for demonstrating the effect of the artefact removal. A dictionary formed from a set of training images was used as a *a priori* knowledge for restoring the compressed image.

METHODS AND MATERIALS

JPEG compression

Let us have a JPEG compressed with $N \times M$ pixels shown in figure 1. The given image is then divided into 8×8 patches as shown in figure 1. In JPEG compression each 8×8 of the given image is *compressed* using the so called 2d discrete cosine transform. The visual information of the image is expressed in terms of the 2d cosine functions.

Let $x \in \mathbb{R}^{64\times1}$ be a single 8×8 image patch taken from the given image in a vectorized form. Let us also denote by $m \in \mathbb{R}^{64\times1}$ the data relating to a JPEG compression of the single image patch x.

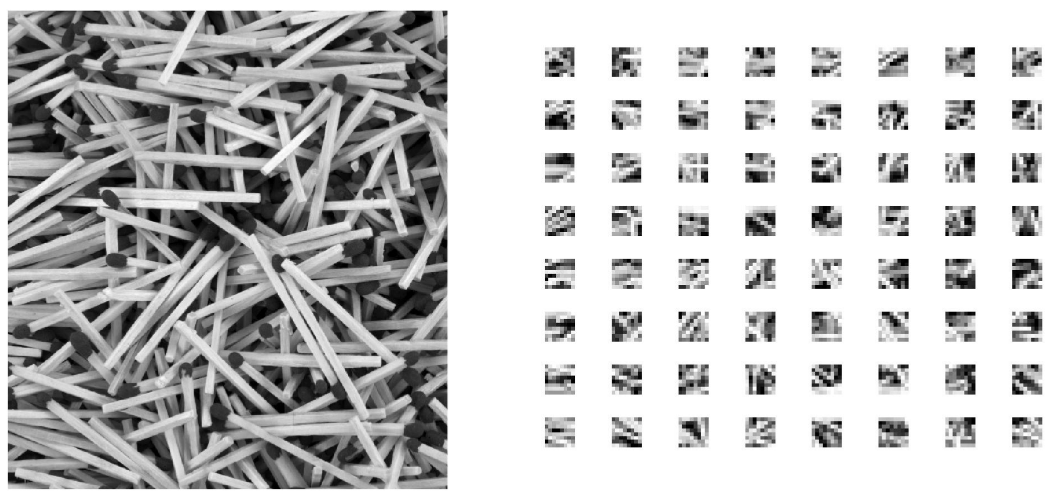

Figure 1: The original test image, V before JPEG compression and test image divided into 8×8 patches.

The JPEG compression of the original 8×8 patch x is formed as follows:

$$m = Jx = QCx, \qquad (1)$$

where matrix $J \in \mathrm{R}^{64\times64}$ is the JPEG compression operator. The operator J in itself is a factor of a two operators, called the quantization operator Q and the 2d discrete cosine transform (2d-DCT) operator C. The vector $m \in \mathrm{R}^{64\times1}$ is the 2d-DCT of x.

The aim is thus to reconstruct x while knowing m.

Dictionary

We make an assumption that our unknown is formed using the so called dictionary matrix $D \in \mathbb{R}^{64\times K}$ as follows

$$x \approx D\alpha, \qquad (2)$$

where $\alpha \in \mathbb{R}^{K\times1}$ gives the coefficients telling how much each column of D is used in x. The columns of D are the so called *atoms* that are interpreted to serve as common features for some chosen set of the so called training images. The training images were chosen to be visually similar to the test image, i.e. randomly oriented matches, see Figure 2. The dictionary D containing the atoms of the set of training images was formed by using K-SVD.

The problem of learning dictionary D can be formulated as minimizing the distance between N image patches $\{p_i\}_{i=1}^N$ of the training image and their corresponding representations $\{D\gamma_i\}_{i=1}^N$ with sparsity constraint:

$$\underset{D,\gamma}{arg\ min} \sum_{i=1}^{N} \|p_i - D\gamma_i\|_2^2 \quad \text{subject to} \quad \|\gamma_i\|_0 \le k_0, \quad 1 \le i \le N, \qquad (3)$$

where the number k_0 is the maximum number of atoms allowed to use per approximation of each p_i.

By denoting $P = [p_1\ p_2\ \cdots\ p_K]$, the data matrix having each 8×8 patch vectorized into its columns, the minimization in (3) becomes

$$\underset{D,A}{arg\ min} \|P - DG\|_2^2 \quad \text{subject to} \quad \|col_i(G)\|_0 \le k_0, \quad 1 \le i \le N, \qquad (4)$$

where ith column of G is the coefficient vector γ_i used to approximate the ith image patch p_i.

In practice the minimization of the distance $\|P - DG\|_2$ is computed first with respect to the coefficient matrix G i.e. $G_{i+1} = \underset{G}{arg\ min}\|P - D_iG\|_2^2$ and then with respect to dictionary D i.e. $D_{i+1} = \underset{D}{arg\ min}\|P - DG_{i+1}\|_2^2$. The first minimization is done by using Orthogonal Matching Pursuit (OMP) and the second minimization (with respect to D) is done by using Singular Value Decomposition. More details of the algorithmic implementation of K-SVD is found in the original paper of Michal Aharon et al. [3].

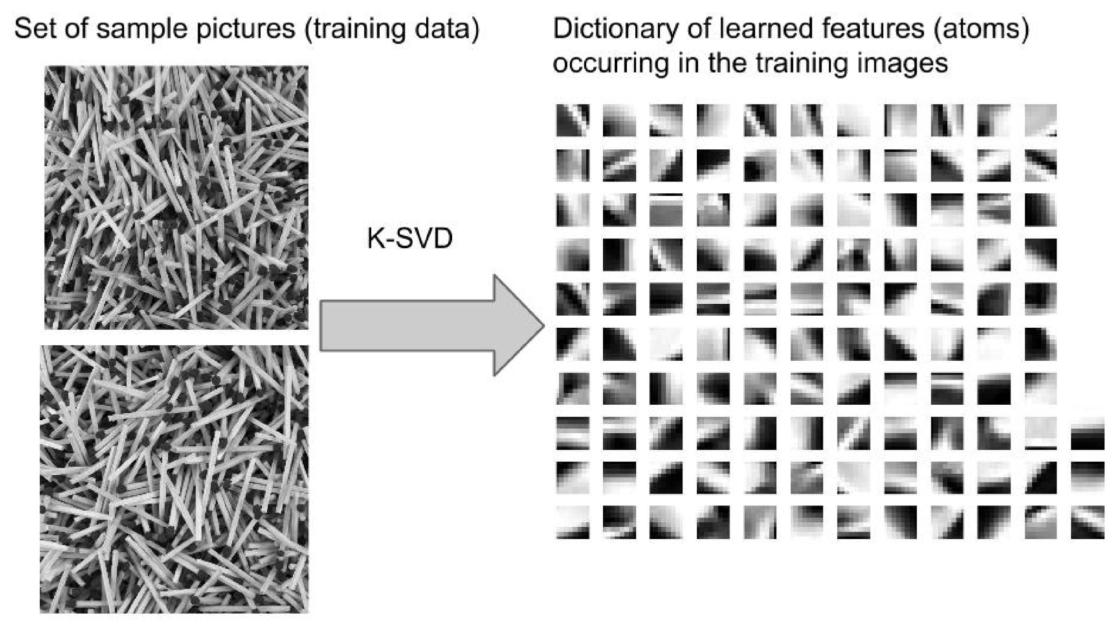

Figure 2: Example of training data set from which the dictionary is learned.

Reconstruction

The problem of reconstructing x is formulated as a minimization problem as follows:
Find such $\alpha \in \mathbb{R}^K$ that minimize the following regularizated function $F_\beta(\alpha) = \|QCD\alpha - m\|_2^2 + \beta\|\alpha\|_1$

As the quantization operation Q does not affect qualitatively in to solution of the minimization problem, one can concentrate on minimizing function

$$F_\beta(\alpha) = \|CD\alpha - m\|_2^2 + \beta\|\alpha\|_1. \qquad (5)$$

The strategy for the minimization problem is to express the objective function (5) in a quadratic form $\frac{1}{2}x^TAx + bx$, as shown in [5]. As $x \approx D\alpha$, where $\alpha \in \mathbb{R}^{K\times1}$ one can express the l_1 -norm in (5) as follows: $\|\alpha\|_1 = \sum_\nu |\alpha_\nu| = \alpha_+^T\mathbf{1} + \alpha_-^T\mathbf{1}$,
where $\alpha_+, \alpha_- \in \mathbb{R}_+^{K\times1}$. As a result, the objective function (5) in a quadratic form is

$$F_\beta(\alpha) = \frac{1}{2}y^THy + h^Ty, \qquad (6)$$

where

$$y = \begin{bmatrix} \alpha \\ \alpha_+ \\ \alpha_- \end{bmatrix}, \qquad (7)$$

$$H = \begin{bmatrix} 2(CD)^TCD & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad (8)$$

and

$$h = \begin{bmatrix} -2(CD)^Tm \\ \beta \cdot \mathbf{1}_{K\times1} \\ \beta \cdot \mathbf{1}_{K\times1} \end{bmatrix}. \qquad (9)$$

For minimizing the quadratic form of $F_\beta(\alpha)$ is used the Matlab's `quadprog.m` built in algorithm. As a result of minimization, one get α that is used for computing the reconstruction of original x.

RESULTS

The reconstruction shown in Figure 3 was computed with parameter value of $\beta = 2 \cdot 10^5$ and using 230 atoms. The effect of the size of dictionary (i.e. number of used atoms) into the quality of the reconstruction was also investigated. Structural Similarity (SSIM) Index between the original test image V and reconstruction V_{rec} with respect to dictionary size is plotted in Figure 5.

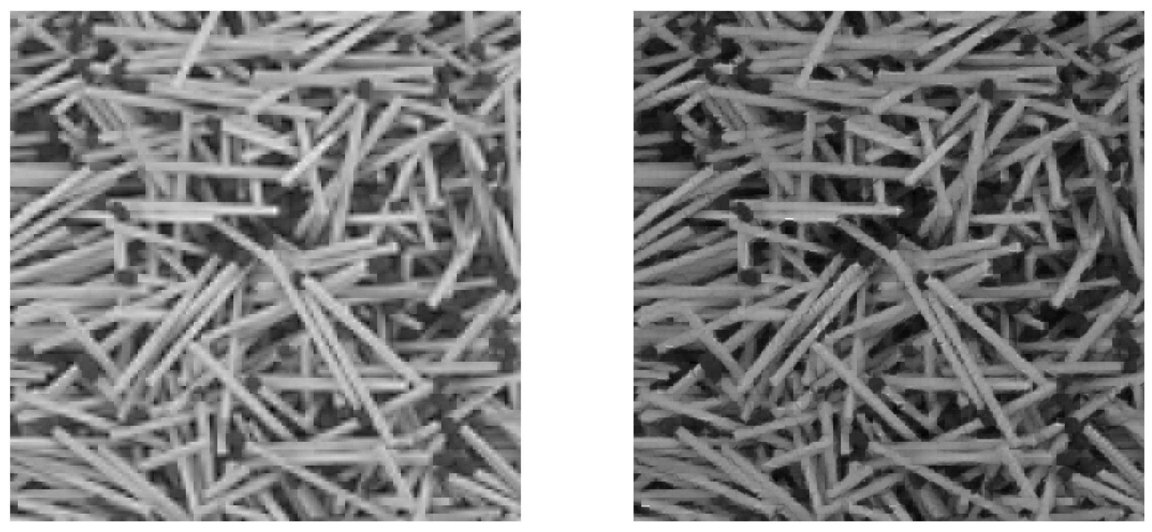

Figure 3: JPEG compression V_{comp} (left), $ssim(V, V_{comp}) = 0.74$ and Reconstruction V_{rec} (right), $ssim(V, V_{rec}) = 0.84$.

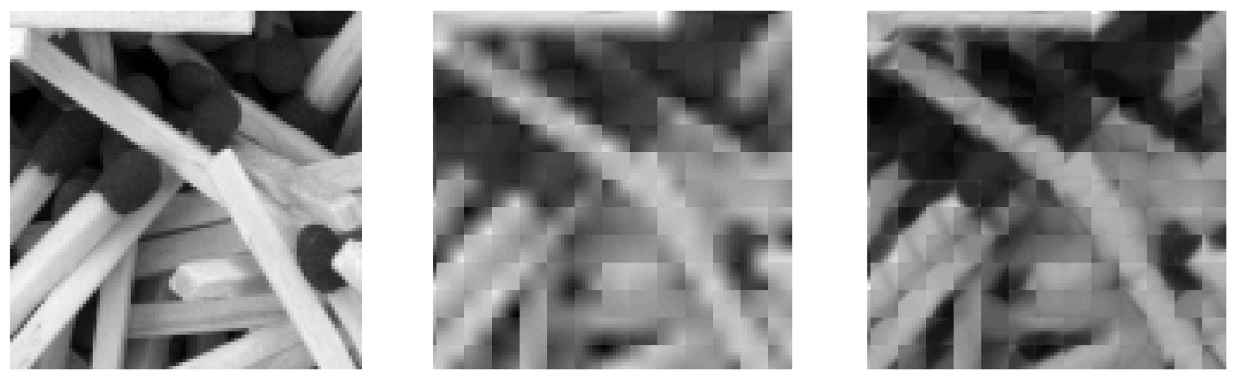

Figure 4: Closeup of the original test image (left), JPEG compression V_{comp} (middle), and Reconstruction V_{rec} (right) of Figure 3.

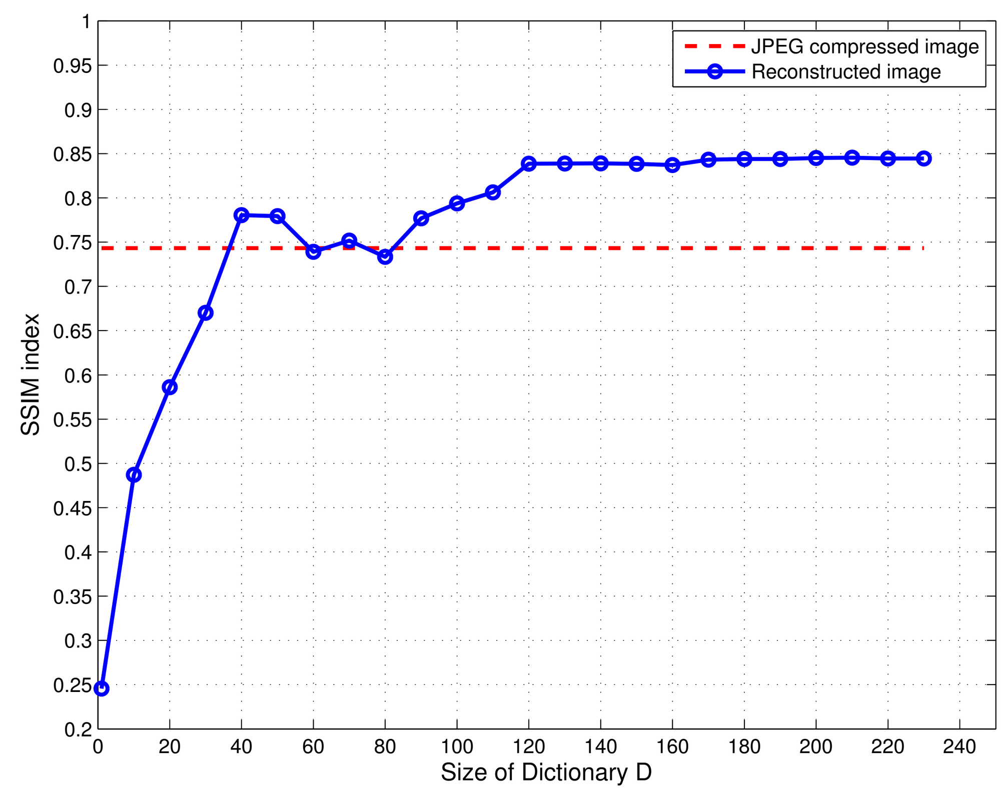

Figure 5: The Structural Similarity Index (SSIM) measure versus size of dictionary used in the reconstruction.

DISCUSSION

DCT in the JPEG compression soften the sharp edges and thus reduce the visual information. The dictionary used to form the reconstruction was found to restore the sharp edges and contrast between adjacent pixels. However the given method does not remove the block artefacts.

From Figure 5 can be seen that the visual quality of the reconstruction does not improve significantly after increasing the size of the given dictionary over 120 atoms.

For achieving optimal result, one could develop and include a penalty term into the objective function that takes into account the discontinuity between adjacent blocks in the reconstruction.

References

[1] http://en.wikipedia.org/wiki/JPEG (31.4.2015)

[2] Chang,Ng,Zeng Reducing Artifact in JPEG Decompression via a Learned Dictionary, Korea Advanced Institute of Science and Technology,National University of Singapore

[3] Aharon, Elad, Bruckstein, K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, *IEEE Transactions on Signal Processing*, Vol. 54, pp. 4311–4322, November 2006

[4] Elad, Michael, Sparse and redundant representations: from theory to applications in signal and image processing, Springer, 2010.

[5] Mueller J L, Siltanen S, *Linear and Nonlinear Inverse Problems with Practical Applications*, SIAM 2012.