Diffusion-Steered Super-Resolution Image Reconstruction

Baraka J. Maiseli

Abstract

For decades, super-resolution has been a widely applied technique to improve the spatial resolution of an image without hardware modification. Despite the advantages, super-resolution suffers from ill-posedness, a problem that makes the technique susceptible to multiple solutions. Therefore, scholars have proposed regularization approaches as attempts to address the challenge. The present work introduces a parameterized diffusion-steered regularization framework that integrates total variation (TV) and Perona-Malik (PM) smoothing functionals into the classical super-resolution model. The goal is to establish an automatic interplay between TV and PM regularizers such that only their critical useful properties are extracted to well pose the super-resolution problem, and hence, to generate reliable and appreciable results. Extensive analysis of the proposed resolution-enhancement model shows that it can respond well on different image regions. Experimental results provide further evidence that the proposed model outperforms.

Keywords: super-resolution, resolution, enhancement, regularization, diffusion

1. Introduction

Before deepening into the super-resolution imaging, let us discuss the term resolution. Most people, particularly those not in the imaging field, define resolution broadly as the physical size of an image. For a two-dimensional digital image, this definition implies an area in the image given as the product of the number of pixels in the horizontal and vertical dimensions (pixel or picture element is the smallest unit of information in a digital image). In this context, therefore, a high-resolution image contains a higher pixel count than a low-resolution image. Figure 1(a) includes features with higher perceptual qualities than those in Figure 1(b), but both images have equal sizes. From the figure, therefore, we see that dimension only seems inadequate to define the resolution of an image.
Resolution, more generally, means the quality of a scene (image or video). Five major types of image resolutions are known: pixel resolution, spectral resolution, temporal resolution, radiometric resolution, and spatial resolution. The use of these variations depends on the application. Pixel resolution refers to the total number of pixels a digital image contains. Hence, both images in Figure 1(a) and (b) possess equal pixel resolutions of $2179 \times 2011$. In other words, each image is approximately 4.4 megapixels ($2179 \times 2011 = 4,381,969$ pixels $\approx 4.4$ megapixels). Unfortunately, pixel count offers fraction of the pieces of information contained in the image. For a colored image with red, green, and blue channels, an individual pixel can only accommodate the details of a single color. Spectral resolution describes the ability of an imaging device to distinguish the frequency (or wavelength) components of an electromagnetic spectrum. Imagine spectral resolution as the degree in which you can uniquely discern two different colors or light sources. Temporal resolution refers to the rate at which an imaging device revisits the same location to acquire data. When dealing with videos, for example, the term implies an average time between consecutive video frames: a standard video camera can record 30 frames per second, implying that every 33 ms, this camera captures an image. In remote sensing, temporal time is usually measured in days to represent time that a satellite sensor revisits a specific location to collect data. Radiometric resolution defines the degree at which an imaging system can represent or distinguish intensity variations on the sensor. Expressed in number of bits (or number of levels), radiometric resolution provides the actual content of information in the image. Spatial resolution explains how an imaging modality can distinguish two objects. In practical situations, spatial resolution describes clarity of an image and defines the resolving power of an image-capturing device. The perceptual quality of an image increases with the spatial resolution. This research presents super-resolution imaging as one of the available techniques to enhance the spatial resolution of an image.

Most people are naturally inclined to high-quality and visually appealing images that contain adequate details. However, this demand is not always achieved because of some imperfections in the imaging process. Therefore, scholars have proposed hardware and software approaches...
to address the challenge. The former approach requires sensor modification, and it may be
achieved by reducing the physical sizes of the pixels—a process that increases pixel density
(number of pixels per unit area) on the surface of the sensor [1]. The hardware approach gives
perfect resolution enhancement, but the technique endures several drawbacks: (1) it introduces
shot noise into the captured images, (2) it makes the imaging device costly and unnecessarily
bulkier, and (3) it lowers the charge transfer rate because of the increased chip size [2]. These
challenges have prompted scholars to search for software techniques, which are cost-effective
and reliable, to improve the spatial resolution of an image without effecting circuitry of the
imaging device. In this case, an image can be captured by a low-cost device and processed to
generate its corresponding high-quality version.

The classical software approach that has gained a considerable attention of scholars is called
super-resolution [3–6], which uses signal processing principles to restore high-resolution
images from at least one low-resolution image. Super-resolution techniques can be put into
two major categories: single-frame-based, which generates a high-resolution image from the
respective single low-resolution image [7, 8], and multi-frame-based, which exploits informa-
tion from a sequence of degraded images to generate a high-quality image [2, 6]. The current
work builds on the multi-frame super-resolution framework, which implicitly encourages
noise reduction from the input low-resolution images. The framework bridges total variation
(TV) [9] and Perona and Malik [10] smoothing functionals and allows for these functionals to
interact in such a way that super-resolution and preservation of critical image features are
simultaneously conducted.

2. Image degradation model

The multi-frame super-resolution framework can better be understood through a conceptual
degradation model, which shows how an unknown high-resolution image, \( u \), undergoes a
variety of degradations to form \( M \) low-quality images, \( y_k \) with \( k = 1, \ldots, M \) denoting positions
of the low-resolution frames (Figure 2). In practice, the degradation process of \( u \) to generate \( y_k \)
involves warping, blurring, decimation (downsampling), and noising, respectively defined in
this work by the operators \( W_k \), \( B_k \), \( D_k \), and \( \eta_k \): warping introduces rotations and translations
into \( u \), hence changing its geometrical properties; blurring reduces sharpness of features in \( u \);
decimation samples \( u \) and lowers its physical size; and noising corrupts \( u \) with noise, assumed
to be additive.

Figure 2 can be transformed into

\[
y_k = W_k B_k D_k u + \eta_k
\]

which explains how the degradation model generates frame \( k \) in a set of low-resolution
images. The goal of the present study is to estimate \( u \) under the degradation conditions, and
one approach to achieve the goal is to re-define Eq. (1) into the minimization problem that aims
to lower \( \eta_k \). Therefore, using the \( L_p \) norm, where \( p \in [1, 2] \) (the range \( 0 \leq p < 1 \) is excluded
because the values of \( p \) contained in this interval lead to nonconvex minimization problems
that are susceptible to unstable solutions), the formulation to optimize \( u \) becomes
min \( u \in \mathbb{R}^M \) \( E(u) = \frac{1}{2M} \sum_{k=1}^{M} \| W_k B_k D_k u - y_k \|_p^p \), \quad (2) 

where \( E \) is modeled as an energy functional that defines noise level in the degraded image. The gradient of the cost of \( E \) in Eq. (2) is

\[
J_p = \frac{\partial}{\partial u} \left[ \frac{1}{2M} \sum_{k=1}^{M} \| W_k B_k D_k u - y_k \|_p^p \right] 
\]

\[
= \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T \text{sign}(W_k B_k D_k u - y_k) \odot |W_k B_k D_k u - y_k|^{p-1}, \quad (4)
\]

where \( D_k^T \) is the upsampling operator, \( B_k^T \) and \( W_k^T \) are the inverse operators for blurring and warping, respectively, and \( \odot \) denotes the Hadamard (element-wise) operator for two matrices. The solution of Eq. (2) can be obtained when \( J_p = 0 \).

For \( p = 1 \), Eq. (4) evaluates to

\[
J_p = \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T \text{sign}(W_k B_k D_k u - y_k) \odot |W_k B_k D_k u - y_k|^{p-1}, \quad (4)\text{(p=1)}
\]
which shows that, after shifting and zero filling, $D_k^T B_k^T W_k^T$ copies values from the low-resolution to the high-resolution images, and $W_k B_k D_k$ reverses the operation [11]. Pixel values are unaffected by these complimentary operations, implying that each entry in $J_1$ is impacted by entries from all low-resolution images. Figure 3 shows the influences of $D$ and that of $D^T$ on the reconstructing image. In their work, Farsiu et al. noted that the $L_1$ minimization in Eq. (5) corresponds to the pixel-wise median, a robust estimator that addresses favorably noise and outliers in the input data. But the $L_1$ norm is nondifferentiable at zero, a property that makes the minimization process unstable and that generates undesirable solutions.

For $p = 2$, Eq. (4) becomes a solution of the $L_2$ norm minimization, or

$$J_2 = \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T (W_k B_k D_k u - y_k) = 0,$$

which was proved in [12] that it represents pixel-wise mean of measurements. The $L_2$ norm is less-robust against erroneous data, but the metric has better mathematical properties: convexity, differentiability, and stability. Therefore, several scholars prefer the $L_2$ objective functions in situations where data contain low noise as in our case.

The super-resolution problem, whether formulated through $L_1$ or $L_2$ norm, has an ill-posedness nature. Given that $r$ is the resolution factor, then for the under-determined case, or for $M < r^2$, and for the square case, or for $M = r^2$, the problem may evaluate to infinitely many undesirable solutions. Also, for the small amount of noise in the data, ill-posed problems tend to introduce larger perturbations in the final solutions. These issues can be effectively addressed through a technique called regularization, which has another advantage of speeding the convergence rate of the evolving solution. This work addresses the super-resolution ill-posedness through regularization functionals from nonlinear diffusion processes, which have been reported that they can preserve important image features (edges, contours, and lines) [13–15]. The proposed regularizer integrates total variation (TV) [9] and Perona and Malik (PM) [10] models that complement one another to generate appealing results.

![Diagram](Image)
3. Hybrid super-resolution model

3.1. Regularization functionals

Considering the super-resolution ill-posedness property, a hybrid framework combining TV and PM regularization kernels has been formulated. The framework includes additional parameters, $\alpha$ and $\beta$, which establish a proper balance between TV and PM during regularization. The objective is to de-emphasize weaknesses of the models and amplify their strengths so that the super-resolved images are superior.

In [9], Rudin et al. established the TV model that explains how noise in the image can be reduced. The model is based on the fact that a noisy image contains a higher total variation, defined by the integral of the absolute gradient of the image or

$$\rho(|\nabla u|) = \int_{\Omega} |\nabla u| dx,$$

where $\rho$ is the TV energy functional, $\Omega$ defines the domain under which $u$ exists, and $x$ denotes the two-dimensional spatial coordinate on $\Omega$. Therefore, reducing noise is equivalent to minimizing $\rho$. Being defined in the bounded variation space, TV functionals allow for discontinuities in the image functions. Hence, regularization through TV promotes recovery of edges, which appear as “jumps” or discontinuous parts of the image, and effective noise removal. But studies have revealed that TV formulations favor piecewise-constant solutions, a consequence that generates staircase effects and introduces false edges [16]. Also, TV regularization tends to lower contrast even in noise-free or flat image regions [17].

In the similar notion of the TV principle, Perona and Malik proposed an energy functional, $\varphi$, defined by

$$\varphi(|\nabla u|) = \frac{K^2}{2} \int_{\Omega} \log \left(1 + \left(\frac{|\nabla u|}{K}\right)^2\right) dx,$$

where $K$ denotes the shape-defining constant, which can be minimized to suppress noise [10]. Minimizing Eq. (8), which originates from robust statistics, produces a nonlinear diffusion equation that embeds a fractional conduction coefficient for preserving edges. The PM energy functional in Eq. (8) is nonconvex for $|\nabla u| > K$, an undesirable property that can generate instabilities in the evolving solution. This work presents a technique that retains the convex portion, $|\nabla u| \leq K$, and complements the nonconvex portion of the PM potential by the TV energy functional.

The regularization process is often supported by the fidelity potentials

$$\psi(u) = \frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx,$$

for additive noise, $f = u + \eta$, and
\[ \phi(u) = \lambda \int_\Omega \left( \log u + \frac{f}{u} \right) \, dx \]  

(10)

for multiplicative noise [18], \( f = u \eta \), where \( f \) is the corrupted image and \( \lambda \) is the fidelity parameter that balances the trade-off between \( u \) and \( f \). The fidelity term is often added to the regularization framework.

### 3.2. Proposed super-resolution model

The hybrid model can be derived from the minimization problem that integrates the corresponding energy functionals from super-resolution, TV, PM, and fidelity. Assuming additive noise and \( L_2 \) estimator for the super-resolution part, the (regularized) minimization super-resolution problem parametrized in \( \alpha \) and \( \beta \) becomes

\[
\min_u \left\{ H(u, |\nabla u|) = \frac{1}{2M} \sum_{k=1}^{M} \| W_k B_k D_k u - y_k \|^2_2 + \alpha \rho(|\nabla u|) + \beta \varphi(|\nabla u|) + \psi(u) \right\}, \tag{11}
\]

where \( \alpha, \beta \in [0, 1] \) and \( \beta = \bar{\alpha} \). Solving Eq. (11) using the Euler-Lagrange equation, and embedding the result into the time-dependent system gives

\[
\frac{\partial u}{\partial t} = \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T (W_k B_k D_k u - y_k) + \text{div} \left( \frac{\alpha}{|\nabla u|} \nabla u \right) + \text{div} \left( \frac{\beta}{1 + (|\nabla u|)^2} \nabla u \right) - \lambda(u - f). \tag{12}
\]

Eq. (12) offers both super-resolution image reconstruction and noise removal capabilities, dictated by TV and PM models. From the equation, as \( t \to \infty \), \( u \) approaches an optimal solution—a stationary function that solves the energy functional, \( H \), in Eq. (11). Eq. (12) has interesting properties for various parts of the image: in flat regions (\(|\nabla u| \to 0\)), Eq. (12) reduces to

\[
\frac{\partial u}{\partial t} = \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T (W_k B_k D_k u - y_k) + (\alpha C + \beta) \Delta u - \lambda(u - f), \tag{13}
\]

where \( C > 0 \) is a constant. This equation has a Laplacian term, \( \Delta u \), which possesses isotropic diffusion characteristics to strongly and uniformly suppress noise in flat regions. In the neighborhood of the edges (\(|\nabla u| \to \infty\)), Eq. (12) becomes

\[
\frac{\partial u}{\partial t} = \frac{1}{2M} \sum_{k=1}^{M} D_k^T B_k^T W_k^T (W_k B_k D_k u - y_k) - \lambda(u - f), \tag{14}
\]

implying protection of edges against smoothing. This automatic interplay between reconstruction and regularization components helps to generate superior super-resolved images.
3.3. Numerical implementation

The solution of the proposed super-resolution model in Eq. (12) was iteratively estimated using the steepest descent method. Therefore, the evolution equation in Eq. (12) can be converted into a numerical system

\[
\begin{align*}
    u_{n+1} & = u_n - \tau \left\{ \frac{1}{2M} \sum_{k=1}^{M} D_k B_k^T W_k^T (W_k B_k D_k u_n - y_k) + \text{div} \left( \frac{\alpha}{\sqrt{\nabla u_n}} \nabla u_n \right) \right. \\
    & \quad + \text{div} \left( \frac{\beta}{1 + \left( \frac{\nabla u_n}{k} \right)^2} \nabla u_n \right) - \lambda (u_n - f_n) \right\},
\end{align*}
\]

(15)

where \( n \) denotes the iteration number that defines the solution space index of \( u \), and \( \tau > 0 \) denotes constant of the step size in the gradient direction. To encourage stability of the evolution equation in (15), the Courant-Friedrichs-Lewy condition, that is \( 0 < \tau \leq 0.25 \), should be satisfied [19]. From the equation, the degradation matrices, namely \( W_k, B_k, \) and \( D_k \), and their corresponding transpose versions may be regarded as direct operators for image manipulations: shifting, blurring, and down sampling, along with the reverse of these operations [11]. With this observation of the matrices properties, implementation of the super-resolution component of Eq. (15) can be achieved using cascaded operators without explicitly constructing the matrices.

Figure 4. Block diagram representation of the proposed super-resolution model. The blocks \( P_k \) and \( Q \) are defined in Figures 5 and 6.
operators as matrices. This implementation strategy helps to boost the algorithmic speed and to optimize hardware resources.

Eq. (15) can be represented in block form by Figure 4. From the Figure, each low-resolution frame, $y_k$, is compared with the current estimate, $u_n$, of the high-resolution image. This process is undertaken by block $P_k$ detailed in Figure 5—an operator that represents the gradient back projection to compare the $k$th degraded frame and the high-resolution estimate at the $n$th

---

**Figure 5.** Extended block diagram representation of the similarity cost derivative, $P_k$, in Figure 4.

---

**Figure 6.** Block diagram representation of the smoothing cost derivative, $Q$, in Figure 4.
iteration of the steepest descent method. Note from Figure 5 that $T(\text{PSF})$, with PSF denoting the point spread function, replaces $B^T_k$ with a simple convolution operator. This block can be implemented by flipping, on the respective axes, rows, and columns of the PSF in the up-down and left-right directions, respectively. Gradient of the regularization term is represented by block $Q$, defined more explicitly in Figure 6, which ensures that the evolution process converges and gives desirable solutions.

4. Experimental methodology

Several experiments were executed to determine performance of the proposed super-resolution model relative to the classical approaches. The methodology and procedures under which the experiments were undertaken can be explained as follows: firstly, high-resolution images of bike, butterfly, flower, hat, parrot, Parthenon, plant, and raccoon (Figure 7) were degraded to generate the corresponding low-resolution images (Figure 8, first column). Note that the original images were downloaded from the public domain with standard test images. These images were selected because they contain detailed features, and hence it would be easier to test the superiority of various super-resolution methods. As an example, the “Raccoon” image contains small-scale features (fine textures or fur) that most super-resolution approaches may find hard to restore. Degradation of the original images was achieved through warping, blurring, decimation, and noise addition to create sequences of 10 low-quality images with consecutive pairs differing by some rotation and translation motions. To void impacts of registration errors on the reconstruction process, the

![Figure 7. Original high-resolution images.](http://www4.comp.polyu.edu.hk/~cslzhang/NCSR.htm)
Figure 8. Super-resolution results from different methods.
warping matrix was fixed. Thus, for 10 multiple low-resolution images, the warping matrix for the horizontal and vertical displacements, respectively denoted by $\Delta x$ and $\Delta y$, was defined as follows:

| $\Delta x$ | 0.56 | 1.03 | 0.85 | 0.32 | −0.45 | −0.43 | 0.92 | 1.23 | 0.93 | 0.64 |
| $\Delta y$ | 0.12 | 0.53 | 0.27 | 0.00 | −0.83 | 1.12 | 1.08 | 0.12 | 0.54 | 1.37 |

Next, super-resolution methods based on a variety of regularizers, namely NC00 [20], TV [9], ANDIFF [21], and Hybrid, were applied on the degraded images to restore their original versions. Lastly, the objective metric, namely feature similarity (FSIM) [22], and the subjective metric were used to compare performances of different methods. FSIM incorporates into its formulation some aspects of the human visual system, and hence the metric is considered superior over several other existing image quality metrics. A visually appealing image has a higher value of FSIM, and vice versa.

5. Results and discussions

Visual results show that the classical methods tend to add undesirable artificial features into the reconstructed images (Figure 8). For instance, NC00 introduces bubble-like features around borders, edges, and corners, which are the critical features that emulate the human visual system. The method, on the other hand, does well on homogeneous image regions. The super-resolution method based on TV produces relatively sharper images, but the method also adds artifacts on homogeneous parts of the final images—an effect that degrades the visual quality of the images. The ANDIFF method generates smoother results that contain little artifacts, but the method underperforms for highly-textured images such as the Raccoon. The proposed hybrid model established a proper balance between smoothness and critical feature preservation (Figure 8, last column). Visually, the reconstructed images by our approach are more natural and are free from obvious artifacts. One may argue about a slight blurriness in our results. However, given the higher capability of the proposed method to preserve sensitive image features, this effect may be ignored. Also, the line graphs (taken near the last row across all columns) further confirm that the proposed method is superior because it generates a one-dimensional curve that closely matches the original one (Figure 9).

Numerical results demonstrate that, in all cases of the input images, the proposed super-resolution method achieves higher quality values (Table 1). These convincing objective observations can be explained well from the new formulation in Eq. (12): the hybrid super-resolution model captures the qualities of both PM and TV, an advantage that may promote higher objective quality results. Besides, our formulation incorporates parameters that give an effective interplay between the regularization functionals.
6. Conclusion

In this work, we have established a hybrid super-resolution framework that combines desirable features of TV and PM models. The framework has been parametrized to mask weaknesses of the models, introduce an automatic interplay between TV and PM regularizations,
and promote appealing results. More emphasis was put on super-resolving low-quality images while retaining their naturalness and preserving their sensitive image features. Experimental results demonstrate that the proposed framework generates superior objective and subjective results.

Author details

Baraka J. Maiseli

Address all correspondence to: barakamaiseli@yahoo.com

Department of Electronics and Telecommunication Engineering, College of Information and Communication Technologies, University of Dar es Salaam, Dar es Salaam, Tanzania

References

[1] Park SC, Park MK, Kang MG. Super-resolution image reconstruction: A technical overview. IEEE Signal Processing Magazine. 2003;20(3):21-36

[2] Maiseli BJ, Elisha OA, Gao H. A multi-frame super-resolution method based on the variable-exponent nonlinear diffusion regularizer. EURASIP Journal on Image and Video Processing. 2015;2015(1):22

[3] Dong C, et al. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016;38(2):295-307

[4] El Mourabit I, et al. A new denoising model for multi-frame super-resolution image reconstruction. Signal Processing. 2017;132:51-65

[5] Peleg T, Elad M. A statistical prediction model based on sparse representations for single image super-resolution. IEEE Transactions on Image Processing. 2014;23(6):2569-2582

[6] Zeng X, Yang L. A robust multiframe super-resolution algorithm based on half-quadratic estimation with modified BTV regularization. Digital Signal Processing. 2013;23(1):98-109

[7] Purkait P, Pal NR, Chanda B. A fuzzy-rule-based approach for single frame super resolution. IEEE Transactions on Image Processing. 2014;23(5):2277-2290

[8] Yang M-C, Wang Y-CF. A self-learning approach to single image super-resolution. IEEE Transactions on Multimedia. 2013;15(3):498-508

[9] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena. 1992;60(1–4):259-268

[10] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990;12(7):629-639
[11] Farsiu S, et al. Fast and robust multiframe super resolution. IEEE Transactions on Image Processing. 2004;13(10):1327-1344

[12] Elad M, Hel-Or Y. A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur. IEEE Transactions on Image Processing. 2001;10(8):1187-1193

[13] Ma H, Nie Y. An edge fusion scheme for image denoising based on anisotropic diffusion models. Journal of Visual Communication and Image Representation. 2016;40:406-417

[14] Tsiotsios C, Petrou M. On the choice of the parameters for anisotropic diffusion in image processing. Pattern Recognition. 2013;46(5):1369-1381

[15] Xu J, et al. An improved anisotropic diffusion filter with semi-adaptive threshold for edge preservation. Signal Processing. 2016;119:80-91

[16] Chen Y, Levine S, Rao M. Variable exponent, linear growth functionals in image restoration. SIAM Journal on Applied Mathematics. 2006;66(4):1383-1406

[17] Chan TF, Esedoglu S. Aspects of total variation regularized L 1 function approximation. SIAM Journal on Applied Mathematics. 2005;65(5):1817-1837

[18] Lv X-G, et al. A fast high-order total variation minimization method for multiplicative noise removal. Mathematical Problems in Engineering. 2013;2013

[19] Courant R, Friedrichs K, Lewy H. On the partial difference equations of mathematical physics. IBM Journal of Research and Development. 1967;11(2):215-234

[20] Pham TQ, Van Vliet LJ, Schutte K. Robust fusion of irregularly sampled data using adaptive normalized convolution. EURASIP Journal on Advances in Signal Processing. 2006;2006(1):083268

[21] Maiseli BJ, Ally N, Gao H. A noise-suppressing and edge-preserving multiframe super-resolution image reconstruction method. Signal Processing: Image Communication. 2015;34:1-13

[22] Zhang L, et al. FSIM: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing. 2011;20(8):2378-2386
