Deviation Based Pooling Strategies For Full Reference Image Quality Assessment

Hossein Ziaei Nafchi, Rachid Hedjam, Atena Shahkolaei and Mohamed Cheriet
École de Technologie Supérieure
{hossein.zi@synchromedia.ca, rachid.hedjam@mcgill.ca, a.shahkolaei@yahoo.com, mohamed.cheriet@etsmtl.ca}

Abstract

The state-of-the-art pooling strategies for perceptual image quality assessment (IQA) are based on the mean and the weighted mean. They are robust pooling strategies which usually provide a moderate to high performance for different IQAs. Recently, standard deviation (SD) pooling was also proposed. Although, this deviation pooling provides a very high performance for a few IQAs, its performance is lower than mean poolings for many other IQAs. In this paper, we propose to use the mean absolute deviation (MAD) and show that it is a more robust and accurate pooling strategy for a wider range of IQAs. In fact, MAD pooling has the advantages of both mean pooling and SD pooling. The joint computation and use of the MAD and SD pooling strategies is also considered in this paper. Experimental results provide useful information on the choice of the proper deviation pooling strategy for different IQA models.

1. Introduction

Automatic image quality assessment (IQA) plays a significant role in many image processing applications. IQA is commonly used for monitoring, benchmarking, image restoration and parameter optimization [14, 18]. Full reference IQAs, which fall within the scope of this paper, evaluate the perceptual quality of a distorted image with respect to its reference image. IQAs mimic the average quality predictions of human observers. This is a non-trivial task because images may suffer from various types and degrees of distortions.

Among IQAs, the mean squared error (MSE) is widely used because of its simplicity. However, in many situations, it does not correlate with human perception of image fidelity and quality [13]. A number of popular and/or high performance IQAs are SSIM [14], MS-SSIM [17], VIF [11], VSNR [2], MAD [4], FSIM [23], GS [7], GMSD [20], VSI [21], and others [5].

Usually, IQAs measure local similarity and produce a similarity score by a pooling strategy. This local quality measurement can be performed on the image, different representations of the original image, or their combination. For example, SSIM and MSSSIM use statistics of smoothed source and distorted images. FSIM [23] uses a phase-derived map and another gradient-based map for quality assessment. GS [7] is a contrast and structure variant metric that utilizes specialized gradient magnitude and image contrast of the image. GMSD [20] also utilizes gradient magnitude. Many of the available full reference IQAs follow this top-down architecture [5]. While average pooling [14, 17, 23, 7] and average weighted pooling [8, 15, 21] are widely used in the literature, GMSD uses standard deviation pooling [20]. For the purpose of this paper, we will take an overview of the FSIM [23] and GMSD [20] indices.

The feature similarity (FSIM) index [23] uses phase congruency (PC) [3] as its main feature, and an image gradient magnitude $G$ as its secondary feature. Phase congruency similarity $S_{PC}$ and gradient magnitude similarity $S_G$ are calculated and then combined as $S_L(x) = [S_{PC}(x)]^\alpha [S_G(x)]^\beta$, where $x$ is an image pixel. For the purpose of pooling, first, the maximum PC of reference and distorted images $PC_m$ is computed. FSIM is then computed by the following mean pooling:

$$FSIM = \frac{\sum_x (S_L(x)PC_m(x))}{\sum_x PC_m(x)}. \quad (1)$$

FSIM is among the leading indices in the literature; however, the high computation cost of the phase congruency makes it an inefficient index.

GMSD uses the Prewitt operator to calculate gradient magnitudes of reference and distorted images, $G_R$ and $G_D$. From these, a gradient magnitude similarity (GMS) is calculated by:

$$GMS(x) = \frac{2G_R(x)G_D(x) + c}{G_R^2(x) + G_D^2(x) + c}. \quad (2)$$

where $c$ is a positive constant that supplies numerical stability. The GMSD is then calculated by a deviation pooling strategy, which is the standard deviation of GMS values. GMSD provides high performance for different datasets and is very efficient.
2. Deviation pooling strategies

Deviation pooling for IQAs is rarely used in the literature, except the standard deviation used in GMSD [20]. Deviation is the variation of data values compared to a measure of central tendency (MCT) such as the mean, median, or mode. A deviation can be seen as the Minkowski distance of order ρ between vector x and a MCT:

\[
\hat{D}(x, \text{MCT})^\rho = \left( \frac{1}{N} \sum_{i=1}^{N} |x_i - \text{MCT}|^\rho \right)^{1/\rho}. \tag{3}
\]

where, \( \rho \geq 1 \) indicates the type of deviation. For the purpose of this paper, it can be seen as:

\[
\hat{D}(x, \text{MCT})^\rho = \left( \frac{1}{N} \sum_{i=1}^{N} |x_i - \text{MCT}|^\rho \right)^{1/\rho}. \tag{4}
\]

Since above equation includes a MCT, it is different than the Minkowski pooling [16], and the Minkowski metric [11]. In the following, the standard deviation pooling strategy introduced in GMSD is revisited. We then suggest using the mean absolute deviation pooling and show that these two pooling strategies can be jointly calculated.

2.1. Standard deviation (SD) pooling

Standard deviation is a simple and very common statistical measure of the spread of scores within a set of data. Let \( \text{LS}^M \) denote the mean of an arbitrary local similarity (LS) map computed by an IQA model. Its standard deviation can be computed by:

\[
\text{LS}^{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{LS}(i) - \text{LS}^M)^2}. \tag{5}
\]

which is equal to equation (4) when \( \rho = 2 \).

2.2. Mean absolute deviation (MAD) pooling

Given the \( \text{LS}^M \), the mean absolute deviation \( \text{LS}^{MAD} \) of an LS map is calculated by:

\[
\text{LS}^{MAD} = \frac{1}{N} \sum_{i=1}^{N} |\text{LS}(i) - \text{LS}^M|. \tag{6}
\]

which is also equal to equation (4) when \( \rho = 1 \). In the experimental results section, the performance of the MAD pooling based indices is evaluated.

2.3. Double deviation (DD) pooling

We also found that a combination of the SD and MAD pooling strategies provides higher performance than either \( \text{LS}^{SD} \) or \( \text{LS}^{MAD} \) for some IQAs. Furthermore, these two poolings can be computed at the same time using the following formulas:

\[
D_i = |\text{LS}(i) - \text{LS}^M|, \tag{7}
\]

\[
\text{LS}^{MAD} = \frac{1}{N} \sum_{i=1}^{N} D_i. \tag{8}
\]

\[
\text{LS}^{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} D_i^2}. \tag{9}
\]

The joint computation of the \( \text{LS}^{SD} \) and \( \text{LS}^{MAD} \) is likewise computationally efficient. We call this combination the double deviation DD pooling strategy and compute it by:

\[
\text{LS}^{DD} = \alpha \text{LS}^{SD} + (1 - \alpha) \text{LS}^{MAD}. \tag{10}
\]

where \( 0 \leq \alpha \leq 1 \) adjusts the relative importance of the \( \text{LS}^{SD} \) and \( \text{LS}^{MAD} \) indices. We are able to combine these two pooling strategies because they share almost the same statistical characteristics and values. The usefulness of this combination is verified in the experiments. It should be noted that the median absolute deviation does not provide satisfactory performance, hence it is not evaluated in this paper.

3. Experimental results

In the experiments, four standard datasets are used. The LIVE dataset [12] contains 29 reference images and 779 distorted images of five categories. The TID 2008 [10]
dataset contains 25 reference images and 1700 distorted images. For each reference image, 17 types of distortions of 4 degrees are available. CSIQ [4] is another dataset that consists of 30 reference images; each is distorted using different types of distortions at four to five levels of distortion. We also used the TID 2013 [9] dataset, which contains 25 reference images and 3000 distorted images. For each reference image, 24 types of distortions of 5 degrees are available. Also, three popular evaluation metrics were used in the experiments: the Spearman Rank order Correlation Coefficient (SRC), the Pearson linear Correlation Coefficient (PCC), and the Root Mean Square Error (RMSE).

MSE, SSIM [14], GS [7, 6], FSIM [23, 22], GMSD [20, 19], and VSI [21] were used in the experiments using two to three deviation pooling strategies. Tables 1 and 2 provide a performance comparison between different indices using different pooling strategies. In reference to Fig. 1 if DD pooling provides higher performance than the others, its performance is added to Tables 1 and 2. Also, some of the state-of-the-art indices are added to the end of Tables 1 and 2. In Figs. 1 and 2, α = 0 postulates to the M AD pooling, and α = 1 postulates to the SD pooling. In fact, Figs. 1 and 2 show SRC and PCC performance variations of different indices utilizing the DD pooling.

From Tables 1 and 2, it can be seen that MSE M AD performs better than the MSE, while the MSE SD shows the lowest performance among them. The original SSIM outperforms its deviation-based versions. The M AD pooling has higher performance than the SD pooling for SSIM. For SSIM, the deviation pooling shows very low performance for some of the distortions, while the M AD pooling is still more robust than the SD pooling.

For GS, the M AD pooling outperforms the others. SD pooling for GS does not provide a high performance because GS uses image contrast, and we had already observed that SD pooling is not a good choice for MSE. While FSIM SD shows overall higher performance than FSIM M AD, its performance for some distortion types is low. For FSIM, DD pooling with α = 0.5 provides higher overall performance than SD pooling. At the same time, DD pooling shows better quality prediction on distortion types than SD pooling. The overall performance of the GMS M AD and GMS SD indices is competitive; however, GMS M AD shows better quality prediction on the distortion types. The overall performance of GMS DD and its performance for

![Figure 1. The impact of the parameter α on the weighted average SRC (Table 1) performance of the double deviation DD pooling strategy.](image)

![Figure 2. The impact of the parameter α on the weighted average PCC (Table 1) performance of the double deviation DD pooling strategy. Note the MSE curve which is nonlinear.](image)

| RQA | TID2013 SRC (distortions) | SRC PCC | avg | min | std |
|-----|--------------------------|--------|-----|-----|-----|
| MSE | 0.6394 0.4765 | 0.7951 0.0766 | 0.2355 |
| MSE SD | 0.6164 0.5118 | 0.7899 0.0952 | 0.2282 |
| MSE M AD | 0.6791 0.7079 | 0.7897 0.0852 | 0.2413 |
| SSIM | 0.7411 0.7895 | 0.8075 0.3775 | 0.1521 |
| SSIM SD | 0.7292 0.7602 | 0.7993 0.0045 | 0.2257 |
| SSIM M AD | 0.7463 0.7806 | 0.8041 0.1471 | 0.1989 |
| VSI | 0.7946 0.8464 | 0.8351 0.3578 | 0.1521 |
| GS SD | 0.7801 0.8232 | 0.8284 0.3139 | 0.1731 |
| GS M AD | 0.8081 0.8613 | 0.8332 0.3344 | 0.1584 |
| FSIM | 0.8015 0.8569 | 0.8219 0.2748 | 0.1662 |
| FSIM SD | 0.8077 0.8602 | 0.8263 0.2126 | 0.1940 |
| FSIM M AD | 0.8051 0.8547 | 0.8327 0.2691 | 0.1706 |
| FSIM DD (α = 0.5) | 0.8118 0.8614 | 0.8293 0.2304 | 0.1852 |
| GMS M | 0.7884 0.8395 | 0.8370 0.3700 | 0.1500 |
| GMSD [20] | 0.8044 0.8590 | 0.8300 0.2948 | 0.1815 |
| GMS M AD | 0.8084 0.8608 | 0.8369 0.3450 | 0.1597 |
| GMS DD (α = 0.5) | 0.8111 0.8658 | 0.8332 0.3132 | 0.1729 |
| VSI SD | 0.8965 0.9000 | 0.8514 0.1713 | 0.1787 |
| VSI M AD | 0.8556 0.8651 | 0.8633 0.3935 | 0.1274 |
| VSI DD (α = 0.1) | 0.8853 0.8965 | 0.8675 0.4564 | 0.1193 |
| MSSSIM [7] | 0.7869 0.8329 | 0.8109 0.4099 | 0.1306 |
| VIF [15] | 0.6769 0.7720 | 0.8267 0.3099 | 0.1464 |
| IWSSIM [13] | 0.7779 0.8319 | 0.7978 0.3717 | 0.1601 |
| MAD [1] | 0.7807 0.8267 | 0.7556 0.0575 | 0.2644 |
distortion types are simultaneously higher than GMS $SD$.

VSI is a high performance similarity index that was recently proposed in [21]. Using $MAD$ pooling, its performance improved considerably on the first three datasets for all of the measures used in this paper. For the TID 2013 dataset, however, its overall performance decreased by 1.2493% for SRC and by 0.3889% for PCC metrics. In turn, it has 1.8910% better average prediction on distortion types. Also, the minimum quality prediction of VSI improved from 0.1713 to the 0.4564. These advantages show that $MAD$ pooling is a good choice for VSI. $SD$ pooling shows the lowest performance for VSI.

Overall, $MAD$ pooling is more robust than $SD$ pooling, especially for assessment of individual distortion types. The low min SRC values for $SD$ pooling in Table 1 show its unreliability in comparison to $MAD$ pooling. In general, higher orders of $\rho$ in equation (4) result in a worst and unstable assessment for distortion types. In other words, the std value in the last column of the Table 1 increases by increasing the $\rho$ value. It is worth noting that this fact may not always be true.

Fig. 3 shows an example in which $SD$ pooling fails in assessment, while $MAD$ pooling provides a true assessment for both the GS and GMS indices. Fig. 4 shows the run time of the three pooling strategies used in this paper. Our experiments were performed on a Core i7 3.4 GHz CPU with 16 GB of RAM running on MATLAB 2013b and Windows 7. $MAD$ is the second fastest after the mean, while the joint calculation of $SD$ and $MAD$ is still efficient. Therefore, GMS $MAD$ is even faster than the highly efficient GMSD. GMS $DD$ is slightly slower than GMSD; however, its improved performance over GMS $SD$ is noticeable.
4. Conclusion

Deviation pooling strategies for full reference image quality assessment were analyzed. The mean absolute deviation (MAD) pooling and the standard deviation (SD) pooling strategies were compared on the basis of their effectiveness, robustness and efficiency. The computation of MAD is faster than SD, and this may be of high interest for designing more efficient indices. While none of them fully outperformed the others, MAD pooling shows a clear advantage of robustness over SD pooling. Furthermore, for some of the image quality assessment models, a combination of these two pooling strategies results in better performing indices. Considering the experimental results, we highly recommend the use of MAD pooling for different image quality assessment purposes.

References

[1] A. Bovik, editor. Handbook of Image and Video Processing. Academic Press, 2000.

[2] D. Chandler and S. Hemami. VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images. IEEE Transactions on Image Processing, 16(9):2284–2298, Sept 2007.

[3] P. Kovesi. Image features from phase congruency. Videre: Journal of Computer Vision Research, 1:1–26, 1999.

[4] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. Journal of Electronic Imaging, 19(1):011006, 2010.

[5] W. Lin and C.-C. J. Kuo. Perceptual visual quality metrics: A survey. Journal of Visual Communication and Image Representation, 22(4):297 – 312, 2011.

[6] A. Liu, W. Lin, and M. Narwaria. Gradient similarity index. Online: http://www.ntu.edu.sg/home/wslin/GSM.zip

[7] A. Liu, W. Lin, and M. Narwaria. Image quality assessment based on gradient similarity. IEEE Transactions on Image Processing, 21(4):1500–1512, April 2012.

[8] A. Moorthy and A. Bovik. Visual importance pooling for image quality assessment. IEEE Journal of Selected Topics in Signal Processing, 3(2):193–201, April 2009.

[9] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Kuo. Color image database TID2013: Peculiarities and preliminary results. In 2013 4th European Workshop on Visual Information Processing (EUVIP), pages 106–111, June 2013.

[10] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics. Advances of Modern Radioelectronics, 10:30–45, 2009.

[11] H. Sheikh and A. Bovik. Image information and visual quality. IEEE Transactions on Image Processing, 15(2):430–444, Feb 2006.

[12] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik. Live image quality assessment database release 2. Online: http://live.ece.utexas.edu/research/quality

[13] Z. Wang and A. Bovik. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. IEEE Signal Processing Magazine, 26(1):98–117, Jan 2009.

[14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, April 2004.

[15] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. IEEE Transactions on Image Processing, 20(5):1185–1198, May 2011.

[16] Z. Wang and X. Shang. Spatial pooling strategies for perceptual image quality assessment. In IEEE International Conference on Image Processing, pages 2945–2948, Oct 2006.

[17] Z. Wang, E. Simoncelli, and A. Bovik. Multiscale structural similarity for image quality assessment. In Thirty-Seven Asilomar Conference on Signals, Systems and Computers, volume 2, pages 1398–1402, Nov 2003.

[18] S. Winkler. Analysis of public image and video databases for quality assessment. IEEE Journal of Selected Topics in Signal Processing, 6(6):616–625, Oct 2012.

[19] W. Xue, L. Zhang, X. Mou, and A. Bovik. Gradient magnitude similarity deviation. Online: http://www4.comp.polyu.edu.hk/~cslzhang/IQA/GMSD

[20] W. Xue, L. Zhang, X. Mou, and A. Bovik. Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index. IEEE Transactions on Image Processing, 23(2):684–695, Feb 2014.

[21] L. Zhang, Y. Shen, and H. Li. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. IEEE Transactions on Image Processing, 23(10):4270–4281, Oct 2014.

[22] L. Zhang, D. Zhang, X. Mou, and D. Zhang. Feature similarity index. Online: http://www4.comp.polyu.edu.hk/~cslzhang/IQA/FSIM

[23] L. Zhang, D. Zhang, X. Mou, and D. Zhang. Fsim: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing, 20(8):2378–2386, Aug 2011.