Exploring the use of gradients in the Structural Similarity image quality measure

Amelia Kunze, Edward R. Vrscay

Abstract In this paper, we investigate if the well-known Structural Similarity image quality measure (SSIM) can be improved by incorporating gradient information. We propose a simple gradient similarity measure which yields results similar to the canonical correlation method. Using the LIVE image database, we show that our proposed gradient-based SSIM exhibits improved performance for degraded images in the mid-to-low quality range.

1 Introduction: Image quality assessment and the MSE

The L^2-based mean squared error (MSE) and its variations continue to be the most widely employed metrics in image processing. This is most probably due to the fact that (1) the MSE is simple to compute and (2) it possesses a number of convenient mathematical properties, including differentiability and convexity. It is well known, however, that these L^2-based measures perform poorly in terms of measuring the visual quality of images. Their failure is partially due to the fact that the L^2 metric does not capture spatial relationships between pixels. This was a motivation for the introduction of the so-called Structural Similarity (SSIM) image quality measure [1] which, along with its variations, continues to be one of the most effective measures of visual quality. The SSIM index measures the similarity between two images by combining three components of the human visual system—luminance, contrast, and structure. In particular, for two corresponding image patches $x, y \in \mathbb{R}^{M \times M}$, the SSIM index is usually defined as the product,

Amelia Kunze
University of Waterloo, Waterloo, Canada e-mail: agkunze@uwaterloo.ca
Current address: Università degli studi dell'Insubria, Varese, Italy e-mail: akunze@uninsubria.it

Edward R. Vrscay
University of Waterloo, Waterloo, Canada e-mail: ervrscay@uwaterloo.ca

$$\begin{aligned}\mathrm{SSIM}(x,y) &= S_1(x,y)\cdot S_2(x,y)\cdot S_3(x,y) \\ &= \frac{2\bar{x}\bar{y}+C_1}{\bar{x}^2+\bar{y}^2+C_1}\;\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\;\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3},\end{aligned} \tag{1}$$

where C_1, C_2, and C_3 are stability constants. Local SSIM values are computed in patches tiling the entire image, then averaged to yield a so-called "mean SSIM" (MSSIM) value. The reader is directed to [1] for a more detailed discussion of the SSIM. For our interests, it is important to note that $S_3(x,y)$ computes the correlation between the patches x and y, and it is our belief, which we thoroughly investigated in [2], that the S_3 term is the most important component of the SSIM.

To illustrate these concepts, we present a particularly compelling failure of the MSE. Consider the so-called "Einstein images"—a set of six 256×256 pixel, 8 bits-per-pixel grayscale images—depicted in Figure 1 below[1]. The images *blur*, *contrast*, *impulse*, *jpg*, and *meanshift* are all perturbations of the reference Einstein image *original*. (A more detailed discussion and analysis of the Einstein images can be found in [2]). For each of these five degradations, the degree of distortion was adjusted to yield nearly equal MSE relative to *original*. Because the degraded images differ significantly and obviously in perceptual quality, they provide a striking example of the failure of the MSE to measure perceptual quality. On the other hand, the MSSIM scores are more indicative of the apparent perceptual quality of the images. The root mean squared error (RMSE) and MSSIM of each of the degraded images relative to *original* are reported alongside the Einstein images in Figure 1.

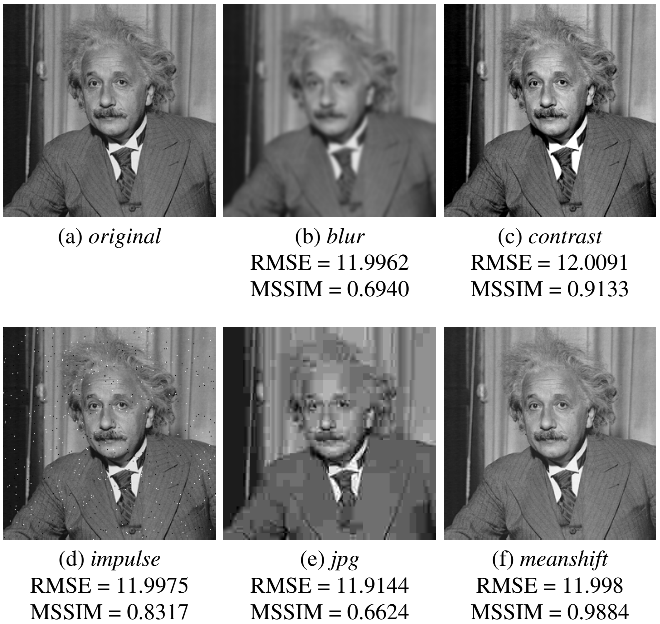
Perhaps more pertinent than the actual MSSIM values are their relative values. The MSSIM scores reported in Figure 1 imply the following ordering of the distorted images in decreasing order of quality:

$$\mathit{meanshift} \;>\; \mathit{contrast} \;>\; \mathit{impulse} \;>\; \mathit{blur} \;>\; \mathit{jpg}. \tag{2}$$

Ideally the ranking in Eq. (2) is consistent with the reader's own subjective preferences of the distorted images in Figure 1.

In this paper, we seek to develop a gradient-based image quality measure. This work is taken from a larger study [2] exploring the use of gradients in imaging, which involves not only image quality measures but also best approximation problems. In the following discussion, many valuable details were necessarily omitted due to space restrictions. Indeed, the interested reader is directed to [2] for a more complete development of our approach. Underlying our study is the question of how sensitive the human visual system could be with respect to changes in the gradient vectors of an image, in terms of either direction or magnitude, or perhaps both. Naturally, we are interested to explore if an image quality measure employing gradients can assess visual quality as well as, or perhaps even better, than current image quality measures, such as the MSSIM, which do not use gradient information.

[1] We are grateful to have obtained these Einstein images from Prof. Z. Wang, Department of Electrical and Computer Engineering, University of Waterloo. To the best of our knowledge, these images were first presented by Wang, Bovik, and Sheikh in [1].



(a) *original*

(b) *blur*
RMSE = 11.9962
MSSIM = 0.6940

(c) *contrast*
RMSE = 12.0091
MSSIM = 0.9133

(d) *impulse*
RMSE = 11.9975
MSSIM = 0.8317

(e) *jpg*
RMSE = 11.9144
MSSIM = 0.6624

(f) *meanshift*
RMSE = 11.998
MSSIM = 0.9884

Fig. 1 The reference Einstein image *original* and its perturbations.

To begin, we first need to decide on the manner in which to compute the image gradients. There are numerous reasonable choices: Indeed, we acknowledge that in the image processing literature, many different formulas—which can be expressed in terms of "filters" operating on the matrices representing the images—are employed. Furthermore, different gradient filters will most probably yield different computational results. That being said, our primary purpose in this paper is only to introduce the idea of using gradients in image quality measures. As such, we will define the gradient using simple forward differences, i.e., the gradient of the image $x \in \mathbb{R}^{N\times M}$ at the $(i,j)^{\text{th}}$ pixel will be defined by,

$$\nabla x_{ij} = (x(i+1,j) - x(i,j), x(i,j+1) - x(i,j)). \tag{3}$$

We assume an even extension of the image, so that at the edges of the image we have $x(N+1,j) = x(N,j)$ and $x(i,M+1) = x(i,M)$. Note that this produces a zero value for the appropriate component of the gradient vector.

The following simple experiment motivates the application of gradient information for image quality assessment. For those Einstein images in Figure 1, we compute the usual L^2 distance *between gradients*. Specifically, for the *original* image x and each of its five perturbations y, we compute the following distances,

$$\|\nabla x - \nabla y\|_2 = \frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\|\nabla x_{ij} - \nabla y_{ij}\|_2^2\right]^{1/2}, \tag{4}$$

where $N = 256$ and the ∇x and ∇y are computed using the forward differences previously described. These results are presented in Table 1. (Because the *meanshift* image is produced from *original* by merely adding a constant to the greyscale values of the latter, the gradients of the two images are identical. Hence, one would expect the *meanshift* RMSE in Table 1 to be exactly 0. The observed deviation is due to a few pixels whole *meanshift* values are restricted by limits on the greyscale range.)

blur	*contrast*	*impulse*	*jpg*	*meanshift*
17.0349	5.5426	23.9110	17.6135	0.1309

Table 1 RMSE between the gradients of the *original* Einstein image and its perturbations.

We can immediately observe a significant deviation in these values. In other words, the structural information encoded in our simplistic gradient allows the MSE, which does not otherwise consider spatial relationships between pixels, to differentiate between the Einstein images. It is clear that computing the MSE *between gradients* already offers a marked improvement over the MSE.

Letting the gradient distances define an ordering of image quality, we obtain,

$$\textit{meanshift} \;>\; \textit{contrast} \;>\; \textit{blur} \;>\; \textit{jpg} \;>\; \textit{impulse}. \tag{5}$$

Unfortunately, this is not in agreement with the ordering dictated by the MSSIM as reported in Eq. (2). However, observe that the only difference between Eq. (2) and Eq. (5) is the placement of the *impulse* image. Indeed, it is easy to understand why impulse noise is particularly bothersome to the gradient distance: Impulse noise strongly affects not only the gradient at each contaminated pixel, but also the gradients of its adjacent neighbours. It is possible that some preprocessing methods, such as blurring or downsampling of the images, could mitigate this effect, perhaps even to the extent that one could retrieve a ranking of gradient distance in agreement with the ordering according to the MSSIM. That being said, we will instead proceed by pursuing other, and hopefully better, ways of computing gradient similarity.

2 Computing gradient similarity

Our foremost source of inspiration on which to model a gradient similarity measure is the correlation. Indeed, in the same way that the SSIM index computes the correlation between two $M \times M$ patches x and y, one might be tempted to compute the correlation between their gradients ∇x and ∇y. However, ∇x and ∇y will each be an $M \times M$ block of vectors, i.e., each component of the $M \times M$ matrix will be a 2-vector as written in Eq. (3). There is a proper way to compute the correlation between vectors of $M \times M$ blocks, namely, the "canonical correlation" method introduced by Hotelling in 1936 [5]. However, this method is somewhat expensive computationally, which led us to examine the effectiveness of simpler methods to compute the correlation. A number of rather simple methods were explored in [2]. Here we shall simply report on the "best" such method, which generally produced results quite comparable to the canonical correlation method.

Our "best" measure of gradient similarity is computed as follows: For two corresponding image patches, let a denote the correlation between the x-components of the gradients and b denote the correlation of the y-component of the gradients. Then compute the "normalized magnitude" of the vector (a,b), i.e., the quantity

$$S_4(a,b) = \frac{1}{\sqrt{2}}(a^2+b^2)^{1/2}, \qquad \text{noting that} \quad S_4(a,b) \in [0,1]. \tag{6}$$

Table 2 reports the values of the normalized magnitude $S_4(a,b)$ for each perturbed Einstein image relative to *original*. In each case, these values correspond to using non-overlapping image patches of size 32 × 32 pixels. If we let these normalized magnitude values dictate the relative quality of the images, we retrieve the ranking imposed by the MSSIM, i.e.,

blur	*contrast*	*impulse*	*jpg*	*meanshift*
0.4391	0.9964	0.5573	0.3571	1.0000

Table 2 Average $S_4(a,b)$ values of the *original* Einstein image and its degradations.

$$\textit{meanshift} \;>\; \textit{contrast} \;>\; \textit{impulse} \;>\; \textit{blur} \;>\; \textit{jpg}. \tag{7}$$

(In [2], we show that the ordering in Eq. (7) is preserved for various patch sizes.) A comparison of the MSSIM values and those S_4 values in Table 2 reveals that the gradient correlation is much more punitive than the usual MSSIM. However, using this small set of Einstein images alone we cannot conclude whether our punitive S_4 is performing "better" than the MSSIM. In the next section, we will use a larger set of images, i.e., the LIVE image database [6], to attempt to answer this question. (In the process, we will need to define what we mean by "better" performance.)

3 Our quest for a gradient-based image quality measure

The LIVE image database contains a total of 982 reference and distorted images. Each image is associated with a unique "difference mean opinion score" (DMOS) indicating its perceptual quality. (Refer to [2] or [7] for a detailed discussion of the database and the method used to obtain the scores.) DMOS $= 0$ indicates perfect perceptual quality, i.e., the associated image has no visible distortions, while DMOS increases as the presence of distortions becomes more visible and bothersome.

In Figure 2 (a), we plot MSSIM against DMOS for all images in the LIVE database. Here, we have adopted the convention in the literature which puts DMOS along the vertical axis. The scatterplot includes a curve of best fit according to the expected nonlinear relationship provided in [7] and which was computed using the Matlab function “fitnlm”. To provide an indication of how well a given image quality measure, e.g., the MSSIM, is performing, we measure the scatter of the points relative to the fitted curve in the following manner: For all images j in the LIVE database, we compute the distance,

$$\mathrm{dist} = \left[\sum_{j=1}^{982} (\mathrm{DMOS}(j) - \mathrm{Quality}(\mathrm{MSSIM}(j)))^2\right]^{1/2}, \tag{8}$$

where $\mathrm{Quality}(x)$ denotes the value of the fitted curve at a point x. Because the curve is obtained through least squares regression, our “dist” measure can be thought of as the variance between the fitted curve and the DMOS values as a function of the algorithm score. This “dist” value is provided in the title of Figure 2 (a).

The data points in Figure 2 (a) are quite concentrated about the curve at the bottom right region of the plot, which may be considered as the “low DMOS” region (i.e., DMOS under 20) or “high MSSIM” region (i.e., MSSIM near 1). As we move leftward and upward, however, the data points are distributed more diffusely about the regression curve. This leads us to think of one possible criterion for “improving” the MSSIM, namely, decreasing the diffusiveness of the data points in the lower DMOS region. With this goal in mind, we propose the following,

$$\mathrm{gradSSIM}(x,y) = \mathrm{SSIM}(x,y) \cdot S_4(x,y) = S_1(x,y) \cdot S_2(x,y) \cdot S_3(x,y) \cdot S_4(x,y), \tag{9}$$

where, in keeping with the spirit of the SSIM, our “normalized magnitude” S_4 has been added as a fourth multiplicative term. Recall that our S_4 term is defined in terms of the correlations (i.e., quotients) a and b. In both a and b, we would like to include a stability constant C_4 to keep the denominator away from 0. This raises the question of which value should be taken by the stability constant C_4.

In the usual SSIM, the stability constants C_1, C_2, and C_3 are included in *both* the numerator and denominator of the S_1, S_2, and S_3 terms, respectively. The suggested values for these constants, provided by Wang et al. in [8], are $C_1 = 6.5$, $C_2 = 58.5$ and $C_3 = 29.2$. In this spirit, one possibility is to include C_4 in both the numerator and denominator of the correlations a and b. Because we think of these correlations as gradient-based analogues of the S_3 component, we set C_4 (practically) equal to C_3, i.e., we let $C_4 = 30$. This reasoning is justified by our observation that the gradient correlations are similar in magnitude to the regular correlations for the Einstein images. This S_4 formulation is plotted against DMOS in Figure 2 (b).

However, [8] describes C_1, C_2 and C_3 as “small” constants to protect against numerical instabilities in the denominator; Given this understanding, their suggested values are significantly larger than we would expect. Moreover, we are unable to find in the literature any clear description of how or why these suggested values were obtained. Our instinct is to include a constant $C_4 \ll 1$ only in the denominators of a and b. The result of using this formulation is shown in Figure 2 (c).

The algorithm scores are generally significantly increased from Figure 2 (a) to Figure 2 (b), which indicates that having the relatively high value $C_4 = 30$ in both the

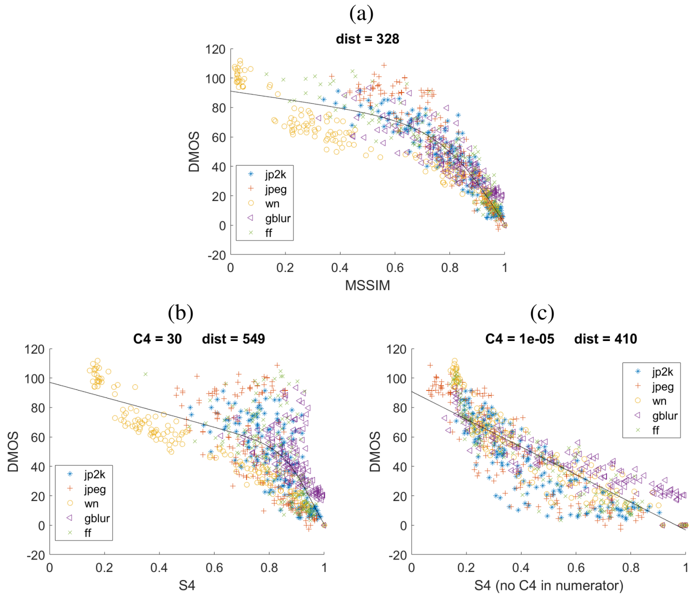


Fig. 2 For all LIVE images, we plot (a) SSIM and our S_4 measure with (b) $C_4 = 30$ in the numerator and denominator and (c) $C_4 = 10^{-5}$ in the denominator only. The legend denotes distortion type: "jp2k" and "jpeg" for JPEG 2000 and JPEG compression, respectively, "wn" for white noise, "gblur" for Gaussian blur, and "ff" for transmission errors in the JPEG 2000 bit-stream.

numerator and denominator artificially pushes the quotients a and b towards 1. On the other hand, the points in Figure 2 (c) are reasonably close to the fitted curve along its entire length. We are inclined to use the S_4 with $C_4 \ll 1$ in the denominator only for our "gradSSIM" measure. This choice, although a departure from conventions set by the SSIM, is further supported by Figure 3 which shows that this S_4 is highly correlated with the aforementioned "rigorous" canonical correlation. Indeed, results virtually identical to those presented below in Figure 4 and Figure 5 are obtained if one replaces our simple S_4 term with the canonical correlation.

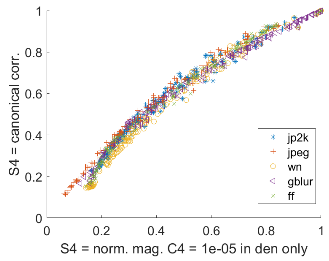


Fig. 3 Our preferred S_4 is highly correlated with the canonical correlation.

The result of computing the corresponding gradSSIM, as defined in Eq. (9), for the LIVE images is shown in Figure 4 (b). In Figure 4 (a) we have included the usual MSSIM for comparison. Here we also flip the orientation of the axes, i.e, we now consider DMOS to be the independent variable. To the best of our knowledge, this understanding has not been adopted elsewhere in the literature. However, because the DMOS scores are obtained from subjective experiments, they are, or at least could be considered to be, the independent values. Given this understanding, it is actually the scatter in the algorithm scores relative to the curve that characterizes goodness of fit. We use a simple quadratic function to fit the "flipped" data and denote the "flipped" distance by "distDMOS" to distinguish it from the "dist" value computed previously.

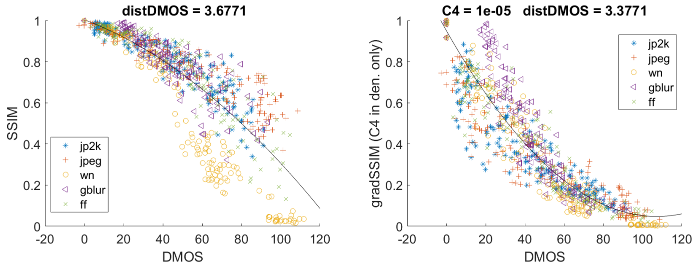


Fig. 4 The "flipped" SSIM and the gradSSIM using our preferred S_4.

The gradSSIM in Figure 4 (b) has a much improved fit in the high DMOS region compared to the MSSIM in Figure 4 (a). Unfortunately, this success is balanced by a loss in the low DMOS region where the gradSSIM is exhibiting its diffuseness. Still, as the "distDMOS" values indicate, there appears to be slightly less overall spread in Figure 4 (b) than in Figure 4 (a). Compared to the MSSIM, the gradSSIM offers a better choice for applications dealing with heavily degraded images. However, according to a subjective evaluation performed in [2], the low-to-mid DMOS region likely captures reasonable distortion levels for most practical applications. In this light, a gain in the high DMOS region may not be worthwhile if it comes at the expense of the good fit for low DMOS.

Looking at Figure 4, it would be most desireable to construct an image quality measure which behaves like the MSSIM for low DMOS and the gradSSIM for high DMOS. In [2], we explored some possible "blended" measures towards that end. Ultimately, we arrive at the following proposal involving SSIM-based exponents: For two image patches x and y, we define the following "blended" local similarity function,

$$\mathrm{gradSSIM1}(x,y) = \mathrm{SSIM}(x,y) \cdot S_4(x,y)^{1-\mathrm{SSIM}(x,y)}. \tag{10}$$

For SSIM near 1, $S_4(x,y)^{1-\mathrm{SSIM}(x,y)} \approx 1$. For SSIM near 0, $S_4(x,y)^{1-\mathrm{SSIM}(x,y)} \approx S_4$. In this way, we reduce the effect of S_4 in the low DMOS region, where it exhibits a great deal of scatter, and increase the effect of S_4 in the high DMOS region, where it is more successful than the MSSIM. If the effect of the S_4 term is not sufficiently strong for larger DMOS, we may prefer the following definition,

$$\mathrm{gradSSIM1}(x,y) = \mathrm{SSIM}(x,y) \cdot S_4(x,y)^{1-\mathrm{SSIM}(x,y)^2}. \tag{11}$$

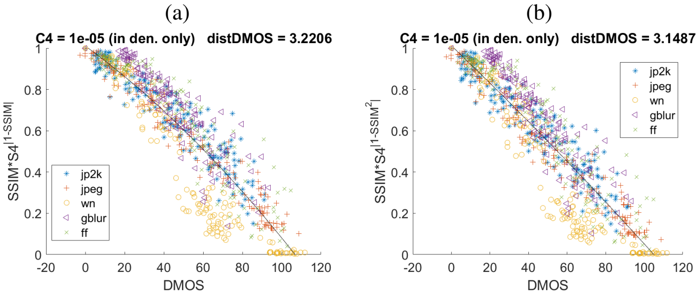


Fig. 5 We plot both versions our the gradSSIM1, as defined in (a) Eq. (10) and (b) Eq. (11).

The results of computing the gradSSIM1 as written in Eq. (10) and Eq. (11) are shown in Figure 5. Both formulations appear to be working well, with little difference between the two plots. According to the "distDMOS" value, Eq. (11) is performing slightly better. In general, the gradSSIM1 improves the scatter in the mid-to-high DMOS without sacrificing the good fit for low DMOS.

For the heavily distorted images with high DMOS, the attempt to cluster the data points may be considered as more of an academic exercise than one of practical value. However, it may still be of practical concern to improve the fit in the mid-DMOS range, i.e., 40 to 80. Indeed, the great majority of mid-DMOS points lie above the fitted curve in the MSSIM plot shown in Figure 4 (a). In [2], we showed that the placement of these points is largely invariant to changes in the stability constants C_1, C_2, and C_3. On the other hand, our gradSSIM1 can produce smaller values for both the mid-DMOS and high DMOS range which, on the basis of our subjective evaluation performed in [2], are warranted.

4 Conclusion

The goal of this paper was to investigate if the MSSIM could be improved by incorporating gradient information. By adding a simple gradient similarity measure, we demonstrated that the resulting "gradSSIM1" improves the fit in the mid-to-high DMOS range without sacrificing much in terms of fit in the low DMOS range. In future studies, we suggest that DMOS be considered as the independent variable.

Other gradient-based similarity measures have been proposed in the literature. In [3], an SSIM-like gradient similarity measure seeks to rate the presence of edges using gradient operators. Edge maps are obtained by Haar wavelet filters and subsequently compared in [4]. However, the rather complicated mathematical machinery employed in both [3] and [4] makes it difficult to understand exactly how the edge data informs the resulting similarity index. It is unclear to what degree these complicated methods of collecting and aggregating gradient information reflect an honest *matching* of the gradient vectors. Our approach based on first principles is mathematically tractable and clearly seeks to *match* the image gradient vectors.

Another novelty in our approach lies in its ability to seamlessly blend the behaviour of different measures using SSIM-based exponents. Our study led us to question the demand that a single mathematical formula should accommodate all types of distortions across the entire quality spectrum. Throughout our work, we speculated that the human visual system may access a variety of processes when judging different aspects of an image. This thought led us to consider a new way of thinking about image quality measures as “blended” formulas which might be more aligned with how the human visual system aggregates information. In this way, our approach introduces a new way of considering image quality measures. It is our hope that this novel “blended” framework will stimulate future research in this area.

Acknowledgments

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant (ERV). AK would also like to thank the Faculty of Mathematics, University of Waterloo for partial financial support.

References

1. Z. Wang, A. C. Bovik, H. R. Sheikh. *Digital Video Image Quality and Perceptual Coding*, Chapter 7: Structural Similarity Based Image Quality Assessment. CRC Press, 2005.
2. A. Kunze. An investigation of the use gradients in imaging, including best approximiton and the Structural Similiarity image quality measure. Master’s Thesis, Department of Applied Mathematics, University of Waterloo, 2023.
3. A. Liu, W. Lin, and M. Narwaria. Image Quality Assessment Based on Gradient Similarity. *IEEE Transactions on Image Processing*, 21(4):1500-1512, 2012.
4. R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand. A Haar Wavelet-based Perceptual Similarity Index for Image Quality Assessment. *Signal Processing: Image Communication*, 61:33-43, 2018.
5. H, Hotelling. Relations Between Two Sets of Variates. *Biometrika*, 28(3-4):321-277, 1936.
6. H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. LIVE Image Quality Assessment Database Release 2. http://live.ece.utexas.edu/research/quality
7. H. R. Sheikh. *Image Quality Assesment Using Natural Scene Statistics.* PhD Thesis, The University of Texas as Austin, 2004.
8. Z. Wang, A. C. Bovik, and H. R. Sheikh, *Digital Video Image Quality and Perceptual Coding*, Chapter 7: Structural Similarity-Based Image Quality Assessment. CRC Press, 2005.