# SUPERFAST SUPERRESOLUTION

*Filip Šroubek, Jan Kamenický* *

UTIA, Academy of Sciences of CR
Pod Vodárenskou věží 4
Prague 8, 182 08, Czech Republic

*Peyman Milanfar*†

University of California Santa Cruz
Electrical Engineering Department
Santa Cruz CA 95064 USA

## ABSTRACT

We propose a fast algorithm for solving the inverse problem of resolution enhancement (superresolution). Robustness is achieved by a non-linear regularizer and a method based on variable splitting is used to obtain an equivalent linear formulation. Special attention is paid to fast implementation using the Fourier transform. In particular, we show that a degradation operator (downsampling) can be implemented in the frequency domain and that all computations can be performed very efficiently without losing robustness. To our knowledge, this is the first attempt towards a very fast SR algorithm, which retains favorable edge-preserving properties of non-linear regularizers.

***Index Terms***— superresolution, deconvolution, total variation, half-quadratic algorithm

## 1. INTRODUCTION

Superresolution (SR) imaging has been an attractive research topic in the last decade. Numerous methods have been proposed in the literature, see a recent book [1] on this topic. In theory, SR allows us to get beyond the resolution of imaging sensors. The fundamental principle that makes SR possible is an aliasing effect. In other words, the image sampling must be coarser than what the Shannon Theorem dictates. To fully recover high frequency information, which is corrupted by the aliasing effect, we need more than one input image. The SR efficiency depends on how the input images differ. Ideally, if the images capture exactly the same scene but shifted by some small sub-pixel translation, reconstruction is usually very good. There are methods, which we do not treat here, that work with only one image. They are referred to as example-based methods or hallucination methods, but they belong to a category of smart interpolation methods (such as LARKs [2]) rather than to true SR.

*Financial support was provided by Czech Ministry of Education under the project 1M0572 (Research Center DAR) and the Grant Agency of the Czech Republic under the project 102/08/1593. F.S. performed the work while at the UCSC supported by the Fulbright Visit Scholar Program.

†This work was supported in part by the US Air Force Grant FA9550-07-1-0365 and the National Science Foundation Grant CCF-1016018

We formulate the problem in the discrete domain and throughout the text we use a vector-matrix notation. Images are assumed to be rectangular, grayscale and represented as column vectors by lexicographically ordering their pixels. Let the $k$-th acquired image $\mathbf{g}_k$ be related to an unknown image $\mathbf{u}$ according to a formation model

$$\mathbf{g}_k = \mathbf{S}\mathbf{H}_k\mathbf{u} + \mathbf{n}_k \,. \tag{1}$$

Matrix $\mathbf{H}_k$ denotes blurring, i.e., convolution with some kernel (point spread function = PSF) $\mathbf{h}_k$, $\mathbf{S}$ models camera sensor functionality and it is called a decimation matrix. Vector $\mathbf{n}_k$ stands for additive noise. Convolution matrices $\mathbf{H}_k$'s model, for example, a camera-motion blur, out-of-focus blur and/or atmospheric turbulence.

It is well known that the problem of estimating $\mathbf{u}$ from $\mathbf{g}_k$'s is ill-posed, thus this inverse problem can only be solved satisfactorily by adopting some sort of regularization (in stochastic terms we call it prior information). A popular recent approach is to let the unknown image $\mathbf{u}$ be represented as a linear combination of few elements of some frame (usually an overcomplete dictionary) and force this sparse representation by using the $l^p$-norm ($0 \leq p \leq 1$). Either we can search for the solution in the transform domain (coefficients of the frame elements), which is referred to as the *synthesis approach*, or regularize directly the unknown image, which is called the *analysis approach*. Analysis versus synthesis approach has been studied earlier [3]. If the frame is an orthonormal basis, both approaches are equivalent. More interesting however is the case of redundant representation (e.g. undecimated wavelet transform), when the two approaches differ. Conclusions presented in [3] suggest that for deconvolution problems (and therefore for SR as well), the analysis approach is preferable, because sparsity should be enforced only on a part of the redundant representation (e.g. high-pass bands) and this can be easily implemented only in the analysis approach.

Using the analysis approach, we formulate SR as a non-linear regularized energy minimization problem. We adopt an additive half-quadratic algorithm [4] and show that all computations (even decimation) can be done in the frequency domain. To our knowledge, this is the first attempt towards

a very fast SR algorithm, which retains favorable edge-preserving properties of non-linear regularizers. The next section describes the proposed method and addresses PSF estimation and fast implementation. Sec. 3 concludes the paper with a real-data experiment.

## 2. REGULARIZED SUPERRESOLUTION

The decimation matrix $\mathbf{S}$ in the formation model (1) consists of convolution with a sensor PSF and downsampling by a factor $\varepsilon$ that corresponds to our desired resolution enhancement. The SR factor is an integer value, typically $\varepsilon = 2, 3$. We model the sensor PSF as a Gaussian function with a known variance.[1] We include the sensor PSF convolution in $\mathbf{H}_k$'s and then the decimation matrix $\mathbf{S}$ is a simple matrix where a single element equals 1 in each row at a position that corresponds to the downsampling factor. For example, for a 1D signal and $\varepsilon = 2$, $\mathbf{S} = [10000\ldots; 00100\ldots; 00001\ldots; \ldots]$.

Let the size of each $\mathbf{g}_k$ be $M \times 1$ (the number of pixels in the input image is $M$), then the size of $\mathbf{u}$ is $N \times 1$, where $N = \varepsilon^2 M$. Let $K$ be the number of input images. We can stack (1) for all $k$'s and write

$$\mathbf{g} = \mathbf{SHu} + \mathbf{n}\,, \tag{2}$$

where the size of $\mathbf{S}$, $\mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_K^T]^T$ and $\mathbf{H} = [\mathbf{H}_1^T, \ldots, \mathbf{H}_K^T]^T$ is $KM \times KN$, $KM \times 1$ and $KN \times N$, respectively. Note that the decimation matrix $\mathbf{S}$ is now a replicated version of the original one in (1) and is equal to $\mathbf{I}_K \otimes \mathbf{S}$, where $\mathbf{I}_K$ is an identity matrix of size $K \times K$ and $\otimes$ denotes the Kronecker product. For the sake of brevity, we keep the same symbols $\mathbf{S}$. Note that if the column-wise sum of $\mathbf{SH}$ has zeros, the above model can be considered also as an inpainting problem.

The analysis formulation applies a regularizer directly to the unknown image and minimizes a functional of the form

$$E(\mathbf{u}) = \frac{\gamma}{2}||\mathbf{SHu} - \mathbf{g}||^2 + Q(\mathbf{u})\,, \tag{3}$$

where $\|\cdot\|$ denotes the $l^2$-norm. The first term is the data term, which is determined by our formation model (2), and $Q(\cdot)$ is a non-linear regularization functional utilizing the $l^p$-norm ($0 \le p \le 1$). Weight $\gamma$ is inversely proportional to the variance of noise $\mathbf{n}$ and it can be also viewed as a regularization parameter.

Arguably, the best known and most commonly used regularizer in the analysis approach is the total variation (TV) norm [5]. An anisotropic version of TV directly translates in our notation to $Q(\mathbf{u}) = \|\mathbf{Du}\|_1$, where $\mathbf{D} = [\mathbf{D}_x^T, \mathbf{D}_y^T]^T$ and $\mathbf{D}_x$, $\mathbf{D}_y$ are matrices performing derivatives with respect to $x$, $y$, respectively. Here we use a superior (isotropic) TV model, which takes the form

$$Q(\mathbf{u}) = \sum_{i=1}^{N} \sqrt{([\mathbf{D}_x\mathbf{u}]_i)^2 + ([\mathbf{D}_y\mathbf{u}]_i)^2} = \sum_{i=1}^{N} \|\mathbf{D}_i\mathbf{u}\|\,, \tag{4}$$

where $[\cdot]_i$ denotes the $i$-th element of a vector and $\mathbf{D}_i$ is a $2 \times N$ matrix having the $i$-th row of $\mathbf{D}_x$ as the first row and the $i$-th row of $\mathbf{D}_y$ as the second row ($\mathbf{D}_i$ calculates the gradient at the $i$-th pixel). The isotropic TV is thus the $l^1$-norm of image gradient magnitudes.

The $l^1$-norm in the regularizer introduces nonlinearity and direct minimization of (3) would be a slow process. A simple procedure that solves this problem is called *variable splitting*, which decouples the $l^2$ and $l^1$ portion of (3) by introducing a new variable and converting the problem to two simpler minimization steps. One can then use the augmented Lagrangian method [6] or split Bregman iterative method [7] to minimize this new problem. Here instead, we adopt a so-called *additive half-quadratic* algorithm [4], which is based on the same idea of variable splitting and show that it minimizes a relaxed form of the original energy (3). Let us replace the TV regularizer by

$$Q_\phi(\mathbf{u}) = \sum_{i=1}^{N} \phi(\mathbf{D}_i\mathbf{u})\,, \tag{5}$$

where

$$\phi(\mathbf{s}) = \begin{cases} \frac{\beta}{2}\|\mathbf{s}\|^2 & \text{if } \|\mathbf{s}\| < \frac{1}{\beta} \\ \|\mathbf{s}\| - \frac{1}{2\beta} & \text{otherwise} \end{cases} \tag{6}$$

Note that $\mathbf{s}$ is a $2 \times 1$ vector and in the case of TV it corresponds to the image gradient. Functional $Q_\phi$ is a relaxed form of the original $Q$ in (4), being quadratic around zero and TV elsewhere. As $\beta \to \infty$, $Q_\phi \to Q$. It is proved in [4] that

$$\phi(\mathbf{s}) = \min_{\mathbf{t}} \left( \frac{\beta}{2}\|\mathbf{s} - \mathbf{t}\|^2 + \|\mathbf{t}\| \right) \tag{7}$$

and the minimum is reached for

$$\mathbf{t} = \frac{\mathbf{s}}{\|\mathbf{s}\|} \max\left( \|\mathbf{s}\| - \frac{1}{\beta}, 0 \right), \tag{8}$$

which is a generalized shrinkage formula for vectors. The relaxed form of TV becomes

$$Q_\phi(\mathbf{u}) = \sum_{i=1}^{N} \min_{\mathbf{v}_i} \left( \frac{\beta}{2}\|\mathbf{D}_i\mathbf{u} - \mathbf{v}_i\|^2 + \|\mathbf{v}_i\| \right) \tag{9}$$

and it is now a quadratic function with respect to $\mathbf{u}$ as oppose to the original formulation in (5). Interestingly, similar derivation can be done for any "$l^p$-norm", $0 \le p \le 1$, with the nonquadratic part of $\phi$ in (6) proportional to $\|\mathbf{s}\|^p$. A closed-form shrinkage formula as in (8) exists only for $p = 0$ and $p = 1$. For any other $p$, a simple and sufficiently accurate approximation in the form of a shrinkage formula exists as well. However, this is outside the scope of this manuscript.

[1] SR is surprisingly stable to this choice. The only requirements are that the sensor PSF must perform averaging of the neighboring pixels and the neighborhood size should be close to the the desired SR factor, e.g., for $\varepsilon = 2$ the size should be $2 \times 2$ or slightly larger.

Substituting the relaxed regularizer in the energy (3), we obtain

$$E(\mathbf{u},\mathbf{v}) = \frac{\gamma}{2}||\mathbf{SHu}-\mathbf{g}||^2 + \sum_{i=1}^{N}\left(\frac{\beta}{2}\|\mathbf{D}_i\mathbf{u}-\mathbf{v}_i\|^2 + \|\mathbf{v}_i\|\right). \quad (10)$$

The energy becomes now a function of the additional variable $\mathbf{v} = \text{vec}([\mathbf{v}_1,\ldots,\mathbf{v}_N]^T)$, due to variable splitting, where $\text{vec}(\cdot)$ generates a column vector by lexicographically ordering elements of its argument. The energy function is now quadratic with respect to $\mathbf{u}$ and the derivatives are linear. Minimization with respect to the image is a solution to a set of linear equations and minimization with respect to $\mathbf{v}$ is given by the shrinkage formula:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}} E(\mathbf{u},\mathbf{v}) \Longleftrightarrow \left(\mathbf{H}^T\mathbf{S}^T\mathbf{SH} + \frac{\beta}{\gamma}\mathbf{D}^T\mathbf{D}\right)\mathbf{u}^* = \mathbf{H}^T\mathbf{S}^T\mathbf{g} + \frac{\beta}{\gamma}\mathbf{D}^T\mathbf{v}, \quad (11)$$

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} E(\mathbf{u},\mathbf{v}) \Longleftrightarrow \mathbf{v}_i^* = \frac{\mathbf{D}_i\mathbf{u}}{\|\mathbf{D}_i\mathbf{u}\|}\max\left(\|\mathbf{D}_i\mathbf{u}\| - \frac{1}{\beta}, 0\right). \quad (12)$$

The restoration algorithm alternates between these two minimization steps until a convergence criterion is met. Since this is a type of EM (Expectation-Maximization) algorithm, convergence to the global minimum is not guaranteed and in theory initialization is important. However, our extensive testing shows that the algorithm is quite stable with respect to initialization and we thus start with $\mathbf{v} = 0$.

### 2.1. Estimation of blurs

Estimation of PSFs $\mathbf{h}_k$ can be carried out in the SR framework as proposed in [8]. However, the computational overhead introduced by minimization with respect to PSFs is excessive and in many practical applications not necessary. SR can hardly recover high-frequency information if this information is decimated by substantial blurring. If input images undergo complex geometric warping, complicated and lengthy registration is necessary. However, SR typically requires registration with sub-pixel accuracy, which is difficult to achieve in practice. We thus consider only the sensor PSF, which we assume to be known, and translation as a geometric transformation. Shifts between images are estimated with sub-pixel accuracy using an optical flow algorithm with hyperbolic numeric as proposed in [9]. We then shift the sensor PSFs accordingly and use them to implement blurring matrices $\mathbf{H}_k$'s.

### 2.2. Fast implementation using FFT

The second minimization step (12) is element-wise and can be computed in $O(N)$ time.

More challenging is the first step (11), which requires inverting a huge $N \times N$ matrix $(\mathbf{H}^T\mathbf{S}^T\mathbf{SH} + \frac{\beta}{\gamma}\mathbf{D}^T\mathbf{D})$. One can apply iterative solvers, such as conjugate gradient (CG), to avoid direct inversion, but we want to do even better and have a one-step solver. In our formulation, both $\mathbf{H}$ and $\mathbf{D}$ are convolution matrices. To avoid any ringing artifacts close to image boundaries, $\mathbf{H}$ and $\mathbf{D}$ should perform "valid" convolution, i.e., the output image is smaller and covers a region where both the input image and convolution kernel are fully defined. If we properly adjust the image borders, by using for example function *edgetaper* in MATLAB, we can replace "valid" convolution with block-circulant one and ringing artifacts will be almost undetectable. In addition, the sparsity regularizer also helps to reduce the artifacts. The 2D discrete Fourier transform (DFT) diagonalizes block-circulant convolution matrices and inversion is thus straightforward. The product $\mathbf{S}^T\mathbf{S}$ becomes a block diagonal matrix under the DFT as also observed in [10]. Its effect on the image spectrum can be visualized as follows: divide the spectrum into non-overlapping $\varepsilon \times \varepsilon$ blocks, calculate a mean block by summing up the blocks element-wise, and replicate the result $\varepsilon \times \varepsilon$ to form a spectrum of the original size. More precisely, let $\hat{u}(\omega_1,\omega_2)$, $1 \le \omega_1,\omega_2 \le \sqrt{N}$ denote the DFT of an image $u(x_1,x_2)$, and $\hat{\mathbf{u}}(\omega_1,\omega_2)$, $1 \le \omega_1,\omega_2 \le \Delta = \sqrt{N}/\varepsilon$ denote a vector of size $\varepsilon^2 \times 1$, which we obtain from $\hat{u}(\omega_1,\omega_2)$ by concatenating $(i,j)$-th elements of each block of $\hat{u}$. For example, for $\varepsilon = 2$ we have $\hat{\mathbf{u}}(\omega_1,\omega_2) = [\hat{u}(\omega_1,\omega_2), \hat{u}(\omega_1+\Delta,\omega_2), \hat{u}(\omega_1,\omega_2+\Delta), \hat{u}(\omega_1+\Delta,\omega_2+\Delta)]^T$.

Under the DFT and some permutation, $(\mathbf{H}^T\mathbf{S}^T\mathbf{SH} + \frac{\beta}{\gamma}\mathbf{D}^T\mathbf{D})$ becomes block diagonal with $\Delta^2$ blocks of size $\varepsilon^2 \times \varepsilon^2$. Each block is given by

$$\frac{1}{\varepsilon^2}\sum_{k=1}^{K}\hat{\mathbf{z}}_k(\omega_1,\omega_2)\hat{\mathbf{z}}_k^T(\omega_1,\omega_2) - \frac{\beta}{\gamma}\text{diag}\{\hat{\mathbf{l}}(\omega_1,\omega_2)\}, \quad (13)$$

where $\hat{\mathbf{z}}_k = \hat{\mathbf{h}}_k^* \odot \hat{\mathbf{h}}_k$, $(\cdot)^*$ denotes complex conjugate and $\odot$ element-wise multiplication. Vector $\hat{\mathbf{l}}$ is constructed from the DFT of discrete Laplacian and $\text{diag}\{\cdot\}$ stands for a diagonal matrix.

To conclude, minimization in (11) is calculated entirely in the DFT and requires $\Delta^2$ inversions of small $\varepsilon^2 \times \varepsilon^2$ matrices. Each DFT can be carried out with $O(N\log N)$ cost using the FFT algorithm. The cost of one inversion is in the worst case $O(\varepsilon^6)$. The overall complexity of minimization with respect to $\mathbf{u}$ is thus $O(N\log N + N\varepsilon^4)$. It is important to note that the inversions are performed only once in the whole algorithm, since the inverted matrices are independent of the updated $\mathbf{v}$. If we used CG iterative minimization and kept all matrix multiplications in the DFT, the complexity would be $O(N^2\log N)$. Our algorithm is thus by a factor of $N$ faster.

## 3. EXPERIMENTS AND CONCLUSION

We recorded two video sequence, one with a compact digital camera ($640 \times 480$) and one with a mobile phone ($320 \times 240$), took five consecutive frames from each and estimated SR images ($\varepsilon = 2$) using the proposed method. Parameters were set as follows, camera data: $\gamma = 500$, $\beta = 10$; and mobile data: $\gamma = 100$, $\beta = 10$. Parameter $\gamma$ is proportional to SNR, $\beta$ is a threshold and depends on the intensity range of input images. The maximum of 10 iterations was sufficient to achieve very good results. Results are in Fig. 1 (a) and (b). Improvement in resolution is clearly visible.

We have proposed superresolution with nonlinear regularization implemented in the Fourier domain, which is $N$-times faster than the equivalent implementation in the image domain. Currently, we use the additive half-quadratic approach. In the future more elaborated algorithms such as augmented Lagrangian will be tested.

## 4. REFERENCES


[1] P. Milanfar, Ed., *Super-resolution Imaging*, CRC Press, 2010.

[2] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, 2007.

[3] I. W. Selesnick and M. Figueiredo, "Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors," in *Proceedings of SPIE*, 2009, vol. 7446.

[4] Junfeng Yang, Wotao Yin, Yin Zhang, and Yilun Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Img. Sci.*, vol. 2, pp. 569–592, May 2009.

[5] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[6] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[7] Tom Goldstein and Stanley Osher, "The split bregman method for l1-regularized problems," *SIAM J. Img. Sci.*, vol. 2, pp. 323–343, April 2009.

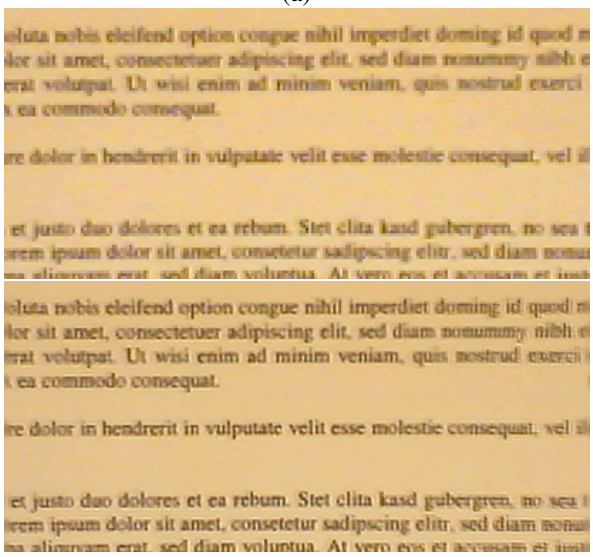
[8] F. Šroubek, G. Cristóbal, and J. Flusser, "A unified approach to superresolution and multichannel blind deconvolution," *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2322–2332, Sept. 2007.

[9] Henning Zimmer, Michael Breuß, Joachim Weickert, and Hans-Peter Seidel, "Hyperbolic numerics for variational approaches to correspondence problems," in *Proceedings of SSVM '09*, Berlin, Heidelberg, 2009, pp. 636–647, Springer-Verlag.

[10] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient fourier-wavelet super-resolution," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2669–2681, 2010.



(a)



(b)

**Fig. 1**. (a): (top) one of five images ($640 \times 480$) acquired with a digital camera, (bottom) the super-resolved image ($1280 \times 960$). (b): (top) one of five images ($320 \times 240$) acquired with a mobile phone, (bottom) the super-resolved image ($640 \times 480$). For better visual comparison, only small sections of the original images are shown.