Unconstrained Structural Similarity-Based Optimization

Daniel Otero(✉) and Edward R. Vrscay

Department of Applied Mathematics, Faculty of Mathematics,
University of Waterloo, Waterloo, ON N2L3G1, Canada
{dotero,ervrscay}@uwaterloo.ca

Abstract. We establish a general framework, along with a set of algorithms, for the incorporation of the Structural Similarity (SSIM) quality index measure as the fidelity, or "data fitting," term in objective functions for optimization problems in image processing. The motivation for this approach is to replace the widely used Euclidean distance, known as a poor measure of visual quality, by the SSIM, which has been recognized as one of the best measures of visual closeness. Some experimental results are also presented.

1 Introduction

Many image processing tasks, e.g., denoising, inpainting, deblurring, are usually carried out by solving an appropriate optimization problem. In most cases, the objective function associated with such a problem is expressed as the sum of a fidelity term (or terms) $f(x)$ and a regularization term (or terms) $h(x)$. The optimization problem then assumes the form

$$\min_x \{f(x) + \lambda h(x)\}, \tag{1}$$

where the constant λ is a regularization parameter.

The role of the fidelity term $f(x)$ is to keep the solution to (1) close to the observed data. A typical choice is $f(x) = \frac{1}{2}\|x - y\|_2^2$, where y is the (corrupted) observation, e.g., a noisy image. The regularization term $h(x)$ has a twofold purpose: (i) It prevents over-fitting to the observed data and (ii) it imposes constraints on the solution based upon prior information or assumptions. For instance, if the optimal solution is assumed to be sparse, a typical regularization term is $h(x) = \|x\|_1$ [5,11,16].

Using the squared Euclidean distance as a measure of closeness is convenient since it is convex, differentiable, and usually mathematically tractable, not to mention easily computed. Furthermore, widely used metrics of visual quality such as Mean Squared Error (MSE) and Peak to Signal Noise Ratio (PSNR) are based on this definition of closeness. Nevertheless, it has been shown that such distortion measures are not the best choice when it comes to quantify visual

A. Campilho and M. Kamel (Eds.): ICIAR 2014, Part I, LNCS 8814, pp. 167–176, 2014.
DOI: 10.1007/978-3-319-11758-4_19

quality [17,18]. For this reason, many measures of visual quality have been proposed in an attempt to model the Human Visual System (HVS). The Structural Similarity (SSIM) image quality measure, originally proposed by Wang *et al.* [18], was based upon the assumption that the HVS evolved to perceive visual errors as changes in structural information. On the basis of subjective quality assessments involving large databases, SSIM has been generally accepted to be one of the best measures of visual quality/closeness.

With these comments in mind, it would seem natural to consider the SSIM as a replacement for the widely-used squared Euclidean distance in the fidelity term $f(x)$ of Eq. (1), given the limitations of the latter to measure visual closeness. Indeed, from a practical point of view, it is easy to make such a replacement since the mathematical expression for the SSIM between x and the observed data y is rather straightforward. One may then be tempted to simply start computing. There is a problem, however, in that the actual mathematical framework behind such an SSIM-based optimization, which would be important for the establishment of existence and uniqueness of solutions, is more complicated due to the fact that the SSIM is not a convex function.

Notwithstanding these obstacles, optimization problems that employ the SSIM as a fitting term have already been addressed. For instance, in [3] the authors find the best approximation coefficients in the SSIM sense when an orthogonal transformation is used (e.g., Discrete Cosine Transform (DCT), Fourier, etc.). Very briefly, a contrast-enhanced version of the best ℓ_2-based approximation is obtained. Based on this result, Rehman *et al.* [13] address the SSIM version of the image restoration problem proposed by Elad *et al.* in [10], where the denoising of images is performed using sparse and redundant representations over learned dictionaries. Furthermore, in [13] the authors also introduce a super-resolution algorithm – also based on the SSIM – to recover from a given low resolution image its high resolution version.

Another interesting application for reconstruction and denoising was proposed in [7]. Here, the authors define the statistical SSIM index (statSSIM), an extension of the SSIM for wide-sense stationary random processes. By optimizing the statSSIM, an optimal filter in the SSIM sense is found. The non-convex nature of the statSSIM is overcome by reformulating its maximization as a quasi-convex optimization problem, which is solved using the bisection method [6,7]. Nevertheless, it is not mentioned that the SSIM – under certain conditions – is a quasi-convex function (see [4]). As a result, it can be minimized using quasi-convex programming techniques, which permits the consideration of a much broader spectrum of SSIM-based optimization problems.

More imaging techniques based on the SSIM can also be found in [14,19]. In these works, optimization of rate distortion, video coding and image classification are explored using the SSIM as a measure of performance.

Note that maximizing $\mathrm{SSIM}(x, y)$ is equivalent to mimizing the function,

$$T(x, y) = 1 - \mathrm{SSIM}(x, y)\,, \tag{2}$$

which may be viewed as a kind of distance function or *dissimilarity* between x and y, i.e., $T(x, y) = 0$ if and only if $x = y$. Many SSIM-based imaging tasks,

including all of the applications mentioned above, may now be expressed in terms of the following optimization problem,

$$\min_x \{T(\Phi(x), y) + \lambda h(x)\}, \tag{3}$$

where Φ is usually a linear transformation. As such, we consider Eq. (3) to define a general set of problems involving *unconstrained SSIM-based optimization*.

In this paper, we introduce a set of algorithms to solve the general problem in (3), in the effort of providing a unified framework as opposed to developing specific methods that address particular applications, which has been the tendency of research literature to date. In particular, we focus our attention on the case in which $h(x)$ is convex. Mathematical and experimental comparisons between ℓ_2 and SSIM approaches are also provided.

Finally, in a future paper we shall address the general complementary problem of *constrained SSIM-based optimization*.

2 The Structural Similarity Index Measure (SSIM)

Structural similarity (SSIM) [18] provides a measure of visual closeness of two images (or local image patches) by quantifying similarities in three fundamental characteristics: luminance, contrast and structure. Luminances are compared in terms of a relative change in means. Contrasts are compared in terms of relative variance. Finally, structures are compared in terms of the correlation coefficient between the two images. The SSIM value is computed by simply taking the product of these changes.

In what follows, we let $x, y \in \mathbb{R}^n$ denote two n-dimensional signal/image blocks. The SSIM between x and y is defined as [18],

$$\text{SSIM}(x, y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right)\left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}\right). \tag{4}$$

Here, μ_x and μ_y denote the mean values of x and y, respectively, and σ_{xy} denotes the cross correlation between x and y, from which all other definitions follow. The small positive constants, C_1, C_2, C_3 provide numerical stability and can be adjusted to accommodate the HVS. Note that $-1 \leq \text{SSIM}(x, y) \leq 1$. Furthermore, $\text{SSIM}(x, y) = 1$ if and only if $x = y$. As such, x and y are considered to be more similar the closer $\text{SSIM}(x, y)$ is to 1.

Setting $C_3 = C_2/2$ leads to the following definition of the SSIM index found in [18] and used in [3] and elsewhere,

$$\text{SSIM}(x, y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right). \tag{5}$$

Since the statistics of images vary greatly spatially, the $\text{SSIM}(x, y)$ is computed using a sliding window of 8×8 pixels. The final result, i.e., the so-called *SSIM index*, is basically an average of the individual SSIM measures.

Definition as a Normalized Metric

In the special case that x and y have equal means, i.e., $\mu_x = \mu_y$, the luminance component of Eq. (5) is unity so that the SSIM becomes

$$\text{SSIM}(x, y) = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \tag{6}$$

A further simplification results when x and y have zero mean, i.e., $\mu_x = \mu_y = 0$. In this special case,

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} x_i y_i \quad \text{and} \quad \sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2. \tag{7}$$

Substitution of these equations into Eq. (6) yields the following simplified formula for the SSIM,

$$\text{SSIM}(x, y) = \frac{2x^T y + C}{\|x\|_2^2 + \|y\|_2^2 + C}, \tag{8}$$

where $C = (n-1)C_2$. For the remainder of this paper, we shall be working with zero mean vectors, so that Eq. (8) will employed in all computations of the SSIM. In this case, the corresponding distance/dissimilarity function $T(x, y)$ in Eq. (2) becomes

$$T(x, y) = 1 - \text{SSIM}(x, y) = \frac{\|x - y\|_2^2}{\|x\|_2^2 + \|y\|_2^2 + C}. \tag{9}$$

Note that $0 \le T(x, y) \le 2$. Furthermore, $T(x, y) = 0$ if and only if $x = y$.

As mentioned earlier, since $\text{SSIM}(x, y)$ is a measure of similarity, $T(x, y)$ can be considered as a measure of dissimilarity between x and y. Eq. (9) is, in fact, an example of a (squared) normalized metric, which has been discussed in [2,4].

3 Unconstrained SSIM-Based Optimization

In [2,4], it was shown that the function $\text{SSIM}(x, y)$ is not convex, but locally quasiconvex. This implies that the unconstrained SSIM-based optimization problem defined in Eq. (3) is, in general, not convex. This, in turn, implies that the existence of a global optimal point cannot be guaranteed. Nevertheless, algorithms that converge to either a local or global minimum can be developed. The algorithm to be used for solving (3) depends on whether the regularizing term $h(x)$ is differentiable or not. We consider these two cases separately below.

3.1 Differentiable $h(x)$

When the regularizing term is differentiable, root-finding algorithms can be employed to find a local zero-mean solution x^* to (3). For example, if Tikhonov regularization is used, we have the following SSIM-based optimization problem,

$$\min_x \{T(Dx, y) + \lambda \|Ax\|_2^2\}, \tag{10}$$

where D is an $m \times n$ matrix. By computing the gradient of (10), we find that the solution x^* must satisfy

$$[(\text{SSIM}(Dx^*, y)D^T D + \lambda(\|Dx^*\|_2^2 + \|y\|_2^2 + C)A^T A]x^* = D^T y. \tag{11}$$

If we define the following function,

$$f(x) = [(\text{SSIM}(Dx, y)D^T D + \lambda(\|Dx\|_2^2 + \|y\|_2^2 + C)A^T A]x - D^T y, \tag{12}$$

then x^* is a (zero-mean) vector in $\mathbb{R}^n$ such that $f(x^*) = 0$.

We may use the Generalized Newton Method [12] to find x^*. From Kantorovich's Theorem, it is known that convergence in any open subset X of Ω, where $\Omega \subset \mathbb{R}^n$, is guaranteed if the initial guess x_0 satisfies the following condition,

$$K\|J_f(x_0)^{-1}\|\|J_f(x_0)^{-1}J_f(x_0)\| \leq \frac{1}{2}. \tag{13}$$

Here, $J_f(\cdot)$ is the Jacobian of $f(\cdot)$, $J_f(\cdot)^{-1}$ denotes its inverse, and $K > 0$ is a constant less or equal than the Lipschitz constant of $J_f(\cdot)$. In fact, it can be proved that for any open subset $X \subset \Omega$, $J_f(\cdot)$ is Lipschitz continuous, that is, there exists a constant $L > 0$ such that for any $x, z \in X$,

$$\|J_f(x) - J_f(z)\|_F \leq L\|x - z\|_2. \tag{14}$$

Here $\|\cdot\|_F$ denotes the Frobenius norm and

$$L = K_1\|D^T D\|_F + \lambda K_2\|A^T A\|_F, \quad K_1, K_2 > 0. \tag{15}$$

From this discussion, and the notation $\mathbf{1} = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$, we propose the following algorithm for solving the problem in Eq. (10).

Algorithm I: Generalized Newton's Method for unconstrained SSIM-based optimization with Tikhonov regularization

---

initialize Choose $x = x_0$ according to (13);
data preprocessing $\bar{y} = \frac{1}{n}\mathbf{1}^T y$, $y = y - \bar{y}\mathbf{1}^T$;
repeat
  $x = x - J_f(x)^{-1}f(x)$;
until stopping criterion is met (e.g., $\|x^{(new)} - x^{(old)}\|_\infty < \epsilon$);
return x, $y = y + \bar{y}\mathbf{1}^T$.

Furthermore, this algorithm can be used for any unconstrained SSIM-based optimization problem by defining $f(\cdot)$ and $J_f(\cdot)$ accordingly.

It is worthwhile to mention that it is not always possible to recover the mean of the non-zero-mean optimal solution $x^\star$. This is because the luminance component of the SSIM is not taken into account. Nevertheless, in some circumstances (e.g., denoising of a signal corrupted by zero-mean additive white

Gaussian noise), the mean of y and $\Phi(x^\star)$ coincide. In this case, we have that $x^\star = x^* + \hat{x}$, where x^* is the zero-mean optimal solution and $\hat{x}$ is a vector such that $D\hat{x} = \bar{y}\mathbf{1}$. If $\Phi(\cdot)$ is any $m \times n$ matrix D, it can be seen that $\hat{x}$ is given by:

$$\hat{x} = \bar{y}(D^T D)^{-1} D^T \mathbf{1}, \tag{16}$$

provided that the inverse of $D^T D$ exists.

3.2 Non-differentiable $h(x)$

In this case, a different approach must be taken. Let us consider the particulary important example $h(x) = \|x\|_1$, i.e., we minimize the following functional

$$\min_x \{T(Dx, y) + \lambda \|x\|_1\}, \tag{17}$$

In this case, the optimal x^* satisfies

$$D^T Dx^* \in \frac{D^T y}{\mathrm{SSIM}(Dx^*, y)} - \lambda \left(\frac{\|Dx^*\|_2^2 + \|y\|_2^2 + C}{2\mathrm{SSIM}(Dx^*, y)} \right) \partial(\|x^*\|_1), \tag{18}$$

where $\partial(\cdot)$ is the sub-gradient operator [11].

To find x^* we employ a coordinate descent approach [15], that is, we minimize (17) along each component of x while the other components are fixed. From (18), for the i-th entry of $x \in \mathbb{R}^n$, the optimal coordinate x_i is given by

$$x_i \in \frac{D_i^T y}{\mathrm{SSIM}(Dx, y)\|D_i^T\|_2^2} - D_i^T Dx_{-i} - \lambda \left(\frac{\|Dx\|_2^2 + \|y\|_2^2 + C}{2\mathrm{SSIM}(Dx, y)\|D_i^T\|_2^2} \right) \partial(|x_i|), \tag{19}$$

where D_i^T is the i-th row of the transpose of D, and x_{-i} is the vector x whose i-th component is set to zero.

The value of x_i can be found by examining the different cases that arise in (19). To begin with, we define

$$\tau_i(x_i) = \lambda \left(\frac{\|Dx\|_2^2 + \|y\|_2^2 + C}{2\mathrm{SSIM}(Dx, y)\|D_i^T\|_2^2} \right) \tag{20}$$

and

$$a_i(x_i) = \frac{D_i^T y}{\mathrm{SSIM}(Dx, y)\|D_i^T\|_2^2} - D_i^T Dx_{-i}\,. \tag{21}$$

Then, $x_i = 0$ if

$$a_i(0) \in \tau_i(0)[-1, 1]. \tag{22}$$

As expected, $x_i > 0$ if $a_i(0) > \tau_i(0)$, so that

$$x_i = a_i(x_i) - \tau_i(x_i). \tag{23}$$

Similarly, we obtain $x_i < 0$ if $a_i(0) < -\tau_i(0)$, in which case x_i is given by

$$x_i = a_i(x_i) + \tau_i(x_i). \tag{24}$$

Notice that when $x_i \neq 0$, we have a result of the form $x_i = g(x_i)$, a non-linear equation that may be solved using either a fixed-point iteration scheme – provided that $g_{\pm}(x_i) = a_i(x_i) \mp \tau_i(x_i)$ is a contraction – or a root-finding algorithm by defining $f_{\pm}(x) = x_i - g_{\pm}(x_i)$. In particular, we follow a fixed-point approach. Moreover, equations (22), (23) and (24) can be combined into the following single operator:

$$\Phi_{\tau_i(0)}(a_i(0)) = \begin{cases} \text{Solve } x_i = a_i(x_i) - \tau_i(x_i), & \text{if } a_i(0) > \tau_i(0), \\ \text{Solve } x_i = a_i(x_i) + \tau_i(x_i), & \text{if } a_i(0) < -\tau_i(0), \\ x_i = 0, & \text{if } |a_i(0)| \leq \tau_i(0)\,. \end{cases} \tag{25}$$

Eq. (25) is an important result since it may be considered an extension of the widely used soft-thresholding (ST) operator [8,16] for the purpose of solving the unconstrained SSIM-based optimization problem (17).

With regard to initial conditions, experimental results show that the optimal ℓ_2 solution of the unconstrained problem $\|Dx - y\|_2^2$ is a good initial guess, i.e., $x_0 = (D^T D)^{-1} D^T y$.

From the above discussion, we introduce the following algorithm to determine the optimal x^* for problem (17).

Algorithm II: Coordinate Descent algorithm for unconstrained SSIM-based optimization with ℓ_1 norm regularization

---

initialize $x = (D^T D)^{-1} D^T y$;
data preprocessing $\bar{y} = \frac{1}{n}\mathbf{1}^T y$, $y = y - \bar{y}\mathbf{1}^T$;
repeat
  for $i = 1$ **to** n **do**
    $x_i = \Phi_{\tau_i(0)}(a_i(0))$;
  end
until stopping criteria is met (e.g., $\|x^{(new)} - x^{(old)}\|_\infty < \epsilon$);
return x, $y = y + \bar{y}\mathbf{1}^T$.

As expected, equation (16) can be used to recover the non-zero mean optimal solution $x^\star$, provided that the means of y and $Dx^\star$ are equal.

4 Experiments

Algorithms I and II can be used for many different SSIM-based applications. In the results presented below, however, we have focussed our attention on the performance of Algorithm II for solving problem (17) when D is an orthogonal transformation. To measure its efficacy, we compare the solutions obtained by the proposed method with the set of solutions of the ℓ_2 version of problem (17), namely,

$$\min_x \left\{ \frac{1}{2}\|Dx - y\|_2^2 + \lambda \|x\|_1 \right\}, \tag{26}$$

which can be solved by means of the ST operator [11,16] if D is an orthogonal matrix.

The experiments reported below were concerned with the recovery of Discrete Cosine Transform (DCT) coefficients. All images were divided into non-overlapping 8×8 pixel blocks, the means of which were subtracted prior to processing. After a block has been processed, its mean is added. Although this procedure is not required for ℓ_2 approaches, it has been performed for the sake of a fair comparison between the two methods.

In Figure 1, the first two plots from left to right corresponds to the average SSIM of all the reconstructions versus the ℓ_0 norm of the recovered coefficients for the test images *Lena* and *Mandrill*. The average SSIM was computed by combining and averaging all the computed SSIMs from all 4096 non-overlapping blocks for both *Lena* and *Mandrill* (both test images have 512×512 pixels). It can be clearly seen that the proposed algorithm outperforms the ℓ_2-based method (ST). This is because minimization of the dissimilarity measure $T(x, y)$ in Eq. (2) is equivalent to maximization of $\mathrm{SSIM}(Dx, y)$, which produces an enhancement in contrast [3]. This effect is demonstrated in the nature of the recovered coefficients. Firstly, the degree of shrinking and thresholding of DCT coefficients by our proposed method is not at strong as ST. Secondly, in some cases, there are DCT coefficients which are thresholded (i.e., set to zero) by the ℓ_2 approach, but kept non-zero by the SSIM-based method for the sake of contrast. These effects are demonstrated in the third and fourth plots in Figure 1. In these two plots, the same block from the image *Lena* was processed, but subjected to two different amounts of regularization.

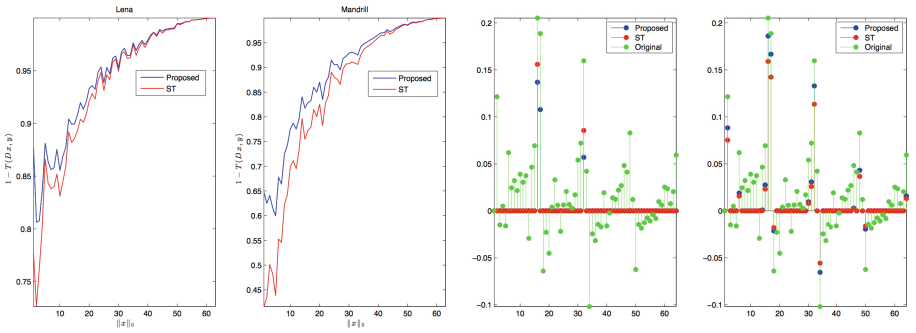

Fig. 1. The first two plots from left to right correspond to the average SSIM versus the ℓ_0 norm of the recovered coefficients for the test images *Lena* and *Mandrill*. In the last two plots, a visual comparison between the original and recovered coefficients from a particular block of the *Lena* image can be appreciated. Regularization is carried out so that the two methods being compared induce the same sparseness in their recoveries. In the two shown examples, the same block was processed but subjected to different amounts of regularization. In particular, the ℓ_0 norm of the set of DCT coefficients that were recovered by both the proposed method and ST is 3 for the first example (third plot), and 15 for the second (fourth plot).

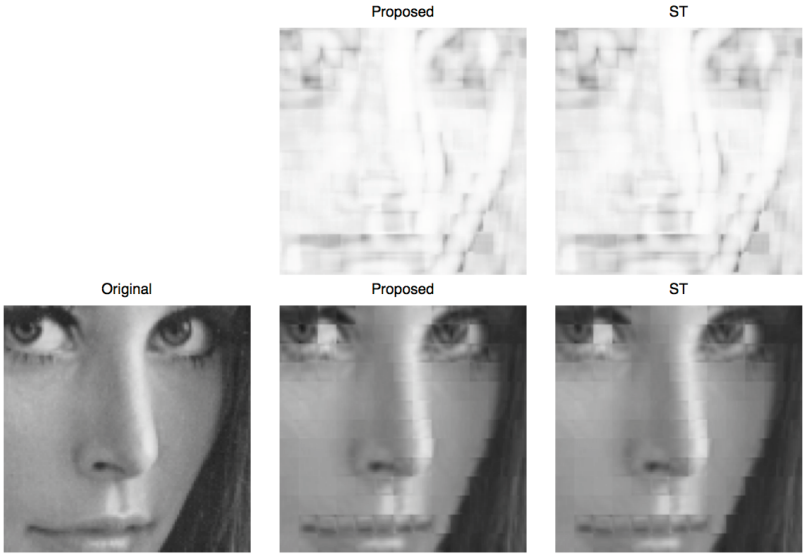

Fig. 2. Visual results for a patch from the test image *Lena*. In all cases, the ℓ_0 norm of the recovered DCT coefficients for each non-overlapping block is 13. In the upper row, the SSIM maps between the reconstructions and the original patch are shown. Reconstructed and original patches can be seen in the lower row. The average $T(Dx, y)$ of all non overlapping blocks for the proposed method is 0.8864, whereas for ST is 0.8609.

In addition, some visual results are shown in Figure 2. In this case, a sub-image from the test image *Lena* was used. The original and recovered images are presented in the bottom row. Regularization was carried out so that the sparsity induced by each method is the same; that is, the ℓ_0 norm of the set of recovered coefficients is 13 in all cases. In the top row of the Figure are shown the SSIM maps that illustrate the similarity between the reconstructions and the original image. The brightness of these maps indicates the degree of similarity between corresponding image blocks – the brighter a given point the greater the magnitude of the SSIM between the retrieved and the original image blocks at that location [18]. It can be seen that the performance of the proposed method and the ℓ_2 approach is very similar. However, the proposed algorithm does perform better than ST in terms of SSIM . This can be seen at some locations in the SSIM maps. For instance, note that the upper left corner of the SSIM map of the proposed method is slightly brighter than the corresponding regions of the other two SSIM maps. This is also evident at other locations. Moreover, the enhancement of contrast is clearly seen when the pupils of the left eyes are compared. With regard to numerical results, the average $T(Dx, y)$ for the ℓ_2 approach is 0.8609, whereas for the proposed method is 0.8864, which is moderately better.

Acknowledgments. This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC).

References

1. Albuquerque, G., Eisemann, M., Magnor, M.A.: Perception-based visual quality measures. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 13–20 (2011)
2. Brunet, D.: A Study of the Structural Similarity Image Quality Measure with Applications to Image Processing. Ph.D. Thesis, Department of Applied Mathematics, University of Waterloo (2012)
3. Brunet, D., Vrscay, E.R., Wang, Z.: Structural similarity-based approximation of signals and images using orthogonal bases. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 11–22. Springer, Heidelberg (2010)
4. Brunet, D., Vrscay, E.R., Wang, Z.: On the mathematical properties of the structural similarity index. Proc. IEEE Trans. Image Processing **21**(4), 1488–1499 (2012)
5. Amir Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences Archive **2**(1), 183–202 (2009)
6. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
7. Channappayya, S.S., Bovik, A.C., Caramanis, C., Heath Jr, R.W.: Design of linear equalizers optimized for the structural similarity index. IEEE Transactions on Image Processing **17**(6), 857–872 (2008)
8. Donoho, D.: Denoising by Soft-Thresholding. IEEE Transactions on Information Theory **41**(3), 613–627 (1995)
9. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least Angle Regression. The Annals of Statistics **32**, 407–451 (2004)
10. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing **15**(12), 3736–3745 (2006)
11. Mairal, J., Bach, F., Jenatton, R., Obozinski, G.: Convex Optimization with Sparsity-Inducing Norms. Optimization for Machine Learning. MIT Press (2011)
12. Ortega, J.M.: The Newton-Kantorovich Theorem. The American Mathematical Monthly **75**(6), 658–660 (1968)
13. Rehman, A., Rostami, M., Wang, Z., Brunet, D., Vrscay, E.R.: SSIM-inspired image restoration using sparse representation. EURASIP J. Adv. Sig. Proc. (2012). doi:10.1186/1687-6180-2012-16
14. Rehman, A., Gao, Y., Wang, J., Wang, Z.: Image classification based on complex wavelet structural similarity. Sig. Proc. Image Comm. **28**(8), 984–992 (2013)
15. Tseng, P., Yun, S.: A Coordinate Gradient Descent Method for Non-smooth Separable Minimization. Journal of Mathematical Programming **117**(1–2), 387–423 (2009)
16. Turlach, B.A.: On algorithms for solving least squares problems under an ℓ_1 penalty or an ℓ_1 constraint. In: Proceedings of the American Statistical Association, Statistical Computing Section, pp. 2572–2577 (2005)
17. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Processing Letters **9**(3), 81–84 (2002)
18. Wang, Z., Bovik, A.C., Sheikh, H.-R., Simoncelli, E.S.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing **13**(4), 600–612 (2004)
19. Wang, S., Rehman, A., Wang, Z., Ma, S., Gao, W.: SSIM-motivated rate-distortion optimization for video coding. IEEE Trans. Circuits Syst. Video Techn. **22**(4), 516–529 (2012)