

CARTOON-LIKE IMAGE RECONSTRUCTION VIA CONSTRAINED ℓ_p -MINIMIZATION

Simon Hawe, Martin Kleinsteuber, and Klaus Diepold

Department of Electrical Engineering and Information Technology,
Technische Universität München, München, Germany.
e-mail: {simon.hawe,kleinsteuber,kldi}@tum.de.

ABSTRACT

This paper considers the problem of reconstructing images from only a few measurements. A method is proposed that is based on the theory of Compressive Sensing. We introduce a new prior that combines an ℓ_p -pseudo-norm approximation of the image gradient and the bounded range of the original signal. Ultimately, this leads to a reconstruction algorithm that works particularly well for Cartoon-like images that commonly occur in medical imagery. The arising optimization task is solved by a Conjugate Gradient method that is capable of dealing with large scale problems and easily adapts to extensions of the prior. To overcome the none differentiability of the ℓ_p -pseudo-norm we employ a Huber-loss term like approximation together with a continuation of the smoothing parameter. Numerical results and a comparison with the state-of-the-art methods show the effectiveness of the proposed algorithm.

Index Terms— Compressive Sensing, ℓ_p minimization, Image Reconstruction, Conjugate Gradient Algorithm

1. INTRODUCTION AND NOTATION

In recent years, Compressive Sensing (CS) [1, 2] has evolved as one of the most active research topics in the signal processing community. Basically, CS is a joint sampling and compression mechanism, which enables perfect signal reconstruction from a very small number of non-adaptively acquired measurements. These measurements encode the entire information about the signal at hand into a very small amount of data, much smaller than the signal's dimension. More precisely, let $\mathbf{s} \in \mathbb{R}^n$ denote the signal of interest. We are computing $m \ll n$ inner products between \mathbf{s} and a set of signal independent measurement vectors $\{\phi_i\}_{i=1}^m$, which can be compactly written as

$$\mathbf{y} = [\phi_1, \dots, \phi_m]^T \mathbf{s} + \mathbf{e} =: \Phi \mathbf{s} + \mathbf{e}. \quad (1)$$

The vector $\mathbf{y} \in \mathbb{R}^m$ contains the measurements, $\mathbf{e} \in \mathbb{R}^m$ models sampling errors, and $\Phi \in \mathbb{R}^{m \times n}$ is the measurement

matrix. Clearly, without further information on the signal, the problem of inferring the signal from the measurements is ill-posed. Prior assumptions on the signal help to well-define the recovery problem: The CS-framework exploits the fact that many interesting signals have a sparse- or compressible representation with respect to some (possibly overcomplete) basis.

Let $\mathbf{x} \in \mathbb{R}^d$ with $d \geq n$ denote the k -sparse representation of the signal, where k -sparse means that only $k \ll d$ entries of \mathbf{x} are nonzero. We write the corresponding linear transformation as $\mathbf{x} = \mathcal{D}\mathbf{s}$, where $\mathcal{D} \in \mathbb{R}^{d \times n}$ is a sparsifying transformation with full rank. Furthermore, let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function that promotes or measures sparsity. We denote the Moore-Penrose pseudo-inverse of \mathcal{D} by \mathcal{D}^\dagger . The recovery of the signal thus leads to the well known *synthesis* approach, cf. [3],

$$\underset{\mathbf{x} \in \mathbb{R}^d}{\text{minimize}} \quad g(\mathbf{x}) \quad \text{subject to} \quad \|\Phi \mathcal{D}^\dagger \mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon, \quad (2)$$

where ϵ is an estimated upper bound on the noise power $\|\mathbf{e}\|_2^2$. Informally speaking, by solving (2), we find the sparsest vector $\hat{\mathbf{x}}$ that is compatible with the acquired measurements. The signal is then recovered by $\mathbf{s}^* = \mathcal{D}^\dagger \hat{\mathbf{x}}$.

Another common procedure is to directly search for a signal $\hat{\mathbf{s}}$ such that $\mathcal{D}\hat{\mathbf{s}}$ is sparse. Formally, $\hat{\mathbf{s}}$ is the solution of

$$\underset{\mathbf{s} \in \mathbb{R}^n}{\text{minimize}} \quad g(\mathcal{D}\mathbf{s}) \quad \text{subject to} \quad \|\Phi \mathbf{s} - \mathbf{y}\|_2^2 \leq \epsilon, \quad (3)$$

which is known as the *analysis* approach, see [3] for the relation between problem (2) and (3). In the field of image processing, problem (3) is favored over (2) due to the lower dimension of the search space. Moreover, for many practically important operators \mathcal{D} the computation of $\mathcal{D}^\dagger \mathbf{x}$ is infeasible. Thus, we will focus on problem (3) here. Throughout the paper, $\mathbf{v}(i)$ denotes the i^{th} entry of the vector \mathbf{v} and $\mathcal{M}(i, j)$ the (i, j) -entry of the matrix \mathcal{M} .

Clearly, with the k -sparsity assumption on \mathbf{x} , the ideal choice for g would be the ℓ_0 -pseudo-norm $\|\mathbf{v}\|_0 := \#\{i : \mathbf{v}(i) \neq 0\}$, which counts the nonzero elements of \mathbf{v} . Unfortunately, solving problem (3) with $g(\mathcal{D}\mathbf{s}) = \|\mathcal{D}\mathbf{s}\|_0$ is computationally intractable as it is combinatorial NP-hard. Instead, it has been shown in [4] that depending on Φ and \mathcal{D}

This work has partially been supported by the Cluster of Excellence *CoTeSys* - Cognition for Technical Systems, funded by the Deutsche Forschungsgemeinschaft (DFG).

the replacement of the ℓ_0 -pseudo-norm by its closest convex surrogate, the ℓ_1 -norm $\|\mathbf{v}\|_1 := \sum_i |\mathbf{v}(i)|$, gives the same solution. In the noise free case, if the number of measurements m is large enough compared to the sparsity factor k , this solution to (3) yields the exact recovery of the signal [5]. The number of measurements required for ideal reconstruction can even be further decreased, if an ℓ_p -pseudo-norm $\|\mathbf{v}\|_p^p := \sum_i |\mathbf{v}(i)|^p$ with $0 < p < 1$ is employed instead of the ℓ_1 -norm, [6]. However, the resulting optimization problem cannot be solved by straightforward linear quadratic programming, but approximations like reweighted least squares [7], or an iterative shrinkage method [8] can be applied.

In this paper we present an algorithm for reconstructing compressively sampled images via ℓ_p -minimization. We introduce an additional prior that copes with the boundedness of the range of images. This prior accelerates the optimization process and drastically improves the reconstruction results.

2. PROBLEM STATEMENT

As stated above, our goal is to reconstruct compressively sampled images $\mathcal{I} \in \mathbb{R}^{h \times w}$, where our focus is on Cartoon-like images, i.e. images that are piecewise flat.

We consider two sampling bases, namely partial discrete Fourier transform (DFT) and the Rudin-Shapiro transformation (RST) [9]. The interest for the DFT arises due to its importance in magnetic resonance imaging (MRI) where Fourier coefficients are directly sampled. The RST, also known as the real valued Dragon-Noiselet-transformation, is used as it can be implemented into a real imaging sensor, and due to its desirable properties for image reconstruction [10].

Following a common approach in image reconstruction, we impose a sparsity prior on the image gradient. The simplest way of approximating the gradient is in terms of finite differences with

$$\frac{\partial \mathcal{I}}{\partial x}(i, j) = \begin{cases} (\mathcal{I}(i, j) - \mathcal{I}(i + 1, j))/\sqrt{2} & \text{if } i < w \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

being the difference between two neighboring pixels in horizontal direction, and $\frac{\partial \mathcal{I}}{\partial y}(i, j)$ the difference between two neighboring pixels in vertical direction defined accordingly. Let $\mathbf{s} := \text{vec}(\mathcal{I}) \in \mathbb{R}^n$ with $n = hw$ be the vectorized image obtained by stacking its columns among each other. We define $\mathcal{D}_x \in \mathbb{R}^{n \times n}$ and $\mathcal{D}_y \in \mathbb{R}^{n \times n}$ as those matrices that realize the approximate image gradients, i.e.

$$\mathcal{D}_x \mathbf{s} = \text{vec} \left(\frac{\partial \mathcal{I}}{\partial x} \right), \quad \mathcal{D}_y \mathbf{s} = \text{vec} \left(\frac{\partial \mathcal{I}}{\partial y} \right). \quad (5)$$

The sparsity assumption on the image gradient yields

$$\mathcal{D} = \begin{bmatrix} \mathcal{D}_x \\ \mathcal{D}_y \end{bmatrix} \in \mathbb{R}^{2n \times n}, \quad (6)$$

as the sparsifying transformation that we consider for our problem at hand.

Regarding the choice of the sparsity measure, we compared the p -pseudo-norms $\|\mathcal{D}\mathbf{s}\|_p^p$ with

$$\|\mathcal{D}\mathbf{s}\|_{\text{TV}_p} := \sum_{i=1}^n ((\mathcal{D}\mathbf{s})(i))^2 + (\mathcal{D}\mathbf{s})(i+n)^2)^{\frac{p}{2}}, \quad (7)$$

which for $p = 1$ are the well known anisotropic, and isotropic *Total Variation* pseudo-norms, respectively. While the p -pseudo-norm enforces sparsity of the gradient in x and y direction separately, the latter (7) enforces the *magnitude* of the gradient to be sparse. This has the nice effect that (7) is invariant under rotations of the underlying picture, consequently, we employ it for further studies here.

Finally, we exploit the fact that the pixel intensities of images are bounded. If we denote the lower bound by τ_l and the upper bound by τ_u , this leads to the constraint optimization problem

$$\begin{aligned} & \underset{\mathbf{s} \in \mathbb{R}^n}{\text{minimize}} && \|\mathcal{D}\mathbf{s}\|_{\text{TV}_p} \\ & \text{subject to} && \|\Phi\mathbf{s} - \mathbf{y}\|_2^2 \leq \epsilon, \tau_l \leq \mathbf{s}(i) \leq \tau_u. \end{aligned} \quad (8)$$

For $\tau_l = 0$ and $\tau_u = +\infty$ this is the well known positivity constraint. Here, as we are dealing with images, we typically chose $\tau_l = 0$ and $\tau_u = 1$ or $\tau_u = 255$ depending on the image format. Our experiments provide evidence that these additional constraints drastically improve the reconstruction quality and accelerate the optimization process.

3. RECONSTRUCTION ALGORITHM

The reconstruction algorithm proposed here is based on a Conjugate Gradient (CG) method that minimizes a smooth approximation of the optimization problem (8). It allows to easily integrate additional priors into the minimization process without severe changes. To enhance legibility, we stick to the matrix-vector notation. However, note that all matrix-vector-products are efficiently implemented via filtering techniques in at least $O(n \log n)$ flops.

The approximation is carried out in three ways. First, the bounding constraint is enforced by the functional $\tau(\mathbf{s}) := \sum_{i=1}^n \tau(\mathbf{s}(i))$ where τ , is a penalty term defined as

$$\tau(x) = \begin{cases} |x - \tau_u|^q & \text{if } x \geq \tau_u \\ |x - \tau_l|^q & \text{if } x \leq \tau_l \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

with $q > 1$. The larger q , the higher the penalty and the tighter the bound. For the presented image reconstruction algorithm $q = 2$ yields very good fit to the range.

Second, since the CG algorithm requires a differentiable cost function, we employ a smooth approximation of (7) that extends the well known Huber loss term to the ℓ_p -pseudo-norm. We define the *p-Huber loss* as

$$h_{p,\mu}(x) = \begin{cases} |x|^p - \kappa_1 & \text{if } |x| \geq \mu \\ \kappa_2 x^2 & \text{otherwise,} \end{cases} \quad (10)$$

with $\kappa_1 = (1 - \frac{p}{2})\mu^p$ and $\kappa_2 = \frac{p}{2}\mu^{p-2}$. Note, that $h_{p,\mu}(x)$ is differentiable for all $p > 0$ and $\mu > 0$. Combining (7) and (10) leads to the following approximation of the sparsity measure for the image gradient

$$\begin{aligned} \|\mathcal{D}\mathbf{s}\|_{\text{TV}_p} &\approx \sum_{i=1}^n h_{\mu,p} \left(\sqrt{(\mathcal{D}\mathbf{s})(i)^2 + (\mathcal{D}\mathbf{s})(i+n)^2} \right) \\ &=: \|\mathcal{D}\mathbf{s}\|_{\text{TV}_{p,\mu}}. \end{aligned} \quad (11)$$

Ultimately, problem (8) is approximated in unconstrained Lagrangian form as

$$\underset{\mathbf{s} \in \mathbb{R}^n}{\text{minimize}} \quad f(\mathbf{s}) = \frac{1}{2} \|\Phi\mathbf{s} - \mathbf{y}\|_2^2 + \tau(\mathbf{s}) + \lambda \|\mathcal{D}\mathbf{s}\|_{\text{TV}_{p,\mu}}. \quad (12)$$

The Lagrange multiplier $\lambda \in \mathbb{R}_0^+$ weighs between the sparsity of the solution and its fidelity to the acquired samples according to $\lambda \sim \epsilon$.

The CG method for minimizing (12) is initiated with $\mathbf{s}_0 = \Phi^\top \mathbf{y}$ and iteratively updates the current solution by

$$\mathbf{s}_{i+1} = \mathbf{s}_i + \alpha_i \mathbf{h}_i. \quad (13)$$

The scalar $\alpha_i \geq 0$ is the line-search parameter or the stepsize, and \mathbf{h}_i is the descent direction at the i^{th} iteration. Various line-search techniques for finding α_i that approximately solve

$$\underset{\alpha \in \mathbb{R}^+}{\text{minimize}} \quad f(\mathbf{s}_i + \alpha \mathbf{h}_i) \quad (14)$$

exist from which we choose *backtracking line-search* [11] as it is conceptually simple and computationally cheap.

Let $\mathbf{g}_i := \nabla f(\mathbf{s}_i)$ be the gradient of the cost function (12) at the i^{th} iteration and let the descent direction be initiated with $\mathbf{h}_0 = -\mathbf{g}_0$. The authors showed in [12] that the Hestenes-Stiefel formula

$$\mathbf{h}_{i+1} = -\mathbf{g}_{i+1} + \frac{\mathbf{g}_{i+1}^\top (\mathbf{g}_{i+1} - \mathbf{g}_i)}{\mathbf{h}_i^\top (\mathbf{g}_{i+1} - \mathbf{g}_i)} \mathbf{h}_i, \quad (15)$$

is well suited for image reconstruction. The algorithm updates \mathbf{s}_i and \mathbf{h}_i as in (13) and (15) until the subsequently defined stopping criterium is met.

Following [13], the algorithm terminates if the relative variation of the regularizing function over the last l iterations

$$\gamma = (|g(\mathcal{D}\mathbf{s}_i) - \bar{g}_i|) / \bar{g}_i \quad (16)$$

with $\bar{g}_i = \frac{1}{l} \sum_{k=1}^l g(\mathcal{D}\mathbf{s}_{i-k})$, falls below a certain threshold δ . Typically, we choose $l \in [10, 20]$ and $\delta \in [10^{-6}, 10^{-10}]$, depending on the required accuracy. The CG-method is summarized in Algorithm 1.

Regarding the smoothing parameter μ and the Lagrange multiplier λ , large values lead to fast convergence but yield biased reconstruction results. To overcome this problem, the reconstruction algorithm consists of repeating Algorithm 1 for

Algorithm 1 CG Algorithm

Input: $\mathbf{y}, \mathbf{s}_0, p, \lambda, \mu, \delta$

Set: $\mathbf{g}_0 \leftarrow \nabla f(\mathbf{s}_0)$

$\mathbf{h}_0 \leftarrow -\mathbf{g}_0$

for $i = 0$ to $\#InnerIter$ **do**

 compute stepsize α_i via backtracking linesearch

$\mathbf{s}_{i+1} \leftarrow \mathbf{s}_i + \alpha_i \mathbf{h}_i$

 If $\gamma < \delta$, then stop (cf. (16))

 Update search direction \mathbf{h}_i as in (15)

end for

Output: \mathbf{s}_i

a predetermined number of times, say N . It is initiated with relatively large μ_0 and λ_0 and updates $\lambda_{k+1} = c_\lambda \lambda_k$, $\mu_{k+1} = c_\mu \mu_k$ such that μ_N and λ_N take a predefined value $\mu_f \ll \mu_0$ and $\lambda_f \ll \lambda_0$. To that end, the continuation parameters are chosen as

$$c_\lambda := \left(\frac{\lambda_f}{\lambda_0} \right)^{1/N}, \quad c_\mu := \left(\frac{\mu_f}{\mu_0} \right)^{1/N}. \quad (17)$$

The complete reconstruction algorithm is summarized in Algorithm 2.

Algorithm 2 Image Reconstruction via CG

Input: $\mathbf{y}, p, \lambda_0, \mu_0, \lambda_f, \mu_f, \delta, N$

Set: $\mathbf{s}_0 \leftarrow \Phi^\top \mathbf{y}$, compute c_λ, c_μ as in (17)

for $k = 0$ to N **do**

$\mathbf{s}^* \leftarrow \text{cg_method}(\mathbf{y}, \mathbf{s}_0, p, \lambda_k, \mu_k, \delta)$

$\lambda_{k+1} \leftarrow c_\lambda \lambda_k, \mu_{k+1} \leftarrow c_\mu \mu_k$

$\mathbf{s}_0 \leftarrow \mathbf{s}^*$

end for

Output: \mathbf{s}^*

4. RESULTS

In this section we present some numerical experiments that reveal the performance of our algorithm. The test images are the Shepp-Logan Phantom (I_1) from [2], a MRI brain image (I_2) Figure 1(a), a cartoon image (I_3) Figure 1(d), and the famous Cameraman (I_4). The intensities of the images have been scaled to $[0, 255]$.

For I_1 and I_2 , DFT is used for sampling, whereas RST is used for I_3 and I_4 . We measure the reconstruction quality in terms of the relative reconstruction error $\text{Rel} = \|\mathbf{s} - \mathbf{s}^*\|_2 / \|\mathbf{s}\|_2$ and in terms of PSNR = $20 \log(255n / \|\mathbf{s} - \mathbf{s}^*\|_2^2) \text{dB}$. Our method is compared with the NESTA algorithm [13]. Certainly, the same set of samples and the same stopping criterium is used for all algorithms.

To illustrate the influence of p , we present the results of our method with $p = 0.9$, $p = 0.7$, and $p = 0.4$. It can be clearly seen in Table 1 that the proposed CG method outperforms NESTA in all situations for the chosen test images, in particular for the noiseless Cartoon-like images I_1 and I_3 but

Algorithm	Phantom, I_1 $m/n = 0.025$		Brain, I_2 $m/n = 0.25$		Cartoon, I_3 $m/n = 0.07$		Cameraman, I_4 $m/n = 0.2$	
	PSNR	Rel	PSNR	Rel	PSNR	Rel	PSNR	Rel
CG $p = 0.9$	37.3	$5.51e^{-2}$	30.0	$1.44e^{-1}$	28.4	$7.2e^{-2}$	29.2	$6.51e^{-2}$
CG $p = 0.7$	136.5	$6.06e^{-7}$	31.1	$1.27e^{-1}$	110.7	$5.85e^{-6}$	29.7	$6.32e^{-2}$
CG $p = 0.4$	173.7	$8.31e^{-9}$	30.8	$1.32e^{-1}$	87.4	$8.52e^{-5}$	28.8	$6.82e^{-2}$
NESTA	24.7	$2.43e^{-1}$	28.8	$1.68e^{-1}$	23.5	$1.39e^{-1}$	28.5	$7.15e^{-2}$

Table 1: Reconstruction quality in PSNR (in dB) and the relative reconstruction error (Rel); m : number of samples; n : signal dimension.

also for the natural image I_4 . We like to mention two further results from the literature on the reconstruction of I_1 from a few Fourier samples employing an ℓ_p like sparsity measure. In [7] the authors report a relative reconstruction error of $\approx 2e^{-3}$ with $m/n = 0.038$ and in [8] an error of $\approx 6.58e^{-10}$ with $m/n = 0.035$. Note that our method yields a reconstruction error of $\approx 8.31e^{-9}$ with only $m/n = 0.025$, i.e. 30% less samples. Finally, Figures 1(b)-(e) present some reconstructed images. It can be seen that our method yields less blurry and smeared images than NESTA. The cartoon 1(d) is quasi perfectly reconstructed.

From numerous experiments with various images we conclude that $p = 0.7$ is a universally good choice. Very good reconstruction results have been observed already after $N = 3$ iterations of Algorithm 1. Considering the smoothing parameter μ , an initial value of $\mu_0 = 0.2$ and final value of $\mu_f = 1e^{-10}$ is well suited for general images. The value of λ depends on the assumption about the noise level in the image. In the noise free case, we recommend $\lambda_k = \mu_k$.

5. CONCLUSION

In this paper we presented an algorithm for reconstructing images. The proposed method relies on the theory of compressive sampling and requires only a few samples compared to the image resolution. It is particularly well suited for Cartoon-like images, which frequently occur in medical imagery. An adaption to any sparsity measure based on the ℓ_p -pseudo-norm, to other sparsifying transformations and also to other signals is possible in a straightforward manner. A Matlab code is available on the webpage www.gol.ei.tum.de.

6. REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] M. Elad, P. Milanfar, and R. Rubinfeld, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 3, no. 3, pp. 947–968, 2007.
- [4] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [5] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [6] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 1–14, 2008.
- [7] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.
- [8] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: Mri reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging*, 2009, pp. 262–265.
- [9] G. Benke, "Generalized rudin-shapiro systems," *Journal of Fourier Analysis and Applications*, vol. 1, no. 1, pp. 87–101, 1994.
- [10] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, 2008.
- [11] J. Nocedal and S. J. Wright, *Numerical Optimization, 2nd Ed.*, Springer, New York, 2006.
- [12] S. Hawe, M. Kleinsteuber, and K. Diepold, "Dense disparity maps from sparse disparity measurements," in *IEEE 13th International Conference on Computer Vision*, 2011.
- [13] S. Becker, J. Bobin, and E. J. Candès, "Nesta: a fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2009.

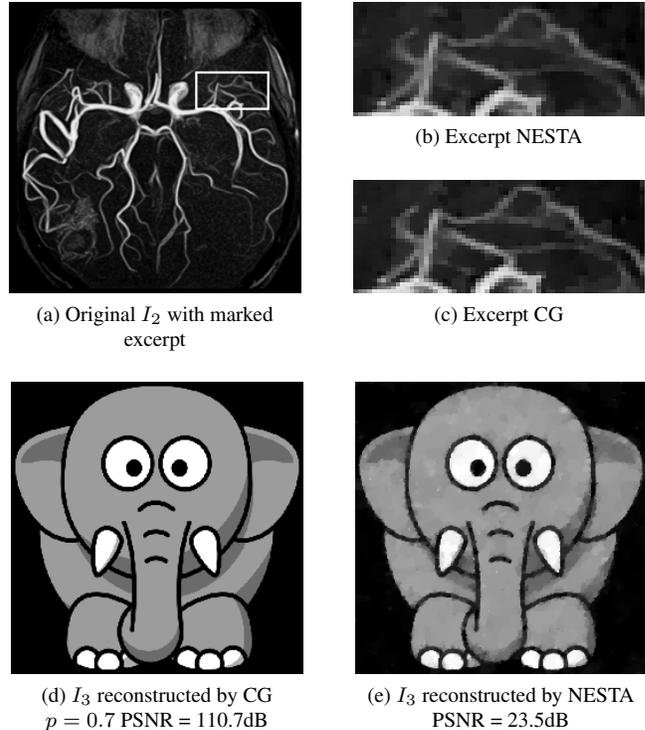


Fig. 1: Reconstruction results for I_2 and I_3