

MYSTERY BEHIND SIMILARITY MEASURES MSE AND SSIM

Gintautas Palubinskas

German Aerospace Center DLR, Remote Sensing Technology Institute
Oberpfaffenhofen, 82234 Wessling, Germany

ABSTRACT

Similarity or distance measures play an important role in various pattern recognition applications such as classification, clustering, change detection, information retrieval, energy minimization and optimization problems. We shall analyze theoretically the two most popular quality measures MSE and SSIM used in image processing by showing their origin, similarities/differences and advantages/drawbacks. Both measures depend on the same parameters: sample means, standard deviations and correlation coefficient. It is shown that SSIM originates from two Dice measures and thus inherit their main drawback - dependence on the absolute mean and standard deviation values. Similarly, MSE depends on the absolute standard deviation values. A new similarity measure Composite quality index based on Means, Standard deviations and Correlation coefficient (CMSC) is proposed inheriting advantages of the both measures but at the same time avoiding their drawbacks.

Index Terms— Similarity, distance, Euclidian, Dice, composite, correlation coefficient, quality index, image

1. INTRODUCTION

Similarity or distance measures are unavoidable in solving various signal/image processing problems such as restoration, de-noising, registration/matching, segmentation, classification, detection and recognition (for a survey see [1]). Usually broadly recognized and accepted measure - Mean Squared Error (MSE) - is used. Recent investigation [2] shows its weakness in some applications e.g. visual perception of images. A new Structural SIMilarity index (SSIM) proposed in [3-4] is spreading fast not only in computer vision community but also other communities such as remote sensing with a very different tasks, e.g. pan-sharpening [5]. The question is arising (despite recent publications [6-9]): can we simply transfer SSIM to other applications requiring mainly relative comparison of data? To answer this question we have looked theoretically what is **really** behind MSE and SSIM measures. This analysis allowed us to detect advantages and drawbacks of these two measures. Moreover, this analysis resulted in the proposal of a new similarity measure - Composite quality index based on sample moments (Mean, Standard deviation) and Correlation coefficient (CMSC), which inherits advantages

of both measures at the same time avoiding their drawbacks. Its variants CMSC_{am} (averaging/multiplication), CMSC_m (multiplication) and CMSC_a (averaging) differ only in how individual similarities are combined.

2. THEORY

First, we shall introduce some notations used in this paper. Let $x = \{x_i | i = 1, \dots, N\}$ and $y = \{y_i | i = 1, \dots, N\}$ denote two images or image patches or more generally signals to be compared, where x_i, y_i - real numbers with a finite range of values $\min \leq x_i, y_i \leq \max$ (e.g. $\min=0$ and $\max=255$ for 8bit images), $R = \max - \min$, N - number of pixels/samples.

Distance or dissimilarity d is defined as a measure indicating how close/far apart are two samples/objects. It exhibits high values for objects which are far from each other and low values for near objects. For example, Euclidian measure is probably the most popular distance measure. The inverse measure to distance d is a similarity measure s , which exhibits high values for similar objects and low values for different objects. Here we can mention a correlation coefficient as a most popular similarity measure. For scaled/normalized measures hold the following relationships. For distance d scaled to interval $0 \leq d \leq 1$, e.g. using $d_{norm} = (d - d_{min}) / (d_{max} - d_{min})$, the corresponding similarity is simply equal to $s = 1 - d$. Analogously, for similarity s normalized to interval $0 \leq s \leq 1$ the corresponding distance is simply equal to $d = 1 - s$. Two or more similarity measures can be combined/composed by averaging, summation and multiplication operators. For example, two distance measures d_1 and d_2 , each normalized to interval $0 \leq d_1, d_2 \leq 1$, first are transformed to individual similarities $s_1 = 1 - d_1$ and $s_2 = 1 - d_2$, then a composite similarity is calculated by averaging $s_{ave} = (s_1 + s_2) / 2 = 1 - (d_1 + d_2) / 2$ or multiplication $s_{mult} = s_1 \cdot s_2 = (1 - d_1) \cdot (1 - d_2)$. Due to $0 \leq d_1, d_2 \leq 1$ the following relation holds $s_{mult} \leq s_{ave}$. Of course, mixed composite measures are possible e.g. $s = 1/2 \cdot (s_1 + s_2) \cdot s_3$. All distances and similarities analyzed in this work are summarized in Table 1.

2.1. MSE

Mean squared error (MSE) is a very popular distance measure (based on original data) and is defined as

$$d_{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2. \text{ We shall use the normalized}$$

version of MSE (nMSE) in this work (2.2, Table 1).

It was shown experimentally in [2] that MSE can be rather poor in some cases especially for visual image quality perception. Some properties of MSE were analyzed, but it is still not clear the real reasons for such behavior of MSE. In this paper we shall try to look theoretically what is really behind MSE.

Using sampled statistics: μ_x, σ_x^2 and correlation coefficient ρ we can rewrite d_{nMSE} (2.2, Table 1) as already proposed in [8] (2.3, Table 1). We see that d_{nMSE} is a sum of two distances: d_1 – normalized squared Euclidian measure for means (2.1, Table 1) and d_2 . Because of $0 \leq (d_1 + d_2) \leq 1$ the similarity for d_{nMSE} can be defined as $s_{nMSE} = 1 - (d_1 + d_2)$.

2.2. SSIM

Structural similarity (SSIM) measure proposed in [3-4] can be written as a composite measure (multiplication) of three similarities: luminance (mean) s_l , contrast (standard deviation) s_2 and correlation coefficient s_3 (1.3, Table 1). Constants for avoiding singular case (zero in denominator) are omitted for simplicity. We see that it is based on the same sample moments and correlation coefficient as MSE. So this is the **first observation/property or mystery revealed about MSE and SSIM: both measures are composed of the same parameters which are only combined in a different way.**

We can see easily that the first two similarities of SSIM are Dice measures (1.1, Table 1), which were independently introduced by botanists Dice [10] and Sørensen and can be easily extended to vector data [1]. The authors of SSIM never mention such origin of SSIM [6-7]. So this is the **second observation/property or mystery revealed about SSIM: SSIM is composed of two Dice measures one for means and another one for standard deviations.**

In Fig. 1 Dice measure values are presented for non-negative data values $0 \leq \mu_x \leq 255$ and $0 \leq \mu_y \leq 255$. So the following two observations about Dice measure are valid: it is unstable around zero point (0,0) and it cannot be used as a similarity measure for data with different signs.

So the **third observation for SSIM is its instability around zero point (0,0) and the fourth one – it can be used only for data of the same sign.** The authors of SSIM solve these problems by introducing small constants and restricting the usage to non-negative data only, respectively.

It can be seen from (1.1, Table 1) directly that e.g. for $\mu_x = 0$ and any $\mu_y = \{0, \dots, 255\}$ the Dice similarity measure is equal to 0. Moreover, it depends not only on the

relative difference of two values (what is expected from the measure) but also on the absolute values as can be seen in Fig. 2 for different line profiles (Fig. 1).

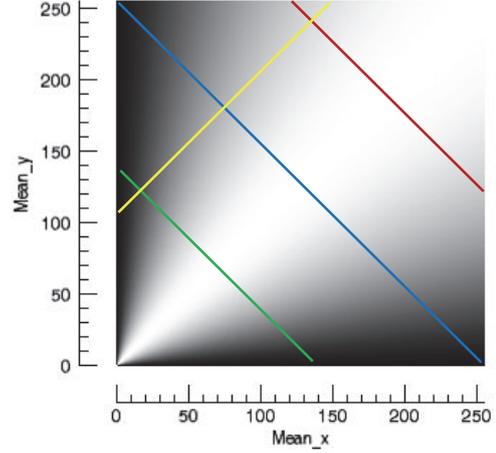


Figure 1. Dice similarity measure for $\mu_x = \{0, \dots, 255\}$ and $\mu_y = \{0, \dots, 255\}$.

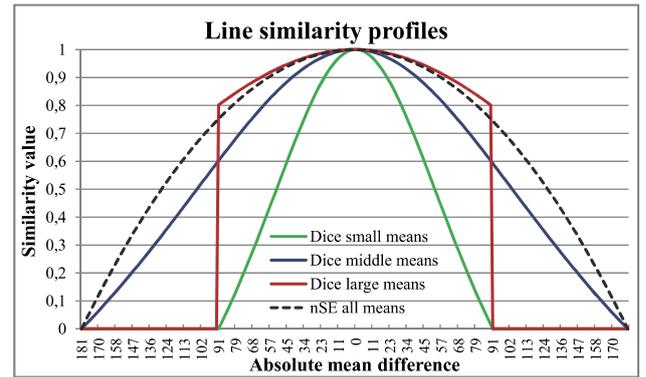


Figure 2. Line profiles of Dice and normalized squared Euclidian (nSE) similarity measures for three lines (see Fig. 1) in dependence of mean difference. Red color stands for large mean values, blue – middle, green – small and black dotted line – all means (nSE).

Line profiles of Dice and normalized squared Euclidian (2.1, Table 1) similarities for constant mean difference $\mu_y - \mu_x = 100$ in dependence on the absolute mean value $\mu_x = \{1, \dots, 155\}$ (yellow line in Fig. 1) are plotted in Fig. 3. As expected nSE exhibits a constant value whereas the Dice measure increases with the increase of the absolute value of one of the parameters.

The **fifth observation for Dice measure and thus for SSIM too is that it depends on the absolute values of input parameters.** First, it is insensitive at all if one of the parameters is equal 0. Secondly, its sensitivity is decreasing by the increase of absolute parameter values.

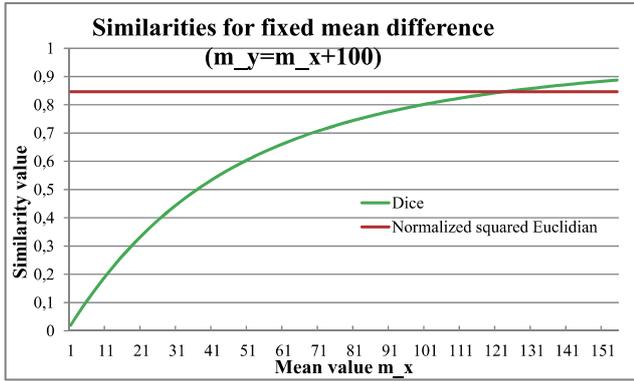


Figure 3. Line profiles of Dice and normalized squared Euclidian (nSE) similarities for constant mean difference $\mu_y - \mu_x = 100$ and $\mu_x = \{1, \dots, 155\}$ (yellow line in Fig. 1).

Usually objective similarity measure should be dependent only on the relative difference of the two parameters and independent of the absolute parameter values.

Normalized squared Euclidian distance for means (2.1, Table 1) e.g. defined as $0 \leq d_{nSE} \leq 1$ and its corresponding similarity measure $s_{nSE} = 1 - d_{nSE}$ depends only on the relative mean difference. Thus it fulfills objective measure requirement what is illustrated in Fig 4. Moreover, it is easily seen that it can be defined for real data (independent of sign).

2.3. DISCUSSION

Here we shall perform analysis of composite similarity measures s_{SSIM}, s_{nMSE} based on the information presented in Table 1.

First, we observe that under the assumptions of $\mu_x = \mu_y, \sigma_x = \sigma_y$ the SSIM similarity is simply reduced to the correlation coefficient $s_{SSIM} = \rho$. Secondly, we see that under following assumptions of $\mu_x = \mu_y, \sigma_x = \sigma_y = 1, R = \sqrt{2}$ the nMSE similarity is also reduced to the correlation coefficient $s_{nMSE} = \rho$ what shows an identity of both measures under special (quite similar) conditions.

Analysis of nMSE second similarity term based on distance d_2 (2.3, Table 1) shows its dependence on the absolute values of two standard deviation values (see Fig. 5). Only for $\rho = 1$ it converges to a squared Euclidian measure as can be seen in Fig. 5 and Table 1.

This last observation has served as an inspiration for a new composite similarity measure – CMSC – which exploits advantages of both MSE and SSIM measures at the same time avoiding their drawbacks.

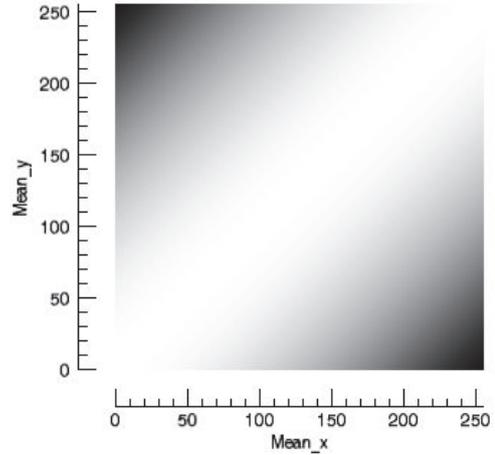


Figure 4. Euclidian similarity measure (nSE) for $\mu_x = \{0, \dots, 255\}$ and $\mu_y = \{0, \dots, 255\}$.

2.4. NEW SIMILARITY MEASURE CMSC

After analyzing two most popular composite similarity measures MSE and SSIM we can propose a new Composite image quality measure based on Means, Standard deviations and Correlation coefficient (CMSC) and consisting of the three components: two normalized squared Euclidian measures and one correlation coefficient (3.1-3.3, Table 1). Depending on the way of combination three versions are possible. CSMCam (3.1) uses averaging and multiplication of individual similarities, CSMCm (3.2) - only multiplication of similarities and CMSCa (3.3) – only averaging of similarities. We have to note, that averaging of similarities gives a possibility for weighting.

It is easy to prove that for the normalization of the second distance d_2 (3.1, Table 1) including standard deviations a two times smaller constant $R/2$ can be used.

All proposed measures are free of drawbacks of MSE and SSIM and thus are more suitable as objective similarity/quality measures not only for the images but any signals.

3. CONCLUSIONS

Theoretical analysis of MSE and SSIM similarity measures is performed, resulting in a new composite similarity measure CMSC. Preliminary experiments on simulated and real data covering various image distortions: mean shift, contrast change, various types of noise (additive Gaussian, multiplicative speckle and impulsive salt&pepper) and blurring support theoretical results. Further research can be conducted towards introducing additional gradient, texture, spectral information for CMSC, similarly as it was already proposed in [11] for SSIM or higher sample moments: skewness and kurtosis.

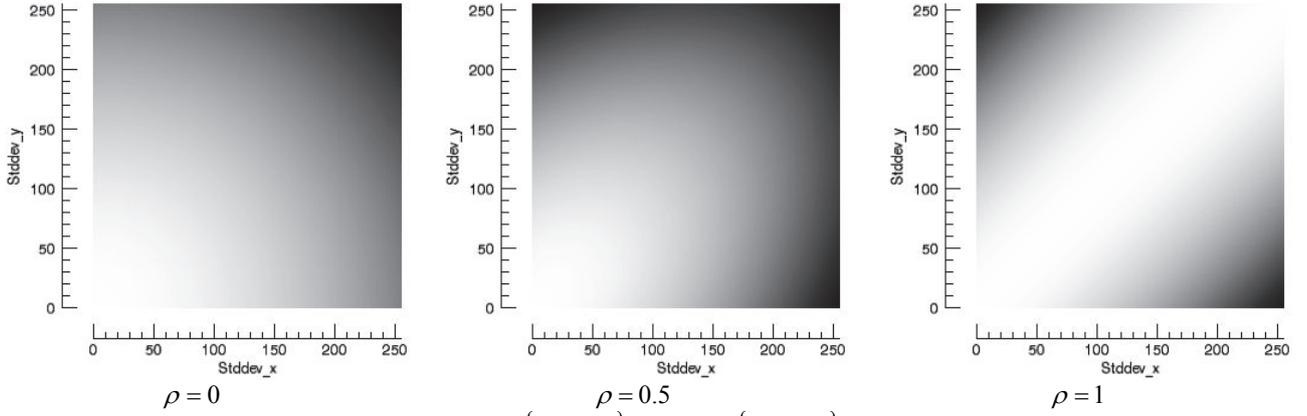


Figure 5. nMSE similarity measure for $\mu_x = \mu_y, \sigma_x = \{0, \dots, 255\}$ and $\sigma_y = \{0, \dots, 255\}$ in dependence of ρ .

Table 1. Summary of distance and similarity measures used in this work. R is a normalization constant e.g. R=255 for 8bit data.

	Distance d_1	Distance d_2	Dist d_3	Comb type	Similarity
1.1 Dice measure	$\frac{(\mu_x - \mu_y)^2}{\mu_x^2 + \mu_y^2}$	-	-	-	$\frac{2 \cdot \mu_x \cdot \mu_y}{\mu_x^2 + \mu_y^2}$
1.2 Correlation coefficient CC $0 \leq \rho \leq 1$	$1 - \rho$	-	-	-	ρ
1.3 Structural similarity SSIM	$\frac{(\mu_x - \mu_y)^2}{\mu_x^2 + \mu_y^2}$	$\frac{(\sigma_x - \sigma_y)^2}{\sigma_x^2 + \sigma_y^2}$	$1 - \rho$	Mult	$\frac{2 \cdot \mu_x \cdot \mu_y}{\mu_x^2 + \mu_y^2} \cdot \frac{2 \cdot \sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2} \cdot \rho$
1.4 SSIM $\mu_x = \mu_y, \sigma_x = \sigma_y$	0	0	$1 - \rho$	Mult	ρ
2.1 Normalized squared Euclidian nSE	$\frac{(\mu_x - \mu_y)^2}{R^2}$	-	-	-	$1 - d_1$
2.2 Normalized MSE (original data) nMSE	$\frac{1}{N \cdot R^2} \sum_{i=1}^N (x_i - y_i)^2$	-	-	-	$1 - d_1$
2.3 Normalized MSE (sample moments) nMSE	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{\sigma_x^2 + \sigma_y^2 - 2 \cdot \sigma_x \cdot \sigma_y \cdot \rho}{R^2}$	-	Sum	$1 - (d_1 + d_2)$
2.4 nMSE $\mu_x = \mu_y, \sigma_x = \sigma_y = 1,$ $R = \sqrt{2}$	0	$\frac{2 \cdot (1 - \rho)}{R^2}$	-	Sum	ρ
2.5 nMSE $\rho = 0$	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{\sigma_x^2 + \sigma_y^2}{R^2}$	-	Sum	$1 - (d_1 + d_2)$
2.6 nMSE $\rho = 1$	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{(\sigma_x - \sigma_y)^2}{R^2}$	-	Sum	$1 - (d_1 + d_2)$
3.1 Composite quality index (Means, Standard deviations, Correlation coefficient) CMSCam	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{(\sigma_x - \sigma_y)^2}{(R/2)^2}$	$1 - \rho$	Ave/ Mult	$\left(1 - \frac{(d_1 + d_2)}{2}\right) \cdot \rho$
3.2 CMSCm	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{(\sigma_x - \sigma_y)^2}{(R/2)^2}$	$1 - \rho$	Mult	$(1 - d_1) \cdot (1 - d_2) \cdot \rho$
3.3 CMSCa	$\frac{(\mu_x - \mu_y)^2}{R^2}$	$\frac{(\sigma_x - \sigma_y)^2}{(R/2)^2}$	$1 - \rho$	Ave	$\frac{2}{3} - \frac{d_1 + d_2}{3} + \frac{\rho}{3}$

4. REFERENCES

- [1] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions", *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4), 300-307, 2007.
- [2] Z. Wang, and A. C. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures", *IEEE Signal Processing Magazine*, 26(1), 98-117, 2009.
- [3] Z. Wang, and A. C. Bovik, "A universal image quality index", *IEEE Signal Processing Letters*, 9(3), 81-84, 2002.
- [4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Processing*, 13(4), 600-612, 2004.
- [5] G. Palubinskas, "Fast, simple and good pan-sharpening method", *Journal of Applied Remote Sensing*, 7(1), pp. 1-12, 2013.
- [6] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: a new image similarity index", *IEEE Trans. Image Processing*, 18(11), 2385-2401, 2009.
- [7] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index", *IEEE Trans. Image Processing*, 21(4), 1488-1499, 2012.
- [8] A. Horé, and D. Ziou, "Image quality metrics: PSNR vs. SSIM", Proc. 20th Int. Conference on Pattern Recognition ICPR, 23-26 August, 2010, Istanbul, Turkey, pp. 2366-2369, 2010.
- [9] R. Dosselmann, and X.D. Yang, "A comprehensive assessment of the structural similarity index", *SIViP*, 5, pp. 81-91, 2011.
- [10] L.-R. Dice, "Measures of the amount of ecologic association between species", *Ecology*, 26, 297-302, 1945.
- [11] E. Blasch, X. Li, G. Chen, and W. Li, "Image quality assessment for performance evaluation of image fusion", Proc. of 11th International Conference on Information Fusion, 30 June - 3 July, 2008, Cologne, Germany, pp. 1-6, 2008.