A New Similarity Measure for Complex Amplitude Holographic Data

Ahar, Ayyoub; Birnbaum, Tobias; Jaeh, Christian; Schelkens, Peter

Published in:
SPIE Optics & Photonics - Applications of Digital Image Processing XL

DOI:
10.1117/12.2274761

Publication date:
2017

Document Version:
Final published version

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

A new similarity measure for complex amplitude holographic data

Ayyoub Ahar
Tobias Birnbaum
Christian Jaeh
Peter Schelkens

SPIE.

A New Similarity Measure for Complex Amplitude Holographic Data

Ayyoub Ahar[ac*], Tobias Birnbaum[ac], Christian Jaeh[b], and Peter Schelkens[ac]

[a]Vrije Universiteit Brussel (VUB), Dept. of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, Belgium
[b]Department of Mathematical Sciences, School of Science, Loughborough University, LE11 3TU, Leicestershire, United Kingdom
[c]imec, Kapeldreef 75, B-3001 Leuven, Belgium

ABSTRACT

In this research, we have adapted our recently proposed Versatile Similarity Measure (VSM) for holographic data analysis. This new measure benefits from nice mathematical properties like boundedness to [0;1], relative error weighting based on the magnitudes of the signals, steerable similarity between original and negative phase; symmetry with respect to ordering of the arguments and the regularity of at least a continuous function. Utilizing its versatile design, here we present a set of VSM constructions specifically tailored to best fit the characteristics of complex wavefield of holograms. Also performance analysis results are provided by comparing the proposed constructions as fast, stand-alone perceptual quality predictors to few available competitors of the field, namely MSE and the average SSIM of the real and imaginary parts of holograms. Comparing their visual quality prediction scores with the mean opinion scores (MOS) of the hologram reconstructions shows a significant gain for all of the VSM constructions proposed in this paper, paving the way towards designing highly efficient perceptual quality predictors for holographic data in the future and also representing the potential of utilizing VSM for other applications working with complex valued data as well.

Keywords: VSM, Digital Holography, Quality Assessment, PQP, Complex analysis

1. INTRODUCTION

Digital Holography provides the prospect to enable a high quality of experience (QoE) in terms of 3D visual perception. In medical imaging for example Digital Holographic Microscopy facilitates a highly accurate axial(depth) resolution which is exploited in a number of applications like study of undisturbed cell processes and their morphology[1] , red blood cell studies[2] and tomography[3] . One setup to record a digital hologram requires the recording of interference patterns at multiple depths of a highly coherent and collimated reference beam and the reflected beam from the object. From this the full information of the electro-magnetic field, which is complex valued, can be inferred in one plane. In Computer Generated Holography this complex valued data can be obtained directly by simulating the light propagation. From this data the object can be visualized by reversing the light propagation numerically or optically. Thereby the focal plane and perspective (within the recorded field of view) can be chosen freely.

One might assume this visual complex data can be handled by repeating a desired operation from regular 2D imagery for the polar or Cartesian coordinate components independently. But even if mathematically equivalent, the polar form was found to be subjected to highly non-linear distortions with respect to its components. This behavior is known from digital image processing[4–7] and persists for holograms, where the phase encodes — amongst others — the depth information. However, for tasks like compression and consequently visual quality assessment of the compressed holograms, even the components in a Cartesian coordinate system are highly correlated and an averaged evaluation on split parts is unfavorable. Over and above, the visual quality assessment of holograms receives an extra level of complexity comparing with regular imaging modalities. This refers to the fact that data processing and channel distortions degrade the hologram while the visual quality of the encoded

*E-mail: aahar@etro.vub.ac.be

Applications of Digital Image Processing XL, edited by Andrew G. Tescher, Proc. of SPIE
Vol. 10396, 103961I · © 2017 SPIE · CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2274761

scene can be judged only after the reconstruction process. On top of that the perceived quality depends on the chosen perspective as well.

An efficient algorithm for perceptual quality prediction of holographic data, should be able to effectively analyze the hologram and predict the perceived quality of the reconstructed object ideally, without performing the reconstruction process. Thus the judgment of the quality of a holographic visualization should be based on its unique complex valued representation. Such algorithms primarily require an efficient measure to compare the complex valued distorted holograms with the reference. Euclidean distance based error measures like Mean Squared Error (MSE) are widely utilized for measuring the dissimilarity of holograms since it is possible to apply them directly on complex data, even though MSE is not bounded and, more importantly, lacks local interpretability. Alternatively, one may use the conventional similarity measures defined for real-valued data to separately evaluate the real-valued components of any representation of complex numbers in holograms. In [8] Structural SIMilarity (SSIM) was used to calculate the similarity of the real and imaginary parts of holograms and then obtain the final score by averaging the two scores. However, that way any correlation between the two parts is discarded and it is also not possible to discriminate between holograms, one with extremely high quality on one channel while the other one is extremely low and another hologram with equally half quality level on both channels.

A successor of the SSIM which is invariant to the size of the matched patterns is the Complex-Wavelet SSIM (CWSSIM)[9] . Although its name may lead one to think that it would be suited for complex valued data, it is designed for real-valued data on a fundamental level. The evaluation of the CWSSIM consists of three stages: 1) Fourier transform, 2) wavelet transform with a complex-valued steerable wavelet pyramid construction, 3) evaluation of a variation of the SSIM on the complex-valued wavelet coefficients. The Fourier transform is used to incorporate low sensitivity to small translations and rotations which manifest itself in the Fourier domain as constant phase shifts. The wavelet transform adds a scale invariance and separates phase from magnitude distortions whilst being insensitive to constant phase shifts. The final component is a version of the SSIM that is defined on complex-valued data. It equals the SSIM with respect to absolute values and *ignores* constant phase shifts. These constant-phase shifts are resulting from small rotations, translations, luminance or contrast changes eventually only after stages 1) and 2), if real-valued input data is assumed. As such the CWSSIM is ill-fit for complex valued data.

Recently in[10] , a new versatile similarity measure (VSM) was introduced specifically designed for error analysis of data containing complex values. In this research, we present a set of VSM constructions that can be used for analysis of complex wavefields. All the new constructions inherit the general characteristics of the VSM framework, which means every construction benefits from nice mathematical properties like boundedness to $[0, 1]$, relative error weighting based on the magnitudes of the signals, steerable similarity between original and negative phase, symmetry with respect to ordering of the arguments, and the regularity of at least a continuous function. In section [2], the VSM framework and the proposed constructions are detailed. In section [3], the experimental set up for performance evaluation of proposed measures is explained. In section [4], the evaluation results are provided. The proposed methods are compared with MSE and SSIM, their differences are analyzed and in the final section we provide the conclusion and outline future work.

2. VERSATILE SIMILARITY MEASURE ORIGINS

The general VSM model is formed as a product of three factors. Given two complex matrices of equal shape, we evaluate per entry:

$$\forall z, w \in \mathbb{C}: \quad \rho(z, w) = \gamma_{\mathrm{M}}(z, w) \cdot \gamma_{\mathrm{r}\phi}(z, w) \cdot \gamma_{\mathrm{a}\phi}(z, w), \tag{1}$$

where γ_{M} is a measure of the error in magnitude, $\gamma_{\mathrm{r}\phi}$ is a measure of the relative phase error, and $\gamma_{\mathrm{a}\phi}$ is a measure of the absolute phase error. This simple design makes it possible to locally interpret the similarity of data and/or to employ any pooling method for a less localized result. Furthermore, the product structure allows for an easy interpretation as it is ~ 0, if and only if any of the components fails. Individual factors may be replaced/adapted to an application independently. In the following, we will motivate the introduction of each component with simple examples, analyze their properties , and provide a set of possible options.

As mentioned earlier, the amplitude and phase parts of complex valued data behave fundamentally different. Hence, it makes sense to judge the quality of visual complex-valued data based on polar coordinates instead of their less sensitive Cartesian pendants. Aside from this we also actively investigated several ideas (not shown here) to extend notions of similarity on real-valued data for Cartesian coordinates and obtained unsatisfying results.

Case 1: Let us consider $z = 1e^{i\phi}$ and $w = \alpha e^{i\phi}$ with $\alpha \in \mathbb{R}, \phi \in [0, 2\pi)$. We will keep the parameter ranges for α, ϕ for the rest of this section. The initial motivation for the VSM was the notion of similarity for real-valued numbers, which sits at the core of the SSIM. This similarity notion is based on a bounded quantity (outside of $(0, 0)$), here simply referred to as s

$$\forall x, y \in \mathbb{R}: \quad s(x, y) = \frac{2xy}{x^2 + y^2} = \frac{1 - (x - y)^2}{x^2 - y^2} \in [-1, 1]\,. \tag{2}$$

Values close to 1 suggest high similarity of x and y, values close to -1 mark high similarity of opposite signs and values close to 0 mark large differences. This notion will suit us whenever we operate in the real domain $\mathbb{R}$. For now, we note that by using $\gamma_{\mathrm{M}}(z, w) = s(|z|, |w|)$ it coincides with eq. (2) on $\mathbb{R}_0^+$ and extends this notion by rotation around the origin. Note, that restricting the matching of $s(x, y)$ to $\mathbb{R}_0^+$ is mandatory because it is an odd function on $\mathbb{R}$ and γ_{M} should stay independent of phase differences. We thus have found one candidate for our first term measuring the magnitude similarities of arbitrary complex numbers. This provides a good foundation for our model by reusing established notions.

Case 2a: We discriminate $z = \alpha e^{i\phi}$ from its negative $-z$ directly through the absolute phase error $\gamma_{\mathrm{a}\phi}$. Intuitively it can be implemented as a (near) step function that solely depends on the absolute values of the phase differences. It assigns distinguishes chosen ranges of those absolute values of phase differences. For example those larger than 90° from those that are smaller. The number of those ranges and the assigned values are rather arbitrary choices and no single good intuition can be established here. For certain applications two intervals, one marking values "close to" z and those "close to" $-z$ with a similarity of -1, as SSIM proposes, seems good but other values such as $0, 0.5, 1$ or more intervals could have their stands too.

Case 2b: The more interesting case of arbitrary phase differences ϕ is analyzed by setting $z = 1$ and $w = e^{i\phi}$. The error in phase is measured best by the deviation of w from the line spanned by z, $-z$ since we could already take care of distinguishing these two points with $\gamma_{\mathrm{a}\phi}$ eventually. Thus we introduce the relative phase error which measures how far off the w is of that very line and assigns a similarity to it. The intuition is that a small phase difference from the line should still make up for a high similarity. The similarity should be low for 90° relative phase difference, that is maximal distance, and again depending on the application possibly other values in the range $(0°, 90°)$. For the sake of simplicity we restrict ourselves in this work to the simple case of minimal similarity for a relative phase difference of 90°, which means two similarity intervals on the unit circle. This notion can be motivated in many applications based on electromagnetic waves since they are known to possess an instantaneous representation by two orthogonal, linear polarization components. These components would correspond to the polarization planes spanned by the direction of propagation including z or its 90° rotated pendant.

Now, all three components are assembled and we facilitate some requirements for each of the components. The first property, we would like to enforce, is scale-invariance to account for mismatches in magnitude:

$$\forall \alpha \in \mathbb{R}^+ \forall z, w \in \mathbb{C} : \rho(z, w) = \rho(\alpha z, \alpha w). \tag{3}$$

Since magnitude is rated only via γ_{M} this is the only term affected by the condition. This requirement is the most intuitive way to handle differences in magnitude. Given for example $z_1 = 1, z_2 = 255, w_1 = 2, w_2 = 256$ we expect that $\gamma_{\mathrm{M}}(z_1, w_1) \ll \gamma_{\mathrm{M}}(z_2, w_2)$. This invariance can be interpreted as assigning similarities proportional to the $1 - \mathrm{dist}$, for example using the Euclidean distance, of the points inverted along the unit circle in the complex domain. This is nothing other than concatenating $\overline{z}^{-1}$ with $1 - \|z - w\|^2$ in case of the Euclidean distance. In practice, this construction will give rise to a series of other problems, such as unboundedness in the origin, but we might still keep the motivation of the relative scaling requirement for γ_{M} and search for similar functions.

One candidate is already delivered by eq. (2) and others will be given shortly. A second required property, which holds for eq. (2) too, is the symmetry with respect to the order of the arguments, i.e.

$$\forall z, w \in \mathbb{C}: \quad \rho(z, w) = \rho(w, z). \tag{4}$$

Note, ρ will be symmetric with respect to the origin in general only, if $\gamma_{\mathrm{a}\phi} \equiv 1$. From the reasoning given for the cases 2a) and 2b) the invariance under constant phase shifts of both arguments already follows

$$\forall \phi \forall z, w \in \mathbb{C} : \rho(z, w) = \rho(z e^{i\phi}, w e^{i\phi}). \tag{5}$$

Fourth, we explicitly aim for functions that are bound for all arguments. This will make normalization before pooling largely unnecessary and ensure a local as well as global interpretability. Finally, we like to use the liberty of selecting each of the components quite freely to obtain a function ρ that is at least continuous. This is a favorable outcome for any possible subsequent optimization tasks. In principle, it would be possible to extend the model, (1), with one exponent per factor, as it is already common for the SSIM. However, we decided not to add these explicit exponents at this stage, as they can be easily included into each of the single components.

These properties are visualized in fig. 1 using one exemplary choice for each of the components, each evaluated with respect to the same point of reference $3 + 0.1i$.

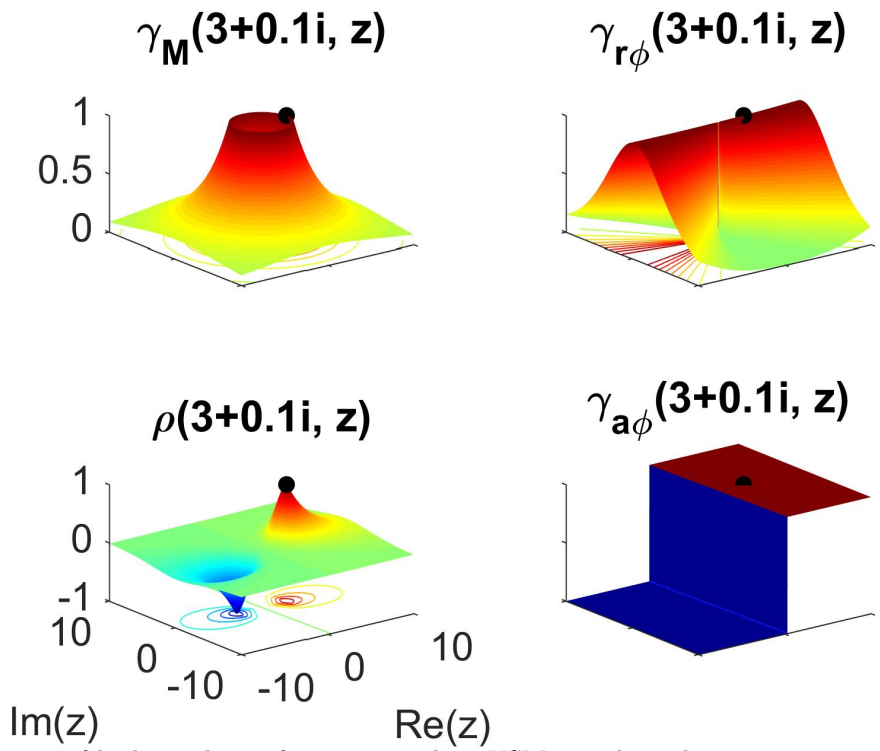
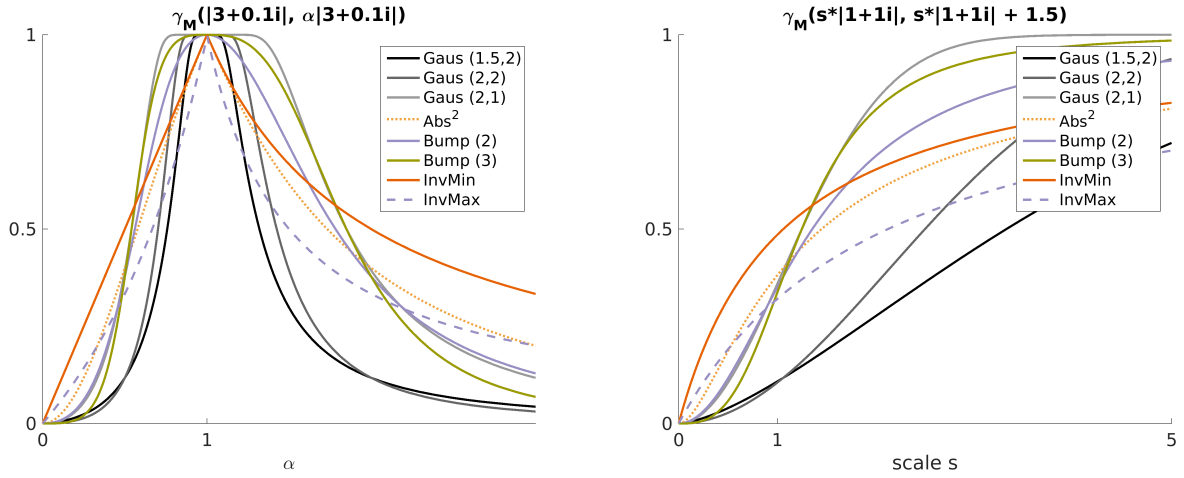


Figure 1: Regions of high similarity for an examplary VSM ρ and its three components: magnitude error γ_{M}, relative $\gamma_{\mathrm{r}\phi}$ and absolute phase error $\gamma_{\mathrm{a}\phi}$ evaluated for varying $z \in \mathbb{C}$.

We close this section with presenting some choices for each of the components. Defining the scale of two complex numbers, z, w, based on the minimum as $s := \min(|z|, |w|)$ and based on the maximum as $t := \max(|z|, |w|)$



(a) Constant reference (b) Constant offset

Figure 2: Magnitude error measure deductions alike $1-\overline{z}^{-1}$ of varying regularity and shape. The values in brackets give the parameter values. For the functions based on the Gaussian distribution the pairs are (σ, p). Shown are the behaviors for a fixed point of reference and a fixed offset.

we have for γ_{M}:

$$\gamma_{\mathrm{M}}\left\{\mathrm{Abs}^{-1}_{Min}\right\}(z,w) := \left(\frac{||z|-|w||}{s}+1\right)^{-1}, \tag{6a}$$

$$\gamma_{\mathrm{M}}\left\{\mathrm{Abs}^{-1}_{Max}\right\}(z,w) := 2\left(\frac{||z|-|w||}{t}+1\right)^{-1}-1 \tag{6b}$$

$$\gamma_{\mathrm{M}}\left\{\mathrm{Abs}\right\}(z,w) := 1-\frac{||z|^p-|w|^p|}{|z|^p+|w|^p} \quad , \; p\in[1,\infty], \tag{6c}$$

$$\gamma_{\mathrm{M}}\left\{\mathrm{Gaus}\right\}(z,w) := 1-\exp\left(-\left(\frac{s}{||z|-|w||}\right)^p \Big/ \left(2\sigma^2\right)\right) \quad , \; p,\sigma\in(0,\infty), \tag{6d}$$

$$\gamma_{\mathrm{M}}\left\{\mathrm{Bump}\right\}(z,w) := \frac{e}{e-1}\exp\left(-\left[1+\left(\frac{s}{||z|-|w||}\right)^p\right]^{-1}\right)-\frac{1}{e-1} \quad , \; p\in[1,\infty). \tag{6e}$$

Examples for the functions above are shown in Fig. 2. All these methods require a regularization in $(0,0)$ to account for $\gamma_{\mathrm{M}}(0,0) := 1$ and smoothen out existing discontinuities near the origin. The regularization is done by imposing a lower bound in form of a cone, of radius ε and height 1, centered in the origin. It is given as:

$$\mathcal{R}_1\left(\gamma_{\mathrm{M}}(x,y),x,y,\varepsilon\right) := \begin{cases} \gamma_{\mathrm{M}}(x,y) & , \; (|x|^2+|y|^2) > \varepsilon, \\ \max\left\{\gamma_{\mathrm{M}}(x,y), 1-\frac{|x|^2+|y|^2}{\varepsilon}\right\} & , \; (|x|^2+|y|^2) <= \varepsilon, \end{cases} . \tag{7}$$

Thereby it should be $\varepsilon \ll 1$, e.g. $\varepsilon = 10^{-8}$.

(6a) follows the initial idea of closely as possible whilst being bound to $[0,1]$ and as such continuous in $\mathbb{C}$. To get a finite upper bound in $\overline{z}^{-1}$, it is changed into $(\overline{z}+1)^{-1}$. If we base the scale on the maximum instead of the minimum absolute value of the two arguments, we arrive at (6b). (6c) extends the relative difference by allowing powers $p \neq 1$. It is continuous and bound to $(0,1)$. (6d) measures the difference in magnitude with respect to a scale corrected as $|z|^{-1}$ within the exponential decay of the Gauss distribution to improve regularity. Its two parameters are the standard deviation, σ, and a power, p, of the $|z|^{-1}$ dependency. A mean value different than 0 is not meaningful as 1 should be assigned only in case of exact coincidence with respect to magnitude. The

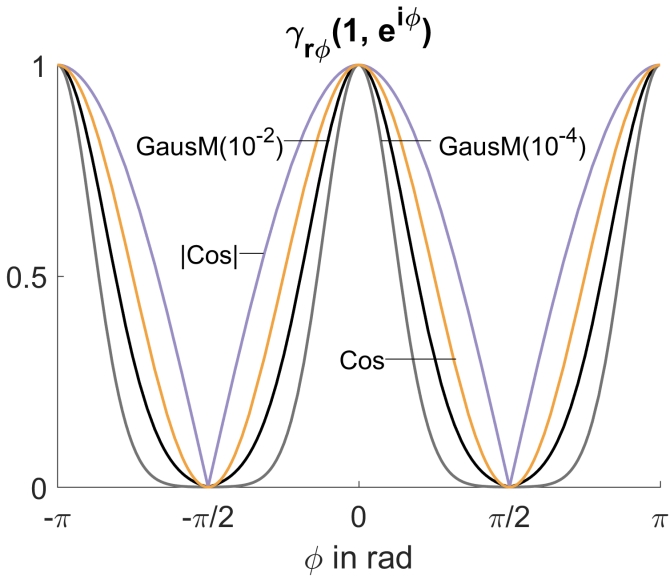


Figure 3: Some examples illustrate the variability in relative phase error measures that yield high values for small angles, decline fast, and are 0 in $\pm\pi/2$. The values in parenthesis are the parameters used.

model is C^∞ and bound to $[0, 1)$. (6e) is based on the so called bump function ($p = 2$). It takes the power p of the relative difference in magnitude as a parameter and is C^∞ as well as bound to $[0, 1)$.

All relative phase differences, $\angle(z, w) \in [0, \pi]$, are measured in terms of the complex phase, $\mathrm{ang}(z) \in [0, 2\pi)$, both defined as:

$$\angle(z, w) := \begin{cases} |\mathrm{ang}(z) - \mathrm{ang}(w)| & ,<= \pi \\ |2\pi - \mathrm{ang}(z) - \mathrm{ang}(w)| & , \text{ otherwise} \end{cases} \tag{8}$$

Some selected options for $\gamma_{r\phi}$ are displayed in Figure. 3. The propositions in this section are all based on exponential or trigonometric functions to achieve high regularity.

$$\gamma_{r\phi}\{|\mathrm{Cos}|\}(z, w) := |\cos(\angle(z, w))|, \tag{9a}$$

$$\gamma_{r\phi}\{\mathrm{Cos}\}(z, w) := \frac{1}{2}(\cos(2\angle(z, w)) + 1), \tag{9b}$$

$$\gamma_{r\phi}\{\mathrm{GausM}\}(z, w) := (1 + d)\exp\left(-a\angle(z, w)^2\right) - d \quad , a := \frac{4}{\pi^2}\log_e\left(\frac{1+d}{d}\right) \quad , d \in (0, 1) \tag{9c}$$

(9a) is a simple reflection of cos at 0, continuous but not differentiable in $\pm\pi/2$. (9b) is a shifted and squeezed cos and hence C^∞ everywhere. (9c) is based on a periodic continuation of a modified Gaussian distribution defined on $[-\pi/2, \pi/2]$. It takes a parameter d that is used to lower and rescale the Gaussian distribution such that it is 1 in 0 and 0 in $\{\pm\pi/2\}$. A mean value $\neq 0$ is not meaningful and a standard deviation $\neq 1$ would lead to non-surjectivity or unboundedness. This model is C^∞ everywhere.

We present only one option to measure the absolute phase error which ties in already perfectly with the physical orthogonality condition on $\gamma_{r\phi}$ still ensuring continuity of ρ.

$$\gamma_{a\phi}(z, w) := \lambda + (1 - \lambda)H(\frac{\pi}{2} - \angle(z, w)), \tag{10}$$

This mirrored and shifted Heaviside function has one height parameter $\lambda \in (0, 1)$, which controls directly the similarity of $\rho_t(z, -z)$. It can be interpreted via $\rho(z, -z) = \gamma_{a\phi}(z, -z) = \lambda \in [0, 1], \rho(z, z) = \gamma_{a\phi}(z, z) = 1$.

3. EXPERIMENTAL SETUP

The goal in these experiments is to investigate the practical relevance of the mentioned functions in previous section for the three modules of VSM (ρ) for specific case of Holographic data. Table 1, shows the selected functions and corresponding parameters settings. A total number of 220 constructions were tested for the VSM based on the 3 $\gamma_{\mathrm{r}\phi}$ and 5 γ_{M} (with parameters as mentioned in Table 1). All 220 constructions were tested on subjectively rated holograms of the INTERFERE-I database which contains 5 references and 96 CGH compressed holograms. More details about tested holograms can be found in [11]. After calculating pairwise VSM values for each Hologram pair, a simple median and average pooling was performed to calculate the overall prediction scores for the distorted hologram. Here after, VSM constructions with median and mean pooling will be represented as VSM_{med} and VSM_{μ} respectively.

We also measured MSE and channel wise SSIM for real and imaginary parts of those holograms. The MSE results were normalized by the square of the Frobenius norm of the reference hologram hereafter will be shown as the Normalized Mean Square Error (NMSE). The SSIM results were averaged over the real and imaginary channels and through the text will be indicated as SSIM_{μ}. Later on the $1-\mathrm{VSM}$ and $1-\mathrm{NMSE}$ were used as similarity measures. The results from all 222 measures (220 VSM's plus NMSE and SSIM_{μ}) were compared against the Mean Opinion Scores (MOS) gathered from the human subjects who rated the quality of the reconstructed holograms. The evaluation criteria used to compare the performance of the above similarity measures are the Mean Absolute Difference (MAD) between MOS scores and the predicted similarity measures, Pearson, Spearman and Kendall correlation coefficients. For the MAD we simply averaged over the absolute difference between 96 MOS for all holograms and the predicted quality scores of each quality measure. The MAD value is used as an indicative of the prediction accuracy, thus lower MAD represents better overall MOS prediction through the whole dataset. The Pearson Correlation (CorrP) shows us how linear the behaviour is for each quality measure compared to MOS. The Spearman rank order correlation (CorrS), while being less sensitive to outliers than CorrP, is used to provide us an estimation of monotonicity for each quality measure compared to the MOS. Kendall's tau correlation (CorrK) is calculated to measure the degree of associativity between MOS and predicted scores.

Table 1: VSM modules, corresponding functions and their parameter settings, overall making 220 combinations to be used as VSM constructions in the experiments.

VSM Module	Function	Parameters
$\gamma_{\mathbf{M}}$	$\gamma_{\mathrm{M}}\left\{\mathrm{Abs}^{-1}_{Min}\right\}(z,w)$ (6a)	-
	$\gamma_{\mathrm{M}}\left\{\mathrm{Abs}^{-1}_{Max}\right\}(z,w)$ (6b)	-
	$\gamma_{\mathrm{M}}\{\mathrm{Abs}\}(z,w)$ (6c)	$p \in \{1,2\}$
	$\gamma_{\mathrm{M}}\{\mathrm{Gaus}\}(z,w)$ (6d)	$(p,\sigma) \in \{(1,1),(1,2),(2,1),(2,2)\}$
	$\gamma_{\mathrm{M}}\{\mathrm{Bump}\}(z,w)$ (6e)	$p \in \{1,2,3\}$
$\gamma_{\mathbf{r}\phi}$	$\gamma_{\mathrm{r}\phi}\{\lvert\mathrm{Cos}\rvert\}(z,w)$ (9a)	-
	$\gamma_{\mathrm{r}\phi}\{\mathrm{Cos}\}(z,w)$ (9b)	-
	$\gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}(z,w)$ (9c)	$d \in \{10^{-4},10^{-3},10^{-2}\}$
$\gamma_{\mathbf{a}\phi}$	$\gamma_{\mathrm{a}\phi}(z,w)$ (10)	$\lambda \in \{-1,0,0.5,1\}$

4. ANALYSIS OF THE RESULTS AND DISCUSSION

In this section, the experimental results from the evaluation test are provided. Figure 4 represents the results for all the VSM constructions in terms of MAD. The corresponding values for NMSE and SSIM_{μ} are depicted in the graph as an straight line (repetition of the same value) for the sake of easier comparison. First, one can see that for almost all VSM_{med} the prediction error is larger than VSM_{μ}, representing less overall accuracy. This can be due to less sensitivity of the Median to the outliers (which in this case will be certain points or regions

Table 2: Best VSM constructions with respect to the evaluation criteria and pooling strategies applied.

Criterion	Pooling	Corresponding Best VSM Construction	Index
MAD (lowest)	Median	VSM1 $= \gamma_{\mathrm{M}}\{\mathrm{Abs}\}_{p=2} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}_{d=10^{-4}} \cdot \gamma_{\mathrm{a}\phi}\{\lambda=-1\}(z,w)$	29
	Mean	VSM2 $= \gamma_{\mathrm{M}}\{\mathrm{Gaus}\}_{(p,\sigma)=(2,1)} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}_{d=10^{-4}} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	89
CorrP (highest)	Median	VSM3 $= \gamma_{\mathrm{M}}\{\mathrm{Gaus}\}_{(p,\sigma)=(2,1)} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}_{d=10^{-4}} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	89
	Mean	VSM4 $= \gamma_{\mathrm{M}}\{\mathrm{Gaus}\}_{(p,\sigma)=(2,1)} \cdot \gamma_{\mathrm{r}\phi}\{\lvert\mathrm{Cos}\rvert\} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	85
CorrS (highest)	Median	VSM5 $= \gamma_{\mathrm{M}}\{\mathrm{Gaus}\}_{(p,\sigma)=(1,2)} \cdot \gamma_{\mathrm{r}\phi}\{\lvert\mathrm{Cos}\rvert\} \cdot \gamma_{\mathrm{a}\phi,\lambda=0.5}(z,w)$	67
	Mean	VSM6 $= \gamma_{\mathrm{M}}\{\mathrm{Bump}\}_{p=3} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{Cos}\} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	161
CorrK (highest)	Median	VSM7 $= \gamma_{\mathrm{M}}\{\mathrm{Abs}\}_{p=1} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}_{d=10^{-4}} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	9
	Mean	VSM8 $= \gamma_{\mathrm{M}}\{\mathrm{Bump}\}_{p=3} \cdot \gamma_{\mathrm{r}\phi}\{\mathrm{Cos}\} \cdot \gamma_{\mathrm{a}\phi,\lambda=-1}(z,w)$	161

on the hologram which are highly degraded) affecting the overall quality of the reconstruction. Apart from this, it can be seen that irrespective to the pooling strategy, almost all of the VSM constructions show less average difference compared to the $NMSE$ values. Also compared to the SSIM_μ, many of the constructions, especially the ones with mean pooling, show better performance. However, in certain cases the constructions show more overall difference than SSIM_μ.

Figure 5 represents the Pearson correlation coefficient for all the tested methods. similar to the Figure 4 values for NMSE and SSIM_μ are represented as straight lines. Interestingly, it turns out all VSM constructions irrespective to the pooling strategy clearly outperform both NMSE and SSIM_μ in terms of linear correlation with MOS. Also it can be seen that SSIM_μ is slightly better than NMSE . In Figure 6, where results for Spearman rank correlation are shown, again a significant gap between all VSM constructions and the other two competitors can be seen. However, here it is the NMSE which shows better performance than SSIM_μ. The Kendall's tau correlations are shown in Figure 7, where similar to the case of CorrS, all VSM constructions clearly perform better than the other two measures and again NMSE stands higher than SSIM_μ. Overall, the results of our experiments demonstrates a consistently better performance of VSM (for all the functions proposed for different modules) over the NMSE and SSIM_μ and this is irrespective to the parameters settings and pooling strategy.

The next goal is to find the best possible construction for VSM for the tested combinations of proposed functions which are summarized in table 1. However, depending on the evaluation criterion and pooling strategy, different constructions show slightly better performance over the others. Due to this, we were unable to find a single construction which consistently performs better than all and for all evaluation criteria. Table 2 shows best constructions with respect to each evaluation criterion and pooling strategy. An interesting conclusion which can be drawn by looking at these constructions, is that for 7 out of 8 constructions, for $\gamma_{\mathrm{a}\phi}(z,w)$ in (10), $\lambda=-1$. This is important because it shows that at least for holographic data modality the best choice(w.r.t to subjective ratings) for maximum phase dissimilarity is equal to phase shift of π. Moreover, wherever $\gamma_{\mathrm{r}\phi}\{\mathrm{GausM}\}$ (9c) appears, $d=10^{-4}$, suggesting that this parameter should be very small.(e.g values smaller than 10^{-4} may slightly improve the performance).

Finally in Figure 8, the correlation values of 8 constructions depicted in Table 2 are shown side by side NMSE and SSIM_μ . As we mentioned previously, it is obvious that any of the tested VSM constructions and especially the ones presented in Table 2 not only outperform the conventional quality measures for all evaluation criteria but also suggest a significant compatibility with the perceptual quality of the reconstructed holograms. Considering the fact that in this research we put the least effort for pooling the final VSM score by simply using the average or Median of all pairwise calculated ρs, even more improvements can be expected to achieve if more advanced pooling strategies utilized in future.

5. CONCLUSION

In this research, The recently introduced Versatile Similarity Measure framework was utilized to propose a set of constructions for visual quality assessment of digital holograms. Every and each of the proposed constructions along with two traditional quality measures of the field namely MSE and SSIM were tested over complex wavefield of a set of subjectively rated computer generated holograms (INTERFERE-I database). The predicted similarity scores of each quality measure was compared with the provided Mean Opinion Scores of reconstructed holograms. The experimental results demonstrated consistently better performance for all of the VSM constructions proposed in this paper over the MSE and SSIM. Such a consistently good performance is suggesting the potential of VSM framework to be utilized for designing future holographic PQPs and also its capability for error analysis of other data modalities where comparison of complex valued data is required.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement n.617779 (INTERFERE).

REFERENCES

[1] Pavillon, N., Benke, A., Boss, D., Moratal, C., Khn, J., Jourdain, P., Depeursinge, C., Magistretti, P. J., and Marquet, P., "Cell morphology and intracellular ionic homeostasis explored with a multimodal approach combining epifluorescence and digital holographic microscopy," *Journal of Biophotonics* **3**(7), 432–436 (2010).

[2] Marquet, P., Rappaz, B., Barbul, A., Korenstein, R., Depeursinge, C., and Magistretti, P., "Red blood cell structure and dynamics explored with digital holographic microscopy," (2009).

[3] Jeong, K., Turek, J. J., and Nolte, D. D., "Fourier-domain digital holographic optical coherence imaging of living tissue," *Appl. Opt.* **46**, 4999–5008 (Aug 2007).

[4] Oppenheim, A. V. and Lim, J. S., "The importance of phase in signals," *Proceedings of the IEEE* **69**, 529–541 (May 1981).

[5] Hsiao, W. H. and Millane, R. P., "Effects of fourier-plane amplitude and phase errors on image reconstruction. i. small amplitude errors," *J. Opt. Soc. Am. A* **24**, 3180–3188 (Oct 2007).

[6] Ni, X. S. and Huo, X., "Statistical interpretation of the importance of phase information in signal and image reconstruction," *Statistics & Probability Letters* **77**(4), 447 – 454 (2007).

[7] Narwaria, M., Lin, W., McLoughlin, I. V., Emmanuel, S., and Chia, L. T., "Fourier transform-based scalable image quality measure," *IEEE Transactions on Image Processing* **21**, 3364–3377 (Aug 2012).

[8] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).

[9] Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., and Markey, M. K., "Complex wavelet structural similarity: A new image similarity index," *IEEE Transactions on Image Processing* **18**, 2385–2401 (Nov 2009).

[10] Ahar, A., Birnbaum, T., Jaeh, C., and Schelkens, P., "A new similarity measure for complex valued data," in [*Digital Holography and Three-Dimensional Imaging*], *Digital Holography and Three-Dimensional Imaging* , Tu1A.6, Optical Society of America (2017).

[11] Ahar, A., Blinder, D., Bruylants, T., Schretter, C., Munteanu, A., and Schelkens, P., "Subjective quality assessment of numerically reconstructed compressed holograms," (2015).

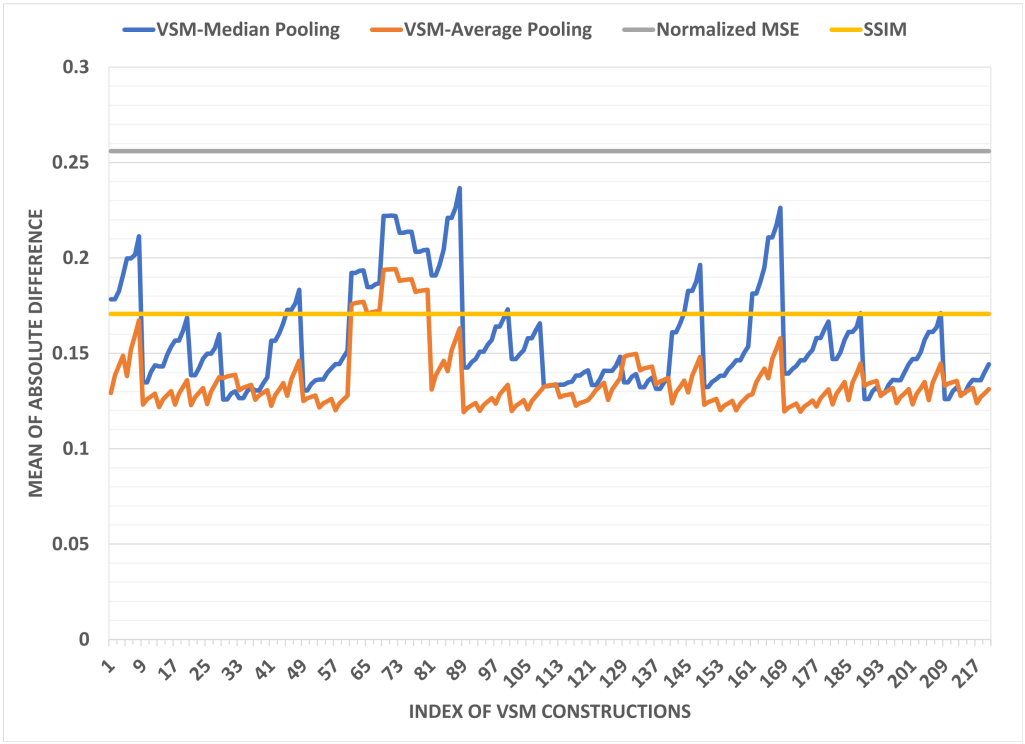


Figure 4: Mean of absolute difference between 96 MOSes and predicted scores for 220 VSM constructions with median and mean pooling. The corresponding value both for NMSE and $SSIM_{\mu}$ are shown as an straight line. Lower value considered better and represents more accurate prediction of MOSes.

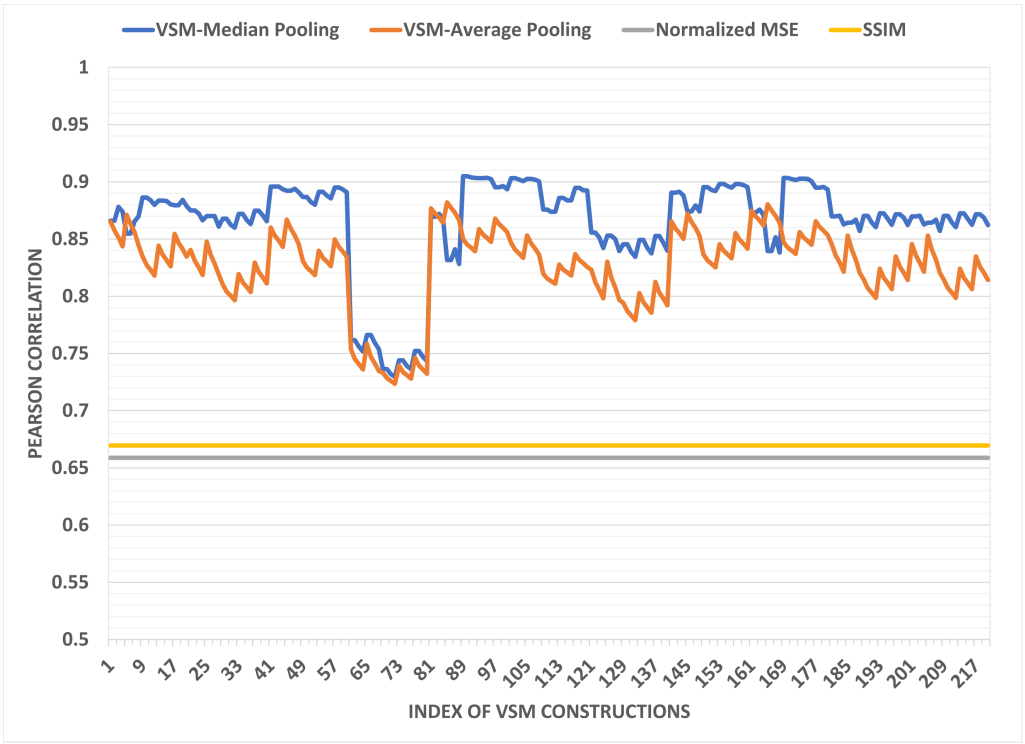


Figure 5: Pearson correlation coefficient calculated for 220 VSM constructions with median and mean pooling. The correlation value both for NMSE and SSIM is shown as a straight line.

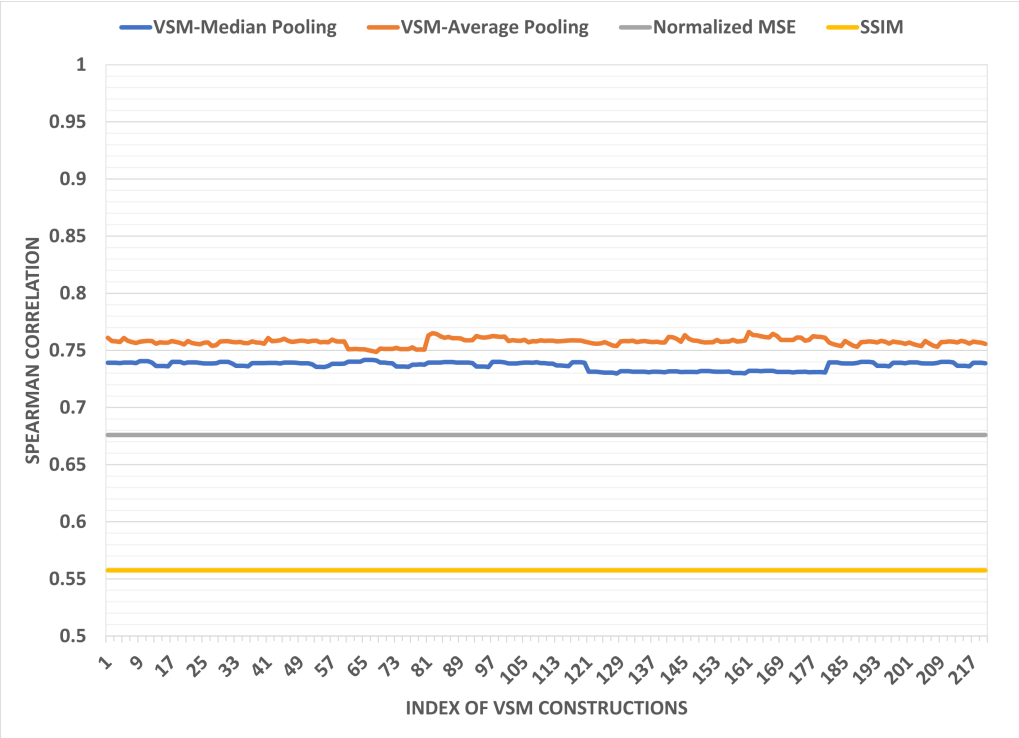


Figure 6: Spearman correlation coefficient calculated for 220 VSM constructions with median and mean pooling. The correlation value both for NMSE and SSIM is shown as a straight line.

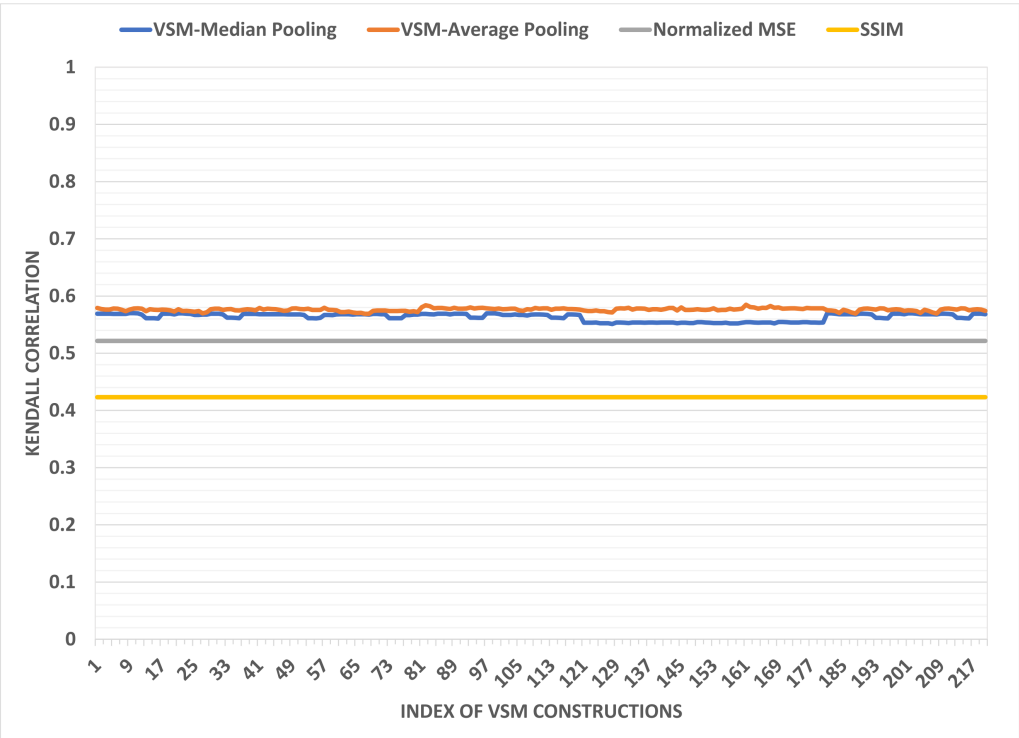


Figure 7: Kendall correlation coefficient calculated for 220 VSM constructions with median and mean pooling. The correlation value both for NMSE and SSIM is shown as a straight line.

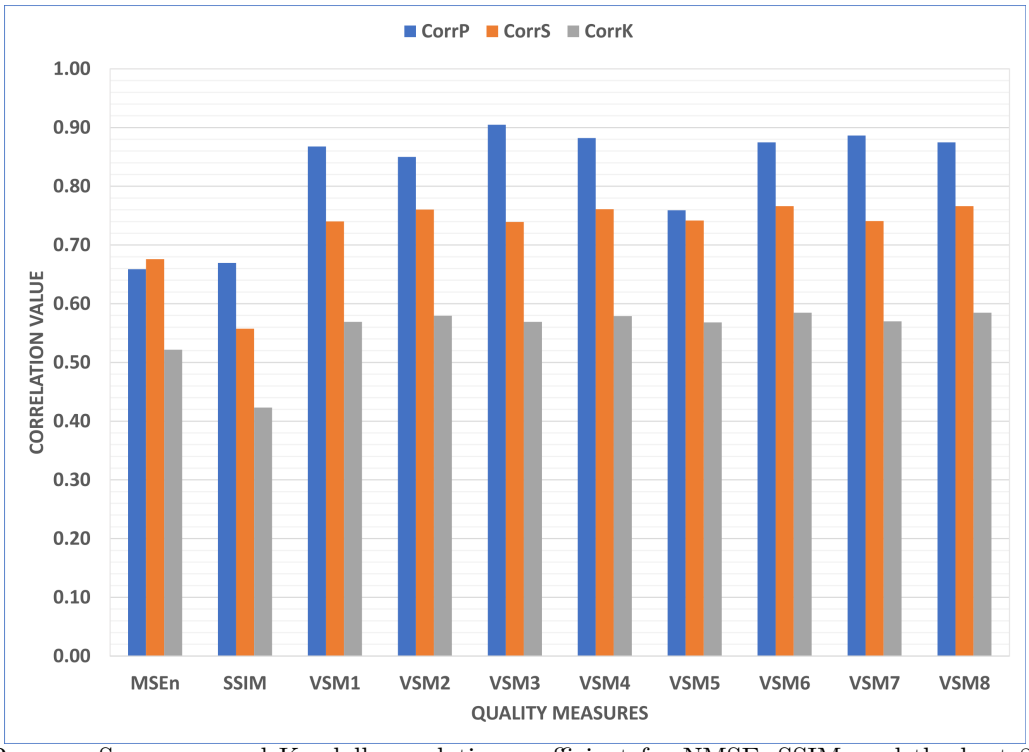


Figure 8: Pearson, Spearman and Kendall correlation coefficient for NMSE, SSIM, and the best 8 VSM constructions.